

МЕТОД ГЕОХРОНОЛОГИЧЕСКОГО ТРЕКИНГА ДЛЯ ПРОВЕРКИ ГИПОТЕЗ ИССЛЕДОВАНИЯ НА БАЗЕ ГИС

Я. А. Ивакин^{1*}, С. Н. Потапычев¹

¹ СПИИРАН, Санкт-Петербург, 199178, Российская Федерация

* Адрес для переписки: ivakin@oogis.ru

Аннотация

Метод геохронологического трекинга есть совокупность процессов накопления и интеграции данных о географическом перемещении в ретроспективе объектов (судов, кораблей и пр.) или личностей за установленный период времени с представлением результатов в виде обобщающего графа в ГИС. **Предмет исследования.** Информационная технология геохронологического трекинга, в прикладном аспекте, представляет собой научно-методический и программный инструментарий для автоматизации определенного класса ретроспективно-исследовательских задач, связанного со слиянием биографических и географических данных на базе геоинформационных систем и соответствующих геоинформационных технологий. Она является примером узкоспециализированных информационных ГИС-технологий, разрабатываемых в интересах специалистов, использующих географическую интерпретацию своей предметной области при решении исследовательских задач ретроспективного характера на морском транспорте. **Метод.** Основной методологической базой данной работы является объектно-ориентированный подход к моделированию в ГИС предметных областей, на базе онтологий, широко применяющейся в современном программировании и разработке сложных программных систем. **Основные результаты.** Гипотезы об устойчивых тенденциях в указанной миграции объектов или личностей за установленный период времени представимы как подграфы указанного графа. **Практическая значимость.** Проверка таких гипотез сведется к поиску и оценке статистической значимости изоморфизма соответствующих графов. Новизна предлагаемых решений заключается в разработке принципов реализации интеграции биографической, пространственно-координатной и географической информации для исследовательской сферы научного знания. Рассмотрению качественно новых возможностей такого подхода и соответствующего математико-алгоритмического аппарата, а также перспектив их применения в сфере информационных технологий работы с пространственными данными, 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.

Информация о статье

УДК 004.896

Язык статьи – русский.

Поступила в редакцию 31.01.18, принята к печати 28.02.18.

Ссылка для цитирования: Ивакин Я. А., Потапычев С. В. Метод геохронологического трекинга для проверки гипотез исследования на базе ГИС // Информационные технологии и телекоммуникации. 2018. Том 6. № 1. С. 72–81.

METHOD OF GEOCHRONOLOGICAL TRACKING FOR HYPOTHESES TESTING BASED IN GIS

Y. Ivakin^{1*}, S. Potapichev¹

¹ SPIIRAS, St. Petersburg, 199178, Russian Federation

* Corresponding author: ivakin@oogis.ru

Abstract—Method of the geochronological tracking is an assembly of processes that accumulate and integrate data about geographic relocation of historical figures for a given time interval and represent the results as a generalizing graph in GIS. **Research subject.** Information technology of the geochronological tracking from the viewpoint of its applied aspect, this information technology is scientific methodological and program tools to automatize certain issues related to biographical and geographical data merge on the basis of geoinformation systems and corresponding geoinformation technologies. It is an example of a specific information technology developed to help employing geographical interpretation of their subject domain when solving research problems. **Method.** The methodology bases on the object-oriented approach to modeling within GIS domains and ontologies widely used in modern programming and complex software systems development. **Core results.** Hypotheses on the stable tendencies in migration could be represented as the above graph's sub-graphs. **Practical relevance.** Such tendencies testing would be reduced to the search and evaluation of the statistical significance for the matching graphs' isomorphism. The novelty of the solution proposed is the principles of biographical and geographical data integration. The proposed paper deals with a consideration of qualitatively new possibilities of such an approach and the corresponding mathematic and algorithmic apparatus.

Keywords—Geographic information systems, GIS – technologies for historic research; geochronological track and tracking; graphs' isomorphism; statistical hypotheses testing; GIS based interdisciplinary research.

Article info

Article in Russian.

Received 31.01.18, accepted 28.02.18.

For citation: Ivakin Y., Potapichev S.: Method of geochronological tracking for hypotheses testing based in GIS // Telecom IT. 2018. Vol. 6. Iss. 1. pp. 72–81 (in Russian).

Введение

В работах [1, 2] представлена и детально описана специализированная информационная технология геохронологического трекинга. Информационная технология геохронологического трекинга есть совокупность методов, моделей, приемов, методик и способов сбора, передачи, обработки, отображения и выдачи потребителю информации об обобщении геохронологических треков исторических личностей. В свою очередь построение геохронологического трека исторической личности (индивида) или исторического объекта на основании геопространственной интерпретации его биографической информации есть интеграция хронологических и географических данных в виде графа, соединяющего географические точки нахождения исторической личности. При этом вершины такого графа имеют строгую историко-географическую привязку, а дуги носят условно-логический характер.

Проведенные апробации и моделирование геохронологического трекинга показали, что итоговый граф для представительной выборки индивидов может иметь сложную и высокосвязную (и даже полностью связную) структуру. Этот факт позволяет рассматривать итоговый граф геохронологического трекинга, как основу для исследования различных миграционных процессов, выявления некоторых частных закономерностей в перемещении исторических личностей или социальных групп, а также проверять статистические гипотезы о характере перемещений. Представительность выборки исторических личностей, принятой для геохронологического трекинга определяет статистическую устойчивость получаемых выводов.

Одним из направлений создания научно-методического инструментария, поддерживающего указанные исследования, является широкое применение и компьютерная интерпретация методов теории графов на базе геохронологического трекинга, и в частности, алгоритмов поиска изоморфизма графов (распознавания изоморфного вложения подграфа в более сложный граф). Предметная интерпретация этих методов в предметной области исследований исторической информатики открывает широкие возможности по анализу различных сетевых структур, и прежде всего, геохронологических треков.

Полнофункциональное развитие компьютерной интерпретации методов теории графов на базе геохронологического трекинга способно обеспечить новое качество исторических исследований с использованием современного ГИС-инструментария. Оно выражается в предоставлении возможности исследователю-историку использовать количественные методы соответствующего логико-аналитического аппарата в своей предметной области. Детализация указанных возможностей с раскрытием существа соответствующих математических и алгоритмических решений, а также описание новых путей их применения в современных исторических исследованиях есть предмет рассмотрения данной статьи.

Предлагаемая статья является более полным развитием описания исследования в [3], выполненного для специалистов гуманитарного профиля.

Оценка статистической значимости принятия гипотез на базе геохронологического трекинга

Выявление фактов историко-географических процессов как устойчивых структурных вложений в структуре соответствующего геохронологического трека обладает очевидными характеристиками риска принятия решения о их наличии. Иными словами, может быть оценена доверительная вероятность в заключении об установлении того или иного факта.

Первоначальное множество исторических личностей, принятой для геохронологического трекинга рассматривается как генеральная совокупность в математическом аппарате проверки статистических гипотез. Выборка числа индивидов, подтвердивших своими перемещениями в географическом пространстве соответствующую гипотезу-подграф, определяет статистическую устойчивость получаемых выводов исследования.

Проверка корректности принятия той или иной гипотезы сводится к оценке статистической значимости вывода о выделении соответствующего изоморфного подграфа в составе геохронологического трека по выше указанной выборке

из генеральной совокупности. При этом задается уровень значимости (риска принятия неправильной гипотезы) как вероятности обратной к доверительной вероятности принятия правильного решения p , т. е.

$$p = 1 - \alpha; \quad (1)$$

Анализ значимости числа индивидов, подтвердивших своими перемещениями в географическом пространстве соответствующую гипотезу-подграф в сравнении с общим числом индивидов, учтенном при геохронологическом трекинге, может быть проведен путем использования методики статистических сравнений, изложенной в [4]. При этом анализируется статистическое отличие разницы Δ_r между указанными числами от 0. Такой анализ значимости, как правило, проводится за семь логических шагов:

Шаг 1. Выдвигается две статистических гипотезы:

$$\left. \begin{array}{l} H_0 : \Delta_r = 0 - \text{незначимо} \\ H_1 : \Delta_r > 0 - \text{значимо} \end{array} \right\}, \quad (2)$$

где H_0 – гипотеза, что полученное значение Δ_r статистически незначимо (нет статистически достаточного числа индивидов для выделения искомого подграфа в составе геохронологического трека); H_1 – гипотеза, что полученное Δ_r – статистически значимо (количество индивидов, подтверждающих своими перемещениями искомым подграф в составе геохронологического трека, статистически значимо и достаточно для его выделения).

Шаг 2. Поскольку результаты Δ_r получены по выборке числа индивидов, подтвердивших своими перемещениями в географическом пространстве соответствующую гипотезу-подграф, то размерность этой выборке принимается за число испытаний n .

Шаг 3. Принимается уровень значимости α . Например, равным 0,05. Тогда доверительная вероятность принятия исследовательской гипотезы, согласно (1), будет 0,95.

Шаг 4. Осуществляется выбор тестовой статистики из табулированных вариантов. Этот выбор определяется размером выборки числа индивидов, подтвердивших своими перемещениями в географическом пространстве соответствующую гипотезу-подграф. Если размерность генеральной совокупности и выборки обеспечивают выполнение статистического закона больших чисел, то в качестве тестовой статистики может быть приняты величины, имеющие широко известные распределения. Например, распределение χ^2 . Для малых (в статистическом понимании) выборок в качестве тестовой статистики традиционно принимаются величины, имеющие t – распределение Стьюдента с $(n - 1)$ степенями свободы.

Шаг 5. Производится определение области принятия статистической гипотезы. Для этого из соответствующих таблиц (базы данных) выбирается критическое значение тестовой статистики $t_{кр}$ при принятом уровне значимости α , что согласно (1) определяет доверительная вероятность γ .

Шаг 6. Формируется итоговое математическое правило проверки статистической гипотезы

$$\left. \begin{array}{l} H_0 : t \leq t_{\text{пр}}; \\ H_1 : t > t_{\text{пр}}. \end{array} \right\} \quad (3)$$

Шаг 7. Выполнение проверки в соответствии с правилом (3) путем расчета текущего значения тестовой статистики t . Формулирование вывода: при выполнении условия H_1 из правила (3) значение Δ_r следует считать значимым при уровне доверия γ .

Таким образом, предлагаемая методика оценки статистической значимости принятия гипотез позволяет ввести вероятностную меру степени доверия к принимаемым исследовательским решениям по интерпретации соответствующих подграфов геохронологического трека.

Представление проверки исследовательских гипотез в виде задачи распознавания изоморфного вложения графа

Результатом реализации функциональности программного компонента «Геохронологический трекинг» является географическая карта, на которой наносится граф обобщающий геохронологические треки индивидов, данные о перемещениях, которых занесены в базу данных [1]. Пример такой реализации показан на рис. 1. Указанный граф далее рассматривается как базовый граф в рамках которого определяется подграф изоморфный заданному. При этом изоморфизм графов понимается как отношение эквивалентности на множестве графов. Изоморфным отображением одного графа на другой называется взаимно однозначное отображение вершин и рёбер одного графа соответственно на вершины и рёбра другого графа, при котором сохраняется отношение инцидентности. Два графа называются изоморфными, если существует изоморфное отображение одного из этих графов на другой [4]. (Инцидентность – геометрический термин, употребляемый для обозначения отношения принадлежности (связи, соединения) между основными объектами геометрии (точками, прямыми, плоскостями и т. д.) [5]).

Граф на изоморфность к которому определяется подграф в составе базового графа геохронологического трекинга графически задает соответствующую гипотезу исследования о миграции (перемещениях в географическом пространстве). Например, пусть выдвигается следующая гипотеза: «Типовыми (традиционным, наиболее принятым и пр.) назначениями (перемещениями по местам прохождения службы) выпускников Санкт-Петербургского морского кадетского корпуса в период с 1870 г. по 1910 г. в чинах младших офицеров было назначения либо в штабы флотов, либо на соответствующие крупные корабельные соединения этих флотов. В дальнейшем при переводах штабная или корабельная специализация сохранялась. Допускались назначения с подчиненных корабельных соединений в соответствующие штабы флотов».

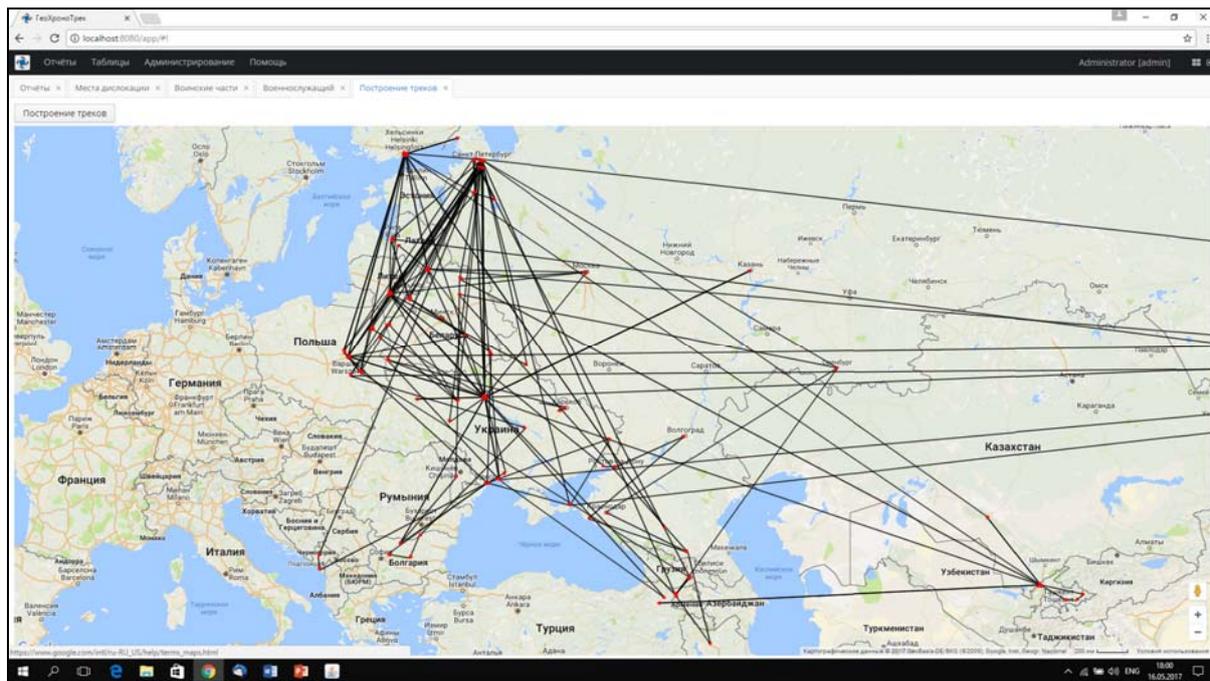


Рис. 1. Пример реализации геохронологического трекинга в ГИС

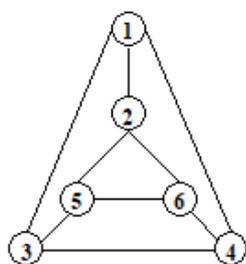
Если в указанный период штабы флотов Российской империи располагались в:

- Санкт-Петербург – Балтийский флот,
- Севастополь – Черноморский флот,
- Порт-Артур – Тихоокеанский флот,

а местами базирования крупных корабельных соединений для соответствующих флотов были, соответственно:

- Гельсингфорс,
- Новороссийск,
- Владивосток,

то интерпретация приведенной гипотезы как ненаправленного графа примет вид, показанный на рис. 2.



1. Санкт-Петербург ;
2. Гельсингфорс;
3. Севастополь;
4. Порт-Артур;
5. Новороссийск;
6. Владивосток.

Рис. 2. Пример графа представляющего исследовательскую гипотезу

Проведенная апробация предлагаемого аналитического аппарата представления и проверки исследовательских гипотез как изоморфного вложения в составе геохронологического трека показала, что графы описывающие миграционные процессы могут носить разнообразную и весьма сложную форму. Так, на рис. 3 приведены некоторые примеры графов, описывающих различные гипотезы – подграфов геохронологического трекинга. В целом, этот рисунок позволяет оценить структурную сложность и вариабельность представительских гипотез в виде графов.

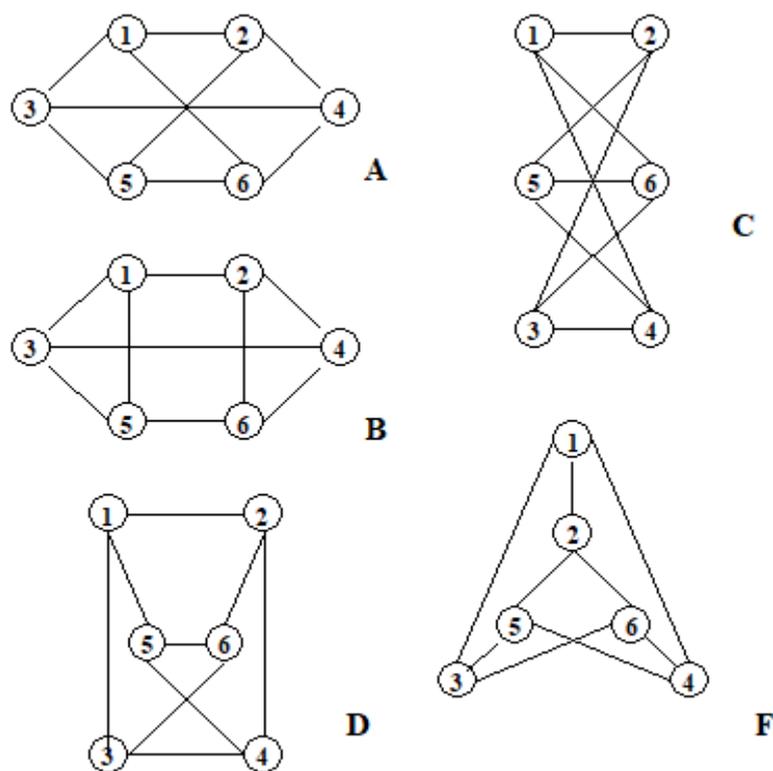


Рис. 3. Варианты графов, описывающих различные гипотезы

Задача распознавания изоморфизма графов является центральной комбинаторной задачей теории графов и имеет корректную интерпретацию в предметной области исторического исследования. На сегодняшний день, согласно [4, 6], разработан целый ряд эффективных алгоритмов распознавания в составе сложного графа, каким является реальный обобщающий геохронологический трек (рис. 4), подграфа изоморфного заданному. При этом условно выделяют две основных группы алгоритмов реализации указанного распознавания. Первая группа связана с реализацией принципа иерархического построения непереборных алгоритмов, рекурсивно улучшающих свою эффективность в смысле полноты (чувствительности) используемых характеристик вершин и (или) ребер, инвариантных относительно изоморфизма графов и называемых инвариантами. Вторая группа алгоритмов связана с реализацией того же принципа иерархического построения алгоритмов на базе обязательной процедуры перебора на одном из этапов поиска изоморфной подстановки. Худшие характеристики имеют алгоритмы, использующие метод полного прямого перебора, они применимы лишь к геохронологическим трекам ограниченной размерности. В целом, само решение задачи изоморфного вложения подграфа в состав графа-геохронологического трека в программно-алгоритмическом плане является тривиальным и описано в работах [4, 6].

Таким образом, представление проверки исторических исследовательских гипотез в виде задачи распознавания (поиска) соответствующего изоморфного подграфа позволяет обнаружить и распознать устойчивые структурные вложения в составе соответствующего геохронологического трека из перечня (базы) заранее определённых структур. Иными словами, это представление может быть применено для автоматизированного поиска устойчивых факторов мигра-

ционных процессов, специфики перемещений отдельных групп исторических личностей, не очевидных фактов кадровой политики, и пр., структура которых заранее известна и описана в виде графов в соответствующей базе данных.

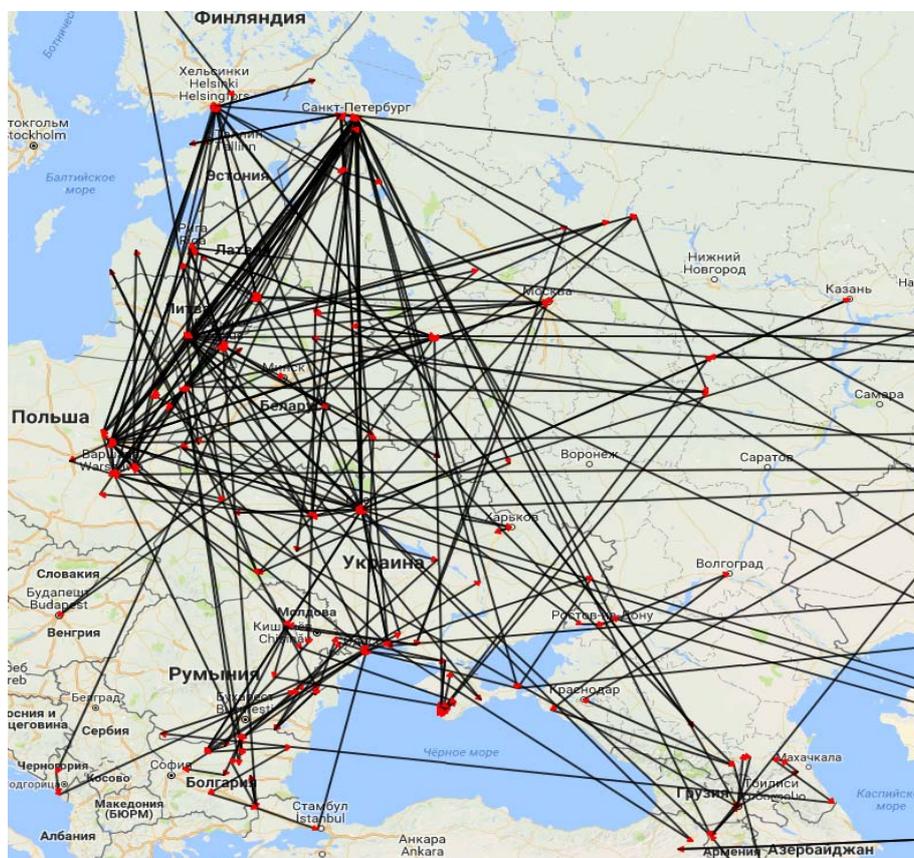


Рис. 4. Геохронологический трек из апробационного исследовательского примера

Проведенный апробационный эксперимент по анализу применимости предлагаемого метода показал, что использование геохронологического трекинга на базе специализированной ГИС для выявления устойчивых фактов историко-графического характера позволяет добиться сокращения временных затрат на проведение работ в 3–4 раза, а также добиться нового качества исследования – введение количественной меры для анализа степени доверия к формулируемым выводам.

Заключение

Применение компьютерной интерпретации методов теории графов на базе геохронологического трекинга в ходе исторических исследований открывает новые горизонты для применения современных математико-аналитических методов в сфере гуманитарного знания. Очевидно, что перспективными направлениями развития количественных методов в гуманитарных исследованиях, применительно к геохронологическому трекингу, является внедрение в соответствующее математическое и программное обеспечение ГИС приложений «мягких» вычислений, аппарата фракталов, методов современной математической топологии, математической теории вероятностей, теории возможностей и пр. Внедрение указанных методов и математического аппарата решения

исследовательских задач в ГИС является востребованным и рассмотрено в целом ряде работ, примеры которых приведены в [7, 8].

Разработка и обоснование конкретизированных интерпретаций указанных методов в предметной области исторических исследований, синтез математических и алгоритмических моделей решения конкретных исследовательских задач составляет существо дальнейших работ и исследований по информационной технологии геохронологического трекинга в ГИС. Вместе с тем, уже сегодня можно констатировать широкую перспективу и прикладную применимость данного подхода к развитию программно-информационных инструментов для гуманитарных исследований.

Поддержка исследований. Работа выполнена при поддержке РФФИ (проект №16-07-00127), а также Государственной Работы – Проведение фундаментальных научных исследований по программам РАН № 0073-2018-0003.

Литература

1. Ивакин Я. А., Потапычев С. Н. Развитие информационной технологии геохронологического трекинга для исторических исследований в ГИС // Историческая информатика. 2017. № 2. С. 85–94.
2. Ивакин Я. А., Потапычев С. Н. Геохронологический трекинг – специализированный ГИС-инструментарий исторического исследования // Историческая информатика. Информационные технологии и математические методы в исторических исследованиях и образовании. 2016. № 1–2. С. 3–11.
3. Ивакин Я. А., Потапычев С. Н., Ивакин В. Я. Проверка гипотез исторического исследования на базе геохронологического трекинга // Историческая информатика. 2018. № 2. С. 96–102.
4. Зыков А. А. Основы теории графов. М.: Вузовская книга. 2004. 664 с.
5. Интеллектуальные географические информационные системы для мониторинга морской обстановки. СПб.: Наука. 2013. 284 с.
6. Нечепуренко М. И., Попков В. К., Майнагашев С. М., Кауль С. Б., Проскуряков В. А., Кохов В. А., Грызунов А. Б. Алгоритмы и программы решения задач на графах и сетях. Новосибирск: Наука. 1990. 515 с.
7. History & Mathematics: Political Demography & Global Ageing. Volgograd: Uchitel Publishing House. 2015. 176 p.
8. Коломеец М. В., Чечулин А. А., Котенко И. В. Обзор методологических примитивов для поэтапного построения модели визуализации данных // Труды СПИИРАН. 2015. Вып. 5 (42). С. 232–257.

References

1. Ivakin Ya., Potapychev S. Development of Information Technology of Geochronological Tracking for Historical Research in GIS // Istoricheskaya informatika. 2017. Iss. 2. pp. 85–94.
2. Ivakin Ya., Potapychev S. Geochronology Tracking – Specialized GIS-Tool for History Researches // Istoricheskaya informatika. Informatsionnye tekhnologii i matematicheskie metody v istoricheskikh issledovaniyah i obrazovanii. 2016. Iss. 1–2. pp. 3–11.
3. Ivakin Ya., Potapychev S., Ivakin V. Method of Geochronological Tracking to Test the Hypothesis of Historiographic Research // Istoricheskaya informatika. 2018. Iss. 2. pp. 96–102.
4. Zykov A. Fundamentals of Graph Theory. M.: Vuzovskaya kniga. 2004. 664 p.
5. Intelligent GIS for Marine Monitoring. SPb.: Nauka. 2013. 284 p.
6. Nechepurenko M., Popkov V., Mainagashhev S., Kaul' S., Proskuryakov V., Kohov V., Gryzunov A. Algorithms and Programs for Solving Problems on Graphs and Networks. Novosibirsk: Nauka. 1990. 515 p.
7. History & Mathematics: Political Demography & Global Ageing. Volgograd: Uchitel Publishing House. 2015. 176 p.

8. Kolomeec M., Chechulin A., Kotenko I. Review of Methodological Primitives for the Phased Construction of Data Visualization Model // Trudy SPIIRAN. 2015. Iss. 5 (42). pp. 232–257.

- Ивакин Ян Альбертович*** – доктор технических наук, ведущий научный сотрудник, профессор, СПИИРАН, Санкт-Петербург, 199178, Российская Федерация, ivakin@oogis.ru
- Потапычев Сергей Николаевич*** – кандидат технических наук, старший научный сотрудник, СПИИРАН, Санкт-Петербург, 199178, Российская Федерация, potapuchev@oogis.ru
- Ivakin Yan*** – Doctor of Engineering Sciences, Leading Researcher, Full Professor, SPIIRAS, St. Petersburg, 199178, Russian Federation, ivakin@oogis.ru
- Potapichev Sergey*** – Candidate of Engineering Sciences, Senior Research Officer, SPIIRAS, St. Petersburg, 199178, Russian Federation, potapuchev@oogis.ru