

Научная статья
УДК 004.853
doi:10.37614/2949-1215.2023.14.7.008

ОНТОЛОГО-ОРИЕНТИРОВАННАЯ ГЕОАНАЛИТИКА ДЛЯ ОПРЕДЕЛЕНИЯ МЕСТ КОНЦЕНТРАЦИИ ДТП НА УЧАСТКАХ УЛИЧНО-ДОРОЖНОЙ СЕТИ

Александр Викторович Смирнов¹, Николай Николаевич Тесля²

^{1, 2}Санкт-Петербургский федеральный исследовательский центр Российской академии наук, Санкт-Петербург, Россия

¹smir@iias.spb.su

²teslya@iias.spb.su, <https://orcid.org/0000-0003-0619-8620>

Аннотация

Геоаналитика является важным инструментом при анализе городской среды. В работе основное внимание уделено развитию инструментов геоаналитики путем онтолого-ориентированного объединения информации из разнородных источников для определения потенциально аварийно-опасных участков дорожной сети, что может быть использовано для построения маршрутов по минимально опасным дорогам. Разнородность вызвана разнообразием типов предоставляемой информации, собираемой как централизованно дорожными и экстренными службами, так и распределенно с помощью волонтеров. В качестве примера рассматривается объединение разнородных источников информации в единой аналитической системе и проведение геоанализа частоты аварийности на отдельных участках дорог для определения наиболее частого типа и причины аварийности. Результат анализа может быть использован для построения маршрута в обход участков дорожной сети с повышенной концентрацией ДТП.

Ключевые слова:

геоаналитика, онтология, анализ аварийности, агрегация информации, места концентрации ДТП

Благодарности:

исследование выполнено в рамках государственного задания Санкт-Петербургского федерального исследовательского центра Российской академии наук FFZF-2022-0005.

Для цитирования:

Смирнов А. В., Тесля Н. Н. Онтолого-ориентированная геоаналитика для определения мест концентрации ДТП на участках улично-дорожной сети // Труды Кольского научного центра РАН. Серия: Технические науки. 2023. Т. 14, № 7. С. 79–85. doi:10.37614/2949-1215.2023.14.7.008.

Original article

ONTOLOGY-ORIENTED GEOANALYTICS FOR DETERMINING CONCENTRATION OF ROAD ACCIDENTS ON AREAS OF THE ROAD NETWORK

Alexander A. Smirnov¹, Nikolay N. Teslya²

^{1, 2}Saint-Petersburg Federal research center of the Russian academy of sciences, St. Petersburg, Russia

¹smir@iias.spb.su

²teslya@iias.spb.su, <https://orcid.org/0000-0003-0619-8620>

Abstract

Geoanalytics is an important tool in analyzing the urban environment. This work focuses on the development of geoanalytics tools by ontology-oriented integration of information from heterogeneous sources to identify potentially accident-prone sections of the road network, which can be used to build routes along minimally dangerous roads. The heterogeneity of a large number of sources is caused by the variety of types of information provided, collected both centrally by road and emergency services, and distributed with the help of volunteers. The novelty of this work is the combination of heterogeneous sources of information in a single analytical system and conducting a geoanalysis of the frequency of accidents on individual road sections, determining the most frequent type and cause of accidents to build a route bypassing sections of the road network with a high concentration of accidents.

Keywords:

geoanalytics, ontology, accident analysis, information aggregation, accident concentration areas

Acknowledgments:

the study was carried out within the framework of the state research of the St. Petersburg Federal Research Center of the Russian Academy of Sciences FFZF-2022-0005.

For citation:

Smirnov A. V., Teslya N. N. Ontology-oriented geoanalytics for determining concentration of road accidents on areas of the road network // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2023. Vol. 14, No. 7. P. 79–85. doi:10.37614/2949-1215.2023.14.7.008.

Введение

Геоаналитика дорожной сети на основе существующих геоинформационных систем и исторических сведений о событиях в дорожной сети, полученных из разнородных источников данных, является в настоящее время актуальной задачей [1, 2]. Проведение подобных аналитических исследований стало возможным благодаря появлению большого количества источников данных со свободным доступом, собираемых как централизованно дорожными и экстренными службами, так и распределенно с помощью волонтеров [3]. Эти данные отображают состояние дорожной сети, основные характеристики отдельных участков с привязкой к пространству и меткам времени, что позволяет делать пространственно-временные срезы заданных областей и проводить их исследование в динамике. Например, на портале статистики ГИБДД доступны карточки дорожно-транспортных происшествий (ДТП), в которых содержится подробная информация о них по всей России начиная с января 2015 г. Эта информация может быть использована для анализа с целью поиска мест концентрации ДТП на участках улично-дорожной сети посредством пространственной кластеризации и последующего изучения типов и причин повышенной аварийности с привлечением дополнительных источников данных.

Целью настоящего исследования является расширение доступной информации об улично-дорожной сети для проведения анализа мест концентрации ДТП на участках улично-дорожной сети. Расширение осуществляется путем объединения нескольких источников данных в единую систему с определением связи между различными источниками (рис. 1). Поскольку используемые источники данных могут быть разнородными (реляционные БД, таблицы, графы знаний), для объединения необходимо преобразовать их в единый формат.

Обзор существующих подходов к описанию пространственно-временных данных

Существуют разные группы подходов к описанию пространственно-временных данных, в частности: вероятностные пространственно-временные модели представления знаний [4, 5]; качественное пространственно-временное обозначение на основе конечных множеств абстракций пространственных и временных отношений с дополнительным использованием ограничений [6]. Отдельно стоит отметить использование онтологий, которое позволяет объединить качественное и количественное описание пространства и времени, делая его пригодным для решения различного рода задач в семантическом вебе [7–9]. При онтологическом представлении пространственно-временных знаний стоит отметить работу [10], в которой проведен глубокий анализ пространственных и временных логических отношений и сформирована алгебра их использования на основе дескрипционной логики. Хотя авторы по большей части рассматривают фреймвые модели представления знаний, модель вычислений и работы с пространством и временем в них может быть адаптирована для задачи слияния знаний, представленных с использованием других формализмов. За счет детальной проработки пространственных и временных отношений именно онтологическое представление является приоритетным для описания данных из разнородных источников для их последующего объединения в задаче геоаналитики.

Для представления геопространственных данных существует стандарт для описания геопространственной онтологии — GeoSPARQL, разработанный открытым геопространственным консорциумом (The Open Geospatial Consortium, OGC) [11], который поддерживает представление данных, имеющих привязку к пространству и времени, в семантической сети, формирование запросов к ним, и определяет словарь для их представления в RDF, расширяя тем самым язык запросов SPARQL.

Для представления времени в онтологии с 2016 г. ведется разработка стандарта OWLTime для языка OWL-DL. К настоящему времени он до сих пор дополняется новыми концептами и связями, позволяющими отразить особенности сущностей, связанных со временем и использовать его в системах представления знаний [12, 13]. Также предпринимались попытки объединить пространственную и временные онтологии в общей задаче для выполнения логического вывода в системе управления контентом [14].

Описание системы геоаналитики для определения мест концентрации ДТП

Инструменты для создания баз знаний из существующих источников данных позволяют формировать онтологическое описание источников на основе их структуры данных. Формирование онтологического описания чаще всего осуществляется путем построения частных онтологий для каждого источника и их объединения в одну общую онтологию, по которой осуществляется запрос. Другим

вариантом может быть использование онтологии для трансляции запроса из формата онтологии в формат базы данных через медиатор (рис. 1). При этом все равно требуется создавать частные онтологии для описания структуры и формата пространственно-временных данных от источника.

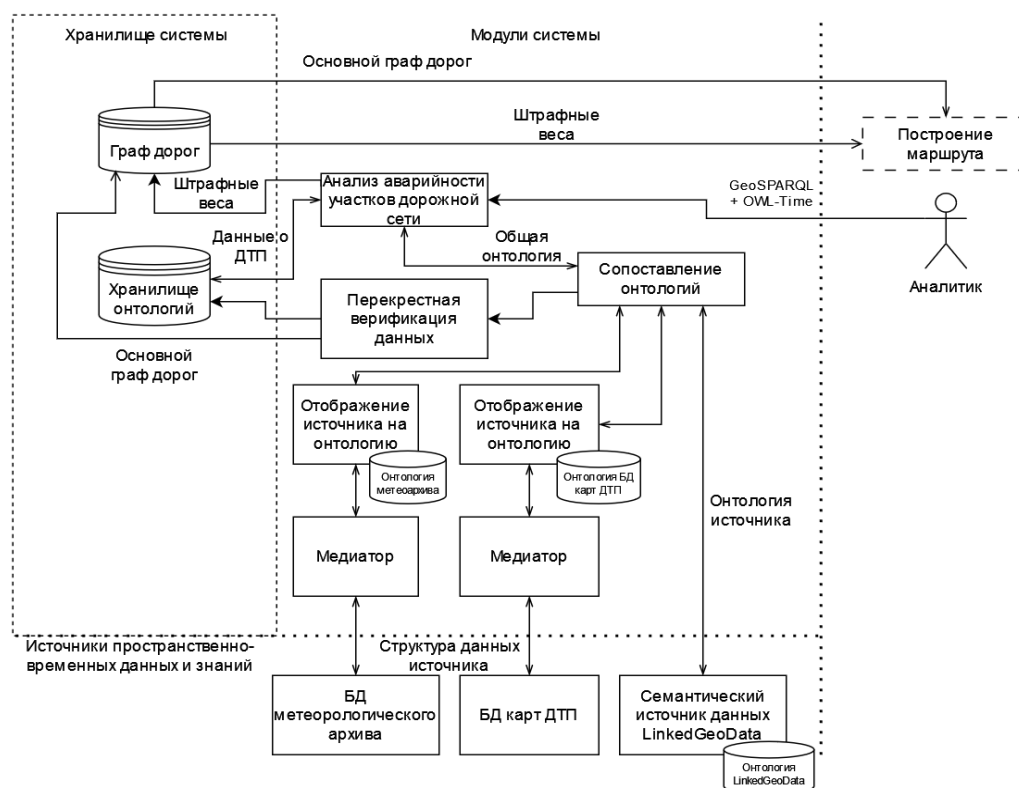


Рис. 1. Концептуальная модель системы геоаналитики для определения мест концентрации ДТП на участках улично-дорожной сети

При проведении геоаналитики дорожной сети для мест концентрации ДТП с использованием данных из разнородных источников предлагается следующая последовательность этапов.

На первом этапе необходимо определить источники данных и методы их обработки. Поскольку готовых открытых источников данных об аварийности нет, то необходимо осуществлять анализ исходных о различных типах ДТП. В настоящей работе основным источником сведений о ДТП является портал статистики ГИБДД. Информацией для анализа являются координаты, адрес в текстовом виде, дата и время, тип ДТП, количество участников, погодные условия. Предварительный анализ источника данных показал, что некоторые данные в карточки занесены с ошибками, в связи с чем необходимо проводить их сверку со сторонними источниками.

Вторым этапом является подключение источника данных к системе и его перекрестная проверка. Поскольку карточки ДТП распространяются в виде файлов XML и JSON и портал статистики ГИБДД имеет ограничение на количество запросов, необходимо создать промежуточное хранилище, в котором карточки ДТП будут преобразованы в базу данных для повышения оперативности доступа к ним и снижения нагрузки на основной источник данных. Для хранения карточек ДТП была создана БД на базе СУБД PostgreSQL, в которую были импортированы все данные с портала статистики ГИБДД начиная с 2015 г. по г. Санкт-Петербург.

Для последующего использования при анализе мест концентрации ДТП на участках улично-дорожной сети база данных должна иметь онтологическое описание. Это делается для сопоставления данных из различных источников в единой системе и определения корректных. Объединение данных в единой системе, их последующий анализ и применение результатов в практических целях встречается в большом количестве исследований [15, 16].

Для верификации данных выбираются источники, которые могут считаться эталонными в силу ручной верификации большим количеством пользователей. В частности, для верификации картографических данных может использоваться их семантическое представление в портале LinkedGeoData. Для верификации данных о погоде могут быть использованы порталы, предоставляющие доступ к метеорологическому архиву такие, как OpenWeatherMap (точность до 1 ч), или «tr5.ru Расписание погоды» (точность до 3 ч).

Все источники (кроме LinkedGeoData) подключались к системе с использованием специального медиатора на языке R2RML [17] и библиотеки ontop [18] для трансформации реляционного источника данных в онтологию. Запросы к источникам осуществляются посредством протокола GeoSPARQL в виде федеративного запроса. При верификации данные из карточек ДТП сравнивались с эталонными по следующим правилам:

1) соответствие координат границ города и адресу в текстовом виде по информации из LinkedGeoData. В случае несоответствия, выставлялись координаты, соответствующие текстовому адресу (геокодирование);

2) соответствие указанных погодных условий историческим сведениям из двух эталонных источников. При совпадении информации из обоих источников и несовпадении в карточке ДТП в нее записывалась согласованная информация о погоде.

Поиск мест концентрации ДТП на участках улично-дорожной сети после верификации данных производится следующим образом:

1) с помощью алгоритма кластеризации DBSCAN [19] определяются кластеры близких точек в радиусе не более 1 км с обязательной привязкой к улице и/или перекрестку. Пример подобного кластера представлен на рис. 2;

2) в найденных кластерах осуществляется статистический анализ типов ДТП и условий, при которых они происходили. Например, «наезд на пешехода» в совокупности с плохой погодой (наличие осадков, туман, темное время суток);

3) создается граф дорог со штрафными весами, на котором весом ребра является количество ДТП на улицах и перекрестках.



Рис. 2. Пример кластеров ДТП и визуализации причин ДТП

Граф со штрафными весами затем может быть использован в алгоритме построения маршрута транспорта. Так, например, маршрут транспорта бригады экстренной службы может быть построен таким образом, чтобы исключить из него места концентрации ДТП с учетом текущих погодных условий для повышения оперативности прибытия и безопасности передвижения бригады.

Заключение

В работе рассматривается задача объединения информации из разнородных источников данных для проведения геоаналитики при поиске мест концентрации ДТП на участках улично-дорожной сети. Предлагается онтолого-ориентированный подход для описания модели информации из источников,

с последующим объединением частных онтологий в одну общую. В качестве экспериментального подтверждения работоспособности метода проведено преобразование источников данных о ДТП и погоде с их объединением в единой системе. По результатам объединения проведен пространственный анализ с выделением мест концентрации ДТП по г. Санкт-Петербургу.

Список источников

1. Mishra R. K., Kumari C. L., Janaki Krishna P. S., Dubey A. Smart Cities for Sustainable Development: An Overview // *Smart Cities for Sustainable Development*. Springer, Singapore, 2022. P. 1–12.
2. Yakimov M. R. Features of the Use of Geoanalytical Data in the Development of Transport Planning Documents // *2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications, SYNCHROINFO 2022 — Conference Proceedings*. Institute of Electrical and Electronics Engineers Inc., 2022.
3. Mashhadi A., Quattrone G., Capra L. The Impact of Society on Volunteered Geographic Information: The Case of OpenStreetMap // *OpenStreetMap in GIScience: Experiences, Research, Applications* / ed. Jokar Arsanjani J. et al. Cham: Springer International Publishing, 2015. P. 125–141.
4. Parisi F., Grant J. Knowledge representation in probabilistic spatio-temporal knowledge bases // *Journal of Artificial Intelligence Research*. 2016. Vol. 55. P. 743–798.
5. Parisi F., Grant J. On repairing and querying inconsistent probabilistic spatio-temporal databases // *International Journal of Approximate Reasoning*. Elsevier Inc., 2017. Vol. 84. P. 41–74.
6. Dylla F., Lee J. H., Mossakowski T., Schneider T., Delden A. V., Ven J. V. D., Wolter D. A Survey of Qualitative Spatial and Temporal Calculi // *ACM Computing Surveys*. 2017. Vol. 50, № 1. P. 1–39.
7. Wolter D., Lee J. H. Connecting qualitative spatial and temporal representations by propositional closure // *IJCAI International Joint Conference on Artificial Intelligence*. 2016. Vol. 2016-Janua. P. 1308–1314.
8. Freksa C., van de Ven J., Wolter D. Formal representation of qualitative direction // *International Journal of Geographical Information Science*. Taylor & Francis, 2018. Vol. 32, № 12. P. 2514–2534.
9. Shrinidhi L., Kadiresan N., Parameswaran L. Ontology Model for Spatio-Temporal Contexts in Smart Home Environments // *IFIP Advances in Information and Communication Technology*. Springer Science and Business Media Deutschland GmbH, 2021. Vol. 611 IFIPAICT. P. 113–124.
10. Кандрашина Е. Ю., Литвинцева Л. В., Поспелов Д. А. Представление знаний о времени и пространстве в интеллектуальных системах / под ред. Д. А. Поспелов. М.: Наука, 1989. 328 с.
11. Lopez X. GeoSPARQL — A geographic query language for RDF data A proposal for an OGC Draft Candidate Standard. 2012. P. 13.
12. Cox S. J. D., Little C. Time Ontology in OWL. 2017.
13. Baratis E., Petrakis E. G., Batsakis S., Maris N., Papadakis N. TOQL: Temporal ontology querying language // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, Berlin, Heidelberg, 2009. Vol. 5644 LNCS. P. 338–354.
14. Nys G.-A., Ruymbeke M. Van, Billen R. Spatio-Temporal Reasoning in CIDOC CRM: An Hybrid Ontology with GeoSPARQL and OWL-Time // *2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage*. 2018.
15. Subramanian A., RR P. K., Vikkurthi M., & Buttigieg P. L. Semantic Harmonisation of Numeric Data from Open Government Data // *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data — CoDS-COMAD '19*. New York, New York, USA: ACM Press, 2019. P. 238–244.
16. Hoffart J., Suchanek F. M., Berberich K., & Weikum G. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia // *Artificial Intelligence*. 2013. Vol. 194. P. 28–61.
17. Calvanese D., Cogrel B., Komla-Ebri S., Kontchakov R., Lanti D., Rezk M., Rodriguez-Muro M., Xiao G. Ontop: Answering SPARQL queries over relational databases // *Semantic Web*. IOS Press, 2017. Vol. 8, № 3. P. 471–487.

18. Bereta K., Xiao G., Koubarakis M. Ontop-spatial: Ontop of geospatial databases // *Journal of Web Semantics*. Elsevier B. V., 2019. Vol. 58. P. 100514.
19. Sander J., Ester M., Kriegel H. P., Xu X. Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications // *Data Mining and Knowledge Discovery*. Springer Netherlands, 1998. Vol. 2, № 2. P. 169–194.

References

1. Mishra R. K., Kumari C. L., Janaki Krishna P. S., Dubey A. Smart Cities for Sustainable Development: An Overview. *Smart Cities for Sustainable Development*. Springer, Singapore, 2022, pp. 1–12.
2. Yakimov M. R. Features of the Use of Geoanalytical Data in the Development of Transport Planning Documents. *2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications, SYNCHROINFO 2022 — Conference Proceedings*. Institute of Electrical and Electronics Engineers Inc., 2022.
3. Mashhadi A., Quattrone G., Capra L. The Impact of Society on Volunteered Geographic Information: The Case of OpenStreetMap. *OpenStreetMap in GIScience: Experiences, Research, Applications* / ed. Jokar Arsanjani J. et al. Cham: Springer International Publishing, 2015, pp. 125–141.
4. Parisi F., Grant J. Knowledge representation in probabilistic spatio-temporal knowledge bases. *Journal of Artificial Intelligence Research*, 2016, vol. 55, pp. 743–798.
5. Parisi F., Grant J. On repairing and querying inconsistent probabilistic spatio-temporal databases. *International Journal of Approximate Reasoning*. Elsevier Inc., 2017, vol. 84, pp. 41–74.
6. Dylla F., Lee J. H., Mossakowski T., Schneider T., Delden A. V., Ven J. V. D., Wolter D. A Survey of Qualitative Spatial and Temporal Calculi. *ACM Computing Surveys*, 2017, vol. 50, no. 1, pp. 1–39.
7. Wolter D., Lee J.H. Connecting qualitative spatial and temporal representations by propositional closure. *IJCAI International Joint Conference on Artificial Intelligence*, 2016, vol. 2016-Janua, pp. 1308–1314.
8. Freksa C., van de Ven J., Wolter D. Formal representation of qualitative direction. *International Journal of Geographical Information Science*. Taylor & Francis, 2018, vol. 32, no. 12, pp. 2514–2534.
9. Shrinidhi L., Kadiresan N., Parameswaran L. Ontology Model for Spatio-Temporal Contexts in Smart Home Environments. *IFIP Advances in Information and Communication Technology*. Springer Science and Business Media Deutschland GmbH, 2021, vol. 611 IFIPAICT, pp. 113–124.
10. Kandrashina E. Yu., Litvinceva L. V., Pospelov D. A. *Predstavlenie znanij o vremeni i prostranstve v intellektual'nyh sistemah* [Representation of knowledge about time and space in intelligent systems]. Moscow, Nauka, 1989, 328 p. (In Russ.).
11. Lopez X. GeoSPARQL — A geographic query language for RDF data. A proposal for an OGC Draft Candidate Standard, 2012, P. 13.
12. Cox S. J. D., Little C. Time Ontology in OWL, 2017.
13. Baratis E., Petrakis E. G., Batsakis S., Maris N., Papadakis N. TOQL: Temporal ontology querying language. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, Berlin, Heidelberg, 2009, vol. 5644 LNCS, pp. 338–354.
14. Nys G.-A., Ruymbeke M. Van, Billen R. Spatio-Temporal Reasoning in CIDOC CRM: An Hybrid Ontology with GeoSPARQL and OWL-Time. *2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage*, 2018.
15. Subramanian A., RR P. K., Vikkurthi M., Buttigieg P. L. Semantic Harmonisation of Numeric Data from Open Government Data. *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data – CoDS-COMAD '19*. New York, New York, USA, ACM Press, 2019, pp. 238–244.
16. Hoffart J., Suchanek F. M., Berberich K., Weikum G. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*, 2013, vol. 194, pp. 28–61.

17. Calvanese D., Cogrel B., Komla-Ebri S., Kontchakov R., Lanti D., Rezk M., Rodriguez-Muro M., Xiao G. Ontop: Answering SPARQL queries over relational databases. *Semantic Web*. IOS Press, 2017, vol. 8, no. 3, pp. 471–487.
18. Bereta K., Xiao G., Koubarakis M. Ontop-spatial: Ontop of geospatial databases. *Journal of Web Semantics*. Elsevier B. V., 2019, vol. 58, pp. 100514.
19. Sander J., Ester M., Kriegel H. P., Xu X. Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery*. Springer Netherlands, 1998, vol. 2, no. 2, pp. 169–194.

Информация об авторах

А. В. Смирнов — доктор технических наук, профессор, главный научный сотрудник;

Н. Н. Тесля — кандидат технических наук, старший научный сотрудник.

Information about the authors

A. V. Smirnov — Doctor of Science (Tech.), professor, chief researcher;

N. N. Teslya — Candidate of Science (Tech.), senior researcher.

Статья поступила в редакцию 14.10.2023; одобрена после рецензирования 01.11.2023; принята к публикации 08.11.2023.
The article was submitted 14.10.2023; approved after reviewing 01.11.2023; accepted for publication 08.11.2023.