METHOD OF GEOHRONOLOGICAL TRACKING TO TEST THE HYPOTHESIS OF HISTORIOGRAPHIC RESEARCH

Ya. A. Ivakin[1], ivakin@oogis.ru,

S. N. Potapychev[1], potapychev@oogis.ru.

[1]St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation.

Information technology of the geochronological tracking is a set of processes that accumulate and integrate data on geographic movement of historical personalities for a given time interval in order to represent the results as a generalizing graph in geographic information systems (GIS). Hypotheses on the stable tendencies in migration can be represented as sub-graphs of this graph. In order to test these hypotheses, we propose estimate statistical significance of the isomorphism of corresponding graphs. We consider qualitatively new possibilities of the proposed approach and the corresponding mathematic and algorithmic apparatus.

Keywords: geographic information systems; GIS-technologies for historic research; geochronological track and tracking; isomorphism of graphs; test of statistical hypotheses; interdisciplinary research based on GIS.

Introduction

The specialized information technology of geochronological tracking is presented and described in detail in [1, 3]. Information technology of geochronological tracking is a set of methods, models and techniques, which allow to gather, pass, process, display and issue to the user an information on the generalization of geochronological tracks of historical personalities. In turn, in order to construct a geochronological track of a historical personality or historical object on the basis of the geospatial interpretation of its biographical information, we represent chronological and geographical data as a graph that connects geographical points of the historical personality location. In this case, vertices of the graph have a strict historical geographic sense, and the arcs have a conventionally logical nature.

For a representative sample of individuals, the carried out approbations and modeling of geochronological tracking showed that the final graph can have a complex and highly connected (and even fully connected) structure. Therefore, we can consider the final graph of geochronological tracking as a basis for investigating various migration processes, revealing some particular patterns in the movement of historical personalities or social groups, as well as test of statistical hypotheses on the nature of movements. The representativeness of the sample of historical personalities using for geochronological tracking determines the statistical stability of the obtained results.

One of the scientific and methodological tools providing these studies is widespread use and computer interpretation of the graph theory methods based on geochronological tracking, and in particular, algorithms to find isomorphism of graphs (to recognize the isomorphic embedding of a subgraph in a more complex graph). An interpretation of these methods in the field of research of historical informatics gives wide possibilities for the analysis of various network structures, and first of all, geochronological tracks.

Full-featured development of computer interpretation of the graph theory methods based on geochronological tracking provides new quality of historical research using modern GIS-tools. Namely, historian can use the quantitative methods of the corresponding logical-analytical apparatus. We consider detailed elaboration of the mentioned possibilities and the corresponding mathematical and algorithmic solutions, as well as a description of new ways in modern historical studies.

The article is a more complete development of the research [10], which was performed for specialists in the humanitarian field.

1. Estimation of the Statistical Significance of the Accepting Hypotheses on the Basis of Geohronological Tracking

There is obvious risk of making a decision about availability of the facts of historical and geographical process as stable structural embedding in the structure of the corresponding geochronological track. In other words, a confidence probability for a particular fact can be estimated.

An initial set of historical personalities used for geochronological tracking is considered as a general population in the mathematical apparatus to test statistical hypotheses. A sample of individuals who confirmed the corresponding hypothesis-subgraph by their movements in geographical space determines the statistical stability of the obtained research results.

In order to make a decision on accepting a hypothesis, we estimate the statistical significance of the conclusion about an existence of the corresponding isomorphic subgraph in the structure of geochronological track by the mentioned sample from the general population. In this case, the level of significance α (the risk of accepting an incorrect hypothesis) is set as the probability, which is inverse to the confidence probability of making the right decision p, i.e.

$$p = 1 - \alpha. \tag{1}$$

In order to analyse a significance of the number of individuals, which confirmed the corresponding hypothesis-subgraph by their movements in geographic space, in comparison with the total number of individuals recorded in geochronological tracking, we can use the statistical comparisons technique described in [4]. Let Δ be a difference between the mentioned numbers. In this case, we analyse the statistical difference between Δ and 0. As a rule, such an analysis of significance is carried out by the following seven logical steps.

Step 1. Propose the following two statistical hypotheses:

$$\begin{cases} H_0 : \ \Delta = 0 - \ \text{insignificant}; \\ H_1 : \ \Delta > 0 - \ \text{significant}, \end{cases} \tag{2}$$

where H_0 is a hypothesis that the obtained value Δ is statistically insignificant (i.e., there is no statistically sufficient number of individuals to identify the desired subgraph in the structure of geochronological track), H_1 is a hypothesis that the obtained value Δ is statistically significant (i.e., the number of individuals, which confirmed the desired subgraph in the geochronological track by their movements, is statistically significant and sufficient to find the subgraph).

Step 2. A dimension of the sample is set to be n (i.e., a number of trials), because the results Δ were obtained from a sample of individuals who confirmed the corresponding hypothesis-subgraph by their movements in geographical space.

Step 3. Set the level of significance α. For example, α is set to be 0.05. According to (1), the corresponding confidence probability p of accepting the research hypothesis is 0.95.

Step 4. Select test statistics from the tabulated variants taking into account size of the sample of individuals who confirmed the corresponding hypothesis-subgraph by their movements in geographical space. If dimensions of the general population and the sample are such that the statistical law of large numbers is fulfilled, then we can take values having widely known distributions as test statistics. For example, the distribution is χ^2. In the case of small (in the statistical sense) samples, we can take values having Student's t-distribution with $(n-1)$ degrees of freedom as test statistics.

Step 5. Determine a region of accepting the statistical hypothesis. To this end, for the accepted level of significance α, we choose the critical value t_0 of the test statistic from the corresponding tables (database). According to (1), the level of significance α determines the confidence probability p.

Step 6. Form the final mathematical rule for testing the statistical hypothesis:

$$\begin{cases} H_0: & t \le t_0; \\ H_1: & t > t_0. \end{cases} \tag{3}$$

Step 7. According to the rule (3), perform the test by calculating the current value t of the test statistic. Formulate the result: if the condition H_1 in the rule (3) is satisfied, then the value Δ is significant at the level of significance α.

Therefore, the proposed method to estimate the statistical significance of the accepting hypotheses allows to introduce a probability measure of the confidence degree in the research decisions taken in order to interpret the corresponding subgraphs of the geochronological track.

2. Test of Research Hypotheses as Recognition of Isomorphic Embedding of a Graph

As a result, the software component «Geochronological tracking» gives a geographic map with the graph, which generalizes the geochronological tracks of individuals such that the data on the movements of the individuals are recorded in the database [1]. An example is shown in Fig. 1. Further, we consider the mentioned graph as the base graph containing a subgraph, which is isomorphic to a given graph. Here the isomorphism of graphs is the equivalence relation on the set of graphs. An isomorphic map of one graph onto another is an one-to-one map of vertices and edges of one graph onto the vertices and edges of another graph, respectively, which saves the incidence relation. Two graphs are said to be isomorphic if there exists an isomorphic map of one of these graphs onto the another [4]. Note that incidence is a geometric term used to denote the relation of connection between the basic geometry objects (i.e., points, lines, planes, etc.) [2].

A graph, which is isomorphic to a subgraph of the basic graph of geochronological tracking, graphically defines the corresponding hypothesis of a research on migration (movements in geographic space). For example, consider the following hypothesis. «Typical
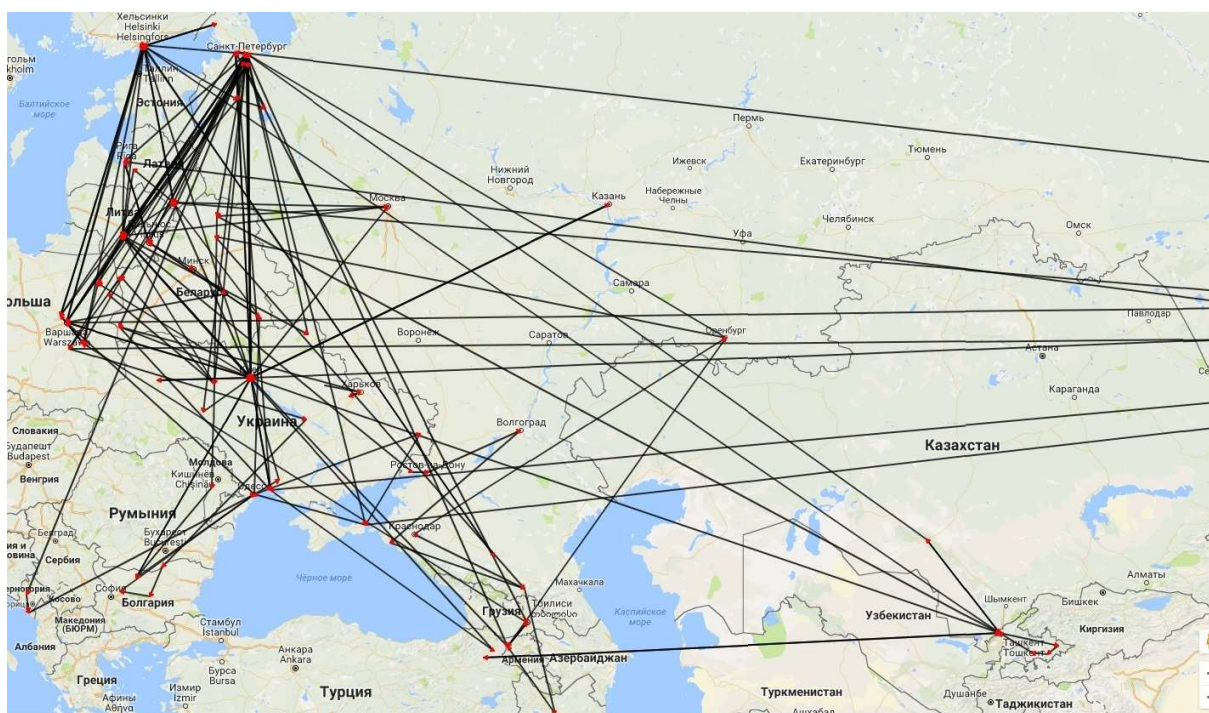
Fig. 1. An example of the implementation of geochronological tracking in GIS.

(traditional, most accepted, etc.) movements between service places for graduates of the St. Petersburg Naval Cadet College in the ranks of junior officers in the period from 1870 to 1910 were connected with either fleet headquarters, or the corresponding large ship formations of these fleets. In the future movements, the staff or ship specialization remained. Movements from subordinate ship formations to the corresponding fleet headquarters were allowed».

Suppose that during the considered period the fleet headquarters of the Russian Empire were located in

- St. Petersburg – Baltic Fleet,

- Sevastopol – Black Sea Fleet,

- Port Arthur – Pacific Fleet,

and basing areas of the large ship formations for the corresponding fleets were

- Helsingfors,

- Novorossiysk,

- Vladivostok,

respectively.

Then, we can interpret the considered hypothesis as an undirected graph shown in Fig. 2.

The considered approbation of the proposed analytical apparatus for presenting and testing research hypotheses as an isomorphic embeddings in the structure of the

1. St. Petersburg
2. Helsingfors
3. Sevastopol
4. Port Arthur
5. Novorossiysk
6. Vladivostok
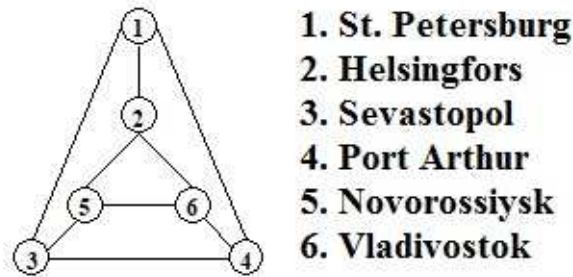
Fig. 2. An example of graph representing the research hypothesis.

geochronological track showed that graphs describing migration processes can have a diverse and highly complex form. For instance, Fig. 3 gives some examples of graphs describing different hypotheses, i.e. subgraphs of geochronological tracking. In general, Fig. 3 allows to estimate the structural complexity and variability of the hypotheses in the form of graphs.
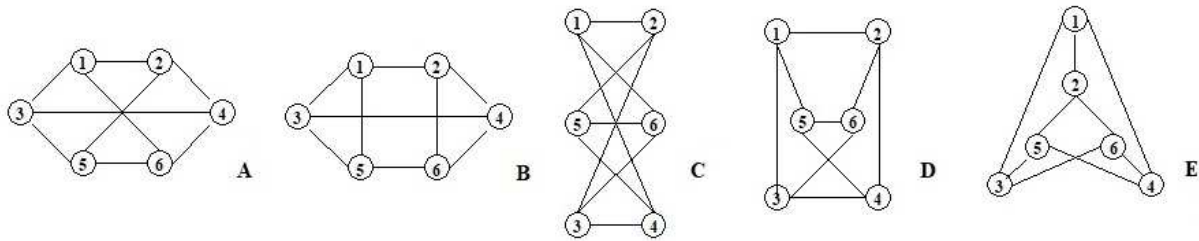
Fig. 3. Variants of graphs describing different hypotheses.

The problem of recognizing isomorphism of graphs is a central combinatorial problem in the graph theory and has a correct interpretation in the historical research. For today, according to [4,5], there is a number of effective algorithms to recognize a subgraph, which is isomorphic to a given subgraph, in the structure a complex graph, and in particular, in the real generalizing geochronological track (Fig. 4). In this case, conventionally, there are the following two main groups of algorithms to recognize a subgraph. The first group uses the hierarchical construction of algorithms that recursively improve their efficiency in the sense of completeness (sensitivity) of the considered characteristics of vertices and (or) edges. These characteristics are invariant under isomorphism of the graphs and are called invariants. The second group of algorithms uses the same hierarchical construction of algorithms based on the obligatory enumeration procedure of the search for an isomorphic substitution at one of the stages. The algorithms that use a direct exhaustive method have the worst characteristics. Such algorithms can be applied only to geochronological tracks having limited dimension. In general, a solution to the problem of isomorphic embedding of a subgraph in the graph (geochronological track) is trivial in the program-algorithmic sense and is described in [4,5].

Therefore, we represent the test of historical research hypotheses in the form of the problem on recognition of the corresponding isomorphic subgraph. Hence, we can to detect and recognize stable structural embeddings in the structure of the corresponding geochronological track using the list of predetermined structures. In other words, we can

use this representation in order to automatically search for facts, whose structure is known in advance and described in the form of graphs in the corresponding database. For example, we can search for stable factors of migration processes, specifics in the movements of historical personalities groups, not obvious facts about the personnel policy, etc.
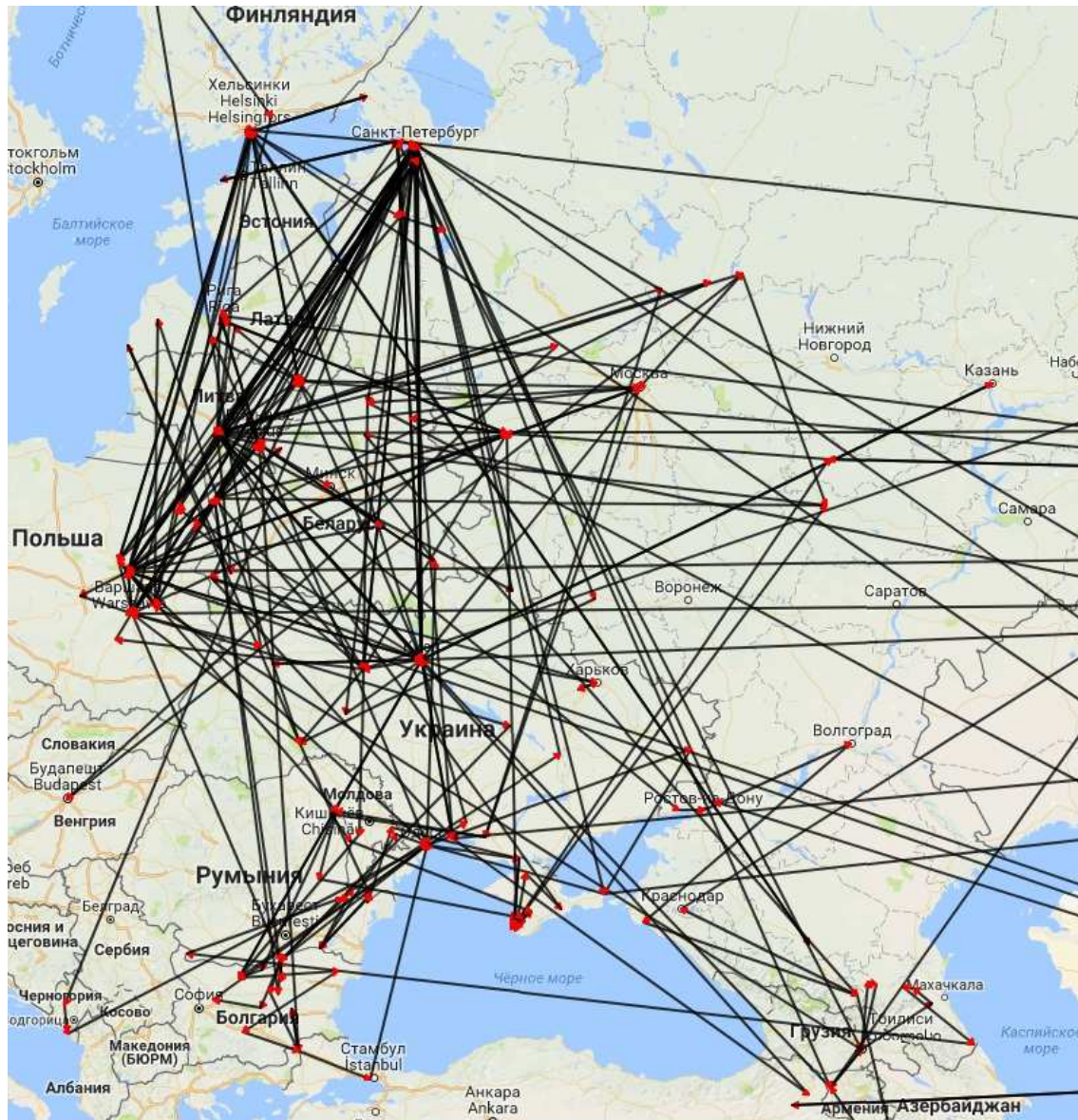

Fig. 4. Geochronological track for the approbation research example.

The carried out approbation experiment on the analysis of the applicability of the proposed method shows the following. In order to recognize stable facts having a historiographic nature, we can use a geochronological tracking on the basis of a specialized GIS. Such tracking allows to achieve a 3-4 times reduction in the time required to carry out the work, as well as to achieve a new quality of the research, i.e. to introduce a quantitative measure in order to analyse the degree of confidence for the obtained results.

Conclusion

In historical research, computer interpretation of graph theory methods based on geochronological tracking provides new applications of modern mathematical and analytical methods in the field of humanitarian knowledge. It is obvious that there are the following promising areas for the development of quantitative methods in humanitarian research, with respect to geochronological tracking: an application of "soft" calculations, apparatus of fractals, methods of modern mathematical topology, mathematical probability theory, theory of possibilities, etc. The use of the considered methods and mathematical apparatus for solving research problems in GIS is in demand and considered in a number of papers, examples of which are given in [6, 7].

The further work and research on the information technology of geochronological tracking in GIS can be devoted to the development and justification of the concretized interpretations of the considered methods in the historical research, as well as the synthesis of mathematical and algorithmic models for solving specific research problems. At the same time, already today we can state a broad perspective and applied applicability of the proposed approach to the development of software and information tools for humanitarian research.

References

1. Ivakin Ya.A., Potapychev S.N. Development of Geochronological Tracking Information Technology for GIS-Based Historical Studies. *Historical Informatics*, 2017, no. 2, pp. 85–94. (in Russian). doi: 10.7256/2585-7797.2017.2.23083.

2. *Intellektual'nye geograficheskie informatsionnye sistemy dlya monitoringa morskoi obstanovki* [*Intelligent Geographic Information Systems for Monitoring the Marine Environment*]. Sankt-Petersburg, Nauka, 2013. (in Russian).

3. Potapychev S.N., Ivakin Ya.A. Geochronology Tracking – Specialized GIS-Tool for History Researches. *Historical Information Science. Information Technology and Quantitative Methods in Historical Research and Education*, 2016, no. 1-2, pp. 3–11. (in Russian).

4. Zykov A.A. *Osnovy teorii grafov* [*The Basics of Graph Theory*]. Moscow, Vuzovskaya kniga, 2004. (in Russian).

5. Nechepurenko M.I., Popkov V.K., Maynagashev S.M., Kaul' S.B., Proskuryakov V.A., Kohov V.A., Gryzunov A.B. *Algoritmy i programmy resheniya zadach na grafah i setyah* [*Algorithms and Programs for Solving Problems on Graphs and Networks*]. Novosibirsk, Nauka, 1990. (in Russian).

6. *History & Mathematics: Political Demography & Global Ageing. Yearbook.* Volgograd, Uchitel, 2015.

7. Kolomeec M.V., Chechulin A.A., Kotenko I.V. Review of Methodological Primitives for the Phased Construction of Data Visualization Model. *SPIIRAS Proceedings*, 2015, no. 5 (42), pp. 232–257. (in Russian).

8. *RGVIA – fond F.400 Glavnogo shtaba voennogo ministerstva* [*Russian State Military-Historical Archive – fund F.400 General Staff of the Military Ministry*]. (in Russian).

9. *RGVIA – fond F.409 Posluzhnie spiski, attestacii i nagradnye listy oficerov russkoy armii* [*Russian State Military-Historical Archive – fund F.409 Track Records, Certifications and Award Lists of Officers of the Russian Army*]. (in Russian).

10. Ivakin Ya.A., Potapychev S.N., Ivakin V.Ya.Verification of Hypotheses for Historical Research on the Basis of Geochronological Tracking. *Historical informatics*, 2018, no. 2, pp. 96–102. (in Russian). doi: 10.7256/2585-7797.0.0.25344.

11. Kochura A.E., Podkolzina L.V., Ivakin Ya.A., Nidziev I.I. Singular Matrix Bundles in a Generalized Symmetric Problem of Eigenvalues. *SPIIRAS Proceedings*, 2013, no. 3 (26), pp. 253–276. (in Russian).

12. Kron G. *Diakoptics: The Piecewise Solution of Large-Scale Systems*. London, Macdonald & Co, 1963.

13. DeMarco T., Lister T. *Waltzing with Bears: Managing Risk on Software Projects*. New York, Dorset House Publishing, 2003.

14. DeMarco T. *The Deadline: A Novel About Project Management*. New York, Dorset House Publishing, 1997.

15. Duvall P.M., Matyas S., Glover A. *Continuous Integration: Improving Software Quality and Reducing Risk*. Addison-Wesley, 2007.

Yan A. Ivakin, DSc (Techn), Professor, Leading Researcher, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (St. Petersburg, Russian Federation), ivakin@oogis.ru.

Sergey N. Potapychev, PhD (Techn), Senior Researcher, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (St. Petersburg, Russian Federation), potapychev@oogis.ru.

МЕТОД ГЕОХРОНОЛОГИЧЕСКОГО ТРЕКИНГА ДЛЯ ПРОВЕРКИ ГИПОТЕЗ ИСТОРИОГРАФИЧЕСКОГО ИССЛЕДОВАНИЯ

Я. А. Ивакин, С. В. Потапычев

Информационная технология геохронологического трекинга есть совокупность процессов накопления и интеграции данных о географическом перемещении исторических личностей за установленный период времени с представлением результатов в виде обобщающего графа в ГИС. Гипотезы об устойчивых тенденциях в миграции представимы как подграфы указанного графа. Проверка таких гипотез сведется к поиску и оценке статистической значимости изоморфизма соответствующих графов. Рассмотрению качественно новых возможностей такого подхода и соответствующего математико-алгоритмического аппарат посвящена данная статья.

Ключевые слова: географические информационные системы; ГИС-технологии для исторических исследований; геохронологический трек и трекинг; изоморфизм графов; проверка статистических гипотез; междисциплинарные исследования на базе ГИС.

Литература

1. Ивакин, Я.А. Развитие информационной технологии геохронологического трекинга для исторических исследований в ГИС / Я.А. Ивакин, С.Н. Потапычев // Историческая информатика. – 2017. – № 2. – С. 85–94.

2. Интеллектуальные географические информационные системы для мониторинга морской обстановки. – СПб.: Наука, 2013.

3. Потапычев, С.Н. Геохронологический трекинг – специализированный ГИС-инструментарий исторического исследования // Я.А. Ивакин, С.Н. Потапычев // Историческая информатика. Информационные технологии и математические методы в исторических исследованиях и образовании. – 2016. – № 1–2 (15–16). – С. 3–11.

4. Зыков, А.А. Основы теории графов / А.А. Зыков. – М.: Вузовская книга, 2004.

5. Нечепуренко, М.И. Алгоритмы и программы решения задач на графах и сетях / М.И. Нечепуренко, В.К. Попков, С.М. Майнагашев, С.Б. Кауль , В.А. Проскуряков, В.А. Кохов, А.Б. Грызунов. – Новосибирск: Наука, 1990.

6. History & Mathematics: Political Demography & Global Ageing. Yearbook. – Volgograd: Uchitel, 2015.

7. Коломеец, М.В. Обзор методологических примитивов для поэтапного построения модели визуализации данных / М.В. Коломеец, А.А. Чечулин, И.В. Котенко // Труды СПИИРАН. – 2015. – Вып. 5 (42). – С. 232–257.

8. РГВИА – фонд Ф.400 Главного штаба Военного министерства.

9. РГВИА – фонд Ф.409 Послужные списки, аттестации и наградные листы офицеров русской армии.

10. Ивакин, Я.А. Проверка гипотез исторического исследования на базе геохронологического трекинга / Я.А. Ивакин, С.Н. Потапычев, В.Я. Ивакин // Историческая информатика. – 2018. – № 2. – С. 96–102.

11. Кочура, А.Е. Сингулярные матричные пучки в обобщенной симметричной проблеме собственных значений / А.Е. Кочура, Л.В. Подкользина, Я.А. Ивакин, И.И. Нидзиев // Труды СПИИРАН. – 2013. – № 3 (26). – С. 253–276.

12. Крон, Г. Исследование сложных систем по частям – диакоптика / Г. Крон. – М.: Наука, 1972.

13. ДеМарко, Т. Вальсируя с медведями. Управление рисками в проектах по разработке программного обеспечения / Т. ДеМарко, Т. Листер. – М.: Издательский дом DH, 2005.

14. ДеМарко, Т. Deadline. Роман об управлении проектами / Т. ДеМарко. – М.: Манн-Иванов-Фербер, 2016.

15. Дюваль, П.М. Непрерывная интеграция. Улучшение качества программного обеспечения и снижение риска / П.М. Дюваль, С. Матиас, Э. Гловер. – СПб.: Символ, 2016.

Ивакин Ян Альбертович, доктор технических наук, профессор, ведущий научный сотрудник, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН) (г. Санкт-Петербург, Российская Федерация), ivakin@oogis.ru.

Потапычев Сергей Николаевич, кандидат технических наук, старший научный сотрудник, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН) (г. Санкт-Петербург, Российская Федерация), potapychev@oogis.ru.