

ОБОГАЩЕНИЕ КОНТЕКСТА ВОПРОСОВ ЗНАНИЯМИ ИЗ CONCEPTNET ДЛЯ  
УЛУЧШЕНИЯ ТОЧНОСТИ ОТВЕТОВ

Смирнов Денис Максимович ([dmsmirnov@hse.ru](mailto:dmsmirnov@hse.ru))

Ильвовский Дмитрий Алексеевич ([dilvovsky@hse.ru](mailto:dilvovsky@hse.ru))

НИУ ВШЭ, Москва, Россия

ENRICHING CONTEXTS OF QUESTIONS WITH CONCEPTNET KNOWLEDGE FOR  
IMPROVING ACCURACY OF ANSWERS

**Abstract:** Modern question answering models can achieve near-human accuracy of answers for factual questions about a given piece of text in English. In the meantime, such models fail to achieve the same performance on datasets of question, which require some background information, not presented in the question context. This paper describes experimental evaluation of simple question context enrichment method based on collecting ConceptNet relations and proposes further direction of work in creating a question answering dataset for Russian language.

**Keywords:** question answering, commonsense knowledge, ConceptNet, BERT

## 1. ВВЕДЕНИЕ

Современные модели, нацеленные на решение задачи поиска ответов на вопросы (question answering), достаточно хорошо подходят для фактических вопросов, когда ответ четко сформулирован в тексте, но они не достигают такой же высокой точности ответов на вопросы, требующие логического вывода или внеконтекстных знаний (commonsense knowledge). Внеконтекстные знания можно неформально определить, как базовые знания об объектах и событиях, встречаемых в повседневной жизни, разделяемые большинством людей. Как правило, такие знания очень редко записываются в явном виде, и потому они часто не попадают в контекст – фрагмент текста, по которому задаётся вопрос. Вопросы, опирающиеся на внеконтекстные знания, обычно просты для людей, но могут быть проблематичным для алгоритма, потому что ответ не может быть получен без внешнего знания о семантических отношениях сущностей, описанных в тексте, поданном на вход алгоритму.

Как и во многих других задачах, связанных с обработкой естественного языка, ключевую роль в создании качественных моделей играют данные для обучения. В случае с задачей поиска ответов на вопросы, требующие внеконтекстных знаний, эта проблема актуальна вдвойне, поскольку нужны как наборы таких вопросов с ответами, так и источники внеконтекстных знаний для модели. Источником внеконтекстных знаний могут быть семантические сети. Примером такой сети является ConceptNet (Robyn Speer et al., 2017). ConceptNet представляет с собой направленный граф, где каждая вершина – это слово или фраза на естественном языке (концепт), а рёбра соответствуют отношениям между концептами. Знания в ConceptNet собраны из множества различных источников, среди которых значатся DBPedia, онтологии OpenCyc, словарь Wiktionary, данные проекта Open Mind Common Sense и лексическая сеть WordNet. Всего в ConceptNet есть 34 вида отношений, некоторые из которых симметричны, такие как *"/r/Antonym"*, показывающее, что концепты являются антонимами или *"/r/DistinctFrom"*, обозначающее, что концепты принадлежат различным множествам, например, это отношение можно обнаружить у концептов *"animal"* и *"person"*. Большая часть отношений однонаправленны, например, *"/r/HasProperty"*, показывающее, что один концепт является свойством другого, или *"Causes"*, используемое для отображения причинно-следственной связи между концептами. ConceptNet поддерживает множество языков, а концепты на разных языках являются различными вершинами графа, связанными между собой отношениями. В данный момент, в ConceptNet есть более 1.8 миллиона английских концептов и более 680 тысяч русских.

Одним из примеров датасета, который содержит вопросы, требующие от алгоритма внеконтекстных знаний, является CommonsenseQA (Talmor et al., 2019), который был собран благодаря масштабируемому подходу для генерации вопросов, разработанному авторами статьи. Представленный набор данных, состоит из более чем 12000 вопросов с множественным выбором из пяти вариантов ответа. Подход для генерации вопросов основывается на применении труда краудсорсерсеров, которые опираются на собственные внеконтекстные знания при составлении вопросов. В качестве исходной информации краудсорсеру предоставлялся подграф, извлечённый из ConceptNet, который содержит целевой концепт и три концепта, которые связаны с целевым одинаковым отношением. Например, это может быть концепт *"river"* и связанные с ним отношением *"/r/AtLocation"* концепты *"waterfall"*, *"bridge"* и *"valley"*. От краудсорсера требовалось написать три вопроса, содержащих в тексте целевой концепт (*"river"*), и имеющих в качестве правильного ответа один из связанных концептов, по одному вопросу на каждый концепт. Пример составленного вопроса для целевого концепта *"river"* и правильного варианта ответа *"bridge"* – *"Where can I stand on a river to see water falling without getting wet?"*. Два незадействованных в этом вопросе концепта *"waterfall"* и *"valley"* используются как неправильные варианты ответа. В последствии, к вопросу добавлялось ещё два других неправильных варианта, чтобы сделать проблему более сложной.

## 2. ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

### 2.1 Постановка эксперимента

В рамках этой работы исследуется влияние добавления внеконтекстных знаний к тексту вопроса (контексту) на примере набора данных CommonsenseQA, описанного в предыдущей секции. Поскольку авторы не опубликовали ответов для тестовой выборки, то в данном исследовании используются только данные обучающей и валидационной выборки.

В качестве внеконтекстных знаний используются извлеченные из ConceptNet отношения между основным концептом вопроса, который изначально указан в исходных данных, и каждым из вариантов ответа. На основе найденных отношений формируются предложения на естественном языке, которые вставляются перед началом текста вопроса. В эксперименте используется одна базовая модель – BERT-base (Devlin et al., 2019), которая независимо обучается на двух наборах одних и тех же вопросов. Первый набор – вопросы из CommonsenseQA без изменений, а второй – вопросы, тексты которых обогащены внеконтекстными знаниями. Настраиваемые параметры обучения, такие как используемый токенизатор, размер батча, параметры оптимизатора и количество эпох обучения, одинаковые для этих двух процедур обучения.

### 2.2 Сбор данных об отношениях

Обучающая выборка исходного датасета содержит 9741 вопрос, а валидационная 1221. Вопросы посвящены 2200 уникальным концептам. Вместе с вариантами ответов эти концепты образуют 26719 уникальных пар, между которыми было проверено наличие отношений в ConceptNet. Отношения в ConceptNet являются направленными рёбрами, так что порядок концептов в паре имеет значение, однако ввиду вычислительной сложности задачи по сбору отношений перестановки концептов в паре не производились. В рассмотренных парах концепт вопроса всегда стоит на первом месте, а вариант ответа на вопрос на втором.

Всего в ConceptNet используется 34 вида отношений, однако отношение *"/r/ExternalURL"*, показывающее, что за пределами ConceptNet можно найти дополнительные связанные данные об этом концепте, явно нерелевантно исследуемой задаче, поэтому оно было исключено из списка проверяемых отношений. Наличие остальных 33 отношений было проверено для каждой из 26719 уникальных пар, что потребовало совершить более 880 000 вызовов API ConceptNet.

Проверка наличия отношения происходила путём вызова API ConceptNet по URL, который был сформирован по следующему шаблону:

*http://api.conceptnet.io/a/[<Идентификатор отношения>,<Идентификатор целевого концепта вопроса>,<Идентификатор концепта варианта ответа>]*

Идентификатор концепта, помимо самого слова, включает в себя код языка (для данного датасета он всегда "en") и опциональный тег части речи, который в рамках данного исследования не подставлялся. Концепт может состоять из нескольких слов, в таком случае в идентификаторе используется нижнее подчеркивание вместо пробела.

Следуя этим правилам формирования URL, на примере пары концептов *"house"* и *"living room"* для отношения *"/r/HasA/"* мы получим следующий URL: *"http://api.conceptnet.io/a/[r/HasA/,c/en/house/,c/en/living\_room/]"*.

По итогам сбора данных через API ConceptNet было выявлено 18272 отношения между концептами вопросов и вариантами ответов. Наиболее часто встречаемыми

отношениями оказались *"/r/AtLocation"* (7318 раз), *"/r/Causes"* (2417 раз) и *"/r/CapableOf"* (1658 раз).

Для 9424 из 9741 вопросов из обучающей выборки и для 1190 из 1221 вопросов валидационной выборки было найдено хотя бы одно отношение. Те вопросы, для которых не было найдено ни одного отношения, были отброшены и не использовались в последующих этапах данного исследования. В среднем, для оставшихся вопросов было найдено 4.5 отношения, а максимальное количество найденных связей – 13.

Большая часть найденных отношений имеет в своем определении поле *"surfaceText"*, в котором содержится его представление на естественном языке. Этот текст был использован как дополнение к вопросу после удаления двойных квадратных скобок вокруг названия концепта. 81 отношение не содержало такого описания, так что для них вместо *"surfaceText"* использовался шаблон из определения отношения в документации ConceptNet, в который были подставлены названия концептов.

Полученные тексты были добавлены перед текстом исходного вопроса. Ниже приведено несколько примеров обогащения контекста вопросов.

*Пример 1.*

**Концепт вопроса:** working

**Добавленная часть:** Something that might happen as a consequence of working is being paid. Something that might happen as a consequence of working is frustration. Sometimes working causes blisters. Something that might happen as a consequence of working is you make money.

**Исходный текст:** Why would you not want to be working with wood?

**Варианты ответа:** A. blisters B. frustration C. being paid D. make money E. money

*Пример 2.*

**Концепт вопроса:** chef

**Добавленная часть:** A chef can prepare food. A chef can thin a sauce. A chef can study french cooking. A chef can cook dinner.

**Исходный текст:** What does a chef do when working in the evening?

**Варианты ответа:** A. cook dinner B. study french cooking C. taste sauce" D. prepare food E. thin sauce

*Пример 3.*

**Концепт вопроса:** killing

**Добавленная часть:** Killing is evil. Killing is wrong. The effect of killing is the death of something. Killing requires the death of something. Killing is bad.

**Исходный текст:** After the guilty verdict in the killing the judge gave a speech, he told the murderer he was pure what?

**Варианты ответа:** A. wrong B. death C. bad D. negative E. evil

## 2.3 Подготовка данных и обучение моделей

В качестве реализации модели использовался класс *BertForMultipleChoice* из библиотеки Hugging Face Transformers (Thomas Wolf et al., 2019). Каждый вопрос преобразовывается в трехмерный тензор. Первое измерение содержит непосредственно текст вопроса и текст вариантов ответа в виде матрицы индексов токенов из стандартного словаря BERT. Чтобы получить эту матрицу, текст вопроса соединяется с вариантами ответа с помощью специальных токенов, как показано на рисунке 1, и подается на вход токенизатору. В конце каждой строки добавляются [PAD] токены до фиксированной длины, которая составляла 80 токенов для небогатых вопросов и 256 для вопросов с добавленными отношениями. Два других измерения - это input attention masks, первая указывает вариант ответа, а вторая - все значащие токены. На выходе модель возвращает вектор из пяти элементов, к которому применяется softmax. Индекс максимального элемента соответствует предсказанному ответу на вопрос.

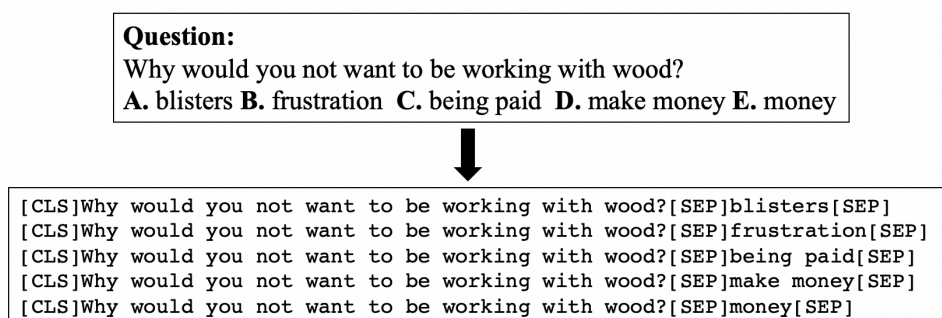


Рисунок 1.

Обучение проводилось в Google Colab на GPU. Ввиду ограничения объема памяти видеокарты, веса модели сохранялись в вещественные переменные половинной точности, а размер батча составил всего 8 вопросов. В качестве оптимизатора использовался Adam. Обучение модели в обоих случаях проводилось в течение 10 эпох без использования ранней остановки.

## 2.4 Результаты эксперимента

График изменения значения функции потерь в процессе обучения показан на рисунке 2. В обоих экспериментах уже после второй эпохи можно наблюдать серьезное расхождение значений на обучающей и валидационной выборках, что может свидетельствовать о переобучении модели.

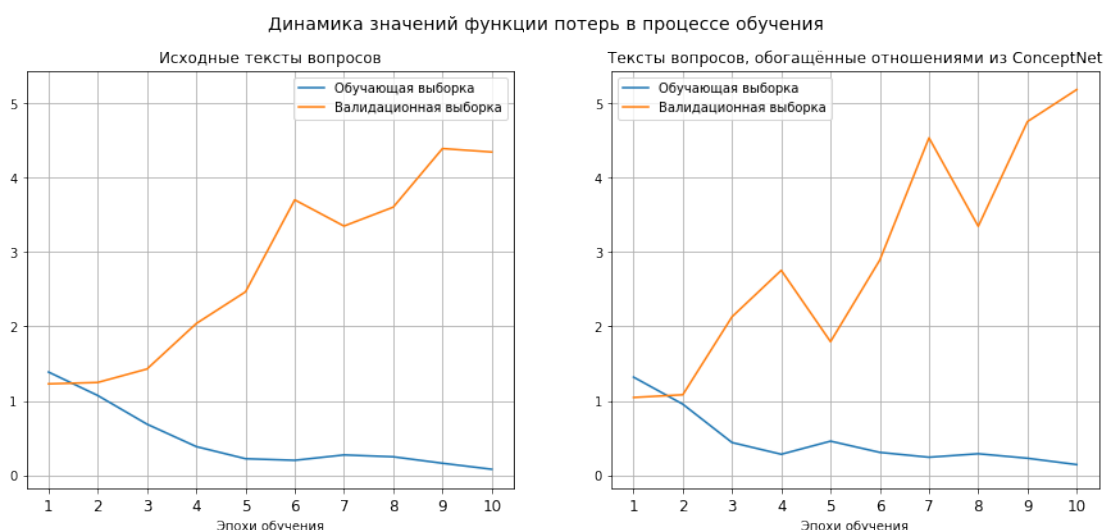


Рисунок 2.

Как видно из графика доли правильных ответов (точности) на валидационной выборке (рисунок 3), модель, обучаемая на обогащенных текстах в течение всего обучения стабильно выдавала более высокий показатель точности: 0.5822 против 0.5176 после первой эпохи и 0.5707 против 0.5299 после второй. На основе этих данных можно предположить, что обогащение текстов внеконтекстной информацией из графа знаний способно улучшить качество ответов на подобные вопросы.

Динамика точности на валидационной выборке в процессе обучения

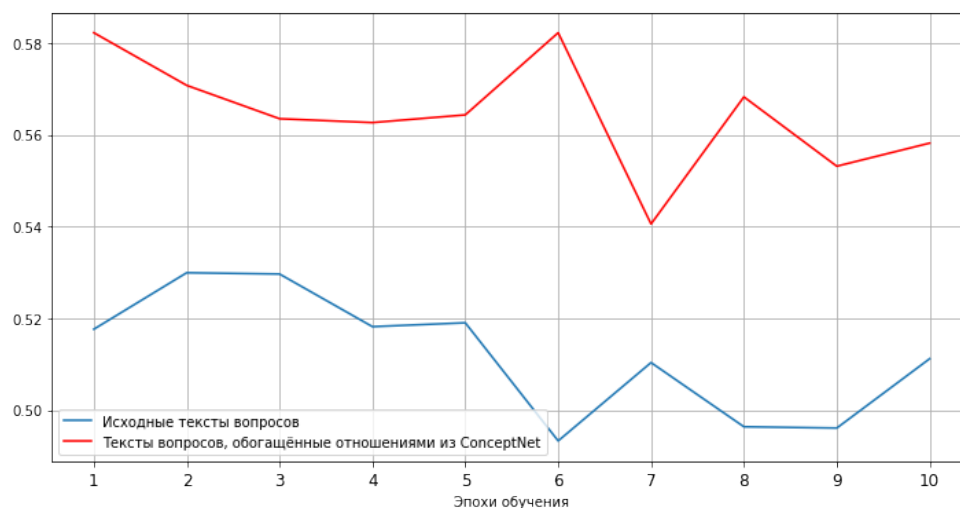


Рисунок 3.

### 3. ДАЛЬНЕЙШЕЕ РАЗВИТИЕ

В этой работе был опробован самый простой способ обогащения контекста вопроса, который не анализировал содержание вопроса. Как можно наблюдать из примеров, добавленный текст не всегда очевидным образом указывал на правильный ответ, поскольку не содержал отношений между вариантами ответа и другими концептами, упомянутыми в вопросе. Так для вопроса из примера 2 *"What does a chef do when working in the evening?"* на правильный ответ в большей степени могло натолкнуть отношение между концептами *"evening"* и *"dinner"*, чем найденные отношения для концепта *"chef"*. Так что есть возможности для улучшения достигнутых результатов.

Большая часть ресурсов, на которых можно обучать и проверять модели для различных задач из области обработки естественного языка, в том числе и для задачи поиска ответов на вопросы, требующие внеконтекстных знаний, создается на английском языке. Подобных наборов данных на русском языке найдено не было, а значит, нет возможности в полной мере использовать самые последние достижения в области обработки естественного языка. Можно возразить, что глубокие предобученные языковые модели уже сразу предобучены для решения задач на многих языках и, например, BERT номинально поддерживает русский язык. Но на примере других языков авторы показали, что моноязычные модели превосходят многоязычные версии (Devlin et al., 2019). Многоязычная модель может быть хорошей инициализацией для дообучения модели под конкретную задачу на другом языке (Kuratov and Arkhipov, 2019), однако для этого так или иначе потребуется иметь данные на этом языке.

Первым шагом к созданию русскоязычного набора вопросов, требующих внеконтекстных знаний, может быть автоматический перевод англоязычных ресурсов. Поскольку ConceptNet хорошо поддерживает русский язык, есть возможность повторить описанный в данной статье эксперимент, используя автоматические переводы оригинальных вопросов на русский язык. В настоящее время ведётся работа, по сбору данных о наличии отношений между концептами на русском языке.

## Библиография

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. 2019. CommonsenseQA: A question answering challenge targeting commonsense knowledge. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, USA.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA. pages 4444–4451.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface’s transformers: State-of-the-art natural language processing. ArXiv abs/1910.03771.

Yuri Kuratov and Mikhail Arkhipov. 2019. Adaptation of deep bidirectional multilingual transformers for russian language. ArXiv abs/1905.07213.