

## ПРИКЛАДНЫЕ ОНТОЛОГИИ ПРОЕКТИРОВАНИЯ

УДК 004.89

Научная статья

DOI: 10.18287/2223-9537-2025-15-1-45-54



## Построение графа знаний по телекоммуникационным данным

© 2025, А.А. Головин<sup>1</sup>✉, Н.А. Жукова<sup>1,2</sup>

<sup>1</sup> Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ «ЛЭТИ»), Санкт-Петербург, Россия

<sup>2</sup> Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН), Санкт-Петербург, Россия

### Аннотация

В работе описывается метод построения графа знаний по телекоммуникационным данным на основе проприетарных и эталонных моделей, используемых в области телекоммуникаций. В качестве эталонных рассматриваются модели, входящие в фреймворк, разработанный консорциумом *TM Forum*. Граф знаний проприетарных моделей предлагается строить с помощью автоматизированной обработки лог-файлов автотестов и таблиц БД биллинговой системы. Актуальность применения графов знаний обусловлена их структурированностью и семантической, а также возможностью последующего применения алгоритмов машинного обучения для генерации рекомендаций по оптимизации телекоммуникационных процессов и систем. Предложено применение метода на основе подхода многошагового рассуждения для создания интерпретируемых рекомендаций по восстановлению отсутствующих связей, путём их прогнозирования в проприетарном графе знаний. Предложенный метод рассматривает многошаговое рассуждение как задачу ответа на вопрос с использованием обработки естественного языка. Применение разработанного решения на основе нейросетевой архитектуры трансформера обеспечило интерпретируемые результаты с сохранением значений метрик, по сравнению с аналогами.

**Ключевые слова:** граф знаний, телекоммуникации, лог-файлы, восстановление структуры, многошаговое рассуждение.

**Цитирование:** Головин А.А., Жукова Н.А. Построение графа знаний по телекоммуникационным данным. *Онтология проектирования*. 2025. Т.15, №1(55). С.45-54. DOI:10.18287/2223-9537-2025-15-1-45-54.

**Финансирование:** работа выполнена при финансовой поддержке гранта, бюджетная тема №FFZF-2022-0006.

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

### Введение

Термин граф знаний (ГЗ) определяется как база знаний, состоящая из сущностей и связей между ними. ГЗ широко используются в обработке естественного языка, вопросно-ответных системах, рекомендательных системах и т.д. [1]. ГЗ содержат структурированные данные о сущностях и связях между ними, представленные в формате триплета  $(h, r, t)$ , где  $h$  — головная сущность,  $r$  — отношение,  $t$  — хвостовая сущность. ГЗ востребованы в различных областях, включая и телекоммуникации [2], где они используются для описания сетевых устройств, сервисов, бизнес-процессов и других сущностей, связанных с телекоммуникационной инфраструктурой [3]. Актуальность применения ГЗ в этой области обусловлена несколькими факторами [4]:

- рост объёма и сложности телекоммуникационных данных, которые требуют эффективного управления, интеграции и обработки.
- необходимость повышения качества и надёжности телекоммуникационных сервисов, а также удовлетворения индивидуальных потребностей клиентов.
- появление новых технологий и стандартов в телекоммуникационной отрасли, которые создают новые возможности для телекоммуникационных операторов.

ГЗ могут помочь в решении различных задач в телекоммуникационной области за счёт использования следующих преимуществ [5]:

- данные в ГЗ являются структурированными и семантическими, что позволяет извлекать, связывать и анализировать информацию из разных источников и в разных форматах;
- для ГЗ возможно применение логического вывода и машинного обучения, что позволяет автоматизировать и оптимизировать процессы принятия решений, обнаружения аномалий, предсказания поведения и др.

Одним из источников данных в области телекоммуникаций являются лог-файлы биллинговых систем. При построении ГЗ на основе данных лог-файлов, в связи с их слабой структурированностью, формируемый граф часто имеет неполную структуру, в которой отсутствует нужная информация. Таким образом, разработка метода эффективного построения ГЗ по телекоммуникационным лог-файлам является перспективной областью исследований.

Целью данной работы является разработка метода построения ГЗ по телекоммуникационным данным по лог-файлам биллинговой системы, а также исследование возможности применения метода на основе подхода многошагового рассуждения для генерации интерпретируемых рекомендаций по восстановлению отсутствующих связей в ГЗ.

## 1 Наборы данных

В ГЗ содержатся данные о сетевых устройствах и информационных системах, функционирующих в телекоммуникационной сети (ТС), и данные, содержащие информацию о бизнес-процессах компаний. ГЗ, используемые в телекоммуникационных системах, включают бизнес-процессные, информационные, функциональные и компонентные модели, а также статистические и оперативные данные о сетях. Данные, представленные в форме ГЗ в сфере телекоммуникаций, могут использоваться для различных целей, таких как анализ, моделирование, оптимизация и управление ТС.

В работах [6, 7] рассматривается телекоммуникационный ГЗ, построенный по данным оператора кабельного телевидения. При построении ГЗ использовались данные о пользователях, сетевых устройствах и услугах оператора. Построение ГЗ позволило провести анализ поведения клиентов кабельного оператора. Важным требованием была возможность обработки большого объёма данных [7].

В [8] представлен ГЗ, построенный на основе стандартизированных для области телекоммуникаций моделей, разработкой и поддержкой которых занимается консорциум *TM Forum*, специализирующийся на стандартизации и управлении телекоммуникационными и цифровыми сервисами. *TM Forum* предоставляет фреймворк (*TM Framework, TMFF*), включающий в себя бизнес-процессную (*Enhanced Telecom Operations Map, eTOM*), информационную (*Shared Information and Data Model, SID*), функциональную модели (*Functional model*) и модель приложений (*Telecom Application Map, TAM*). Модель *TMFF* и связи между её элементами на примере задачи заказа услуг показаны на рисунке 1.

При построении ГЗ ТС операторов, как правило, используются многие источники, которые предоставляют информацию о ТС — это лог-файлы, различные текстовые описания, включая пользовательскую документацию, документацию программного кода и т.д. Объёмы

таких данных достаточно большие, в частности, в [7] при построении проприетарных ГЗ (ПГЗ, закрытый граф, разработанный для решения задач конкретной сферы или предприятия) использовались лог-файлы сетевых устройств, в которые могло быть записано более миллиона сообщений в день. Однако, эти данные часто являются неполными, что приводит к нарушению полноты структуры формируемого ГЗ.

ГЗ на основе *TMFF* может использоваться как эталонная модель, с которой сопоставляются модели, построенные по реальным данным телекоммуникационных компаний. При формировании рекомендаций по восстановлению структуры предлагается применять подход на основе многошагового рассуждения, который обеспечит интерпретируемые прогнозы. Эталонная модель *TMFF* при этом используется в качестве обучающего и валидационного набора данных.

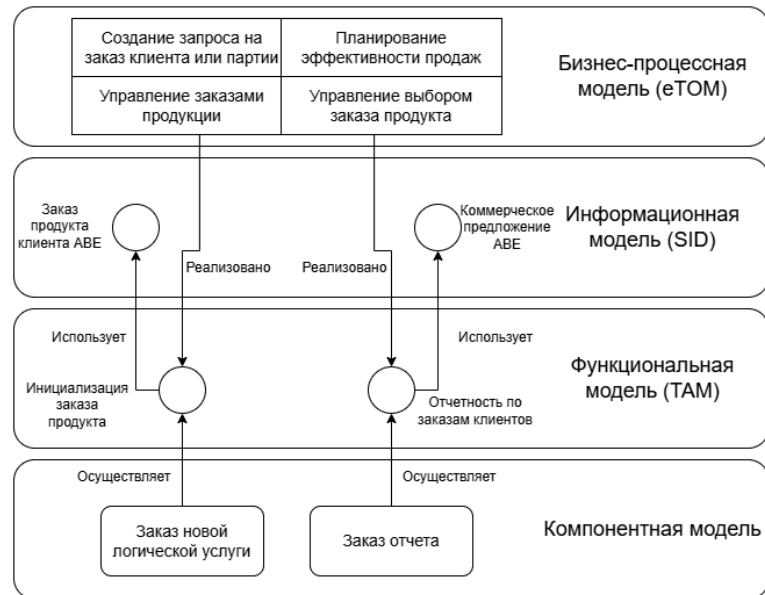


Рисунок 1 – Модель *TMFF* и зависимости между её элементами

## 2 Метод решения

Предлагаемый метод предусматривает построение ГЗ, содержащего данные о ТС и реализуемых в них процессах, на основе имеющихся о сетях данных, получаемых из различных источников, а также восстановление структуры графа с использованием многошагового рассуждения и ГЗ, построенного на основе эталонных моделей *TM Forum*.

### 2.1 Построение ГЗ по телекоммуникационным данным

Эталонный ГЗ был построен на основе модели *TMFF*, включающей в себя модели *eTOM*, *SID*, *TAM* и компонентную модели. С использованием этих моделей были сгенерированы файлы *CSV* и выполнен маппинг ГЗ в формате *JSON* для СУБД *Neo4j*. Данный ГЗ экспортирован и помещён в открытый доступ на платформе *GitHub* [8].

ПГЗ построен по телекоммуникационным данным, полученным из биллинговой системы. В состав ГЗ были включены данные об информационных объектах (для формирования информационной составляющей ГЗ), данные из спецификаций программного кода (для формирования бизнес-процессной и функциональной составляющих ГЗ) и данные об автотестах (для формирования компонентной составляющей ГЗ).

Данные об информационных объектах получены с помощью реляционной базы данных (БД) в результате разбора документации, который позволил выявить связи между элементами данных из БД и компонентами биллинговой системы.

По данным, хранящимся в БД, построена проприетарная онтология, которая в дальнейшем интегрирована в единый ПГЗ. Для преобразования данных из БД в онтологию использовался редактор онтологий *Protégé*, в частности функция импорта *R2RML*. Для создания файлов маппинга сущностей БД использовались следующие классы: имя БД (*Database name*), имена таблиц (*Table names*), имена столбцов (*Column names*) и имена ключей (*Key names*). Иерархия классов онтологии представлена на рисунке 2.

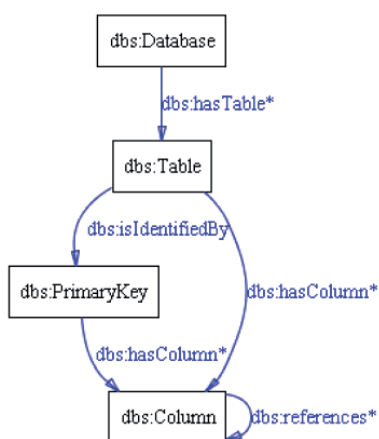


Рисунок 2 – Иерархия классов проприетарной онтологии

Для построения компонентной модели, включаемой в состав ПГЗ, использовались лог-файлы, полученные в результате выполнения автотестов. Были проанализированы лог-файлы и сформированы CSV файлы, содержащие информацию об иерархии компонентов биллинговой системы. Информация о бизнес-процессах и функциях была получена путём автоматической обработки исходного кода. На основе этой информации был построен ПГЗ с использованием СУБД Neo4j.

Связи между элементами функциональной и информационной моделей в ПГЗ показаны на рисунке 3.

Примеры связанных сущностей, соответствующих функциям «Извлечение данных баланса» из проприетарной функциональной модели и «Централизованное управление правилами заказа» из эталонной функциональной модели, а также элементы из связанных с ними сущностей информационной модели показаны на рисунках 4 и 5.

На рисунке 6 отображены связи между сущностью «Заказы логических услуг», соответствующей проприетарной информационной модели, и сущностями, соответствующими SID.



Рисунок 3 – Фрагмент графа знаний, показывающий связи между элементами функциональной и информационной моделей



Рисунок 4 – Пример связанных сущностей, соответствующих функции «Извлечение данных баланса» из проприетарной функциональной модели

## 2.2 Восстановление заказа структуры ГЗ

Дальнейшим этапом является восстановление структуры полученного ПГЗ. Восстановление структуры ГЗ – это задача прогнозирования недостающих связей или фактов в ГЗ [9], необходимость решения которой обуславливается тем, что значительная часть создаваемых ПГЗ является неполными и зашумленными, что ограничивает их полезность и надёжность [10]. Для восстановления структуры ГЗ предлагается использовать методы, способные вывести новые факты из существующих и заполнять пробелы в ГЗ, в частности, метод многошагового рассуждения.



Рисунок 5 – Пример связанных сущностей, соответствующих функции «Централизованное управление правилами заказа» из эталонной функциональной модели



Рисунок 6 – Фрагмент графа знаний с связанными сущностями, соответствующими информационным моделям эталонного и проприетарного графов знаний

### 2.2.1 Многошаговое рассуждение

Многошаговое рассуждение – это подход, позволяющий предсказывать новые связи для неполных ГЗ, который заключается в поиске и комбинировании нескольких логических шагов, связывающих сущности и отношения [11]. При многошаговом рассуждении вместо того, чтобы предсказывать хвостовую сущность  $t$  для данного триплета запроса  $(h, r)$ , обеспечивается предоставление доказательного пути от  $h$  до  $t$ , который демонстрирует процесс вывода [12].

Многошаговое рассуждение помогает восстановить отсутствующие элементы ГЗ, используя косвенные связи и выполняя интерпретируемые логические выводы на основе доступных данных, что особенно актуально для телекоммуникационных ГЗ, построенных по лог-файлам и другим данным телекоммуникационных операторов, которые имеют неполную структуру [7]. Интерпретируемость полученных результатов позволяет обосновывать прогнозы и принимать взвешенные решения, что важно при решении прикладных задач.

Применение подхода многошагового рассуждения для ПГЗ в связке с эталонным ГЗ, построенным по модели *TM Forum*, позволяет предоставить рекомендации по восстановлению структуры ГЗ телекоммуникационного оператора. Для этого модель на основе многошагового рассуждения обучается на эталонном ГЗ. Рекомендации формируются за счёт предсказания недостающих связей в ПГЗ.

Многошаговое рассуждение может быть реализовано с помощью различных методов, которые предполагают использование моделей, основанных на обучении с подкреплением [13, 14], нейросетевой архитектуре трансформера и других [15]

### 2.2.2 Предложенное решение

Предлагаемый метод основан на многошаговом рассуждении, которое использует архитектуру трансформера кодер-декодер для преобразования запроса в доказательный путь («*sequence-to-sequence*»).

Использование такой архитектуры обеспечивает существенные преимущества [16]:



- метод не полагается на существующие связи для создания доказательного пути и обладает гибкостью для восстановления отсутствующих связей (особенно эффективно при работе с разреженными ГЗ);
- доказательные пути, генерируемые предлагаемым методом, обеспечивают интерпретируемость результатов.

Разработанный метод рассматривает многошаговое рассуждение как задачу преобразования последовательностей из одного домена последовательности в другой. Запрос является исходной последовательностью, а доказательный путь — целевой. Кодер-трансформер отображает запрос и предыдущую последовательность путей в контекстуализированное представление, и это представление затем используется для авторегрессионного декодирования выходного пути, токен за токеном.

Идея предлагаемого метода состоит в том, чтобы свести задачу многошагового рассуждения к задаче ответа на вопрос с использованием обработки естественного языка. Различные сущности и отношения интерпретируются как разные слова, контекстуальные вложения которых обучаются на графе.

Процесс обучения предлагаемого метода:

$$q = (h, r), \quad \tau = (r_1, t_1, \dots, r_n, t, [\text{eos}]), \quad (1)$$

где [eos] – конец токена последовательности,  $q$  и  $\tau$  обозначают запрос и путь.

Разложение вероятности:

$$p(\tau|q) = \prod_{k=1}^{|\tau|} p(\tau_k|q, \tau_{<k}), \quad (2)$$

где  $\tau_k$  обозначает  $k$ -й токен выходного пути, а  $|\tau|$  – количество токенов в пути.

Обучается матрица встраивания токенов  $E \in \mathbb{R}^{V \times d}$ , где  $d$  — размерность векторного представления. Размер словаря равен  $V = |\mathcal{E}| + |\mathcal{R}| + |\mathcal{S}|$ , поскольку словарь токенов включает в себя набор сущностей  $\mathcal{E}$ , набор отношений  $\mathcal{R}$  и набор специальных токенов  $\mathcal{S}$ . При декодировании  $\tau_k$  маска позволяет кодировщику обрабатывать только токены запроса ( $h, r$ ) и предыдущие токены  $\tau_{<k}$  в выходном пути, предотвращая раскрытие будущей информации.

Далее метод вычисляет распределение вероятностей для  $k$ -го токена как:

$$p(\cdot | q, \tau_{<k}) = \text{Softmax}(MLP(Enc(h, r, \tau_{<k})) \cdot E), \quad (3)$$

где  $MLP(\cdot)$  — многослойный перцептрон,  $Enc(\cdot)$  — кодировщик (*Transformer Encoder*) предлагаемого «*sequence-to-sequence*» метода (рисунок 7).

В процессе обучения метод выбирает конкретный путь для каждого триплета  $(h, r, t) \in T$  из всех возможных путей между  $h$  и  $t$  с максимальной длиной  $N = 3$ . Значение  $N$  было выбрано на основе наблюдения, что большинство пар сущностей в наборах данных ГЗ связаны максимум за 3 шага. Этот выбор также обеспечивает корректное сравнение с аналогами, которые также ограничивают максимальную длину пути до 3 [15]. В качестве функции потерь используется перекрёстная энтропия с дополнительным сглаживанием меток.

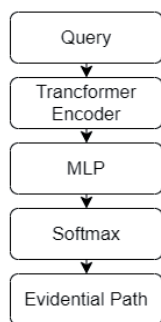


Рисунок 7 – *Sequence-to-sequence* метод

### 3 Исследование свойств решения

В ходе эксперимента построены эталонный ГЗ и ПГЗ. Параметры построенных ГЗ приведены в таблице 1.

Для оценки качества предлагаемого метода на основе многошагового рассуждения рассчитаны следующие метрики: средний обратный ранг (*MRR*) и количество попаданий в  $N$ . *MRR* — это показатель, используемый для оценки эффективности системы ранжирования в задачах поиска информации и рекомендаций [17]. Он измеряет среднее качество ранжиро-

ванных результатов, учитывая ранговую позицию первого релевантного элемента. *MRR* рассчитывается путём взятия обратной величины ранга первого релевантного элемента и последующего усреднения этих обратных рангов по всем запросам:

$$MRR = \frac{1}{|Q|} \sum_{q=1}^{|Q|} \frac{1}{rank_q}, \quad (4)$$

где  $|Q|$  — общее количество запросов, а  $rank_q$  — позиция наиболее релевантного для запроса элемента в первых  $N$  результатах.

Количество попаданий в  $N$  ( $H@N$ ) — метрика оценки, используемая в системах информационного поиска и рекомендаций [18]. Она измеряет долю запросов, по которым хотя бы один релевантный элемент появляется в первых  $N$  результатах:

Таблица 1 – Ключевые параметры и значения построенного эталонного и проприетарного графов знаний

Параметры	Значения эталонного графа знаний	Значения проприетарного графа знаний
Количество сущностей	10458	2882
Количество связей	30708	3040

$$H@N = \frac{|\{q \in Q \mid rank_q \leq N\}|}{|Q|}. \quad (5)$$

В таблице 2 представлены метрики, рассчитанные на построенном эталонном ГЗ для существующих решений и предлагаемого метода.

Таким образом, разработанный метод практически достигает тех же результатов, что и методы, основанные на векторных представлениях. Результаты метрик предложенного метода «*sequence-to-sequence*» превысили результаты метода *KG-BERT*. При этом использование предложенного метода обеспечило интерпретируемость, в то время как результаты, получаемые с использованием других методов, интерпретируемыми не являются.

Таблица 2 – Результаты прогнозирования отношений на эталонном графе знаний

Метод	MRR	H@1	H@3	H@10
TransE	0,426	0,321	0,471	0,633
ConvE	0,437	0,343	0,432	0,627
RotatE	0,423	0,314	0,473	0,626
TuckER	0,452	0,358	0,494	0,635
ConE	0,448	0,342	0,492	0,641
KG-BERT	–	0,041	0,301	0,422
Seq-to-seq	0,439	0,324	0,462	0,619

### Заключение

В результате проведённого исследования построены эталонные ГЗ и ПГЗ. Для создания эталонного ГЗ применялись модели ТМ *Forum*, а ПГЗ строился за счёт автоматизированной обработки лог-файлов автотестов и таблиц БД биллинговой системы.

Разработанный на основе многошагового рассуждения «*sequence-to-sequence*» метод для восстановления структуры ГЗ рассматривает триплет запроса и доказательный путь как последовательности и использует архитектуру трансформера. Предложенный метод имеет хорошую сходимость и возможность интерпретации результата.

### Авторский вклад

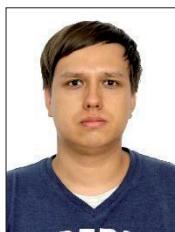
Головин А.А. построил ГЗ, провел исследование, участвовал в написании рукописи. Жукова Н.А. спланировала исследование и структуру ГЗ, участвовала в написании рукописи.

### Список источников

- [1] *Li J., Hou L.* Review of knowledge graph research. *Natural Sci. Ed.* 2017. Vol.40(3). P.454-459. DOI: 10.13451/j.cnki.shanxi.univ(nat.sci.).2017.03.008.
- [2] *Mier F.J.Z.* Applications of Knowledge Graphs in Telecommunication Systems Management. *IEEE Internet Computing.* 2023. Vol.27(3). P.29-34. DOI: 10.1109/MIC.2023.3253305.

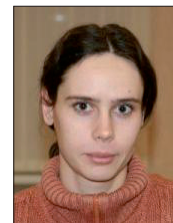
- [3] **Krinkin K., Kulikov I., Vodyaho A., Zhukova N.** Architecture of a Telecommunications Network Monitoring System Based on a Knowledge Graph. 26th *Conference of Open Innovations Association (FRUCT)*. 2020. P.231-239. DOI: 10.23919/FRUCT48808.2020.9087429.
- [4] **Lei T., Wang Y.** Telecom Operator Knowledge Graphs Reasoning for Question Answering // 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI). 2021. P.981-984. DOI: 10.1109/CISAI54367.2021.00196.
- [5] **Zou X.** A Survey on Application of Knowledge Graph // *Journal of Physics: Conference Series*. 2020. Vol.1487. DOI: 10.1088/1742-6596/1487/1/012016.
- [6] **Krinkin K., Vodyaho A., Kulikov I., Zhukova N.** Models of Telecommunications Network Monitoring Based on Knowledge Graphs. 9th *Mediterranean Conference on Embedded Computing (MECO)*. 2020. P. 1-7. DOI: 10.1109/MECO49872.2020.9134148.
- [7] **Zhukova N., Kulikov I., Utkin N.** The Method for Searching Patterns In Log Files of Telecommunication Devices for Monitoring their State. 2021 *XXIV International Conference on Soft Computing and Measurements (SCM)*. 2021. P. 124-126. DOI: 10.1109/SCM52931.2021.9507157.
- [8] **TMF Business Process Framework Dataset for Neo4j.** <https://github.com/AIgor/TMF-Business-Process-Framework-Dataset-for-Neo4j>.
- [9] **Chen Z., Wang Y., Zhao B., Cheng J., Zhao X., Duan Z.** Knowledge Graph Completion: A Review. *IEEE Access*. 2020. Vol.8. P.192435-192456. DOI: 10.1109/ACCESS.2020.3030076.
- [10] **Issa S., Adekunle O., Hamdi F., Cherfi S. S. -S., Dumontier M., Zaveri A.** Knowledge Graph Completeness: A Systematic Literature Review. *IEEE Access*. 2021. Vol.9. P.31322–31339. DOI: 10.1109/ACCESS.2021.3056622.
- [11] **Wan G., Du B.** Gaussian Path: A Bayesian Multi-Hop Reasoning Framework for Knowledge Graph Reasoning. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. Vol.35(5). P.4393-4401. DOI: 10.1609/aaai.v35i5.16565.
- [12] **Das R., Dhuliawala S., Zaheer M., Vilnis L., Durugkar I., Krishnamurthy A., Smola A., McCallum A.** Go for a Walk and Arrive at the Answer: Reasoning Over Paths in Knowledge Bases using Reinforcement Learning. *ICLR 2018*. 2018. P.1-18. DOI: 10.48550/arXiv.1711.05851.
- [13] **Bai L., Yu W., Chen M., Ma X.** Multi-hop reasoning over paths in temporal knowledge graphs using reinforcement learning. *Applied Soft Computing*. 2021. Vol.103. P.1-9. DOI: 10.1016/j.asoc.2021.107144.
- [14] **Cui H., Peng T., Xiao F., Han J., Han R., Liu L.** Incorporating anticipation embedding into reinforcement learning framework for multi-hop knowledge graph question answering. *Information Sciences*. 2023. Vol.619. P.745-761. DOI: 10.1016/j.ins.2022.11.042.
- [15] **Bai Y., Lv X., Li J., Hou L., Qu Y., Dai Z., Xiong F.** SQUIRE: A Sequence-to-sequence Framework for Multi-hop Knowledge Graph Reasoning. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 2022. P.1649-1662. DOI: 10.18653/v1/2022.emnlp-main.107.
- [16] **Zou J., Wan J., Zhang H., Zhang Y.** Multi-hop Path Query Answering Model for Knowledge Graph based on Neighborhood Aggregation and Transformer. *Journal of Physics: Conference Series*. 2023. Vol.2560. P.1-15. DOI: 10.1088/1742-6596/2560/1/012049.
- [17] **Caballero M.** A Brief Survey of Question Answering Systems // *International Journal of Artificial Intelligence & Applications* (IJAIA). 2022. Vol. 12(5). P. 1-7.
- [18] **Kim B., Hong T., Ko Y., Seo J.** Multi-Task Learning for Knowledge Graph Completion with Pre-trained Language Models. *Proceedings of the 28th International Conference on Computational Linguistics*. 2020. P.1737-1743. DOI: 10.18653/v1/2020.coling-main.153.

## Сведения об авторах



**Головин Алексей Андреевич**, 1999 г. рождения. Окончил магистратуру СПбГЭТУ «ЛЭТИ». Аспирант кафедры математического обеспечения и применения ЭВМ СПбГЭТУ «ЛЭТИ». ORCID: 0009-0004-4687-8393; Author ID (РИНЦ): 1214796; Author ID (Scopus): 57442656600; [golovin99@icloud.com](mailto:golovin99@icloud.com). ✉.

**Жукова Наталья Александровна**, 1983 г. рождения. Окончила СПбГЭТУ «ЛЭТИ» в 2006 году, д.т.н. (2020). Профессор кафедры математического обеспечения и применения ЭВМ СПбГЭТУ «ЛЭТИ», ведущий научный сотрудник СПб ФИЦ РАН. В списке научных трудов более 100 работ в области обработки и анализа данных. ORCID: 0000-0001-5877-4461; Author ID (РИНЦ): 222968; Author ID (Scopus): 56406142300. [nazhukova@mail.ru](mailto:nazhukova@mail.ru).



Поступила в редакцию 14.10.2024, после рецензирования 20.11.2024. Принята к публикации 23.12.2024.





## Building a knowledge graph from telecommunication data

© 2025, A.A. Golovin<sup>1</sup>✉, N.A. Zhukova<sup>1,2</sup>

<sup>1</sup> Saint Petersburg Electrotechnical University «LETI» (ETU «LETI»), Saint-Petersburg, Russia

<sup>2</sup> St. Petersburg Federal Research Centre of the Russian Academy of Sciences (SPCRAS), Saint-Petersburg, Russia

### Abstract

The paper presents a method for building a knowledge graph from telecommunication data using proprietary and reference models commonly employed in the telecommunications domain. Reference models are based on those included in the framework developed by the TM Forum consortium. The proposed approach involves building a knowledge graph for proprietary models through automated processing of autotest log files and billing system database tables. The relevance of knowledge graphs stems from their structured and semantic nature, as well as their potential for applying machine learning algorithms to generate recommendations for optimizing telecommunication processes and systems. A method based on a multi-step reasoning approach is proposed for creating interpretable recommendations by predicting and restoring missing links in a proprietary knowledge graph. This approach treats multi-step reasoning as a question-answering task using natural language processing techniques. The implementation of the proposed solution, leveraging a transformer-based neural network architecture, yielded interpretable results while maintaining metric values comparable to existing methods.

**Keywords:** knowledge graphs, telecommunication, log files, structure recovery, multi-step reasoning.

**For citation:** Golovin AA, Zhukova NA. Building a knowledge graph from telecommunication data [In Russian]. *Ontology of designing*. 2025; 15(1): 45-54. DOI:10.18287/2223-9537-2025-15-1-45-54.

**Financial support:** This work was supported by the state budget, project No. FFZF-2022-0006.

**Conflict of interest:** The authors declare no conflict of interest.

### List of figures and tables

Figure 1 - TMFF model and dependencies between its elements

Figure 2 - Proprietary ontology class hierarchy

Figure 3 - A fragment of the knowledge graph that shows the relations between the elements of the functional and information models

Figure 4 - Examples of related entities corresponding to the "Balance Data Retrieval" function from proprietary functional model

Figure 5 - Examples of related entities corresponding to the "Centralized Ordering Rules Management" function from the reference functional model

Figure 6 - A fragment of the knowledge graph with related entities corresponding to the information models of the reference and proprietary knowledge graphs

Figure 7 - Sequence-to-sequence method overview

Table 1 - Key parameters of the constructed reference and proprietary knowledge graphs

Table 2 - Results of relation prediction on the reference knowledge graph

### References

- [1] **Li J, Hou L.** Review of knowledge graph research. *Natural Sci. Ed.* 2017; 40(3): 454-459. DOI: 10.13451/j.cnki.shanxi.univ(nat.sci.).2017.03.008.
- [2] **Mier FJZ.** Applications of Knowledge Graphs in Telecommunication Systems Management. *IEEE Internet Computing*. 2023; 27(3): 29-34. DOI: 10.1109/MIC.2023.3253305.
- [3] **Krinkin K, Kulikov., Vodyaho A, Zhukova N.** Architecture of a Telecommunications Network Monitoring System Based on a Knowledge Graph. 26th *Conference of Open Innovations Association (FRUCT)*. 2020. P.231-239. DOI: 10.23919/FRUCT48808.2020.9087429.

- [4] **Lei T, Wang Y.** Telecom Operator Knowledge Graphs Reasoning for Question Answering. 2021 *International Conference on Computer Information Science and Artificial Intelligence (CISAI)*. 2021. P.981-984. DOI: 10.1109/CISAI54367.2021.00196.
  - [5] **Zou X.** A Survey on Application of Knowledge Graph. *Journal of Physics: Conference Series*. 2020; 1487. DOI: 10.1088/1742-6596/1487/1/012016.
  - [6] **Krinkin K, Vodyaho A, Kulikov I, Zhukova N.** Models of Telecommunications Network Monitoring Based on Knowledge Graphs. 9th *Mediterranean Conference on Embedded Computing (MECO)*. 2020. P.1-7. DOI: 10.1109/MECO49872.2020.9134148.
  - [7] **Zhukova N, Kulikov I, Utkin N.** The Method for Searching Patterns In Log Files of Telecommunication Devices for Monitoring their State. 2021 *XXIV International Conference on Soft Computing and Measurements (SCM)*. 2021. P.124-126. DOI: 10.1109/SCM52931.2021.9507157.
  - [8] TMF Business Process Framework Dataset for Neo4j. <https://github.com/Algord/TMF-Business-Process-Framework-Dataset-for-Neo4j>.
  - [9] **Chen Z, Wang Y, Zhao B, Cheng J, Zhao X, Duan Z.** Knowledge Graph Completion: A Review. *IEEE Access*. 2020; 8: 192435-192456. DOI: 10.1109/ACCESS.2020.3030076.
  - [10] **Issa S, Adekunle O, Hamdi F, Cherfi SS-S, Dumontier M, Zaveri A.** Knowledge Graph Completeness: A Systematic Literature Review. *IEEE Access*. 2021; 9: 31322-31339. DOI: 10.1109/ACCESS.2021.3056622.
  - [11] **Wan G, Du B.** Gaussian Path: A Bayesian Multi-Hop Reasoning Framework for Knowledge Graph Reasoning. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021; 35(5): 4393-4401. DOI: 10.1609/aaai.v35i5.16565.
  - [12] **Das R, Dhuliawala S, Zaheer M, Vilnis L, Durugkar I, Krishnamurthy A, Smola A, McCallum A.** Go for a Walk and Arrive at the Answer: Reasoning Over Paths in Knowledge Bases using Reinforcement Learning. *ICLR 2018*. 2018. P.1-18. DOI: 10.48550/arXiv.1711.05851.
  - [13] **Bai L, Yu W, Chen M, Ma X.** Multi-hop reasoning over paths in temporal knowledge graphs using reinforcement learning. *Applied Soft Computing*. 2021; 103: 1-9. DOI: 10.1016/j.asoc.2021.107144.
  - [14] **Cui H, Peng T, Xiao F, Han J, Han R, Liu L.** Incorporating anticipation embedding into reinforcement learning framework for multi-hop knowledge graph question answering. *Information Sciences*. 2023; 619: 745-761. DOI: 10.1016/j.ins.2022.11.042.
  - [15] **Bai Y, Lv X, Li J, Hou L, Qu Y, Dai Z, Xiong F.** SQUIRE: A Sequence-to-sequence Framework for Multi-hop Knowledge Graph Reasoning. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 2022. P.1649-1662. DOI: 10.18653/v1/2022.emnlp-main.107.
  - [16] **Zou J, Wan J, Zhang H, Zhang Y.** Multi-hop Path Query Answering Model for Knowledge Graph based on Neighborhood Aggregation and Transformer. *Journal of Physics: Conference Series*. 2023; 2560: 1-15. DOI: 10.1088/1742-6596/2560/1/012049.
  - [17] **Caballero M.** A Brief Survey of Question Answering Systems. *International Journal of Artificial Intelligence & Applications (IJAIA)*. 2022; 12(5): 1-7.
  - [18] **Kim B, Hong T, Ko Y, Seo J.** Multi-Task Learning for Knowledge Graph Completion with Pre-trained Language Models. *Proceedings of the 28th International Conference on Computational Linguistics*. 2020. P.1737-1743. DOI: 10.18653/v1/2020.coling-main.153.
- 

## About the authors

**Aleksei Andreevich Golovin** (b. 1999) graduated from the Master's program at the ETU «LETI». Postgraduate student of the Department of Software Engineering and Computer Applications at ETU «LETI». ORCID: 0009-0004-4687-8393; Author ID (RSCI): 1214796; Author ID (Scopus): 57442656600; [golovin99@icloud.com](mailto:golovin99@icloud.com). ✉

**Nataly Aleksandrovna Zhukova** (b. 1983) graduated from the ETU «LETI» in 2006, Ph.D (2020). Associate Professor at the ETU. Leading researcher at SPCRAS. She is a co-author of more than 100 publications in the field of data processing and analysis. ORCID: 0000-0001-5877-4461; Author ID (Scopus): 56406142300. [nazhukova@mail.ru](mailto:nazhukova@mail.ru).

---

Received October 10, 2024. Revised November 21, 2024. Accepted December 23, 2024.

---