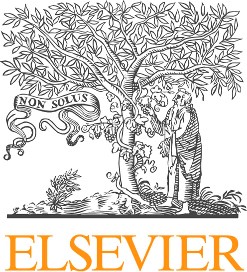
[Искусственный интеллект в медицине 102 (2020) 101767](https://doi.org/10.1016/j.artmed.2019.101767)



Списки содержимого доступны по адресу

НаукаДирец

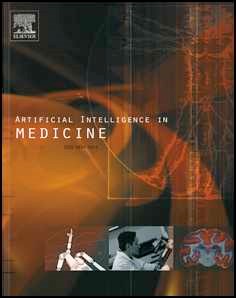
[t](http://www.sciencedirect.com/science/journal/09333657)

Искусственный интеллект в медицине

домашняя страница журнала:

www.elsevier.com/locate/artme

[d](https://www.elsevier.com/locate/artmed)



SemBioNLQA: Семантическая биомедицинская система ответов на вопросы для получения точных и идеальных ответов на вопросы естественного языка



[T](http://crossmark.crossref.org/dialog/?doi=10.1016/j.artmed.2019.101767&domain=pdf)

Мурад Саррутия,\*, Саид Уатик Эль Алауиб,c

a *Национальный центр биомедицинских коммуникаций Листер Хилл, Национальная медицинская библиотека США, Национальные институты здравоохранения США, Bethesda, MD* b *Национальная школа прикладных наук, Университет Ибн Тофаил, Кенитра, Марокко* c *Лаборатория информатики и моделирования, FSDM, Университет Сиди Мохаммеда Бен Абдалла, Фес, Марокко*

А Р Т И К Л Е И Н Ф ОА Б С Т Р А С Т

|  |  |
| --- | --- |
| *Ключевые слова:*  Ответы на биомедицинские вопросы  Поиск информации  Извлечение прохода  Обработка естественного языка  Обучение машины  Биомедицинская информатика  БиоАСК | *Предыстория и цель:* Ответы на вопросы (QA), выявление коротких точных ответов на вопросы пользователей, написанные на естественном языке, является давней проблемой, широко изучаемой в течение последних десятилетий в открытой области. Тем не менее, это по-прежнему остается реальной проблемой в биомедицинской области, поскольку большинство существующих систем поддерживают ограниченное количество типов вопросов и ответов, а также по-прежнему требуют дальнейших усилий для повышения их производительности с точки зрения точности поддерживаемых вопросов. Здесь мы представляем семантическую биомедицинскую систему контроля качества под названием SemBioNLQA, которая имеет возможность обрабатывать виды «да/ нет», «фактоид», «список» и «резюме» вопросов на естественном языке.  *Методы:* В данной работе описывается архитектура системы и оценка разработанной сквозной биомедицинской системы контроля качества под названием SemBioNLQA, которая состоит из модулей классификации вопросов, поиска документов, извлечения пассажей и извлечения ответов. Он принимает вопросы естественного языка в качестве входных данных и выводит как короткие точные ответы, так и резюме в качестве результатов. Система SemBioNLQA, работающая с четырьмя типами вопросов, основана на (1) созданных вручную лексико-синтаксических шаблонах и алгоритме машинного обучения для классификации вопросов, (2) сходстве PubMed и UMLS для извлечения документов, (3) модели BM25, стволовых словах и концепциях UMLS для извлечения пассажей и (4) метатезавре UMLS, синонимах BioPortal, анализе настроений и метрике частоты терминов для извлечения ответов.  *Результаты и выводы:* По сравнению с современными биомедицинскими системами контроля качества, SemBioNLQA, полностью автоматизированная система, имеет потенциал для работы с большим количеством типов вопросов и ответов. SemBioNLQA быстро извлекает информационные потребности пользователей, возвращая точные ответы (например, «да», «нет», название биомедицинской сущности и т. Д.) И идеальные ответы (т. Е. Резюме соответствующей информации размером с абзац) на вопросы «да/ нет», «фактоид» и «список», тогда как он предоставляет только идеальные ответы на краткие вопросы. Кроме того, экспериментальные оценки, проведенные по биомедицинским вопросам и ответам, предоставленным в рамках конкурса BioASQ, особенно в 2015, 2016 и 2017 годах (в рамках нашего участия), показывают, что SemBioNLQA достигает хороших результатов по сравнению с самыми современными современными системами и позволяет практическую и конкурентоспособную альтернативу, чтобы помочь искателям информации найти точные и идеальные ответы на свои биомедицинские вопросы. Исходный код SemBioNLQA находится в открытом доступе по адресу [https://github.com/sarrouti/sembionlqa.](https://github.com/sarrouti/sembionlqa) |

# Знакомство

Получение высококачественных кратких ответов на заданный вопрос на естественном языке из постоянно растущего объема биомедицинской литературы является ключом к созданию высококачественных систематических обзоров, которые поддерживают практику доказательной медицины [1–3] и улучшают качество ухода за пациентами [4,5]. Однако взрывной рост количества научной литературы в биомедицинской области затруднил усвоение всей соответствующей информации даже для экспертов в интересующей их области.

|  |  |
| --- | --- |
| |  | | --- | | ⁎ Автор-корреспондент. Тел: +1 (301) 518-1033.  *Адрес электронной почты:* sarrouti.mourad@gmail.com (М. Саррути).  <https://doi.org/10.1016/j.artmed.2019.101767> |   Получено 4 мая 2018 года; Получено в пересмотренном виде 19 ноября 2019 года; Принято 19 ноября 2019  0933-3657/ Опубликовано Elsevier B.V. |

Поэтому, чтобы предоставить искателям информации соответствующие ответы, необходимы более сложные технологии [6,7] и все чаще исследуются автоматические методы биомедицинского анализа текста [8,2,9,6,10,11]. Недавние доклады, такие как доклад Христовского и др. [12] подчеркнули, что наиболее широко используются специализированные системы поиска информации (IR), такие как PubMed. Однако в нынешних условиях набор извлеченных документов, потенциально релевантных, представляет собой размер ответа, который все еще слишком велик, чтобы легко идентифицировать точную информацию. Более того, в классических ИК-системах пользователям часто приходится сталкиваться с бременем изучения и фильтрации возвращенных цитат своих запросов, чтобы найти точную информацию, которую они искали. В этом контексте оценочное исследование, представленное Эли и его коллегами [13], показало, что врачи тратили в среднем менее двух минут на поиск информации для ответа на клинические вопросы, хотя многие из их вопросов остались без ответа. В другом исследовании Hersh et al. [14], в среднем студентам-медикам и медсестрам требуется не менее 30 минут, чтобы ответить на клинические вопросы с помощью MEDLINE.

В отличие от IR, системы ответов на вопросы (QA) имеют потенциал для преодоления вышеупомянутых недостатков. Они нацелены на непосредственное создание и предоставление точных ответов, а не целых документов на вопросы пользователей, путем автоматического анализа тысяч статей, в идеале менее чем за несколько секунд. Такие системы, которые могут помочь пользователям быстро находить полезную информацию, нуждаются в лингвистической и семантической обработке как вопросов пользователей, так и источников данных для извлечения соответствующей информации. Например, для биомедицинского вопроса «Как называется анорексия у гимнасток?», система контроля качества предоставит в качестве ответа конкретное название, то есть Anorexia Athletica, для анорексии у гимнасток. Типичную систему контроля качества можно рассматривать как конвейер, состоящий из трех основных этапов обработки [15–18], а именно обработки вопросов, обработки документов и обработки ответов, каждый из которых должен иметь дело с конкретными проблемами и проблемами.

Хотя QA в открытой области является давней проблемой, широко изученной в течение последних десятилетий [19-22], она по-прежнему требует дальнейших усилий в биомедицинской области, чтобы позволить пользователю иметь больше шансов найти точный ответ на свой вопрос о естественном языке человека. Открытые доменные системы отвечают на вопросы из различных областей, а предметно-ориентированные системы - сосредоточены на специализированной области - используют конкретные лингвистические ресурсы, которые позволяют давать более точные ответы. Athenikos и Han [16] сообщили о следующих характеристиках биомедицинского домена QA: (1) большие текстовые корпуса, (2) очень сложная терминология, специфичная для предметной области, и (3) предметно-ориентированный формат и типология вопросов. С другой стороны, Bauer et al. [23] показали, что количество поддерживаемых типов вопросов и ответов является наиболее важными задачами и особенностями для сравнения систем контроля качества .

Несмотря на важность ответов на вопросы в биомедицинской области, до сих пор существует лишь несколько интегральных систем, подобных описанным в [15,24–26,18,27], которые могут получить ответы на биомедицинские вопросы, написанные на естественном языке. Хотя эти системы оказались довольно успешными в ответах на биомедицинские вопросы, они предоставляют ограниченное количество типов вопросов и ответов, например, некоторые из них [15,24,26] обрабатывают только вопросы определения или возвращают исключительно короткие резюме в качестве ответов на все типы вопросов, а большинство других не имеют дела с вопросами «да/ нет», которые являются одним из самых сложных типов вопросов для ответа, поскольку они ищут. для четкого ответа «да» или «нет». Кроме того, эти системы по-прежнему требуют дальнейших усилий для повышения их производительности с точки зрения точности до поддерживаемых в настоящее время типов вопросов и ответов.

С другой стороны, с момента запуска конкурса BioASQ теории и методы в области биомедицинского контроля качества продолжают развиваться, чтобы лучше удовлетворять потребности пользователей [[1]](#footnote-1)[28–32], благодаря многочисленным изданиям конкурса BioASQ. Однако эти системы страдают от ограничений. Например, в таких системах пользователи должны вручную задавать или выбирать тип вопроса для каждого заданного вопроса. Кроме того, для данного биомедицинского вопроса эти системы используют документы и отрывки, выпущенные вызовом BioASQ, чтобы получить ответ. Таким образом, биомедицинская система контроля качества, предназначенная для оказания помощи пользователям в автоматическом поиске соответствующих ответов в биомедицинской области, должна быть в состоянии автоматически определять соответствующие документы, соответствующие отрывки и типы вопросов и, следовательно, отвечать на различные типы вопросов.

В этой статье мы представляем полностью автоматизированную систему SemBioNLQA Semantic Biomedical Natural Language Question Answering - которая имеет возможность обрабатывать виды вопросов «да/ нет», фактоид

вопросы, список вопросов и краткие вопросы, которые обычно задают в биомедицинской области.

* Вопросы «да/нет»: вопросы, требующие ответа «да» или «нет».
* Фактоидные вопросы: вопросы, которые требуют определенного названия сущности (например, болезни, лекарства или гена), числа или аналогичного короткого выражения в качестве ответа.
* Список вопросов: вопросы, которые ожидают список названий сущностей (например, список названий генов, список названий лекарств), цифры или аналогичные короткие выражения в качестве ответа.
* Сводные вопросы: вопросы, на которые в качестве ответа ожидаются краткие резюме.

Целью данного исследования является описание процесса, лежащего в основе проектирования и разработки SemBioNLQA. SemBioNLQA является производным от наших ранее установленных методов в (1) классификации вопросов [33], (2) извлечении документов [34], (3) извлечении отрывков [35] и (4) системе извлечения ответов, которая была одним из победителей2в издании конкурса BioASQ 2017 года [36]. Мы разработали систему SemBioNLQA на основе интеграции этих методов и техник.

SemBioNLQA сначала принимает в качестве входных данных биомедицинский вопрос на естественном языке и включает в себя предварительную обработку вопроса, идентификацию типа вопроса и ожидаемого формата ответа на основе лексо-синтаксических шаблонов ручной работы и машины опорных векторов (SVM), а также построение запроса с использованием сущностей UMLS. Затем, основываясь на сходстве PubMed и UMLS, он извлекает документы, удовлетворяющие запросу, из базы данных MEDLINE. После этого он извлекает отрывки с самым высоким рейтингом из документов с самым высоким рейтингом, основанных на модели BM25, стволовых словах и концепциях UMLS. Наконец, он генерирует и возвращает как «точные» (в зависимости от ожидаемого ответа на каждый тип вопроса), так и «идеальные» ответы из этих отрывков на основе метатезавра UMLS, синонимов BioPortal, SENTIWORDNET и метрики частоты терминов. По сравнению с предыдущими работами, наш подход имеет следующие вклады:

1. Насколько нам известно, ни одна из предыдущих систем не может автоматически предоставить конкретные точные ответы для каждого типа вопросов.
2. SemBioNLQA, комплексная биомедицинская система контроля качества, предназначена для того, чтобы иметь возможность принимать различные вопросы на естественном языке и генерировать соответствующие ответы на естественном языке, предоставляя как точные, так и идеальные ответы.
3. SemBioNLQA предоставляет точные ответы «да» или «нет» на вопросы «да/нет», биомедицинские именованные сущности для фактоидных вопросов и список биомедицинских именованных сущностей для вопросов списка. В дополнение к точным ответам на вопросы «да/ нет», фактоидные вопросы и вопросы списка, SemBioNLQA также возвращает идеальные ответы, в то время как он получает только идеальные ответы на краткие вопросы.
4. Наконец, обширные эксперименты с эталонными наборами данных, предоставленными задачами BioASQ, показывают, что SemBioNLQA является более эффективным и конкурентоспособным по сравнению с современными системами.

Остальная часть документа организована следующим образом. Во-первых, соответствующая работа, касающаяся основных систем контроля качества с особым акцентом на биомедицинскую область, рассматривается в разделе 2. Затем подробная информация об общей архитектуре системы SemBioNLQA представлена в разделе 3. Несколько комплексных экспериментов проводятся для оценки эффективности разработанной системы в разделе 4, где представлены экспериментальные условия, оценочные метрики, наборы эталонных данных, результаты и обсуждение. Заключение и футурная работа, наконец, представлены в разделе 5.

# Связанная работа

Ответы на вопросы получили большое внимание со стороны исследовательского сообщества в последние годы. Область исследований QA в открытых доменах выиграла от Конференции по поиску текста (TREC) [37], которая проводится регулярно каждый год с 1999 года. В противном случае биомедицинский контроль качества был проблемой последних нескольких лет. Не было никакого значительного прогресса в этом направлении из-за сложности естественного языка в биомедицинской области. В этом разделе мы обсудим основные интегральные системы контроля качества с особым акцентом на биомедицинскую область.

В этом контексте Lee et al. [15] разработали медицинскую систему контроля качества под названием MedQA, которая состоит из пяти компонентов, включая (1) классификацию вопросов, (2) генерацию запросов, (3) извлечение документов, (4) извлечение ответов и (5) обобщение текста. В MedQA, во-первых, компонент классификации вопросов автоматически классифицирует медицинские вопросы по категориям таксономии, созданной Ely et al. [13] на основе контролируемых подходов машинного обучения. Затем компонент извлечения документов использует термины запроса для извлечения соответствующих документов из веб-документов с помощью Google или локально индексированных корпусов MEDLINE с помощью Lucene. Затем компонент извлечения ответов определяет из полученных документов соответствующие предложения, которые отвечают на вопросы на основе лексико-синтаксических паттернов. Наконец, компонент обобщения текста удаляет избыточные предложения и сжимает предложения в связное резюме, которое рассматривается как ответ. Хотя система MedQA возвращает краткие резюме, которые потенциально могут ответить на медицинские вопросы, текущие возможности MedQA ограничены: она дает ответы только на вопросы определений.

Круше и др. [24] создали биомедицинскую систему контроля качества под названием HONQA, которая извлекает предложения из веб-сайтов, сертифицированных Health On the Net Foundation (HON), и предоставляет их в качестве ответов на биомедицинские вопросы. HONQA основан на подходе к обучению для определения типа вопроса и семантических ресурсах, таких как UMLS, для руководства системой, особенно в выборе ответов, но в публикации не представлено никаких подробностей. В своем нынешнем виде он не в состоянии дать точные ответы на другие типы вопросов, например, да/нет и фактоидные вопросы [23]..

Гобейл и др. [25] разработали биомедицинскую систему контроля качества EAGLi, целью которой является извлечение ответов на биомедицинские вопросы из документов MEDLINE. Задавая вопрос на естественном языке, EAGLi сначала анализирует вопрос, чтобы найти тип вопроса и построить запрос на основе набора шаблонов. Затем он извлекает набор соответствующих документов из MEDLINE с помощью PubMed или EasyIR. Наконец, система извлекает и вычисляет оценку для каждой из концепций, выраженных в наиболее релевантных документах, и, наконец, выводит ранжированный список ответов кандидатов. Возможности нынешнего ЕАГПи ограничены вопросами wh-типа, поскольку они охватывают только вопросы определения и фактоида.

Цао и др. [26] представили клиническую систему контроля качества под названием AskHERMES, которая возвращает краткие резюме в качестве ответов на специальные клинические вопросы. AskHERMES был разработан с помощью следующих основных этапов: (1) анализ вопросов, (2) поиск документов, (3) извлечение отрывков и (4) обобщение и представление ответов. На этапе анализа вопроса авторы сначала классифицировали клинические вопросы по общим темам (например, устройство, диагноз и т. Д.) На основе классификатора SVM, а затем определили ключевые слова, которые захватывают наиболее важное содержание вопроса, используя условные случайные поля. На этапе извлечения документов модель BM25 использовалась для извлечения соответствующих документов. После этого они извлекли отрывки-кандидаты на основе динамически генерируемых границ прохождения и оценили их на основе сходства как на уровне слов, так и на уровне последовательности слов на этапе извлечения отрывка. Наконец, ответ был сгенерирован на основе структурной кластеризации с использованием содержательных терминов. Система AskHERMES возвращает отрывки (короткие тексты), которые потенциально могут ответить на все типы клинических вопросов. Тем не менее, он возвращает большое количество результатов, что, как правило, противоречит намерению системы контроля качества уменьшить объем информации, которая должна быть прочитана. Более того, система поддерживает только один тип ответа в виде нескольких отрывков предложений для всех типов вопросов [23]..

Абача и Цвайгенбаум [18] представили семантическую медицинскую систему контроля качества под названием MEANS. MEANS состоит из трех основных этапов: (1) аннотация корпусов, (2) анализ и классификация вопросов и (3) поиск ответов. Авторы применили методы NLP, распознавание именованных сущностей и извлечение связей для построения RDF-аннотаций исходных документов и SPARQL-запросов, представляющих вопросы пользователей. Они также определили онтологию MESA для представления концепций и отношений между ними, чтобы построить SPARQL переводы вопросов естественного языка. Для извлечения ответов были выполнены запросы SPARQL для опроса триплов RDF, построенных на шаге корпус-аннотация. Несмотря на это, авторы рассмотрели четыре типа вопросов, они сосредоточились на вопросах фактоида и да / нет, поскольку для решения сложных вопросов все еще требуются более конкретные процессы (например, почему, когда).

Христовский и др. [12] внедрили биомедицинскую систему контроля качества, SemBT, основанную на семантических отношениях, извлеченных из биомедицинской литературы. SemBT состоит из трех основных этапов обработки: (1) предварительная обработка, (2) обработка вопросов и (3) обработка ответов. На этапе предварительной обработки авторы сначала извлекли семантические отношения с помощью системы обработки естественного языка SemRep из предложений, извлеченных из цитат MEDLINE, а затем сохранили их в базе данных. На этапе обработки вопроса авторы построили запрос для поиска в базе данных извлеченных семантических отношений. Наконец, на этапе обработки ответов они представили результирующие семантические отношения в виде ответов сверху вниз, сначала семантические отношения с агрегированной частотой возникновения, а затем конкретные предложения, из которых извлекаются семантические отношения. Система SemBT возвращает ответы в виде семантических отношений и отдельных предложений, из которых извлекаются семантические отношения. Однако в своей текущей реализации вопросы должны быть в виде SubjectRelation-Object, а значит, не позволяют задавать вопросы в формате естественного языка, например, вопрос на естественном языке «Какие препараты можно использовать для лечения диабета?» можно задать в SemBT как «phsu лечит диабет», где «phsu» расшифровывается как «фармакологическое вещество», а «treats» — название семантического отношения (этот пример взят из [12]).

Краус и др. [38] разработали систему Olelo для интуитивного исследования биомедицинской литературы. Система Olelo состоит из трех основных модулей: (1) обработка вопросов, (2) поиск документов/отрывков и (3) обработка ответов. В Olelo модуль обработки вопросов основан на системе, описанной в [32].. Во втором модуле Olelo сначала использует токены и соответствующие термины для формулирования запроса к базе данных таким образом, чтобы получить рефераты извлеченных документов, а затем ранжирует полученные рефераты в соответствии с вхождением и важностью искомых токенов. Наконец, ответ возвращается пользователю в зависимости от типа вопроса третьим модулем. Хотя Олело оказался довольно успешным в ответах на биомедицинские вопросы, в настоящее время Олело поддерживает только три типа вопросов, включая фактоидные, списочные и сводные вопросы. Действительно, он не поддерживает вопросы «да/ нет», которые являются одним из самых сложных вопросов для ответа, поскольку они ищут четкий ответ «да» или «нет». Более того, согласно результатам оценки Olelo [27] в конкурсе BioASQ 2017 года, в котором мы также участвовали [36], система по-прежнему требует дальнейших усилий для повышения своей производительности с точки зрения точности до поддерживаемых в настоящее время типов вопросов.

С момента запуска трека биомедицинского контроля качества на конкурсе BioASQ различные подходы были направлены на обеспечение качества в биомедицинской области. Основываясь на подходах, основанных на обучении, Yang et al. [28] описали разрушение биомедицинской системы контроля качества , которая возвращает только точные ответы на фактоидные и списочные вопросы. Они разработали два контролируемых алгоритма, которые разделяются обоими типами вопросов: (1) прогнозирование типа ответа и (2) оценка ответа кандидата. Они использовали несколько признаков (например, лемму, семантический тип каждого понятия и т. д.) и логистическую регрессию в качестве многоклассового классификатора для определения семантического типа ответа на данный вопрос. С другой стороны, авторы рассматривали концепции как ответы-кандидаты и оценивали их на основе метода умножения, чтобы объединить оценку принуждения типа и количество случаев, а также логистическую регрессию для изучения функции оценки. Чжан и др. [29] описали биомедицинскую систему контроля качества , которая извлекает исключительно точные ответы на фактоидные и списочные вопросы. Система сначала использовала PubTator [39] для генерации ответов-кандидатов, а затем ранжировала их, используя метрики частоты терминов. Choi [30], с другой стороны, разработала биомедицинскую систему контроля качества, основанную на ключевых словах, которая извлекает только идеальные ответы на биомедицинские вопросы. Напротив, Neves [31] предложил биомедицинскую систему контроля качества, основанную на базе данных памяти (IMDB) и ее встроенных функциях анализа текста, чтобы сформулировать как точные, так и идеальные ответы на биомедицинские вопросы. Он состоит из трех основных этапов обработки: (1) обработка вопросов, (2) обработка документов и пассажей и (3) обработка ответов. Сначала они построили запрос к заданному вопросу на этапе обработки вопроса. Затем они извлекли до 200 документов с самым высоким рейтингом и извлекли заголовок и аннотацию для каждого PMID с помощью поисковой системы PubMed и веб-сервисов BioASQ соответственно. Затем они извлекли отрывки-кандидаты на основе встроенных функций поиска информации, доступных в IMDB, которые основаны на приближенном сходстве строк, чтобы сопоставить термины из запроса со словами в документах, и ранжировали их с помощью метрик TFIDF. Они имели дело с четырьмя типами вопросов, то есть вопросами «да/нет», фактоидными вопросами, вопросами списка и краткими вопросами. Действительно, решение о ответах «да» или «нет» было основано на прогнозах анализа настроений, предоставленных IMDB. Они извлекли фактоидные и списочные ответы, основанные также на встроенных прогнозах, предоставленных IMDB, более конкретно, на аннотациях существительных фраз и тем. Для кратких вопросов они построили резюме для идеальных ответов, основанных на фразах, которые содержат чувства. Шульце и др. [32] представили приложение к биомедицинскому обеспечению качества, которое касалось исключительно идеальных ответов. Последний основан на алгоритме LexRank [40], но который использовал исключительно именованные сущности для функции подобия.

Хотя разработка систем контроля качества представляет собой прогресс с момента запуска задачи BioASQ, системы, тем не менее, страдают от ограничений. Многие из них были ограничены очень конкретными контекстами и охватывали ограниченный тип вопросов и ответов. Более того, большинство из этих систем не являются сквозными системами контроля качества.

В этой работе наша цель состоит в том, чтобы выйти за рамки предыдущих биомедицинских систем контроля качества и разработать систему контроля качества с возможностью автоматической обработки большого количества типов вопросов, включая вопросы «да/ нет», фактоидные вопросы, вопросы списка и краткие вопросы, которые обычно задаются в биомедицинской области [41]. Кроме того, по сравнению со сквозными системами контроля качества, такими как Olelo, HONQA, AskHERMES, EAGLi, MEANS и MedQA, наша система не только способна принимать различные типы вопросов, но и способна дать конкретные ответы на каждый тип вопросов. В таблице 1 приведены размеры и сравнения вышеупомянутых сквозных систем контроля качества и SemBioNLQA.

# Методика

Inthispaper,weproposeabiomedicalQAsystembasedona

**Таблица 1**

семантический подход с использованием документов PubMed. Система имеет возможность работать с четырьмя типами биомедицинских вопросов: вопросы «да/нет», фактоидные вопросы, вопросы-списки и сводные вопросы [41]. На фиг.1 показана архитектура SemBioNLQA и ее основные компоненты, которыми являются: (1) классификация вопросов и переформулировка запросов, (2) извлечение документов, (3) извлечение отрывков и (4) извлечение ответов.

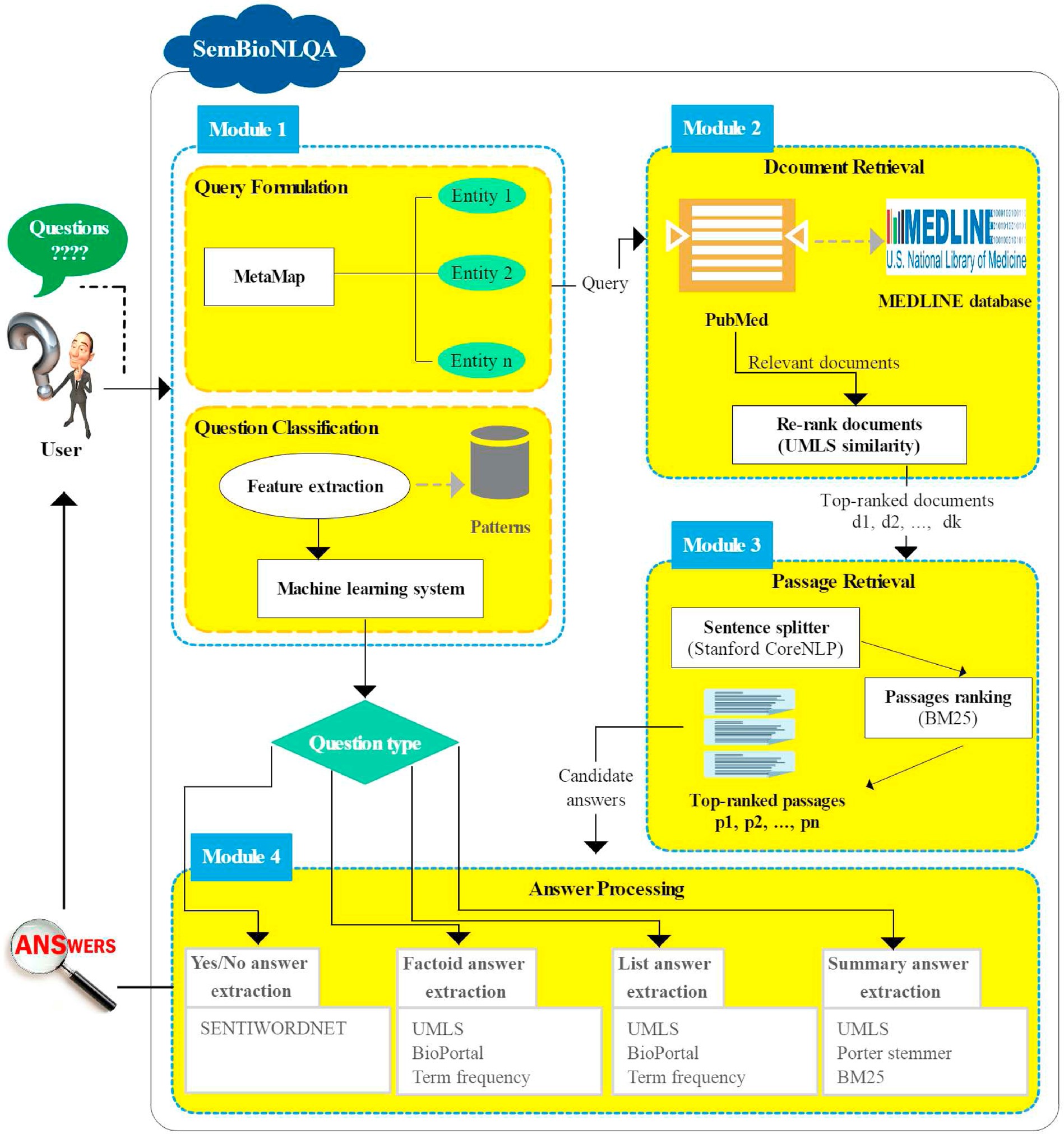
Первый компонент получает входные данные, введенные пользователем, т.е. вопрос на естественном языке (например, что, почему, где и т.д.), и включает в себя предварительную обработку вопроса, идентификацию типа вопроса и ожидаемого формата ответа, который должен потребоваться, а также построение запроса, который является входом для извлечения документа, второй компонент. Система поиска документов используется для извлечения документов, удовлетворяющих запросу. После этого отрывки с самым высоким рейтингом извлекаются из документов с самым высоким рейтингом путем поиска отрывков, третьего компонента. Вывод этого компонента представляет собой набор отрывков с самым высоким рейтингом, который используется в качестве набора ответов-кандидатов и в качестве входных данных для последнего компонента, извлечения ответа. На этом этапе ответы кандидатов сопоставляются с ожидаемым типом ответа, генерируемым первым компонентом, и ранжируются по тому, насколько хорошо они удовлетворяют вопросу пользователя, используя соответствующий алгоритм извлечения ответов. Наконец, пользователю показываются ответы кандидатов с самым высоким рейтингом и необработанные тексты, из которых были извлечены ответы. В следующих разделах (см. разделы 3.1, 3.2, 3.3 и 3.4) мы предоставим подробное описание каждого компонента.

## Классификация вопросов и переформулировка запросов

Для данного биомедицинского вопроса задача классификации вопросов в SemBioNLQA направлена на определение типа вопроса и, следовательно, определение ожидаемого формата ответа, чтобы увидеть, должен ли ответ быть названием биомедицинской сущности, кратким резюме, «да» или «нет» и т. Д. Формат ожидаемого ответа позволяет использовать алгоритмы извлечения ответов для конкретного типа и отклонять возможные ответы неправильного типа [42–46].. Соответственно, мы классифицировали биомедицинские вопросы в одну из следующих четырех категорий, определенных вызовом BioASQ [41]: (1) да/нет, (2) фактоид, (3) список и (4) резюме. Предлагаемый метод классификации типов вопросов в SemBioNLQA, представленный и оцененный в [33], использует наш предопределенный набор лексо-синтаксических паттернов ручной работы (например, «[Что] + [VBZ] + [\*] + [определение]+ [\*] +?», где «+», VBZ и «\*» обозначают конкатенацию, любые термины и глагол 3-го лица единственного числа присутствуют соответственно) для извлечения особенностей обучения для алгоритмов машинного обучения. Мы использовали обучающие вопросы 810 BioASQ для обучения классификатора. Процесс классификации состоит из двух этапов. Для данного вопроса он сначала представлен вектором терминов. Эти термины извлекаются с использованием нашего набора лексическо-синтаксических паттернов ручной работы. Затем алгоритм машинного обучения классифицирует данный вопрос в одну из четырех вышеупомянутых категорий. Поскольку биомедицинский вопрос может быть отнесен к одному из четырех классов, была использована многоклассовая классификация. Более того, мы исследовали несколько алгоритмов обучения, таких как наивные байесы, дерево решений и машины опорных векторов (SVM), показав 4-классный SVM с линейным ядром, достигшим наилучшего результата. Результаты показали, что линейное ядро превосходит другие ядра, такие как ядро RBF, древовидное ядро и составное ядро. Тем

|  |  |  |  |
| --- | --- | --- | --- |
| Системы контроля качества | Формат вопроса | Типы целевых вопросов | Типы ответов |
| Медка [15] | Естественный язык | Определение | Резюме |
| ХОНКА [24] | Естественный язык | Определение, фактоид | Предложение |
| ЕАГЛи [25] | Естественный язык | Определение, фактоид | Отрывки из нескольких фраз и список отдельных Сентити |
| АскГЕРМЕС [26] | Естественный язык | - | Отрывки из нескольких предложений |
| СРЕДСТВА [18] | Естественный язык | Определение, да/нет, фактоид | Предложение, именованная сущность, "да" или "нет" |
| СемБТ [12] | Субъект-Отношение-Объект | - | Семантические отношения, предложения |
| Олело [38] | Естественный язык | фактоид, список, резюме | Термин MeSH, список терминов Mesh, краткие резюме |
| SemBioNLQA | Естественный язык | Да/нет, фактоид, список, резюме | "Да" или "нет", сущность UMLS, список организаций UMLS, краткие резюме |

Системы сравнения характеристик между вышеупомянутыми системами и SemBioNLQA. Символ "-" означает, что система не включала классификацию типов целевых вопросов.



**Рис. 1.** Общая архитектура предлагаемой биомедицинской системы ответа SemBioNLQA.

наилучшее значение параметра C равно 1,01, которое было зафиксировано после 5-кратной перекрестной проверки. Наша оценка пяти партий аннотированных вопросов BioASQ показала, что использование многоклассового классификатора SVM и нашего набора лексо-синтаксических паттернов ручной работы приводит к высочайшей точности 89,40%.

С другой стороны, чтобы построить запрос на извлечение документов, мы сначала использовали MetaMap [47] для сопоставления терминов вопроса с метатезавром унифицированной системы медицинского языка (UMLS) для извлечения названий биомедицинских сущностей. Затем мы объединили извлеченные сущности с оператором "+". UMLS, который состоит из трех ресурсов знаний: (1) метатезавра, (2) семантической сети и (3) лексики SPECIALIST, является хранилищем биомедицинских концепций, разработанных Национальной медицинской библиотекой США (NLM). Он объединяет более 2 миллионов названий для около 900 000 понятий из более чем 60 семейств биомедицинских словарей, а также 12 миллионов отношений между этими понятиями [48].

## Извлечение документов

Поиск документов, который заключается в извлечении набора соответствующих документов, которые, вероятно, будут содержать ответ на заданный запрос, построенный предыдущим модулем [49], является одним из существенных и важных компонентов предлагаемой системы контроля качества. Правильные ответы могут быть найдены только тогда, когда набор извлеченных документов определен правильно. В этой работе подход к извлечению соответствующих документов PubMed примерно аналогичен описанному в нашей недавней системе поиска документов, представленной и оцененной в [34].. В этой задаче мы сначала отправили сконструированный запрос в поисковую систему PubMed, вызвав веб-сервис E-utilities из PubMed для извлечения соответствующих документов из базы данных MEDLINE. Эта база данных, поддерживаемая NLM США, считается авторитетным источником медицинских доказательств для медицинских работников, биомедицинских исследователей и многих других пользователей [[2]](#footnote-2)[50]. Затем мы переранжировали возвращенные документы, используя предложенный алгоритм [34], который основан на подобии UMLS4[51]. Идея состоит в том, чтобы вычислить сумму баллов семантического сходства между биомедицинскими концепциями данного вопроса и каждым названием возвращенных документов. Наконец, после повторного ранжирования возвращенных документов *документы D* с высшим рейтингом сохраняются и используются в качестве входных данных следующего компонента, извлечения пассажа. Чтобы вычислить семантическое сходство между заголовками возвращенных документов и вопросом, мы сначала сопоставили как заголовки, так и вопрос с концепциями метатезавра UMLS, чтобы извлечь концепции UMLS, на которые они ссылаются. Затем мы вычислили семантическое сходство между концепциями пар, используя сходство UMLS. Длина пути была использована в качестве меры подобия, где оценка подобия обратно пропорциональна количеству узлов вдоль кратчайшего пути между понятиями вонтологии Mesh5. Рис. На фиг.2 показан пример симантических показателей сходства между биомедицинским вопросом «Является ли туберозный склероз генетическим заболеванием?» и названием «Комплекс туберозного склероза, диагностированный по поражениям полости рта» документа PubMed (PMID=24310804).

## Извлечение прохода

Как только документы d с самым высоким рейтингом (d1, *d2*, .. ., *dd*) извлекаются для заданного вопроса, модуль извлечения отрывков извлекает и исправляет отрывки P с самым высоким рейтингом (фрагменты текстов) из извлеченных документов. Эти отрывки с самым высоким рейтингом подаются в качестве кандидатов на ответы, и система SemBioNLQA извлекает из них ответ. Несколько исследований, таких как исследование, о котором сообщалось в [52], подчеркнули, что правильный ответ на данный биомедицинский вопрос может быть найден с высокой вероятностью, когда он уже существует в одном из извлеченных отрывков. Соответственно, мы разработали новый подход к извлечению соответствующих отрывков из документов PubMed к данному биомедицинскому вопросу. В предлагаемом подходе к поиску пассажей, представленном и оцененном в [35], мы сначала направили рефераты документов *d* с самым высоким рейтингом в Stanford CoreNLP [53] разделитель предложений, чтобы разделить их на предложения. В частности, мы определяем отрывок в SemBioNLQA как одно предложение. Затем мы предварительно обработали полученный набор предложений (т.е. набор отрывков-кандидатов), включая токенизацию, удаление стоп-слов6и применение stemmer Портера [54] для извлечения стволовых слов. Кроме того, мы использовали программу MetaMap для отображения как биомедицинских вопросов, так и переходов кандидатов к концепциям UMLS, чтобы извлечь биомедицинские концепции. Кроме того, исправлена система дизамбигов смысла слова MetaMap [55] для устранения двусмысленностей в текстах путем выявления значения неоднозначных терминов. Используя стволовые слова и концепции UMLS в качестве признаков, мы, наконец, ранжировали набор отрывков-кандидатов и сохранили *P* с самым высоким рейтингом, используя BM25 в качестве модели извлечения. Вывод этого компонента представляет собой набор ответов-кандидатов, т.е. отрывков P с самым высоким рейтингом (p1, p2, .., *pp*), который является входом для извлечения ответа, последнего шага. Подробности этой работы приведены в [35]..

## Извлечение ответа

|  |  |
| --- | --- |
| 1. <http://maraca.d.umn.edu/cgi-bin/umls_similarity.cgi> 2. <https://www.nlm.nih.gov/mesh/> 3. <http://www.textfixer.com/resources/common-english-words.txt> | 1. <http://participants-area.bioasq.org/results/5b/phaseB/> 2. <http://www.bioasq.org/participate/fifth-challenge-winners> 3. <http://sentiwordnet.isti.cnr.it/> |

Извлечение ответов является наиболее сложной задачей SemBioNLQA, поскольку именно тогда точный ответ должен быть извлечен из ответов кандидатов, извлеченных и выбранных компонентом извлечения пассажа. Вывод из этого модуля представляет собой короткий точный ответ на вопрос пользователя. Здесь соответствующий метод извлечения ответа, который извлекает окончательный ответ из ответов-кандидатов, выбирается в соответствии с типом вопроса, который был автоматически обнаружен модулем классификации вопросов. В частности, после получения ответов кандидата P и определения категории вопросов для данного биомедицинского

вопрос, система SemBioNLQA применяет соответствующий метод извлечения ответов для извлечения окончательных ответов. Поскольку SemBioNLQA имеет дело с четырьмя типами вопросов, то есть вопросами «да/ нет», фактоидными вопросами, вопросами списка и краткими вопросами, поэтому мы разработали новые методы представления ответов для каждого типа вопроса. Эти методы получены из нашей ранее установленной системы извлечения ответов [36], которая показала хорошие результаты в рейтинге BioASQ7 2017 года [56,57] в топ-уровнях всех команд и была одним из пятых победителей конкурса8. Предлагаемая система предоставляет точные ответы (например, «да», «нет», биомедицинская сущность и т. д.) и идеальные ответы (резюме) на вопросы «да/нет», «фактоид» и «список», в то время как она извлекает только идеальные ответы на краткие вопросы. В следующих разделах (см. разделы 3.4.1, 3.4.2, 3.4.3 и 3.4.4) мы предоставим подробное описание каждого метода извлечения ответа.

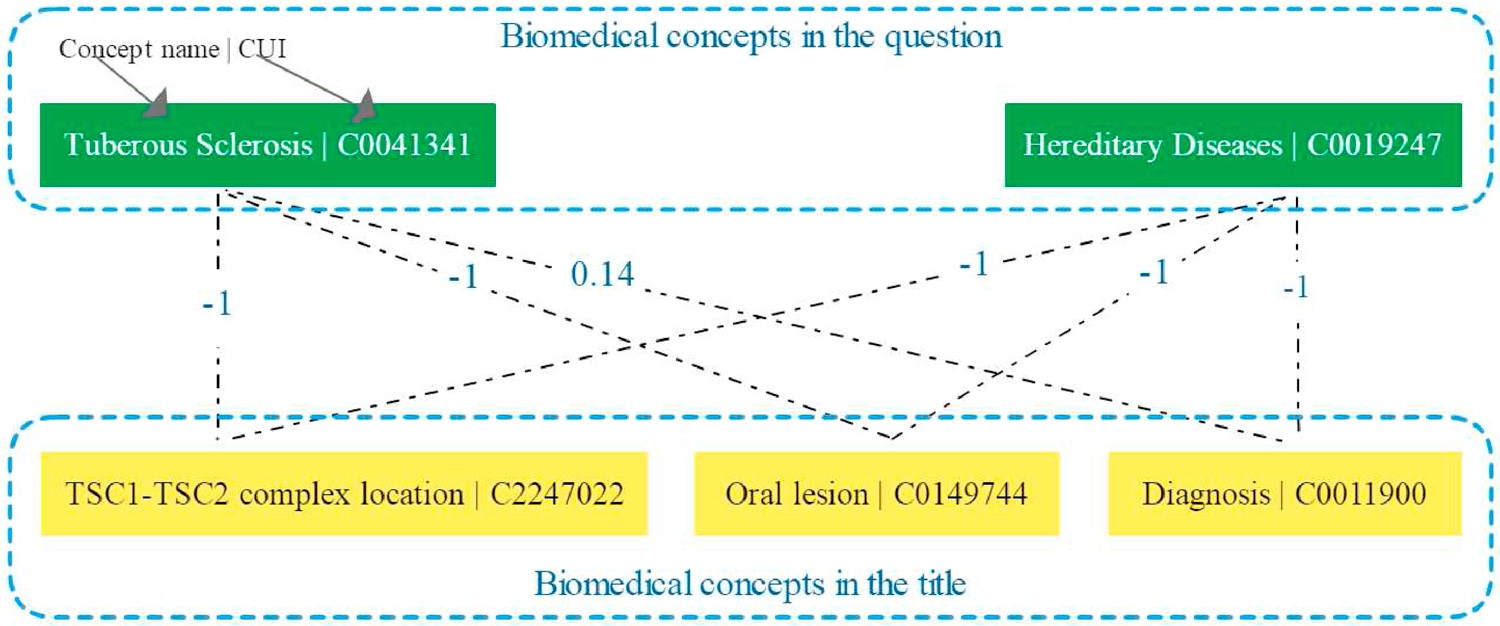
### Вопросы да/нет

После получения ответов кандидата P на заданный вопрос «да/нет» решение по ответу «да» или «нет» принимается с использованием предложенного нами метода извлечения ответа «да/нет», представленного и оцененного в [58]. Последний основан на SENTIWORDNET9: лексическом ресурсе для анализа настроений и интеллектуального анализа мнений [59].. В этой задаче сначала инструменты Stanford CoreNLP [53] использовались для токенизации и пометки части речи один за другим, *когда P* извлекал ответы кандидатов. Затем каждому слову ответа кандидата присваивается оценка SENTIWORDNET. Наконец, решение о выводе «да» или «нет» зависит от количества положительных или отрицательных ответов кандидата: «да» для положительных итоговых ответов кандидата и «нет» для отрицательного. Алгоритм 1 дополнительно иллюстрирует, как работает предложенный метод извлечения ответа «да/нет».

**Алгоритм 1** Биомедицинский генератор ответов «да/нет » [58]

|  |  |  |
| --- | --- | --- |
| 1: | **Ввод**: *да/нет вопрос Q и набор ответов кандидата P* |  |
| 2: | **Выход**: *ответ: "да" или "нет"* |  |
| 3: | *постив* ⟵0; *отрицательный* ⟵0 | ▷ количество положительных и отрицательных ответов кандидатов |
| 4: | *i* ⟵1 |  |
| 5: | **функция** Препроцессирование(*p*:*ответ condidate*) |  |
| 6: | *ТОКЕН*[1... *m*] ⟵ *ТОКЕНИЗАЦИЯИПОМЕТКИ*(p) |  |
| 7: | **вернуть** *ТОКЕН* |  |
| 8: | **конечная функция** |  |
| 9: | **do** |  |
| 10: | *В*[1. .. *m*] ⟵ *ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКАP*[*i*] | ▷ получить набор слов и их POS-теги ответа кандидата |
| 11: | *оценка* ⟵0. 0; *j* ⟵1 |  |
| 12: | **do** |  |
| 13: | *оценка* ⟵ *оценка* + *SENTIWORDNETW*[j] |  |
| 14: | *д* ⟵ *д* + 1 |  |
| 15: | **в то время как** *J* ≤ *m* | ▷ m - размер набора слов W |
| 16: | **Если** *оценка* ≥ 0, **то** |  |
| 17: | положительный ⟵ *положительный* + 1 |  |
| 18: | **еще** |  |
| 19: | отрицательный ⟵ *отрицательный* + 1 |  |
| 20: | **окончание , если** |  |
| 21: | *в* ⟵ *в* + 1 |  |
| 22: | **в то время как** *я* ≤ *np* | ▷ np - количество ответов кандидатов |
| 23: | **если** *положительный* ≥ *отрицательный,* **то** | ▷ решение для ответов «да» или «нет» основано на количестве положительных и отрицательных ответов кандидата |
| 24: | *выход* ⟵"*да*" |  |

|  |
| --- |
| **Рис. 2.** Пример оценки семантического сходства между биомедицинскими понятиями, извлеченными с помощью инструмента MetaMap. АПИ указывает уникальный идентификатор концепции UMLS. -1 |

означает, что между парными понятиями UMLS нет семантической связи.

25: **иначе**

26: *выход* ⟵"*нет*"

27: **конец, если**

### Фактоидные вопросы

Чтобы ответить на фактоидные вопросы в SemBioNLQA, мы предложили метод извлечения фактоидных ответов, основанный на метатезавре UMLS, синонимах BioPortal и метриках частоты терминов. Для этого ответы кандидата *P* сначала сопоставляются с метатезавром UMLS (источник знаний 2016AA) с помощью программы MetaMap [[3]](#footnote-3)[47], чтобы извлечь набор биомедицинских имен сущностей *Es*. Далее полученный набор биомедицинских названий сущностей ранжируется на основе терминов частотных метрик *TF*(*ei*, *Es*), количество раз, когда имя сущности *ei*появлялось в наборе биомедицинских названий сущностей Es. Мы изучили несколько методов взвешивания терминов, таких как TFIDF и BM25, показав, что частота терминов достигла наилучшего результата для этой задачи. Мы предполагаем, что ответы расположены в ответах первого и второго кандидатов. Затем синонимы для каждого из имен сущностей с самым высоким рейтингом T извлекаются с помощью веб-служб из BioPortal. Наконец, [[4]](#footnote-4)имена биомедицинских сущностей с самым высоким рейтингом T и их лучшие синонимы T отображаются в качестве ответов, исключая сущности, также упомянутые в вопросе. Идея исключения сущностей, упомянутых в вопросе, заключается в том, что после анализа обучающего набора вопросов и ответов, выпущенных организаторами BioASQ, мы обнаружили, что большинство сущностей, которые появляются в вопросах, не являются частью ответов. Например, ответ на фактоидный вопрос «Как называется ингибитор тирозинкиназы Брутона, который может быть использован для лечения хронического лимфоцитарного лейкоза?», который содержит несколько сущностей (например, «Хронический лимфоцитарный лейкоз»), — это «Ибрутиниб», который не является частью вопросных сущностей. Как описано в конкурсе BioASQ, фактоидный вопрос имеет один правильный ответ, но допускается до пяти ответов-кандидатов и их синонимов. Инжир. На фиг.3 приведен пример всего процесса, т.е. сопоставление с метатезавром UMLS и извлечение синонимов для кандидата на фактоидный вопрос «Для какого типа рака легких используется афатиниб?».

### Список вопросов

Предлагаемый метод, используемый для ответа на вопросы списка в SemBioNLQA, аналогичен методу, описанному для фактоидных вопросов. Как показывает вызов BioASQ, основное различие между фактоидными и списковыми вопросами заключается в том, что первые требуют единого списка ответов, в то время как вторые ожидают списка списков имен сущностей, номеров или аналогичных коротких выражений. Каждая сущность может сопровождаться списком синонимов. Поэтому предлагаемая биомедицинская система контроля качества должна предоставлять список списков для каждого вопроса. Другими словами, точный ответ одинаков для фактоидных вопросов, но интерпретация отличается для вопросов списка: все сущности с самым высоким рейтингом считаются частью одного и того же ответа на вопрос списка, а не как кандидаты.

### Краткие вопросы

Поскольку краткие вопросы не имеют точных ответов, на них просто отвечают в SemBioNLQA путем формулирования кратких резюме (т. Е. Идеальных ответов) соответствующей информации. В этой работе идеальные ответы на вопросы формируются путем объединения двух отрывков с самым высоким рейтингом, которые были извлечены с помощью предложенного подхода к поиску пассажей, основанного на длине предложения Stanford CoreNLP как длине прохода, стволовых словах и концепциях UMLS в качестве признаков для модели BM25. Сначала мы направили рефераты из *N* соответствующих документов по данному биомедицинскому вопросу в Стэнфордский разделитель предложений CoreNLP, чтобы разделить их на предложения. Затем мы предварительно обработали полученный набор отрывков из ответов кандидатов, включая токенизацию, удаление стоп-слов и применение stemmer Портера для извлечения стволовых слов. Кроме того, мы использовали программу MetaMap для отображения как биомедицинских вопросов, так и переходов кандидатов к концепциям UMLS, чтобы извлечь биомедицинские концепции. Кроме того, система дизамбигов смысла слова MetaMap была использована для устранения двусмысленностей в текстах путем определения значения неоднозначных терминов. Используя стволовые слова и концепции UMLS в качестве признаков, мы, наконец, ранжировали набор отрывков ответов кандидатов, используя BM25 в качестве модели извлечения, и объединили два отрывка из ответов кандидатов на первое место.

В частности, в дополнение к точным ответам, возвращенным на предыдущие вопросы (т.е. да/нет, фактоид и список), система также предоставляет идеальные ответы. Таким образом, SemBioNLQA предоставляет как точные ответы, так и идеальные ответы на вопросы «да/нет», «фактоид» и «список», тогда как он дает только идеальные ответы на краткие вопросы.

# Экспериментальные результаты и обсуждение

В этом разделе мы сначала сообщаем о систематической оценке системы SemBioNLQA на контрольных наборах данных, предоставленных конкурсом BioASQ, чтобы сравнить с системой Olelo [38], которая, к счастью, оценивается на конкурсе BioASQ 2017 года [27], в котором мы также участвовали [36]. Затем мы представляем ручную оценку с использованием обучающих вопросов и ответов BioASQ для сравнения с системами AskHERMES [26], EAGLi [25], поскольку, насколько нам известно, они не были недавно оценены по комплексным биомедицинским критериям качества, таким как те, которые предоставляются BioASQ. Наконец, мы обсуждаем сравнение SemBioNLQA с этими системами.

|  |
| --- |
| **Рис. 3.**  Пример сопоставления с метатезавром UMLS и извлечением синонимов. |

Конкурс BioASQ в рамках издания 2017 года [56] состоял из трех задач: (1) Задача 5a по крупномасштабной онлайн-биомедицинской семантической индексации, (2) Задача 5b по биомедицинскому семантическому обеспечению качества и (3) Задача c по финансированию извлечения информации из биомедицинской литературы. Целью задачи 5b является оценка производительности систем контроля качества на разных этапах процесса контроля качества. Он подразделяется на две фазы: фазу A и фазу B. BioASQ создал и выпустил контрольные наборы данных биомедицинских вопросов на английском языке вместе с ответами золотого стандарта. Было четыре типа вопросов: да/нет, фактоид, список и резюме вопросов. На этапе А участники должны были ответить биомедицинскими концепциями, соответствующими документами, соответствующими отрывками и тройками RDF. На этапе B участникам было предложено ответить точными ответами и идеальными ответами (резюме размером с абзац). Точные ответы требуются только в случае да/нет, фактоида, списка, в то время как идеальные ответы, как ожидается, будут возвращены для вопросов. Выпущенные вопросы сопровождались их видами и правильными ответами на требуемые элементы (документы и отрывки) первого этапа.

## Метрики оценки

Эффективность предлагаемой биомедицинской системы контроля качества оценивается с использованием оценочных показателей, описанных в конкурсе BioASQ [41].. В случае вопросов «да/нет» точные ответы SemBioNLQA должны были быть либо «да», либо «нет». Таким образом, точность является основным показателем оценки, используемым для оценки ответов на вопросы «да/нет». Пусть *n* будет, количество вопросов «да/нет», и *k* — количество правильно ответивших вопросов «да/нет», точность вычисляется следующим образом:

*k*

Точность *=*

*н* (1)

Для фактоидных вопросов средний взаимный ранг (MRR) является основным баллом, используемым для оценки фактоидных систем контроля качества. Предполагая, что существует *n* фактоидных вопросов, MRR определяется следующим образом, где *ri*является самой верхней позицией, содержащей имя золотой сущности (или один из его синонимов) в возвращенном списке возможных ответов.

1н 1

MRR *=* \*

*n i=*1 *ri* (2)

С другой стороны, для оценки точных ответов на вопросы списка используются средние показатели точности, среднего среднего отзыва и среднего среднего значения Fmeasure, которые вычисляются путем усреднения точности (P), отзыва (R) и F1-меры (F1) по вопросам списка. Эти показатели рассчитываются следующим образом, где TP - количество возможных ответов, которые включены как в возвращенный, так и в золотой список; FP – количество возможных ответов, которые упоминаются в возвращенном, но не в золотом списке; и FN — это количество возможных ответов, которые включены в золотой, но не в возвращенный список. Действительно, средняя средняя F-мера является официальным баллом, используемым вызовом BioASQ для вопросов списка.

 ТПТП

*P= R=* F1 *=* 2\* TP *+* FP TP *+* FN(3)

Наконец, идеальные ответы на вопросы (все типы вопросов: да/нет, фактоид, список и резюме) были автоматически оценены с использованием мер Rouge. Более подробная информация об этих метриках оценок приведена в [60]..

## Наборы оценочных данных

Для оценки производительности SemBioNLQA используется общедоступная инфраструктура оценки биомедицинского контроля качества, разработанная BioASQ challenge [41], которая включает в себя наборы эталонных данных. Контрольные наборы данных, содержащие биомедицинские вопросы на английском языке и ответы золотого стандарта, были построены и вручную аннотированы командой биомедицинских экспертов BioASQ. Было четыре типа вопросов: да/нет, фактоид, список и резюме вопросов. На этапе B, задача b конкурса BioASQ, в рамках каждого издания организаторы выпустили обучающий набор пар вопросов-ответов и пять пакетов тестовых вопросов. Каждая партия содержит около 100 биомедицинских вопросов, а также соответствующие типы, золотые документы и золотые отрывки. В этом исследовании наборы данных BioASQ 2015, 2016 и 2017 годов используются для оценки эффективности системы SemBioNLQA. Ниже приведены некоторые примеры биомедицинских вопросов из набора обучающих данных BioASQ:

* Какие лекарства сравнивались в испытании ROCKET AF? (идентификатор 56bb616dac7ad10019000008)
* Какие ферментымедoesMLN4924inhibit? (идентификатор

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Таблица 2**  Общие результаты предложенных методов извлечения ответов в SemBioNLQA и сравнение с современными системами на пяти партиях тестовых наборов данных, предоставленных BioASQ 3b 2015. Символ "-" заменяет оценки систем, которые не проводили оценку по этой партии или не занимались этой задачей, в то время как "nr" указывает, что результаты не представлены для этой оценочной меры. P, R и F1 указывают на точность, отзыв и F1-меру соответственно.  Наборов данных  Имя системы  Точные ответы  Идентификационные ответы  Да/Нет  Фактоид  Список   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  | Точность | МРР | P | R | Формула-1 | Красный-2 | Красный-СУ4 | | Партия 1 | SemBioNLQA | 0.6970 | 0.1692 | 0.1545 | 0.2409 | 0.1830 | 0.2716 | 0.2860 | |  | Чжан и др. [29] | - | 0.1423 | Нет | Нет | 0.0756 | - | - | |  | Невес [31] | 0.6667 | - | 0.0292 | 0.0603 | 0.0364 | 0.1884 | 0.2008 | |  | Цой [30] | - | - | - | - | - | - | 0.3071 | | Партия 2 | SemBioNLQA | 0.6250 | 0.1776 | 0.1929 | 0.2714 | 0.2127 | 0.3123 | 0.3364 | |  | Чжан и др. [29] | - | 0.0859 | Нет | Нет | 0.1160 | - | - | |  | Невес [31] | 0.5625 | - | 0.0714 | 0.0161 | 0.0262 | 0.2026 | 0.2227 | |  | Цой [30] | - | - | - | - | - | - | 0.3710 | | Партия 3 | SemBioNLQA | 0.8621 | 0.1840 | 0.2353 | 0.2927 | 0.2524 | 0.3879 | 0.4078 | |  | Чжан и др. [29] | - | 0.0846 | Нет | Нет | 0.1319 | - | - | |  | Ян и др. [28] | - | 0.1615 | 0.0539 | 0.6933 | 0.0969 | - | - | |  | Невес [31] | 0.6207 | - | - | - | - | 0.1934 | 0.2189 | |  | Цой [30] | - | - | - | - | - | - | 0.3941 | | Партия 4 | SemBioNLQA | 0.7600 | 0.2960 | 0.2783 | 0.2713 | 0.2588 | 0.3917 | 0.4108 | |  | Чжан и др. [29] | - | 0.2299 | Нет | Нет | 0.2192 | - | - | |  | Ян и др. [28] | - | 0.5155 | 0.3836 | 0.3480 | 0.3168 | - | - | |  | Невес [31] | 0.5600 | 0.0345 | 0.1522 | 0.0473 | 0.0689 | 0.2504 | 0.2724 | |  | Цой [30] | - | - | - | - | - | - | 0.3906 | | Партия 5 | SemBioNLQA | 0.6071 | 0.1568 | 0.0583 | 0.0736 | 0.0625 | 0.3440 | 0.3533 | |  | Чжан и др. [29] | - | 0.2500 | Нет | Нет | 0.1340 | - | - | |  | Ян и др. [28] | - | 0.2727 | 0.1704 | 0.2573 | 0.1875 | - | - | |  | Невес [31] | 0.3571 | 0.0909 | 0.0625 | 0.0292 | 0.0397 | 0.1694 | 0.1790 | |  | Цой [30] | - | - | - | - | - | - | 0.3665 | |

56ed03862ac5ed1459000004)

## Результаты

Чтобы оценить эффективность системы SemBioNLQA и сравнить с текущими интегральными биомедицинскими системами контроля качества, мы представляем две различные оценки: (1) систематическую / автоматическую оценку наборов эталонных данных, предоставляемых вызовами BioASQ, и (2) ручную оценку с точки зрения качества ответов с использованием учебных вопросов и ответов BioASQ.

### Систематическая оценка

В этом эксперименте мы представляем систематическую оценку биомедицинских вопросов, предоставленных вызовом BioASQ, чтобы сравнить с системами участников BioASQ. Как мы уже отмечали ранее, задачи BioASQ на этапе B (т.е. точные идеальные ответы) задачи b предоставляют тестовый набор биомедицинских вопросов вместе с их золотыми документами, золотыми фрагментами и типами вопросов [61,62,56], а системам участников [29,31,30,28,32] было предложено ответить точными ответами и идеальными ответами с использованием золотых документов, золотых фрагментов, и типы золотых вопросов. На каждый вопрос каждая участвующая система может дать идеальный ответ, т.е. резюме соответствующей информации размером с абзац. В случае вопросов «да/нет», «фактоид» и «список» системы также могут возвращать точные ответы; для кратких вопросов точные ответы не будут возвращены. Поэтому для сравнения с такими системами мы также использовали отрывки золотого стандарта и типы вопросов, предоставленные BioASQ, вместо тех, которые извлекаются предлагаемой системой.

В таблице 2 и таблице 3 показаны экспериментальные результаты предложенных методов извлечения ответов в SemBioNLQA и сравнение с современными исследованиями, представленными в [29,31,30,28,32] на пяти партиях тестовых наборов данных, предоставленных конкурсом BioASQ в 2015 и 2016 годах соответственно.

Кроме того, в Таблицах 4 и 5 представлены результаты нашего участия в Фазе B, Задаче 5b конкурса BioASQ 2017 года с использованием только предложенных методов извлечения ответа в SemBioNLQA. Значения внутри параметров указывают наш текущий ранг, общее количество представлений и общее количество команд, участвующих в задаче. Обратите внимание, что в отличие от проблем BioASQ 2015 и 2016 годов, биомедицинским системам контроля качества больше не разрешается предоставлять собственный список синонимов в задаче 2017 года. Полные результаты этапа B задачи 5b доступны онлайн. Название нашей системы для подачи было "sarrouti".[[5]](#footnote-5)

В частности, мы также сообщаем об общих результатах оценки сквозной системы SemBioNLQA на наборах данных BioASQ, предоставленных вызовом в изданиях 2015 и 2016 годов, чтобы продемонстрировать эффективность SemBioNLQA, влияние его различных компонентов на общую производительность, а также облегчить новые сравнения. В этой оценке мы использовали соответствующие документы, соответствующие отрывки и типы вопросов, возвращенные предлагаемой системой. Прямое сравнение с современными методами, представленными в [31,29,32,28,30] и оцененными по задачам BioASQ 2015 или 2016 года, непросто, поскольку авторы использовали тестовый набор биомедицинских вопросов вместе с их золотыми документами, золотыми фрагментами и типами вопросов, выпущенными задачами BioASQ. Как уже было сказано ранее, на данный биомедицинский вопрос SemBioNLQA сначала извлекает документы *с* самым высоким рейтингом D, затем находит отрывки *с* самым высоким рейтингом P и, наконец, применяет соответствующее извлечение ответа в соответствии с типом вопроса, обнаруженным модулем классификации вопросов. В частности, мы решили использовать *D* = 10 документов с самым высоким рейтингом и *P* = 10 самых рейтинговых проходов, поскольку только 10 первых из полученного списка разрешены для тестирования в испытаниях BioASQ 2015 и 2016 годов. В таблице 6 приведены результаты по наборам данных BioASQ 3b 2015 и BioASQ 4b 2016 с использованием системы BioASQ Oracle13, онлайн-системы, позволяющей загружать файлы результатов JSON и получать оценки в любое время. Все ответы, возвращенные системой SemBioNLQA на наборах данных BioASQ 3b 2015 и BioASQ 4b 2016, доступны для загрузки. С другой стороны, рис. [[6]](#footnote-6)4, рис. 5 и рис. На фиг.6 показаны выходные данные SemBioNLQA для четырех биомедицинских вопросов, которые взяты из учебных вопросов BioASQ.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Таблица 3**  Общие результаты предложенных методов извлечения ответов в SemBioNLQA и сравнение с текущими современными методами по пяти партиям тестовых наборов данных, предоставленных BioASQ 4b 2016. Символ "-" заменяет оценки систем, которые не справлялись с этой задачей, в то время как "nr" указывает, что результаты не представляются для этого показателя оценки. P, R и F1 указывают на точность, отзыв и F1-меру соответственно.  Наборов данных  Имя системы  Точные ответы  Идентификационные ответы  Да/Нет  Фактоид  Список   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  | Точность | МРР | P | R | Формула-1 | Красный-2 | Красный-СУ4 | | Партия 1 | SemBioNLQA | 0.8214 | 0.0726 | 0.2182 | 0.3939 | 0.2756 | 0.4772 | 0.4918 | |  | Шульце и др. [32] | - | - | - | - | - | Нет | 0.2231 | | Партия 2 | SemBioNLQA | 0.8750 | 0.1452 | 0.2381 | 0.2505 | 0.2349 | 0.5021 | 0.5115 | |  | Шульце и др. [32] | - | - | - | - | - | Нет | 0.2240 | | Партия 3 | SemBioNLQA | 0.8400 | 0.1218 | 0.2381 | 0.3627 | 0.2812 | 0.4978 | 0.5061 | |  | Шульце и др. [32] | - | - | - | - | - | Нет | 0.2559 | | Партия 4 | SemBioNLQA | 0.8095 | 0.1129 | 0.1467 | 0.2231 | 0.1702 | 0.5192 | 0.5231 | |  | Шульце и др. [32] | - | - | - | - | - | Нет | 0.2280 | | Партия 5 | SemBioNLQA | 0.8148 | 0.1136 | 0.1900 | 0.2353 | 0.1963 | 0.4979 | 0.5027 | |  | Шульце и др. [32] | - | - | - | - | - | Нет | 0.3233 |   **Таблица 4**  Получены результаты нашего участия в «Точных ответах», Фаза B, Задача 5b конкурса BioASQ 2017 года с использованием предложенных методов извлечения ответов. Первое значение внутри параметров указывает на наш текущий ранг и общее количество представлений для задачи, а второе указывает на наш текущий ранг и общее количество участвующих команд.  Наборов данных  Да/Нет  Фактоид  Список   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | Точность | МРР | Точность | Вспоминать | F-мера | | Партия 1 | 0.7647 | 0.1833 (6/15) (4/9) | 0.1909 | 0.2539 | 0.2037 (4/15) (3/9) | | Партия 2 | 0.7778 | 0.0887 (11/21) (5/9) | 0.2400 | 0.3922 | 0.2920 (7/21) (3/9) | | Партия 3 | 0.8387 (1/21) (1/10) | 0.2212 (9/21) (4/10) | 0.2000 | 0.4117 | 0.2625 (6/21) (3/10) | | Партия 4 | 0.6207 (2/27) (2/11) | 0.0970 (13/27) (5/11) | 0.1077 | 0.2013 | 0.1369 (12/27) (5/11) | | Партия 5 | 0.4615 | 0.2071 (9/25) (3/11) | 0.2182 | 0.3178 | 0.2529 (11/25) (6/11) | |

### Ручная оценка

Эта оценка направлена на сравнение SemBioNLQA с комплексными биомедицинскими системами контроля качества, включая AskHermes, EAGLi, Olelo. Для этого мы использовали ту же стратегию, что и в [27]. Действительно, мы случайным образом выбрали 30 вопросов из набора обучающих данных BioASQ и задали их четырем системам - AskHermes, EAGLi, Olelo и SemBioNLQA. Эта оценка проводилась вручную, и поэтому нам нужно было ограничить количество вопросов и типов. Мы решили ограничить его фактоидными, списковыми и «да/нет» вопросами, учитывая, что эти типы ответов легче проверить вручную, чем резюме. Эта последовательность из 30 вопросов, которые перечислены в Приложении А, содержит 10 фактоидных вопросов, 11 вопросов списка и 9 вопросов «да/нет». В нашей оценке ответ считается правильным, если первая возвращенная биомедицинская сущность (для фактических вопросов), по крайней мере одна из первых пяти возвращенных биомедицинских сущностей (для вопросов списка) или булева значение, то есть «да» или «нет», является правильным. Действительно, мы вручную проверили результаты, возвращенные каждой системой, чтобы найти правильные стандартные ответы, предоставленные вызовом BioASQ. В таблице 7 представлены и сопоставлены результаты работы вышеупомянутых систем и SemBioNLQA. Все ответы, возвращаемые системами, доступны для скачивания. Обратите внимание, что обновления базы данных Medline могут привести к различным извлеченным документам, фрагментам или ответам.[[7]](#footnote-7)

## Обсуждение

В то время как QA с открытой областью широко изучается, немногие интегральные системы, подобные описанным в [25,26,38], в настоящее время способны автоматически отвечать на вопросы из постоянно растущего объема рецензируемых научных статей в биомедицинской области. В этой работе мы рассмотрели недостатки этих систем, такие как ограниченное удобство использования и производительность с точки зрения точности для поддерживаемых в настоящее время типов вопросов и ответов. В отличие от этих систем, наша разработанная биомедицинская система контроля качества SemBioNLQA имеет возможность обрабатывать большое количество вопросов и ответов, таких как да / нет, фактоид, список и резюме вопросов, которые могут охватывать все типы вопросов.

Результаты систематической оценки показали, что SemBioNLQA является более конкурентоспособным по сравнению с современными системами. Как показано в таблице 2, по сравнению с двумя подходами, предложенными в [31,29], один из которых основан на базе данных в памяти, а другой использует PubTator, наш подход дает лучшие результаты. Представленная система значительно превзошла вышеупомянутые системы в извлечении как точных ответов, так и идеальных ответов на вопросы «да/нет», «фактоид», «список» и «резюме» биомедицинских вопросов. Более того, повышенная производительность была статистически значимой (P-значение составляет 7,6e-05, результат значителен при p < 0,01). Кроме того, предлагаемая система является конкурентоспособной по сравнению с подходом, представленным в [30], который касался только идеальных ответов (0,2860 против 0,3071, 0,3364 против 0,3710, 0,4078 против 0,3941, 0,4108 против 0,3906 и 0,3533 против 0,3665 Rouge-SU4 в партии 1, партии 2, партии 3, партии 4 и партии 5 соответственно). С другой стороны, как показано в таблице 3, SemBioNLQA по-прежнему достигает хорошей производительности по сравнению с победившей системой 2016 года, разработанной Schulze et al. [32], который касался исключительно идеальных ответов на вопросы. Последний основан на алгоритме LexRank [40], но который использовал исключительно именованные сущности для функции подобия. Здесь важно отметить, что наша система значительно превосходит систему Шульце и др. [32] во всех партиях тестовых наборов данных. Наибольшая разница в терминах Rouge-SU4 между SemBioNLQA и вышеупомянутой системой составила 0,2951 (0,5231 0,2280 в партии 4), что ясно указывает на то, что наш метод не только эффективен, но и надежен в извлечении идеальных ответов. В рамках нашего участия [36] в фазе B (точные и идеальные ответы), Задание 5B

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Таблица 5**  Получены результаты нашего участия в «Идеальных ответах», Фаза B, Задача 5b конкурса BioASQ 2017 года с использованием предложенных методов извлечения ответов. Первое значение внутри параметров указывает наш текущий ранг и общее количество команд, участвовавших в выполнении задачи.  Наборов данных  Автоматические оценки  Ручные оценки   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  | Красный-2 | Красный-СУ4 | Удобочитаемость | Вспоминать | Точность | Повторение | | Партия 1 | 0.5087 (2/9) | 0.5247 (2/9) | 3.65 (2/9) | 4.42 (2/9) | 3.90 (2/9) | 3.89 (2/9) | | Партия 2 | 0.4823 (2/9) | 0.4828 (2/9) | 3.68 (2/9) | 4.59 (2/9) | 4.01 (2/9) | 3.91 (2/9) | | Партия 3 | 0.5658 (2/10) | 0.5729 (2/10) | 3.91 (2/10) | 4.64 (2/10) | 4.07 (2/10) | 4.00 (2/10) | | Партия 4 | 0.5667 (3/11) | 0.5640 (3/11) | 3.86 (3/11) | 4.51 (3/11) | 4.02 (2/11) | 3.95 (3/11) | | Партия 5 | 0.5616 (3/11) | 0.5595 (3/11) | 3.82 (3/11) | 4.53 (2/11) | 3.91 (3/11) | 3.90 (3/11) |   **Таблица 6**  Общие результаты оценки системы SemBioNLQA по пяти партиям биомедицинских вопросов предоставлены BioASQ 3b 2015 и BioASQ 4b 2016.  Наборов данных  Пакетов  Точные ответы  Идентификационные ответы  Да/Нет  Фактоид  Список   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  | Точность | МРР | Точность | Вспоминать | F1-мера | Красный-2 | Красный-СУ4 | | BioASQ 3b 2015 | Партия 1 | 0.7273 | 0.0128 | 0.0364 | 0.0682 | 0.0462 | 0.1039 | 0.1397 | |  | Партия 2 | 0.6875 | 0.0339 | 0.0714 | 0.1232 | 0.0891 | 0.1044 | 0.1413 | |  | Партия 3 | 0.8621 | 0.0641 | 0.0353 | 0.0368 | 0.0352 | 0.1269 | 0.1557 | |  | Партия 4 | 0.6800 | 0.0586 | 0.0696 | 0.0880 | 0.0751 | 0.1467 | 0.1702 | |  | Партия 5 | 0.6786 | 0.0795 | 0.0083 | 0.0208 | 0.0119 | 0.0915 | 0.1205 | | BioASQ 4b 2016 | Партия 1 | 0.8214 | 0.0534 | 0.1091 | 0.1545 | 0.1268 | 0.1552 | 0.1855 | |  | Партия 2 | 0.7188 | 0.0495 | 0.0476 | 0.0477 | 0.0462 | 0.1378 | 0.1720 | |  | Партия 3 | 0.8800 | 0.1186 | 0.0857 | 0.1667 | 0.1122 | 0.1430 | 0.1730 | |  | Партия 4 | 0.8095 | 0.0253 | 0.0400 | 0.0667 | 0.0500 | 0.1097 | 0.1289 | |  | Партия 5 | 0.8519 | 0.0687 | 0.0300 | 0.0517 | 0.0368 | 0.1609 | 0.1836 | |

2017 BioASQ challenge, предложенные методы извлечения ответа в SemBioNLQA, показали хорошие результаты в рейтинге задач в командах высшего уровня, как показано в таблице 4 и таблице 6 . В общей сложности 16, 22, 21, 27 и 25 прогонов были представлены для партии 1, партии 2, партии 3, партии 4 и партии 5 соответственно. Обратите внимание, что многие забеги были представлены одними и теми же командами. Предложенная система стала одним из победителей конкурса.[[8]](#footnote-8)

С другой стороны, как показано в таблице 7 , в которой представлены и сравниваются результаты ручной оценки SemBioNLQA, EAGLi, AskHERMES, а также Olelo с точки зрения количества распознанных вопросов и правильных ответов, SemBioNLQA получает лучшие результаты и преуспел в ответах на большинство случайно выбранных вопросов BioASQ. Мы вручную проанализировали ответы, предоставленные на биомедицинские вопросы каждой системой. В отличие от SemBioNLQA, который оказался довольно успешным в извлечении точных ответов в зависимости от ожидаемого ответа на каждый тип вопроса, Olelo вернул резюме в качестве ответа на большинство вопросов, а AskHERMES вернул отрывок из нескольких предложений в качестве ответа на все вопросы. Действительно, SemBioNLQA смог обнаружить тип вопросов, которые были фактоидными, списковыми или да/нет типами, и, таким образом, сгенерировал точные ответы в зависимости от ожидаемого ответа на каждый тип вопроса. Это указывает на то, что интеграция нашего метода классификации вопросов дает SemBioNLQA возможность понимать и правильно распознавать информационные потребности пользователей. Напротив, несмотря на то, что Olelo был разработан для работы с фактоидными, списковыми и сводными вопросами, он не смог обнаружить типы для заданных вопросов и, таким образом, сгенерировал резюме для всех вопросов, и, следовательно, пользователи должны были прочитать эти резюме, чтобы найти точные ответы. В частности, он возвращает точные ответы только при обнаружении как заголовка, так и семантического типа, в дополнение к тому, что ответы-кандидаты относятся к этому же семантическому типу. С другой стороны, как показано в таблице 7, и SemBioNLQA, и Olelo удалось вернуть ответы на все вопросы, в то время как AskHERMES и EAGLi не смогли дать ответы на большинство вопросов, вместо этого только следующие сообщения «Ничего не найдено! Пожалуйста, уточните ваш вопрос» в первом и «ЕАГЛи не понял ваш вопрос. Попробуйте популярный пример или перейдите в ручной режим." в последнем.

Хотя система SemBioNLQA может эффективно ответить на различные биомедицинские вопросы, мы обнаружили, что все еще есть некоторые ошибки, которые предлагаемая система не может исправить. Например, мы обнаружили, что нынешняя форма предлагаемой системы не смогла дать ответы на некоторые вопросы, особенно для тех, которые ожидают числа в качестве ответа вместо биомедицинских сущностей. Например, ответы на биомедицинский вопрос «Какова распространенность синдрома короткого интервала QT?» (идентификатор 52fb78572059c6d71c000067) и «Каково количество генов, кодирующих белок, в геноме человека?» (идентификатор 535d3c069a4572de6f000006), собранных из учебных вопросов BioASQ, являются «0,01%-0,1%», «Между 20 000 и 25 000» соответственно. Такие вопросы представляются достаточно сложными и требуют более конкретных методов извлечения информации. С другой стороны, из таблицы 6 ясно, что различные компоненты системы SemBioNLQA оказывают значительное влияние на задачу извлечения ответа и, следовательно, на общую производительность SemBioNLQA, поскольку, если набор извлеченных документов, отрывков и тип данного вопроса не определены правильно, дальнейшие этапы обработки для извлечения ответов также неизбежно потерпят неудачу. Например, на вопрос «Какова связь спермидина с нейротоксичностью α-синуклеина?» (идентификатор 56c073fcef6e394741000020) из партии 1 тестового набора вызова BioASQ 2016 года возвращаемый тип вопроса является «резюме», тогда как в корпусе тип вопроса —«фактоид». Таким образом, система SemBioNLQA неизбежно не сможет извлечь и вывести правильный ответ, поскольку извлечение ответа на фактоидный вопрос, который запрашивает биомедицинскую сущность, не то же самое, что извлечение ответа на сводный вопрос, который ищет только идеальный ответ.

|  |
| --- |
| **Рис. 4.** Вывод SemBioNLQA для биомедицинского вопроса «Какие симптомы характеризуют синдром Мюнке?» (идентификатор 52bf1d3c03868f1b060000d). |

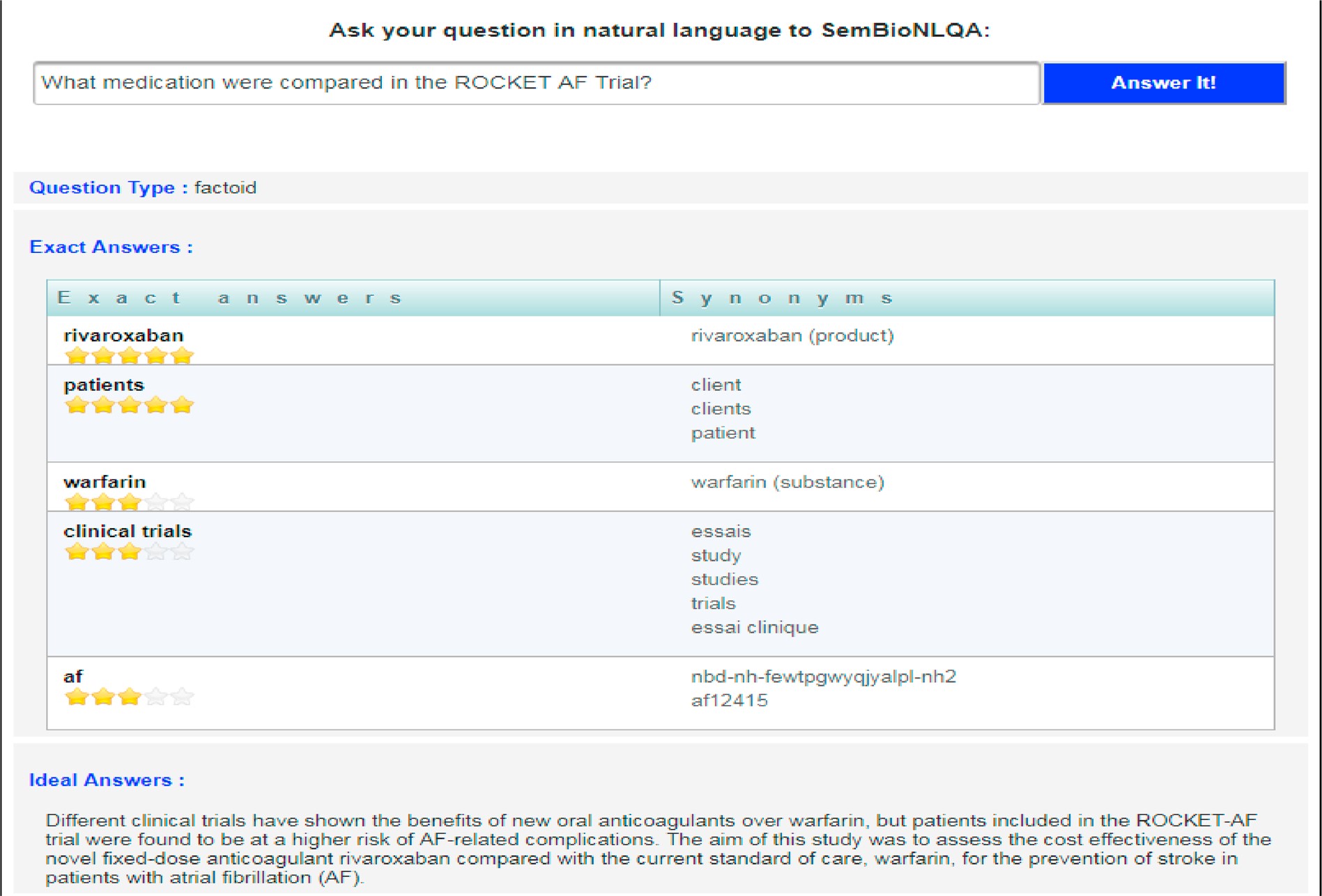
В целом, SemBioNLQA имеет ряд преимуществ перед современными системами. Во-первых, интеграция нашего метода классификации типов вопросов, который он предлагает, имеет явное преимущество перед Olelo в том, что он возвращает точные ответы, которые зависят от ожидаемого ответа каждого типа вопроса. Во-вторых, SemBioNLQA обеспечивает непревзойденное преимущество перед AskHERMES, EAGli и Olelo в том, что он обрабатывает большое количество типов вопросов, включая да / нет, фактоид, список и сводные вопросы. В-третьих, результаты систематических и ручных оценок показали, что SemBioNLQA более эффективен по сравнению с вышеупомянутыми системами.

Таким образом, биомедицинское обеспечение качества является очень сложной задачей, поскольку оно позволяет задавать запросы как вопросы на естественном языке и дает точные ответы, а не только представлять потенциально релевантные документы путем интеграции различных ресурсов. Поэтому ни одна нынешняя система не всегда может хорошо работать по множеству вопросов, которые могут быть заданы ей. SemBioNLQA предоставляет практическую и конкурентоспособную альтернативу, чтобы помочь пользователям найти точные и идеальные ответы.

# Заключение и будущая работа

В этой статье мы рассмотрели полностью автоматическую систему контроля качества в биомедицинской области, SemBioNLQA, которая имеет возможность решать четыре типа биомедицинских вопросов, включая вопросы «да/ нет», фактоидные вопросы, вопросы списка и сводные вопросы. SemBioNLQA в настоящее время может предоставить точные ответы (например, «да», «нет», биомедицинский

|  |
| --- |
| **Рис. 5.**  Результаты SemBioNLQA для биомедицинского вопроса «Какие лекарства сравнивались в исследовании ROCKET AF?» (идентификатор 56bb616dac7ad10019000008).    **Рис. 6.** Вывод SemBioNLQA для биомедицинского вопроса «Связаны ли мутации в гене C9orf72 с дегенерацией желтого пятна?» (идентификатор |

58e11bf76fddd3e83e0000c).

**Таблица 7**

Сравнение результатов SemBioNLQA, EAGLi, AskHERMES и Olelo по количеству признанных вопросов и правильным ответам

|  |  |  |
| --- | --- | --- |
| Системы | Количество признанных вопросов | Количество правильных ответов |
| ЕАГЛи [25] | 7/30 | 3/30 |
| АскГЕРМЕС [26] | 12/30 | 2/30 |
| Олело [38] | 30/30 | 6/30 |
| SemBioNLQA | 30/30 | 18/30 |

сущность и т.д.) и идеальные ответы размером с абзац (резюме соответствующей информации) на вопросы «да/нет», «фактоид» и «список», в то время как он извлекает только идеальные ответы на краткие вопросы. Система опиралась на (1) лексо-синтаксические шаблоны ручной работы и подход машинного обучения для классификации вопросов, (2) сходство поисковой системы PubMed и UMLS для извлечения документов, (3) модель BM25, стволовые слова и концепции UMLS для извлечения пассажей и (4) метатезавр UMLS, синонимы BioPortal, анализ тональности и метрику частоты терминов для поиска ответов.

Благодаря оценке стандартных бенчмарков, предоставленных конкурсом BioASQ 2017 года во время нашего участия в фазе B Задачи 5b, мы отметили, что SemBioNLQA достигла многообещающих результатов по сравнению с текущей интегральной биомедицинской системой контроля качества. Кроме того, результаты ручной оценки показали, что SemBioNLQA достигла хороших результатов с точки зрения количества признанных вопросов и правильных ответов по сравнению с самой современной системой, а также предоставляет практическую и конкурентоспособную альтернативу, чтобы помочь искателям информации найти точные и идеальные ответы на свои биомедицинские вопросы.

Для повышения производительности системы SemBioNLQA запланированы три основные перспективы для будущей работы:

* В настоящее время мы рассматриваем несколько способов включения глубокого обучения в SemBioNLQA, поскольку модели глубокого обучения становятся современными для обеспечения качества, особенно в открытой области. Тем не менее, модели глубокого обучения обычно применяются для QA фактоидного типа. Более того, такие подходы требуют большого количества пар ответивших на вопросы для этапа обучения.
* Мы постоянно работаем над некоторыми конкретными методами НЛП, чтобы включить отношения между биомедицинскими сущностями, чтобы улучшить рейтинг и лучшие ответы на поддерживаемые в настоящее время типы вопросов.
* Мы намерены начать работу с полным текстом (например, с PubMed Central), чтобы увидеть, сможем ли мы улучшить производительность SemBioNLQA с помощью целенаправленного взгляда на полный текст. MEDLINE, который индексирует более 25 миллионов ссылок, содержит ссылку на бесплатный полный текст статьи, заархивированной в PubMed Central, которая содержит 5,1 миллиона статей.

# Вклад автора

Все авторы разработали содержание рукописи, внедрили систему, способствуют написанию различных итераций статьи и дают свое окончательное одобрение представленной рукописи.

**Конкурирующие интересы**

Авторы заявляют об отсутствии конфликта интересов.

# Итоговые моменты

* Разработка автоматической биомедицинской системы ответов на вопросы SemBioNLQA для получения точных и идеальных ответов на биомедицинские вопросы из рецензируемых научных статей.
* Он основан на (1) лексико-синтаксических паттернах ручной работы и классификаторе машины опорных векторов (SVM) для классификации вопросов, (2) сходстве PubMed и UMLS для извлечения документов, (3) модели BM25, стволовых словах и концепциях UMLS для извлечения пассажей и (4) метатезавре UMLS, синонимах BioPortal, SENTIWORDNET и метрике частоты терминов для извлечения ответов.
* Он способен принимать различные вопросы на естественном языке и генерировать соответствующие ответы.
* Он предоставляет точные ответы «да» или «нет» на вопросы «да/нет», сущности UMLS для фактоидных вопросов и список сущностей UMLS для вопросов списка. В дополнение к точным ответам на эти типы вопросов, SemBioNLQA также возвращает идеальные ответы (резюме размером с абзац), в то время как он извлекает только идеальные ответы для сводных вопросов.
* Экспериментальные оценки, проведенные по биомедицинским вопросам, предоставленным вызовом BioASQ, показывают, что SemBioNLQA достигает хороших показателей по сравнению с текущими интегральными системами состояния.

# Наличие

Исходный код SemBioNLQA доступен по адресу [https://github.com/ sarrouti/sembionlqa.](https://github.com/sarrouti/sembionlqa)

# Приложение A. Перечень вопросов BioASQ, используемых для ручной оценки

1. Какой ген чаще всего мутирует при болезни Тея-Сакса? (идентификатор 536e46f27d100faa09000012)
2. Какие лекарства сравнивались в испытании ROCKET AF? (идентификатор 56bb616dac7ad10019000008)
3. Какие ферментымедoesMLN4924inhibit? (id

56ed03862ac5ed1459000004)

1. Где находитсяпротеинПаннексин1? (идентификатор56af9f130a360a5e45000015 )
2. Вакцина RTS S AS01 была разработана для предотвращения какого заболевания? (идентификатор 56bc77a3ac7ad10019000015)
3. Для какого типа миеломы оценивается иксазомиб? (идентификатор56ed0ffe2ac5ed1459000008)
4. WhatenzymeisinhibiedbyOpicapone? (id

56c1d857ef6e394741000033)

1. Чтогенехасбеимплифицирован в MajeedSyndrome? (id 56f7c15a09dd18d46b000012)
2. Какой ген чаще всего ассоциируется с тяжелой циклической нейтропенией врожденного потомства? (идентификатор 5503133ae9bde6963400001d)
3. Что является рецептором для иммуносупрессивного препарата циклоспоринА (CsA)? (идентификатор 56f6c11109dd18d46b00000e)
4. Какие белки участвуют в формировании комплекса активации транскрипции Notch? (идентификатор 54fb6fb5d176fff445000004)
5. Что измеряется акселерометром у пациентов с болями в спине? (идентификатор 533f9df0c45e133714000016)
6. Список ингиборов, нацеленных на переходную пору митохондриальной проницаемости (идентификатор 5717cdd2070aa3d072000001)
7. ListsymptomsoftheIFAPsyndrome(identifier

56c1f038ef6e394741000051)

1. Что является функциональнымпротеинDrp1? (идентификатор 5717dbfe7de986d80d800001)
2. Какие рецепторы связаны сТасимелтеоном? (идентификатор 56c1f043ef6e394741000057)
3. Какие фенотипы заболеваний связаны с мутациями PRPS1? (идентификатор 5713b0a51174fb17550000e)
4. Какие гены, как считается, участвуют в развитии медуллобластомы? (идентификатор 5539029cbc4f83e828000012)
5. Какие микроРНК могут быть использованы в качестве потенциальных биомаркеров эпителиаловарианского рака? (идентификатор 553fa78b1d53b76422000007)
6. Какиеинтерлейкины являются ограниченными поДупилумабу? (идентификатор 56c1f005ef6e39474100003a)
7. Какие гены, как считается, регулируются EWS/FLI? (идентификатор 552faa43bc4f83e828000004)
8. Показан ли ингибитор АПФ для лечения рака легких? (идентификатор530cf4fe960c95ad0c000005)
9. IsPTENinvolvedinfollicularthyroidcarcinoma? (идентификатор 55031650e9bde 69634000026)
10. Представлена ли анемия Фанкони как генетически и клинически гетерогенная болезнь? (идентификатор 54ede5-

c394afd61504000006)

1. Можно ли использовать технологию iPS-клеток в терапии анемии Фанкони? (идентификатор 54edef0594afd6150400000d)
2. Улучшает ли операция по поводу эндометриом яичников фертильность? (идентификатор 54f088ee94afd61504000015)
3. Синдром раздраженного кишечника чаще встречается у женщин с эндометриозом? (идентификатор 54f08d4a94afd61504000016)
4. Является ли регуляция эндонуклеазы Vsr независимой от фазы роста бактерий? (идентификатор 553fbe9fe00431e071000001
5. Вызывает ли мутация гена TRIM37 нанизм Мулибри? (идентификатор56c1f03cef6e394741000054)
6. Является ли ген MAOA эпигенетически модифицированным метилированием? (идентификатор 56cf50253975bb303a00000b)

# Ссылки

1. Sarker A, Mollá D, Paris C. Query-ориентированное извлечение доказательств для поддержки практики доказательной медицины. Журнал биомедицинской информатики 2016;59:169–84. [https://doi.org/10.1016/j.jbi.2015.11.](https://doi.org/10.1016/j.jbi.2015.11.010) [010. URL-https://doi.org/10.1016%2Fj.jbi. 2015.11.010г.](https://doi.org/10.1016/j.jbi.2015.11.010)
2. Ji X, Ritter A, Yen P-Y. Использование семантического сходства на основе онтологии для облегчения процесса скрининга статей для систематических обзоров. Журнал биомедицинской информатики 2017;69:33–42. [https://doi.](https://doi.org/10.1016/j.jbi.2017.03.007) [org/10.1016/j.jbi.2017.03.007. URL https://doi.org/ 10.1016%2Fj.jbi.2017.03.007.](https://doi.org/10.1016/j.jbi.2017.03.007)
3. Frasca M. Gene2disco: Ген к болезни с использованием общих черт болезни. Искусственный интеллект в медицине 2017;82:34–46. [https://doi.org/10.1016/j.artmed.2017.](https://doi.org/10.1016/j.artmed.2017.08.001)

[08.001г. URL-https://doi.org/10.1016%2Fj.artmed.2017.08.001.](https://doi.org/10.1016/j.artmed.2017.08.001)

1. Копаница Г. Интеграция больничной информации и систем поддержки принятия клинических решений для обеспечения возможности повторного использования данных электронных медицинских карт. Методы инфармации в медицине 2017;56(3):238–47. [https://doi.org/10.3414/me16-01-0057. URL https://doi.org/10.3414%2Fme16-01-0057.](https://doi.org/10.3414/me16-01-0057)
2. Kropf S, Krücken P, Mueller W, Denecke K. Структурирование отчетов о патологии наследия с помощью архетипов openEHR для обеспечения семантических запросов. Методы информации в медицине 2017;56(2):230–7. [https://doi.org/10.3414/me16-01-0073. URL-https:// doi.org/10.3414%2Fme16-01-0073.](https://doi.org/10.3414/me16-01-0073)
3. Каннампаллил ТГ, Патель ВЛ. Специальный выпуск о методах когнитивной информатики для интерактивных клинических систем. Журнал биомедицинской информатики 2017;71:207–10. [https://doi.org/10.1016/j.jbi.2017.06.003. URL-https://doi.org/10.1016%2Fj.jbi. 2017.06.003г.](https://doi.org/10.1016/j.jbi.2017.06.003)
4. Wei L, Xing P, Zeng J, Chen J, Su R, Guo F. Улучшено прогнозирование белково-белковых взаимодействий с использованием новых отрицательных образцов, признаков и ансамблевого классификатора.

Искусственный интеллект в медицине 2017;83:67–74. [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.artmed.2017.03.001)

[artmed.2017.03.001. URL-https://doi.org/10.1016%2Fj.artmed.2017.03.001.](https://doi.org/10.1016/j.artmed.2017.03.001)

1. Wang L, Fiol GD, Bray BE, Haug PJ. Создание словарей лечения заболеваний из цитат MEDLINE. Журнал биомедицинской информатики 2017;65:46–57. [https://doi.org/10.1016/j.jbi.2016.11.004. URL https://doi.org/ 10.1016%2Fj.jbi.2016.11.004.](https://doi.org/10.1016/j.jbi.2016.11.004)
2. Chen J, Yu H. Неконтролируемый ансамбль ранжирования терминов в электронных медицинских картах на основе их важности для пациентов. Журнал биомедицинской информатики

2017;68:121–31. [https://doi.org/10.1016/j.jbi.2017.02.016. URL-адрес https://doi.org/ 10.1016%2Fj.jbi.2017.02.016.](https://doi.org/10.1016/j.jbi.2017.02.016)

1. Лами Дж-Б. Owlready: Онтологически ориентированное программирование на python с автоматической классификацией и высокоуровневыми конструкциями для биомедицинских онтологий. Искусственный интеллект в медицине 2017;80:11–28. [https://doi.org/10.1016/j.artmed.2017.](https://doi.org/10.1016/j.artmed.2017.07.002)

[07.002г. URL-https://doi.org/10.1016%2Fj.artmed.2017.07.002.](https://doi.org/10.1016/j.artmed.2017.07.002)

1. Гольдштейн А., Шахар Ю., Оренбух Е., Коэн М.Д. Оценка автоматизированной основанной на знаниях текстовой системы обобщения продольных клинических данных в области интенсивной терапии. Искусственный интеллект в медицине 2017;82:20–33. [https:// doi.org/10.1016/j.artmed.2017.09.001. URL-адрес http](https://doi.org/10.1016/j.artmed.2017.09.001) [s://doi.org/10.1016%2Fj.artmed.](https://doi.org/10.1016/j.artmed.2017.09.001)  [2017.09.001г.](https://doi.org/10.1016/j.artmed.2017.09.001)
2. Д. Христовский, Д. Диневский, А. Кастрин, Т.C. Риндфлеш, Биомедицинские ответы на вопросы с использованием семантических отношений, BMC Bioinformatics 16 (1). doi:10.1186/s12859014-0365-3. URL-https://doi.org/10.1186%2Fs12859-014-0365-3.
3. Эли Дж. Таксономия родовых клинических вопросов: классификационное исследование. BMJ 2000;321(7258):429–32. [https://doi.org/10.1136/b](https://doi.org/10.1136/bmj.321.7258.429) [mj.321.7258.429. URL-https://doi.org/10.1136%2Fbmj.321.7258.429.](https://doi.org/10.1136/bmj.321.7258.429)
4. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Friedman CP, Tidmarsh P, Mosbaek C, Kraemer D. Факторы, связанные с успехом в поиске медлайна и применении доказательств для ответа на клинические вопросы. Журнал Американской ассоциации медицинской информатики 2002;9(3):283–93. [https://doi.org/10.1197/jamia.M0996. arXiv:http://jamia.oxfordjournals.org/content/9/3/283.full.pdf, URL http://jamia.](https://doi.org/10.1197/jamia.M0996)  [oxfordjournals.org/](https://doi.org/10.1197/jamia.M0996) [контент/9/3/283.](https://doi.org/10.1197/jamia.M0996)
5. [Lee M, Cimino JJ, Zhu HR, Sable C, Shanker V, Ely JW, Yu H. Beyond information retrieval-medical question answering. Материалы](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0075) [ежегодного симпозиума](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0075) [AMIA, Vol.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0075)  [2006 2006:469–73.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0075)
6. Athenikos SJ, Han H. Биомедицинский вопрос ответ: Опрос. Компьютерные методы и программы в биомедицине 2010;99(1):1–24. [https://doi.org/10.1016/j.cmpb.](https://doi.org/10.1016/j.cmpb.2009.10.003)

[2009.10.003г. URL-https://doi.org/10.1016%2Fj.cmpb.2009.10.003.](https://doi.org/10.1016/j.cmpb.2009.10.003)

1. Невес М., Лесер У. Ответы на вопросы по биологии. Методы 2015;74:36–46. [https://doi.org/10.1016/j.ymeth.2014.10.023. URL-https://doi.org/10.1016%2Fj. ymeth.2014.10.023.](https://doi.org/10.1016/j.ymeth.2014.10.023)
2. Abacha AB, Zweigenbaum P. MEANS: Медицинская система вопросов-ответов, сочетающая методы НЛП и семантические веб-технологии. Обработка и управление информацией 2015;51(5):570–94. [https://doi.org/10.](https://doi.org/10.1016/j.ipm.2015.04.006) [1016/j.ipm.2015.04.006.](https://doi.org/10.1016/j.ipm.2015.04.006)

[URL-https://doi.org/10.1016%2Fj.ipm.2015.04.006.](https://doi.org/10.1016/j.ipm.2015.04.006)

1. Грин БФ, Вольф АК, Хомский С, Смех К. Бейсбол. Материалы западной совместной компьютерной конференции IRE-AIEE-ACM, Лос-Анджелес, 9-11 мая 1961 года. [https:// doi.org/10.1145/1460690.1460714. URL-https://doi.org/. doi: 10.1145/1460690.](https://doi.org/10.1145/1460690.1460714)  [1460714.](https://doi.org/10.1145/1460690.1460714)
2. Вудс ВА. Прогресс в понимании естественного языка. Материалы национальной компьютерной конференции и экспозиции - AFIPS, 4-8 июня 1973 года. [https://doi.org/10.](https://doi.org/10.1145/1499586.1499695)  [1145/1499586.1499695. URL-https://doi.org/. doi: 10.1145/1499586.1499695.](https://doi.org/10.1145/1499586.1499695)
3. Katz B, Felshin S, Yuret D, Ibrahim A, Lin J, Marton G, McFarland AJ, Temelkuran B. Omnibase: Равномерный доступ к разнородным данным для ответов на вопросы. Обработка естественного языка и информационные системы 2002. С. 230–4. [https://doi.org/10.](https://doi.org/10.1007/3-540-36271-1_23)  [1007/3-540-36271-1\_23. URL-https://doi.org/. doi: 10.1007/3-540-36271-1\_23.](https://doi.org/10.1007/3-540-36271-1_23)
4. Кайзер М. Демонстрация ответов на вопросы QuALiM. Материалы 46-го ежегодного собрания Ассоциации компьютерной лингвистики по человеческому языку

Демонстрационная сессия по технологиям - HLT 08 2008. [https://doi.org/10.3115/1564144. 1564153. URL](https://doi.org/10.3115/1564144.1564153) [https://doi.org/.](https://doi.org/10.3115/1564144.1564153)  [doi: 10.3115/1564144.1564153.](https://doi.org/10.3115/1564144.1564153)

1. Бауэр М.А., Берлеант Д. Юзабилити-исследование биомедицинских систем ответов на вопросы. Геномика человека 2012;6(1):17. [https://doi.org/10.1186/1479-7364-6-1](https://doi.org/10.1186/1479-7364-6-17) [7. URL-https://doi.org/10.1186%2F1479-7364-6-17.](https://doi.org/10.1186/1479-7364-6-17)
2. [Круше С., Гаудинат А., Риндфлеш Т., Бойер К. Как насчет доверия к миру ответов на вопросы. Ежегодный симпозиум AMIA 2009 2009.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0120)
3. [Gobeill J, Patsche E, Theodoro D, Veuthey A-L, Lovis C, Ruch P. Ответы на вопросы для биологии и медицины. 9-я Международная конференция по информационным технологиям и приложениям в биомедицине 2009:1–5.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0125)
4. Cao Y, Liu F, Simpson P, Antieau L, Bennett A, Cimino JJ, Ely J, Yu H. AskHERMES: Онлайн-система ответов на вопросы для сложных клинических вопросов. Журнал биомедицинской информатики 2011;44(2):277–88. [https://doi.org/10.1016/j.jbi.2011.01.](https://doi.org/10.1016/j.jbi.2011.01.004)

[004. URL-https://doi.org/10.1016/j.jbi.2011.01.004.](https://doi.org/10.1016/j.jbi.2011.01.004)

1. Neves M, Eckert F, Folkerts H, Uflacker M. Оценка производительности olelo, биомедицинского приложения для ответа на вопросы в режиме реального времени. BioNLP 2017, Ассоциация компьютерной лингвистики 2017. [https://doi.org/10.18653/v1/w17-2344. URL-https://doi.org/10.18653%2Fv1%2Fw17-2344 .](https://doi.org/10.18653/v1/w17-2344)
2. [Yang Z, Gupta N, Sun X, Xu D, Zhang C, Nyberg E. Учимся отвечать на вопросы о биомедицинских фактах и списках: OAQA на BioASQ 3B. CLEF (Working Notes) 2015.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0140)
3. [Чжан Ю, Пэн С, Ю Р, Се З, Ван Б, Чжу С. Участие fudan в конкурсе bioasq 2015: Крупномасштабная биомедицинская семантическая индексация и ответы на вопросы. В рабочих записках к Конференции и лабораториям Форума по оценке (CLEF) 2015.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0145)
4. [Чхве С. Снамединфо на треке CLEF QA BioASQ 2015. Рабочие записки для Конференции и лабораторий Форума по оценке (CLEF), Тулуза, Франция 2015.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0150)
5. [Система ответов на вопросы Neves M. HPI в конкурсе BioASQ 2015. Рабочие записки для конференции и лабораторий Форума по оценке (CLEF) 2015.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0155)
6. Schulze F, Schuler R, Draeger T, Dummer D, Ernst A, Flemming P, Perscheid C, Neves M. HPI система ответов на вопросы в BioASQ 2016. Труды

Четвертый семинар BioASQ на конференции Ассоциации компьютерной лингвистики 2016:38–44. [https://doi.org/10.18653/v1/w16-3105. URL-https://doi. org/10.18653%2Fv1%2Fw16-3105.](https://doi.org/10.18653/v1/w16-3105)

1. M. Sarrouti, S.O. E. Alaoui, Метод машинного обучения для классификации типов вопросов в биомедицинских ответах на вопросы, Методы информации в медицине 56 (3). doi:10.3414/me16-01-0116. URL-https://doi.org/10.

3414%2Fme16-01-0116.

1. Саррути М., Алауи ГП. Общая структура поиска документов, основанная на сходстве UMLS для биомедицинской системы ответов на вопросы. Технологии интеллектуального принятия решений 2016 2016. с. 207–16. [https://doi.org/10.1](https://doi.org/10.1007/978-3-319-39627-9_18) [007/978-3-319-39627-9\_ 18. URL-https://doi.org/10.1007%2F978-3-319-39627-9\_18.](https://doi.org/10.1007/978-3-319-39627-9_18)
2. Саррути М., Алауи ГП. Метод поиска пассажей, основанный на вероятностном поиске информации и концепциях UMLS в биомедицинских ответах на вопросы. Журнал биомедицинской информатики 2017;68:96–103. [https://doi.org/10.1016/j.jbi.2017.03. 001. URL-https://doi.org/10.1016%2Fj.jbi.2017.03.001.](https://doi.org/10.1016/j.jbi.2017.03.001)
3. Саррути М., Алауи ГП. Биомедицинская система ответов на вопросы в BioASQ 2017. BioNLP 2017, Ассоциация компьютерной лингвистики 2017. [https://doi.org/10.](https://doi.org/10.18653/v1/w17-2337)

[18653/v1/w17-2337. URL-https://doi.org/10.18653%2Fv1%2Fw17-2337.](https://doi.org/10.18653/v1/w17-2337)

1. Vicedo JL, Gómez J. TREC: Эксперимент и оценка в области поиска информации. Журнал Американского общества информационных наук и технологий 2007;58(6):910–1. [https://doi.org/10.1002/asi.20583. URL-https://doi.org/10.](https://doi.org/10.1002/asi.20583)

[1002%2Фази.20583.](https://doi.org/10.1002/asi.20583)

1. Kraus M, Niedermeier J, Jankrift M, Tietbohl S, Stachewicz T, Folkerts H, Uflacker M, Neves M. Olelo: веб-приложение для интуитивного исследования биомедицинской литературы. Исследование нуклеиновых кислот 2017;45(W1):W478–83. [https://doi.org/10.1093/ nar/gkx363. URL-https://doi.org/10.1093%2Fnar%2Fgkx363.](https://doi.org/10.1093/nar/gkx363)
2. Wei C-H, Kao H-Y, Lu Z. PubTator: веб-инструмент для интеллектуального анализа текста для помощи в биокурации. Исследование нуклеиновых кислот 2013;41(W1):W518–22. [https://doi.org/10.](https://doi.org/10.1093/nar/gkt441)

[1093/nar/gkt441. URL-https://doi.org/10.1093%2Fnar%2Fgkt441.](https://doi.org/10.1093/nar/gkt441)

1. [Эркан Г., Радев ДР.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0200)  [Lexrank: Лексическая центральность](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0200) [на основе графов как значимость в обобщении текста. Journal of Artificial Intelligence Research 2004;22:457–79.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0200)
2. Цацаронис Г., Баликас Г., Малакасиотис., Парталас I, Зшунке М., Алверс МР, Вайсенборн Д., Критара А., Петридис С., Полихронопулос Д., Альмирантис Ю.,

Павлопулос Ж., Баскиотис Н., Галлинари., Артьер Т., Нгомо А-КН, Хейно Н., Гауссье Э., Баррио-Альверс Л., Шредер М., Андруцопулос I, Палиурас Г. Обзор крупномасштабной биомедицинской семантической индексации BIOASQ и конкурса ответов на вопросы. BMC Биоинформатика 2015;16(1):1–28. [https://doi.org/10.1186/s12859-015](https://doi.org/10.1186/s12859-015-0564-6) [0564-6. URL-https://doi.org/10.1186/s12859-015-0564-6.](https://doi.org/10.1186/s12859-015-0564-6)

1. Молдавский D, Paşca M, Harabagiu S, Surdeanu M. Проблемы производительности и анализ ошибок в системе ответов на вопросы с открытым доменом. ACM Transactions on Information Systems (TOIS) 2003;21(2):133–54. [https://doi.org/10.1145/763693.](https://doi.org/10.1145/763693.763694)  [763694. URL-https://doi.org/10.1145%2F763693.763694.](https://doi.org/10.1145/763693.763694)
2. Cao Y-g, Cimino JJ, Ely J, Yu H. Автоматическое извлечение информационных потребностей из сложных клинических вопросов. Журнал биомедицинской информатики 2010;43(6):962–71. [https://doi.org/10.1016/j.jbi.2010.07.007. URL-https://doi.org/10.1016%2Fj.jbi. 2010.07.007г.](https://doi.org/10.1016/j.jbi.2010.07.007)
3. Патрик Дж., Ли М. Онтология для клинических вопросов о содержании записей пациента. Журнал биомедицинской информатики 2012;45(2):292–306. [https://doi.org/10. 1016/j.jbi.2011.11.008. URL-https://doi.org/10.1016/j.jbi.2011.11.008.](https://doi.org/10.1016/j.jbi.2011.11.008)
4. [Робертс К., Киличоглу Х., Фишман М., Демнер-Фушман Д. Автоматическая классификация типов вопросов для вопросов здоровья потребителей. AMIA Анну Симп Прок 2014:1018–27.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0225)
5. Саррути М., Лачкар А. Новый и эффективный метод, основанный на синтаксических зависимостях, предназначен для специальной классификации клинических вопросов. Международный журнал исследований и приложений биоинформатики 2017;13(2):161–77. [https://doi.org/10.](https://doi.org/10.1504/ijbra.2017.10003490)

[1504/ijbra.2017.10003490. URL-https://doi.org/10.1504%2Fijbra.2017. 10003490.](https://doi.org/10.1504/ijbra.2017.10003490)

1. [Аронсон А.Р. Эффективное отображение биомедицинского текста на метатезавр UMLS:](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0235) информационные системы 1980;14(3):130–7. [https://doi.org/10.1108/eb046814.](https://doi.org/10.1108/eb046814) [URL-адрес программы MetaMap. Материалы симпозиума AMIA. 2001. С. 17.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0235)  [https://doi.org/10.1108/eb046814.](https://doi.org/10.1108/eb046814)
2. Боденрейдер О. Единая система медицинского языка (UMLS): интеграция биома-[55] Aronson AR, Lang F-M. Обзор MetaMap: историческая перспектива и современная дикальная терминология. Исследование нуклеиновых кислот 2004;32(90001):267D–70D. [https://doi.](https://doi.org/10.1093/nar/gkh061) авансы. Журнал Американской ассоциации медицинской информатики [org/10.1093/nar/gkh061.](https://doi.org/10.1093/nar/gkh061)  [URL-https://doi.org/10.1093%2Fnar%2Fgkh061.](https://doi.org/10.1093/nar/gkh061) 2010;17(3):229–36. [https://doi.org/10.1136/jamia.2009.002733. URL https://](https://doi.org/10.1136/jamia.2009.002733) [49] Ханауэр Д.А., Ву ДТ, Ян Л., Мэй К., Мурковски-Штеффи КБ, Видишваран В.В., Чжэн [doi.org/10.1136%2Fjamia.2009.002733.](https://doi.org/10.1136/jamia.2009.002733)

К. Разработка и эмпирическая ориентированная на пользователя оценка семантически обоснованных [56] Nentidis A, Bougiatiotis K, Krithara A, Paliouras G, Kakadiaris I. Результаты пятого запроса рекомендации для поисковой системы электронных медицинских карт. Журнал издания конкурса BioASQ. БиоНЛП 2017 2017. [https://doi.org/10.18653/v1/](https://doi.org/10.18653/v1/w17-2306) Биомедицинская информатика 2017;67:1–10. [https://doi.org/10.1016/j.jbi.2017.01.013.](https://doi.org/10.1016/j.jbi.2017.01.013)  [w17-2306. URL-https://doi.org/10.18653%2Fv1%2Fw17-2306.](https://doi.org/10.18653/v1/w17-2306)

[URL-https://doi.org/10.1016%2Fj.jbi.2017.01.013.](https://doi.org/10.1016/j.jbi.2017.01.013)  [57] [Коэн КБ, Демнер-Фушман Д., Ананиаду С., Цудзии Дж.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0285)  [Труды 16-го числа](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0285)

1. Демнер-Фушман Д., Лин Дж. Ответы на клинические вопросы с помощью основанных на знаниях и [семинара по БиоНЛП. Ассоциация компьютерной лингвистики; 2017. Статистические](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0285) методы URL. Компьютерная лингвистика 2007;33(1):63–103. [https://doi.](https://doi.org/10.1162/coli.2007.33.1.63) [http://www.aclweb.org/anthology/W17-2300.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0285)

[org/10.1162/coli.2007.33.1.63. URL-https://doi.org/10.1162%2Fcoli.2007.33.](https://doi.org/10.1162/coli.2007.33.1.63) [58] Саррути М., Алауи СОЭ. Генератор ответов «да»/«нет» на основе оценок тональности и слов

[1.63.в](https://doi.org/10.1162/coli.2007.33.1.63) биомедицинских ответах на вопросы. Международный журнал медицинской информации

1. [Макиннес БТ, Педерсен Т, Пахомов СВ. UMLS-интерфейс и UMLS-сходство: открытые](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0255) системы и информатика 2017;12(3):62–74. [https://doi.org/10.4018/ijhisi.](https://doi.org/10.4018/ijhisi.2017070104)  [исходное программное обеспечение для измерения путей и семантического сходства. Ежегодная 2017070104 AMIA](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0255)[.](https://doi.org/10.4018/ijhisi.2017070104)  [URL-https://doi.org/10.4018%2Fijhisi.2017070104.](https://doi.org/10.4018/ijhisi.2017070104)

[Труды симпозиума, том](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0255) [2009 2009:431.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0255) [59] [Баччанелла С., Эсули А., Себастьяни Ф.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0295)  [Sentiwordnet 3.0: Расширенный лексический ресурс](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0295) [52] Otterbacher J, Erkan G, Radev DR. Biased LexRank: извлечение пассажей с использованием случайных [для анализа тональности и анализа мнений.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0295)  [LREC, Vol. 10 2010:2200–4.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0295)

ходит с вопросами, основанными на приорах. Обработка и управление информацией[60] [Balikas G, Partalas I, Kosmopoulos A, Petridis S, Malakasiotis P, Pavlopoulos I,](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0300) 2009;45(1):42–54. [https://doi.org/10.1](https://doi.org/10.1016/j.ipm.2008.06.004) [016/j.ipm.2008.06.004. URL-https://doi.](https://doi.org/10.1016/j.ipm.2008.06.004) [Андруцопулос I, Баскиотис Н., Гауссье Е., Артьер Т., Галлинари.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0300)  [Evaluation](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0300) [fraorg/10.1016%2Fj.ipm.2008.06.004.](https://doi.org/10.1016/j.ipm.2008.06.004)  [Технические характеристики](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0300) [mework. 2013. результат](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0300) [проекта d4.1, 05/2013.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0300)

1. Мэннинг К., Сурденау М., Бауэр Дж., Финкель Дж., Бетхард С., Макклоски Д. Стэнфорд[61] [Баликас Г., Космопулос А., Критара А., Палиурас Г., Какадиарис И. Результаты](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0305) инструментария обработки естественного языка CoreNLP. Материалы 52-го ежегодного [собранияBioASQ задачи лаборатории ответов на вопросы на CLEF 2015. CLEF 2015:2015.](http://refhub.elsevier.com/S0933-3657(18)30275-6/sbref0305)  Ассоциации компьютерной лингвистики: демонстрации систем 2014. [62] Критара А., Нентидис А., Палиурас Г., Какадиарис И. Итоги 4-го издания [https://doi.org/10.3115/v1/p14-5010. URL-https://doi.org/. doi: вызов 10.3115/v1/](https://doi.org/10.3115/v1/p14-5010) BioASQ. Материалы Четвертого семинара BioASQ 2016. [https://doi.](https://doi.org/10.18653/v1/w16-3101)  [С14-5010.](https://doi.org/10.3115/v1/p14-5010) [org/10.18653/v1/w16-3101. URL-https://doi.org/10.18653%2Fv1%2Fw16-3101.](https://doi.org/10.18653/v1/w16-3101)
2. Портер М. Алгоритм снятия суффиксов. Программа: электронная библиотека и

1. <http://www.bioasq.org/> 2 <http://www.bioasq.org/participate/fifth-challenge-winners> [↑](#footnote-ref-1)
2. <https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?> [↑](#footnote-ref-2)
3. <http://metamap.nlm.nih.gov/> [↑](#footnote-ref-3)
4. <http://data.bioontology.org/documentation> [↑](#footnote-ref-4)
5. <http://participants-area.bioasq.org/results/5b/phaseB/>13 <http://participants-area.bioasq.org/oracle/> [↑](#footnote-ref-5)
6. <https://sites.google.com/site/mouradsarrouti/datasets> [↑](#footnote-ref-6)
7. <https://sites.google.com/site/mouradsarrouti/datasets> [↑](#footnote-ref-7)
8. <http://www.bioasq.org/participate/fifth-challenge-winners> [↑](#footnote-ref-8)