Журнал биомедицинской информатики 44 (2011) 277–28

[8](http://dx.doi.org/10.1016/j.jbi.2011.01.004)



Списки содержимого доступны по адресу

НаукаДирец

[t](http://www.sciencedirect.com/science/journal/15320464)

Журнал биомедицинской информатики

домашняя страница журнала:

www.elsevier.com/locate/yjbi

[n](http://www.elsevier.com/locate/yjbin)



AskHERMES: Онлайн-система ответов на вопросы для сложных клинических вопросов

Юнган Цао а,1, Фэйфань Лю а, Пиппа Симпсон б, Ламонт Антио а, Эндрю Беннетт к,д, Джеймс Дж.

a Департамент медицинских наук, Университет Висконсин-Милуоки, Милуоки, Висконсин, США b Департамент педиатрии, Медицинский колледж Висконсина, Милуоки, Висконсин, США c Департамент психиатрии, Медицинский колледж Висконсина, Милуоки, Висконсин, Висконсин, США d Госпиталь по делам ветеранов, Милуоки, Висконсин, США e Клинический центр, Национальные институты здравоохранения, Бетесда, доктор медицинских наук, США

f Департамент семейной медицины, больницы и клиники Университета Айовы, Айова-Сити, США

gДепартамент электротехники и компьютерных наук, Университет Висконсин-Милуоки, Милуоки, штат Висконсин, США

|  |  |  |
| --- | --- | --- |
| а р т и к л е | до n f o | а б с т р а с т |

|  |  |
| --- | --- |
| История статьи:  Получено 20 апреля 2010 г.  Доступно онлайн 21 января 2011  Ключевые слова:  Ответы на клинические вопросы  Анализ вопросов  Извлечение прохода  Уплотнения  Презентация ответов | Цель: Клинические вопросы часто являются длинными и сложными и принимают различные формы. Мы создали клиническую систему ответов на вопросы под названием AskHERMES для выполнения надежного семантического анализа сложных клинических вопросов и вывода в качестве ответов экстрактивных резюме, ориентированных на вопросы.  Дизайн: В данной статье описывается архитектура системы и предварительная оценка AskHERMES, которая реализует инновационные подходы к анализу вопросов, обобщению и представлению ответов. В этой системе были проиндексированы пять типов ресурсов: рефераты MEDLINE, полнотекстовые статьи PubMed Central, документы eMedicine, клинические рекомендации и статьи Википедии.  Измерение: Мы сравнили систему AskHERMES с Google (Google и Google Scholar) и UpToDate и попросили врачей оценить три системы по простоте использования, качеству ответа, затраченному времени и общей производительности.  Результаты: AskHERMES позволяет врачам вводить вопрос естественным образом с минимальной формулировкой запроса и позволяет врачам эффективно перемещаться между всеми предложениями ответов, чтобы быстро удовлетворить свои информационные потребности. Напротив, врачи должны формулировать запросы для поиска информации в Google и UpToDate. Развитие системы AskHERMES все еще находится на ранней стадии, а ресурс знаний ограничен по сравнению с Google или UpToDate. Тем не менее, результаты оценки показывают, что производительность AskHERMES сопоставима с другими системами. В частности, отвечая на сложные клинические вопросы, он демонстрирует потенциал превзойти системы Google и UpToDate. Выводы: AskHERMES, доступный в [http://www.AskHERMES.org](http://www.AskHERMES.org/), может помочь врачам практиковать доказательную медицину и улучшить качество ухода за пациентами. |

2011 Elsevier Inc. Все права защищены.

# 1. Введение

Врачи генерируют до шести вопросов для каждой встречи с пациентом [1–6], и эти вопросы могут быть различных типов. Хотя для врачей важно удовлетворять свои информационные потребности, исследования показали, что многие из их вопросов остаются без ответа. Например, Эли и его коллеги заметили, что врачи не искали ответы на 45% из 1062 вопросов, поставленных в клинических условиях, часто потому, что они сомневались, что смогут быстро найти хорошие ответы, а для тех, на кого они искали ответы, они потерпели неудачу.

|  |
| --- |
| ⇑ Автор-корреспондент. Адрес: 2400 E Hartford Ave., Room 939, Milwaukee, WI 53211, США. Факс: +1 414 229 5100.  Адрес электронной почты: hongyu@uwm.edu (H. Yu). 1 Нынешний адрес: Amazon.com, Inc., Сиэтл, Вашингтон, США.  1532-0464/$ - см. front matter 2011 Elsevier Inc. Все права защищены.  doi:[10.1016/j.jbi.2011.01.004](http://dx.doi.org/10.1016/j.jbi.2011.01.004) |

найти ответы на 41% из них [7].. В результате более 67% клинических вопросов, поставленных медиками, остались без ответа.

Одним из способов удовлетворения информационных потребностей является обращение к опубликованной литературе для получения соответствующих клинических данных [8]. Хотя оригинальные исследовательские статьи, которые являются как научно строгими, так и клинически значимыми, появляются в высоких концентрациях только в нескольких избранных журналах (например, The New England Journal of Medicine, Annals of Internal Medicine, JAMA и Archives of Internal Medicine), многие клинические доказательства появляются в широком спектре других биомедицинских журналов [9] . Даже с развитием поисковых систем для облегчения поиска соответствующей биомедицинской литературы потребности врачей все еще не могут быть удовлетворены должным образом, поскольку оценочное исследование показало, что поставщику медицинских услуг требуется в среднем более 30 минут для поиска ответа от MEDLINE, что сделало «этот тип поиска информации практичен только «в нерабочее время», а не в клинических условиях» [10].

Поисковые системы в Интернете (например, Google) предоставляют еще одно решение для врачей, ищущих ответы на свои вопросы [11–13].. Однако успех поиска в Интернете часто зависит от квалифицированных врачей [14], а поиск в Интернете неизбежно создает проблемы с точки зрения связанности и качества информационного контента [15–24]. Кроме того, традиционные поисковые системы (например, Google) возвращают длинные списки статей, а не автономные ответы на конкретные вопросы, и многие из этих статей оказываются нерелевантными для конкретных вопросов из-за неизбежной неоднозначности запросов поисковых систем с открытым доменом. Например, в недавнем исследовании [25] PubMed показал себя лучше, чем Google Scholar, в поиске релевантных и важных литературных статей для ответа на конкретные вопросы, связанные с наркотиками.

Важность ответов на вопросы врачей, связанные с уходом за пациентами, побудила к разработке многих клинических ресурсов (например, UpToDate, Thomson Reuters, eMedicine, National Guideline Clearinghouse) для предоставления высококачественных резюме клинически значимой информации. Эти резюме, однако, написаны экспертами в этой области, которые вручную просматривают литературу, касающуюся конкретных медицинских тем. Как таковые, эти ресурсы могут быть ограничены по объему и своевременности. Оценочное исследование показало, что при предоставлении 10 наиболее часто используемых клинических ресурсов без ограничения по времени врачи смогли ответить только на 70% из 105 вопросов, случайно выбранных из 1062 вопросов, собранных Эли и его коллегами [26]. Другое исследование показало, что, несмотря на то, что UptoDate является главным целевым сайтом, используемым врачами, только 10,8% врачей используют его для проведения исследований редких заболеваний [27]. Фактически, Сообщалось, что UpToDate используется нечасто в ряде оценочных исследований [13,28–31].. Кроме того, по мере усложнения баз данных требуется соответственно больше времени для поиска ответа даже в коммерческих клинических базах данных. Например, одно оценочное исследование [32] показало, что поиск ответов в UpToDate занимает более четырех минут. Однако исследования показали, что, когда поиск занимает более двух минут, он, вероятно, будет оставлен [10,26].

Системы ответов на вопросы (QA) имеют потенциал для преодоления этих недостатков. Во-первых, чтобы максимизировать охват и повысить своевременность, они могут автоматически извлекать соответствующие знания из нескольких источников и обобщать результаты для формирования ответов на основе важных концепций, встроенных в вопрос. Во-вторых, для повышения эффективности они могут предоставлять краткие ответы, а не целые документы, что может помочь пользователям быстро определить полезную информацию. Однако из-за трудностей, с которыми сталкиваются машины при понимании текста, современные подходы к обеспечению качества в основном сосредоточены на ответах на фактоидные вопросы, основанные на извлечении фактов или конкретных типах вопросов, таких как вопросы определения. К сожалению, из-за потенциала для неточных результатов майнинга в целом, такие фактоидные системы контроля качества здравого смысла обеспечивают более низкое удобство использования по сравнению с появляющимися базами данных фактов, управляемыми вручную, такими как Wikipedia, Answers, Freebase и т. Д.

Наша цель состоит в том, чтобы выйти за рамки таких фактоидных систем и разработать систему контроля качества с возможностью обработки сложных вопросов, которые обычно задаются в клинической области, с помощью структурированной предметно-ориентированной онтологии. Предметно-ориентированные знания могут быть использованы для расширения возможностей системы автоматически отвечать на вопросы, ориентированные на сложные проблемы, а не просто на обнаружение фактов, что жизненно важно для ответа на вопросы, задаваемые в клинической области. Мы предполагаем, что знания предметной области могут значительно улучшить вычислительную модель для поиска информации (IR) на основе машинного обучения, а технологии интеллектуального анализа текста могут быть объединены с IR для представления семантически присущего обобщения ответов. Наше внимание сосредоточено на улучшении качества ответов, особенно в ответ на сложные клинические вопросы, так что вместо предоставления одного факта или списка документов система уменьшит человеческие усилия, извлекая наиболее подходящую информацию по данному вопросу из большого количества литературы в клинической области.

Наша полностью автоматизированная система AskHERMES – Help physicians Извлечение и считывание мультимедийной информации из литературы для ответа на их специальные клинические quEstionS [33–43] – автоматически извлекает, извлекает, анализирует и интегрирует информацию из нескольких источников, включая медицинскую литературу и другие информационные онлайн-ресурсы, для формулирования ответов в ответ на специальные медицинские вопросы.

Флетчер [9] определил три основных навыка, необходимых врачам для управления своими информационными потребностями: (1) найти потенциально релевантную информацию, (2) судить о лучшем из гораздо большего объема менее достоверной информации и (3) судить, обеспечивает ли лучшая полученная информация достаточные доказательства для принятия клинических решений. AskHERMES обращается к первым двум компонентам, находя и фильтруя клиническую информацию. Ранее мы обнаружили, что AskHERMES превосходит несколько других систем (например, PubMed) для ответа на вопросы определений [38,39].. В настоящее время AskHERMES пытается ответить на все типы клинических вопросов, и в этой статье сообщается о разработке, внедрении и оценке системы AskHERMES.

# 2. Предыстория

Ответы на вопросы можно считать продвинутой формой поиска информации. В 1990-х годах Text REtrieval Conference (TREC) поддерживала исследования в сообществе по поиску информации, предоставляя инфраструктуру, необходимую для крупномасштабной оценки методологий поиска текста. В 1999 году TREC представил трек ответов на вопросы (QA), и самые ранние экземпляры трека QA были сосредоточены на ответах на фактоидные вопросы (например, «Сколько калорий в Биг Маке?»). С 2003 года TREC занимается вопросами сценариев (например, вопросами определения, такими как «Что такое X?»), которые требуют длинных и сложных ответов. TREC Genomics ввела поиск пассажа для ответов на вопросы в области геномики [44,45]..

Разработка в области ответов на вопросы в основном была сосредоточена на улучшении производительности извлечения базовых ответов, особенно по сравнению со стандартным набором вопросов, что, например, было целью трека обеспечения качества в TREC [46]. Однако усовершенствования таких экспериментов с пакетным запуском могут не привести к реальным выгодам для конечных пользователей [47].. На сегодняшний день немногие системы ответов на вопросы сосредоточены на разработке эффективных интерфейсов. Многие существующие коммерческие поисковые системы, такие как Google, Yahoo и Bing, возвращают только длинный ранжированный список релевантных документов, интерфейс, который аналогичным образом используется PubMed в биомедицинской области. Как обсуждалось ранее, таких интерфейсов недостаточно для предоставления кратких и релевантных ответов, что является особенно актуальным вопросом для врачей, у которых мало времени для просмотра длинных списков извлеченных документов. Кластеризация результатов поиска, техника визуализации, впервые представленная в системе Scatter–Gather [48], пытается предоставить пользователю существенную информацию о структуре тем в полученных результатах, и аналогичный подход был применен в поиске медицинской литературы [49] и усовершенствован несколькими поисковыми сайтами, включая Vivisimo, iBoogie и систему Carrot. Однако выходные данные этих систем по-прежнему основаны на традиционных результатах поиска, и пользователи должны читать ранжированный список документов даже для запроса по одной теме.

Решение вопросов, отвечающих на клинические вопросы, было активным усилием биомедицинского сообщества. Чимино и др. [50] пометил клинические вопросы семантически, чтобы сделать их общими (например, «Вызывает ли аспирин язвы» стал «Вызывает ли <лекарство> вызывает <диза>»). Цвайгенбаум [51,52] исследовал возможность ответов на вопросы в биомедицинской области. Ринальди и его коллеги [106] адаптировали систему ответов на вопросы с открытой областью для ответа на геномные вопросы (например, «Где наблюдался спонтанный апоптоз?»). Проект EpoCare (Evidence at Point of Care) предложил основу для предоставления врачам наилучшей доступной медицинской информации как из литературы, так и из клинических баз данных [53,54]. Infobuttons [8,31,55–62] служили медицинским порталом для внешних информационно-поисковых систем (например, PubMed) и баз данных (например, UpToDate). Проект CIQR (Context Initiated Question Response) [63] фокусируется на анализе типов вопросов, задаваемых клиницистом при поиске ссылок, что позволяет говорить в клинических условиях [64]. Чтобы включить конкретную информацию о пациенте в поиск релевантных и актуальных доказательств, система PERSIVAL (Personalized Retrieval and Summarization of Image, Video and Language Resource) [65–67] была разработана для обеспечения персонализированного доступа к распределенной цифровой библиотеке.

Автоматический анализ клинических вопросов является важным шагом на пути к ответам на клинические вопросы. Врачи часто задают сложные и многословные вопросы, включающие в себя широкий спектр типов. Типология типов вопросов с репрезентативными примерами, собранными в четырех исследованиях [7,26,68,69], показана в таблице 1. Существует множество исследований, предлагающих способы категоризации таких специальных вопросов. Эли и его коллеги вручную сопоставили 1396 клинических вопросов [70] с набором из 69 типов вопросов (например, «Какова причина симптома X?» и «Какова доза препарата X?») и 63 медицинских тем (например, лекарство или кардиология). Чимино и его коллеги [50] предварительно определили набор общих типов вопросов (например, «Что такое лечение болезни?»), а затем сопоставили специальные клинические вопросы с этими типами. Соль и его коллеги [71] выделили четыре основных типа вопросов: лечение, диагностика, этиология и прогноз. Такие типологии предлагают различные решения для автоматизированных систем для преодоления широкого спектра изменчивости в формах, которые могут принимать клинические вопросы. Другие исследователи применили популярную структуру «Население, вмешательство, сравнение и исход» (PICO) как способ борьбы с изменчивостью клинических вопросов [53,54, 72–75].

Информационно-поисковый компонент может быть интегрирован с анализом вопросов в системе ответов на вопросы для поиска соответствующих документов. Для поиска информации были разработаны различные модели, в том числе булевы модели [76], векторное пространство

Таблица 1

модели [77],онтологические подходы [78], латентная семантическая индексация [79,80] и языковые модели [81]..

Кроме того, многие существующие системы обращаются к внешним знаниям для поддержки более глубокого семантического анализа в ответах на вопросы. SemRep [82,83] сопоставляет биомедицинский текст с концепциями Единой системы медицинского языка (UMLS) [84] и представляет концептуальные отношения с семантическими отношениями UMLS (например, TREATS, Co-OCCURS\_WITH и OCCURS\_IN). Система обобщения SemRep конденсирует понятия и их семантические отношения для создания краткого резюме [83]. Essie - это механизм поиска информации, разработанный и используемый в Национальной медицинской библиотеке США (NLM), который включает в себя расширение запросов на основе знаний и эвристическое ранжирование [85]. CQA-1.0 [72] разработан как клиническая система ответов на вопросы. В отличие от AskHERMES, который обрабатывает специальный вопрос на естественном языке и включает в себя в основном статистические подходы и подходы машинного обучения, CQA-1.0 требует, чтобы пользователь вводил вопрос с помощью структуры PICO (пациент, вмешательство, сравнение и результат), и обеспечивает семантический анализ на уровне документа и текста, выявляя элементы PICO и документы, которые имеют клиническое значение. Снейдерман и др. [82] Интегрировал три системы (SemRep, Essie и CQA-1.0) для достижения наилучшей системы поиска информации (которая превзошла каждую из трех систем) в ответ на клинические вопросы.

Однако ни одна из вышеупомянутых систем не доступна онлайн для тестирования. Хотя ответы на вопросы являются активной областью исследований, большинство онлайн-систем контроля качества, как показано в таблице 2, не применимы к клинической области. Вклад системы AskHERMES включает в себя:

1. Индивидуальная модель машинного обучения, которая автоматически извлекает информационные потребности из сложных клинических вопросов.
2. Динамическая модель для иерархической кластеризации предложений в качестве ответов и новая функция ранжирования предложений.
3. Новая модель представления ответов, в которой ответы, а не списки документов, организованы по ключевым словам, ориентированным на вопросы.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Типология типов вопросов с репрезентативными примерами, собранная Эли и его коллегами в четырех исследованиях. Левая колонка представляет собой долю общих типов вопросов, с которыми можно сопоставить 4654 вопроса; вопросы, начинающиеся с вопросов «Что», «Как», «Сделать» и «Может», составляют 2231 (или 48%), 697 (или 15%), 320 (или 7%) и 187 (или 4%) вопросов соответственно. Репрезентативные примеры находятся в правой колонке.   |  |  |  |  | | --- | --- | --- | --- | | Общий тип вопроса  (и в процентах) | Примеры вопросов | | | | «Что... '' (48%) | 1. В чем причина и лечение стоматита этого старика? 2. Что делать с кем-то, кто не поправляется от эпикондилита после того, как физиотерапия и нестероидные противовоспалительные препараты не сработали? | | | | «Как... '' (15%) | 3. Как долго следует оставлять пациента на Кумадине и гепарине? | | | | ''Делайте... '' (7%) | 4. Работают ли ингибиторы ангиотензина II как обычные ингибиторы ангиотензинпревращающего фермента для сохранения функции почек при легком диабете? | | | | «Может... '' (4%) | 5. Может ли Лорабид вызывать головные боли? | |  | | Другие (25%) | 6. Интересно, может ли этот пациент иметь вращательную манжету? | |  | | Таблица 2  Список онлайн-неклинических систем ответов на вопросы. | |  |  | | Система | | Домен | Характеристики | | AnswerBus ([http://www.answerbus.com/index.shtml)](http://www.answerbus.com/index.shtml) | | Открыть домен | Возвращает соответствующие документы из WWW в ответ на специальный вопрос. | | Спросить ([http://www.ask.com/)](http://www.ask.com/) | | Открыть домен | Возвращает соответствующие абзацы в ответ на конкретный вопрос. | | БрейнБуст ([http://www.answers.com/bb/)](http://www.answers.com/bb/) | | Открыть домен | Возвращает предложения, относящиеся к специальному вопросу. | | ЕАГЛи ([http://eagl.unige.ch/EAGLi/)](http://eagl.unige.ch/EAGLi/) | | Геномика domain | Возвращает документы MEDLINE в ответ на специальный вопрос по геномике. | | Старт (<http://start.csail.mit.edu/>) | | Открыть домен | Возвращает короткую фразу в ответ на фактоидный вопрос. | | Почему-Вопрос ([http://lands.let.ru.nl/cgi-bin/retrieve\_wikidoc.pl/)](http://lands.let.ru.nl/cgi-bin/retrieve_wikidoc.pl/) | | Открыть домен | Возвращает соответствующие абзацы в Википедии в ответ на вопрос типа «почему» | | НоуИТАлл (<http://www.cs.washington.edu/research/knowitall/>) | | Открыть домен | Возвращает список извлеченных отношений в ответ на предикативный запрос. | | Вольфрам Альфа ([http://www.wolframalpha.com](http://www.wolframalpha.com/)) | | Открыть домен | Вычислительный механизм знаний на основе внутренней базы данных | |

1. Наконец, AskHERMES является единственной онлайн-системой, которая пытается автоматически ответить на весь спектр сложных клинических вопросов.

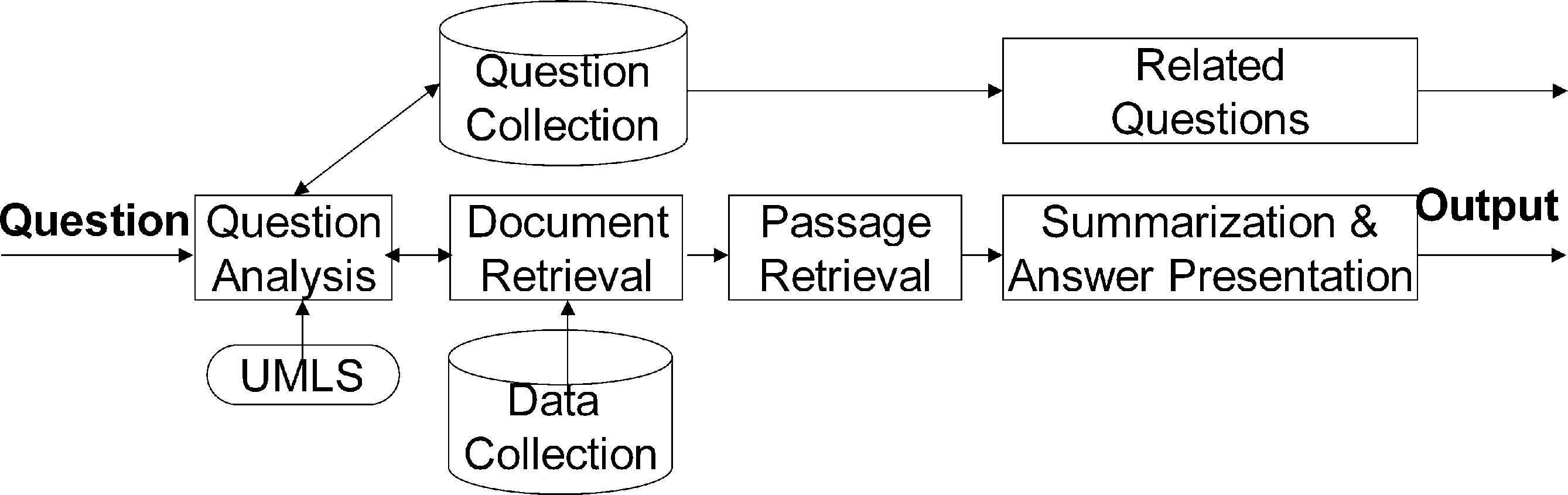


Рис. 1. Архитектура AskHERMES.

# 3. Методы

1 показана системная архитектура AskHERMES, которая принимает в качестве входных данных клинический вопрос естественного языка. Анализ вопросов автоматически извлекает информационные потребности из вопроса и выводит список терминов запроса. Ресурс знаний UMLS используется для расширения термина запроса. Модуль Извлечение связанных вопросов возвращает список похожих вопросов. Функция извлечения информации возвращает соответствующие документы, которые были локально проиндексированы. Извлечение информации идентифицирует соответствующие отрывки. Summaryation & Answer Presentation агрегирует отрывки из ответов, удаляет избыточную информацию, автоматически генерирует структурированные резюме и представляет резюме пользователю, задающему вопрос. В следующих разделах мы предоставим подробное описание каждого компонента.

## 3.1. Источники данных и предварительная обработка

### 3.1.1. Сбор данных

На момент оценки AskHERMES проиндексировал более 17 миллионов рефератов MEDLINE (1966–2008); 2732 электронных медицинских документа (загружено в 2008 году); 2254 клинических руководства (загружено в 2008 году); 167 000 полнотекстовых статей (загружено с PubMed Central в 2008 году); и 735 200 документов Википедии. Всего насчитывается 15 046 596 статей, содержащих 3 миллиона уникальных словесных токенов.

### 3.1.2. Предварительная обработка для сохранения смыслового содержания

Большинство подходов НЛП сосредоточены на добыче повествовательных текстов в различных документах или статьях, которые, однако, теряют семантическую информацию, встроенную в таблицы и списки. Это происходит довольно часто, особенно в клинически значимых статьях, таких как электронная медицина, клинические рекомендации и Википедия. Например, на рисунке 2 показана частичная таблица в статье из eMedicine. Клетки (например, «1994», «год», «2,5%» и «смерть») в таблице сами по себе не имеют большого значения, но вместе содержание может помочь правильно ответить на такие вопросы, как «Какова смертность от острых коронарных синдромов в 1994 году?» Аналогичным образом, группировка пунктов из списка на рисунке 3 позволяет AskHERMES ответить на такие вопросы, как «Что включает в себя навыки выживания для пациентов с диабетом?» Поэтому мы реализовали вручную курируемые правила для сохранения семантической информации, содержащейся в каждой таблице и списке. В частности, для каждой таблицы вся текстовая информация в каждой строке или столбце (в зависимости от расположения заголовка) вместе с соответствующим заголовком и текстом заголовка будет сформирована как отдельный отрывок, который будет индексироваться системой; для каждого списка мы рассматривали его как дерево, и каждый узел ветви без листьев (включая корневой узел) генерировал отдельный отрывок для индексации, сворачивая все узлы, которые он содержит, а затем объединяя соответствующую текстовую информацию с текстом заголовка. Тем временем мы удалили шумные короткие якорные тексты (например, «Свяжитесь с нами») из необработанного текста перед индексацией.

Другая стратегия, которую мы приняли при предварительной обработке, заключается в объединении всех названий разделов с любым предложением в соответствующем разделе, предполагая, что заголовки разделов в статьях несут важное семантическое содержание, которое не обязательно может быть явно описано в повествовательных абзацах. Несмотря на простоту такого подхода, он улучшил производительность AskHERMES. Например, объединив название «Лечение диабета 1 типа» с предложением «Одновременная трансплантация поджелудочной железы и почки является многообещающим решением, показывающим аналогичные или улучшенные показатели выживаемости только по сравнению с трансплантацией почки», объединенный текст может быть признан AskHERMES как ответ на вопрос «Каково лекарство от диабета 1 типа?», который в противном случае был бы пропущен, потому что одно предложение не включает слово «лечение», и само по себе название не дает описания потенциального излечения.

## 3.2. Анализ вопросов

В открытой области распространенным подходом к анализу вопросов является сопоставление вопросов в предопределенный шаблон вопроса (например, ''What-type'' и ''How-type'') [86]. Такой подход также был внедрен в большинство существующих онлайн-поисковых систем контроля качества (например, AnswerBus). Такие шаблонные подходы имеют значительные ограничения, поскольку они не могут справиться с вариациями, которые изобилуют в клинических вопросах. Например, три вопроса, показанные ниже, относятся к трем разным шаблонам, но требуют одного и того же ответа:

1. Как лечить полименорею у 14-летней девочки?
2. Каково лечение (или терапия) полименореи для 14-летней девочки?
3. Кто может подсказать мне лечение полименореи для 14-летней девочки?

Кроме того, многие клинические вопросы не могут быть сопоставлены с конкретным шаблоном, как показано в примере ниже:

Максимальная доза эстрадиола валерата составляет 20 миллиграммов каждые 2 недели. Мы используем 25 мг каждый месяц, что, кажется, контролирует ее приливы. Но достаточно ли этого для профилактики остеопороза и сердечно-сосудистых заболеваний?»

|  |
| --- |
| Рис. 2. Выдержка из частичной таблицы, появляющейся в статье eMedicine. |

Поэтому клиническая система контроля качества должна иметь возможность решать широкий спектр сложных вопросов, на многие из которых невозможно ответить подходами, зависящими от предопределенных шаблонов. Соответственно, мы разработали новые подходы к автоматическому извлечению информационных потребностей из сложных вопросов [40]. Во-первых, мы классифицируем вопрос на 12 общих тем, чтобы облегчить поиск информации. Эти темы включают устройство, диагностику, эпидемиологию, этиологию, историю, управление, фармакологию, физическое обнаружение, процедуру, прогноз, тест и лечение и профилактику, которые имеют

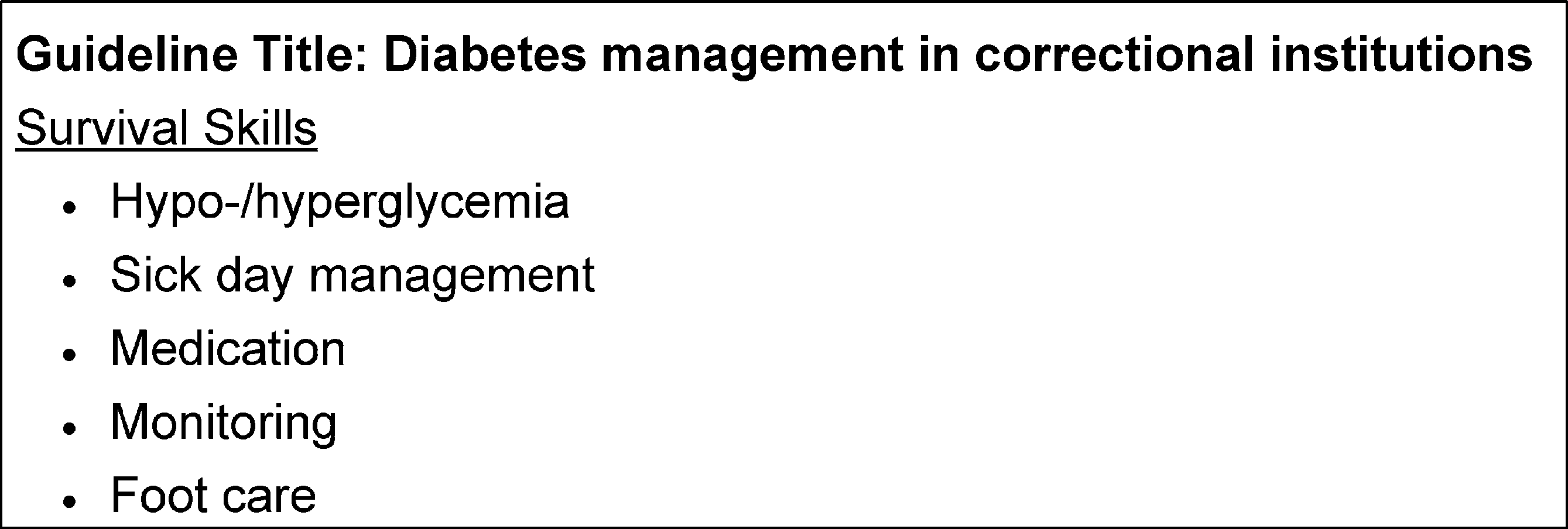


Рис. 3. Выдержка из перечня руководящих принципов в Национальном информационно-координационном центре руководящих принципов.

был использован для аннотирования 4654 клинических вопросов [40] клиницистами, которые записали вопросы. Например, приведенный выше вопрос представляет собой лекарство, и поэтому мы можем определить фармакологическую базу данных (например, Thomson Reuters) как лучший ресурс для потенциальной извлечения ответа. Во-вторых, ключевые слова, которые захватывают наиболее важное содержание вопроса, определяются автоматически. В том же примере с вопросом ключевыми словами являются «эстрадиол валерат» и «остеопороз и профилактика сердечно-сосудистых заболеваний». Ключевые слова могут использоваться в качестве терминов запроса для извлечения соответствующих документов, а также в качестве якорных терминов для извлечения ответов.

Мы разработали контролируемые подходы машинного обучения для автоматической классификации вопроса по общим темам и автоматического определения ключевых слов. Для классификации вопросов мы исследовали несколько алгоритмов обучения, показав, что машины опорных векторов (SVM) [87] достигли наилучшего результата. Поскольку вопрос может быть назначен нескольким темам, мы разработали двоичный классификатор (Да или Нет) для каждой из 12 тем. Для идентификации ключевых слов мы сформулировали ее как задачу маркировки последовательностей с использованием модели условных случайных полей ( CRFs) [88]. В дополнение к основным лексическим признакам (например, униграмма, биграмма) и синтаксическим особенностям (например, частям речи), мы включили лексический инструмент MMTx, реализацию MetaMap [89], для сопоставления текста с концепциями и семантическими типами UMLS в качестве функций обучения для обеих задач. Используя аннотированную коллекцию из 4654 клинических вопросов в качестве данных обучения и тестирования, наши результаты показывают в среднем 76,5% F1-баллов для классификации вопросов и 58% F1-баллов для извлечения ключевых слов. Подробности этой работы приведены в [40]..

## 3.3. Извлечение документов

Мы интегрировали последнюю версию вероятностной модели релевантности BM25 [90] с системой AskHERMES для поиска документов, поскольку она оказалась наиболее эффективной системой [90,91] для таких задач, как на недавней конференции Text REtrieval Conference (TREC), и обладали преимуществом простоты, интерпретации и скорости вычислений. Мы эмпирически настроили модель поиска в нашей системе.

## 3.4. Извлечение прохода

Предыдущая работа показала, что пользователи QA предпочитают, чтобы ответы были отрывками, а не предложениями [92]. Это особенно верно в нашей задаче, поскольку много важного содержания с точки зрения дискурсивных отношений (например, причинно-следственных и временных) отсутствует, если группы изолированных предложений извлекаются в качестве ответов. Большая часть работы, проделанной ранее, определила отрывок как естественный пункт [93] или фиксированное окно [94]. Однако отрывки-кандидаты, определяемые таким определением, иногда будут слишком многословными, чтобы иметь практическую ценность для ответов на вопросы. Поэтому мы разработали подход, который динамически генерирует границы прохождения.

В частности, мы определяем отрывок в AskHERMES как одно или несколько смежных предложений, в которых каждое предложение включает в себя один или несколько терминов запроса из вопроса. Наш подход отличается от TextTiling [95], популярного метода сегментации из нескольких абзацев, тем, что поставленный вопрос играет важную роль для распознавания пассажей в нашей системе.

В рамках этой работы мы определили новую функцию оценки для измерения сходства между предложением и вопросом (SS), которая интегрирует сходство между вопросом и предложением на уровне слов и последовательности слов в отрывке ответа кандидата, как показано в:

01

Ss 1/4 Sd TFq UTq B@~~qffiffiffiffiLCS22AC~~; s 2 dð1Þ

Lq þ Lp

Sd обозначает сходство вопрос-документ на основе сходства BM25, TFq - общее количество терминов запроса, которые появляются в предложении, UTq - уникальное количество терминов запроса в предложении, а LCS - сходство между предложением и всем вопросом, основанное на самой длинной общей подпоследовательности (LCS) [96].

LCS - это алгоритм, который идентифицирует самую длинную подпоследовательность, общую для всех последовательностей в наборе последовательностей (обычно только две) и признан очень важной для измерения сходства для многих приложений обработки текста, например, оценка суммирования (оценка ROUGE [97]). Включение оценки LCS в функцию оценки предложения может захватить более подробную информацию о зависимостях, чем просто мешок слов или даже биграммы. Например, учитывая вопрос «Как мне относиться к опоясывающему герпесу этого человека?», ответы-кандидаты, представленные предложениями (1) и (2) ниже, имеют те же слова и частоту, которые сопоставляются с извлеченными терминами запроса («лечить», «герпес» и «зостер»), что означает, что подход без LCS будет ранжировать два ответа одинаково. Тем не менее, LCS присваивает предложению (1) значение 3, а предложению (2) значение 2, давая предложению (1), которое является лучшим ответом для удовлетворения потребностей вопроса, более высокий рейтинг, чем предложение (2).

1. Кортикостероиды использовались для лечения опоясывающего герпеса гораздо дольше, чем противовирусные препараты, но влияние кортикостероидов на ПГН, по-видимому, не является последовательным.
2. У значительной части пожилых людей с опоясывающим герпесом развивается постгерпетическая невралгия (ПГН), хроническое состояние, которое трудно поддается лечению.

Как только оценка релевантности для каждого предложения получена, оценка прохождения Sp определяется эмпирическими метриками, показанными в:

( нннн

maxedSs1Это минс1Что; maxðSs1-й < 2 мин1Это

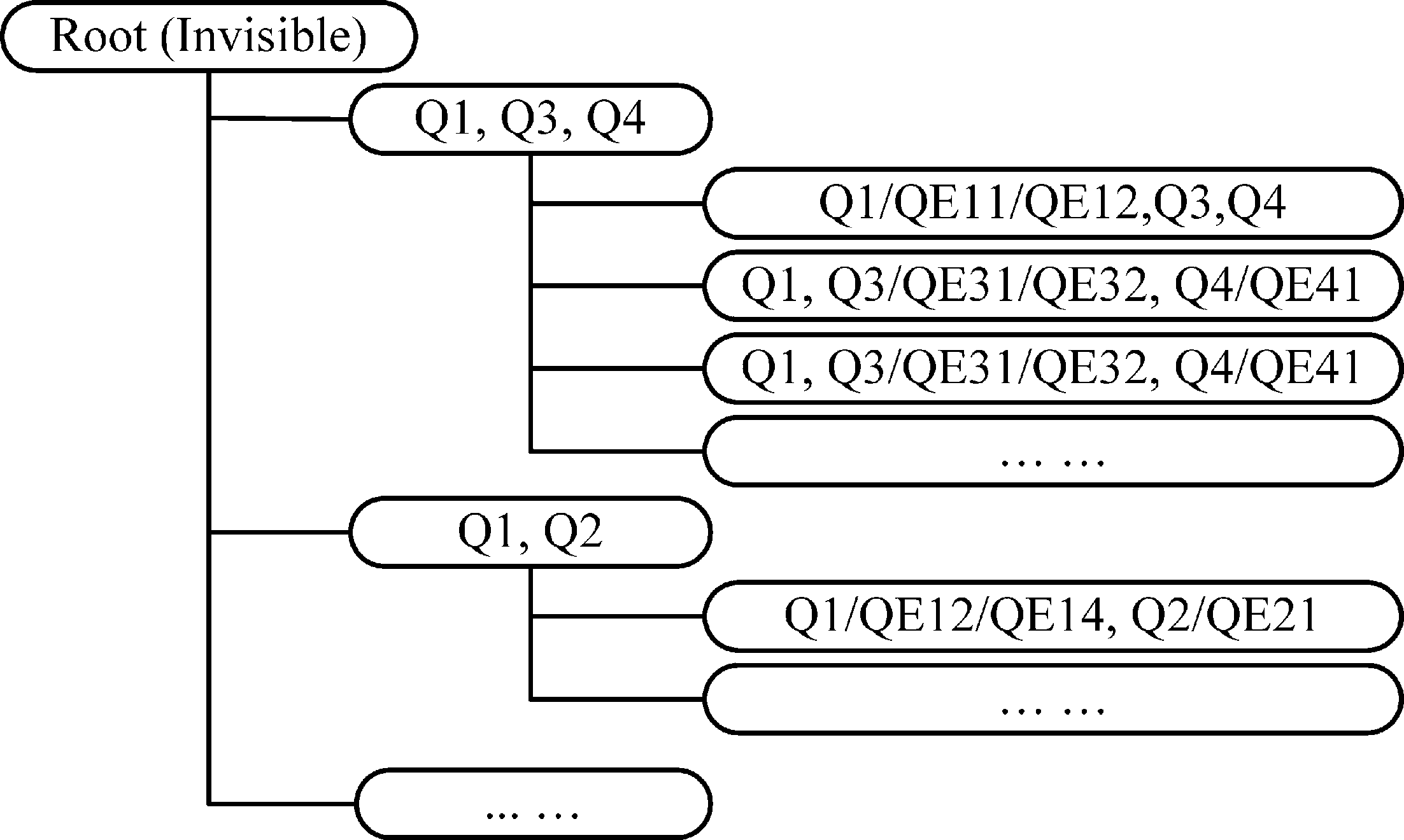
Sp 1/4 nð2Что максс1Что; иначе

где n — количество предложений в этом отрывке, max(Ss) — максимальный балл релевантности среди всех предложений, а min(Ss) — минимальный балл среди всех предложений.

## 3.5. Обобщение и представление ответов

Мы разработали новый подход к обобщению вопросов и представлению ответов, основанный на технике кластеризации. Как указывалось ранее, клинические вопросы, как правило, длинные и многословные и часто относятся к нескольким темам. Наша модель автоматического извлечения ключевых слов эффективно извлекает богатые содержанием ключевые слова из специальных вопросов, и такие ключевые слова затем можно использовать для иерархической структуры обобщенных ответов.

Например, в вопросе «Как мне лечить полименорею у 14-летней девочки?» термины «лечить», «полименорея», «14-летняя» и «девочка» являются четырьмя важными терминами содержания, и идеальный ответ включал бы все четыре термина. Однако в действительности большинство отрывков из ответов содержат меньше терминов содержания, а иногда и

содержат только один из четырех терминов. Мы предполагаем, что пользователи смогут более эффективно идентифицировать ответ, если ответы могут быть сгруппированы по терминам контента.

Еще одним преимуществом этой структуры является то, что если бы врачи были заинтересованы в поиске общего лечения «полименореи», они могли бы изучить группу ответов, содержащую «лечить» и «полименорею» без каких-либо возрастных терминов.

В ответ на эту мотивацию мы разработали новую систему обобщения, основанную на структурной кластеризации с использованием содержательных терминов, которая обеспечивает более удобный интерфейс представления ответов, чтобы помочь врачам быстро и эффективно просматривать кластеры ответов.

Рис. 5. Иллюстрация иерархической структуры кластеризации на основе терминов запроса.

|  |  |
| --- | --- |
| 3.5.1 . Тематическая кластеризация, ранжирование и иерархическое предисловие ответов  Чтобы расширить дух кластеризации результатов поиска, в этой статье мы предлагаем инновационный иерархический интерфейс представления ответов, в котором все соответствующие отрывки сгруппированы по различным темам на основе двухуровневой кластеризации. Мы представили отрывки ответов, связанные с запросами, которые были структурно сгруппированы на основе содержательных терминов запроса, а не просто предоставляли ранжированный список документов в качестве выходных данных. Метки тем назначаются каждому кластеру в AskHERMES с использованием терминов запроса и расширенных терминов из UMLS.  Используя тематическую древовидную структуру, созданную в результате кластеризации первого уровня, врачи могут легко найти интересующую информацию, прежде чем углубляться в более подробную информацию. Кроме того, кластеризация второго уровня предоставляет более совершенные категории для многогранных ответов. Каждый листовой узел представляет собой небольшой отрывок, а не документ, что облегчает просмотр.  Для кластеризации на основе терминов запроса мы используем исходные термины запроса (Q), которые появляются в вопросе, и термины расширения запроса UMLS (QE). Сначала мы группируем все синонимы (термин запроса и его расширенные термины представлены с помощью соответствующего термина запроса), а затем генерируем корневые кластеры, каждый из которых содержит различные комбинации этих синонимических понятий. Более формально мы предполагаем, что qi является i-м термином запроса в Q и имеет  Синонимы Mi qei1, qei2, qei3, ..., qeiMi. Под «группой вместе» мы подразумеваем, что мы используем qiдля обозначения всех синонимов в кластеризации, то есть i-го понятия. Отрывки, содержащие различные комбинации этих понятий в корневом узле, делятся на разные кластеры. | Различные варианты из каждой комбинации синонимов дополнительно приводят к иерархическим подкластерам.  4 дополнительно иллюстрирует, как работает алгоритм кластеризации на основе ковша, что приводит к типу иерархической структуры кластеризации, показанной на рисунке 5. Все ведра/кластеры могут содержать несколько прикрепленных проходов.  Все эти иерархические ковши генерируются динамически, что предотвращает взрыв комбинации, что особенно актуально для сложных вопросов. Для более компактного представления ответов мы игнорируем корневые узлы, к которым напрямую не прикреплены предложения, и продвигаем их дочерние ветви на один уровень выше. 5, в случаях, когда в узле ''Q1, Q3, Q4'' нет предложения, содержащего все три набора синонимических терминов, отображаются только его дочерние ветви.  Мы ранжируем кластеры на основе терминов запроса, отображаемых в кластере. Мы используем ту же стратегию ранжирования, что и взвешивание запроса в разделе 3.2 , и суммируем все веса термина запроса, которые встречаются как оценка ранжирования этого кластера. Мы также ранжируем сгенерированные корневые корзины, суммируя значения IDF концепции запроса, которую он охватывает. Поскольку в корневых корзинах для каждой концепции запроса существует несколько синонимов, нам нужно было найти способ выбрать значение IDF для каждой свернутой концепции. Мы выбрали минимальное значение IDF среди синонимов, основываясь на наблюдении, что некоторые общие слова редко используют синонимы с очень высокими значениями IDF.  Чтобы еще больше облегчить просмотр результатов, для кластеров с верхним рейтингом (не корневых кластеров), если количество проходов в них ex- |

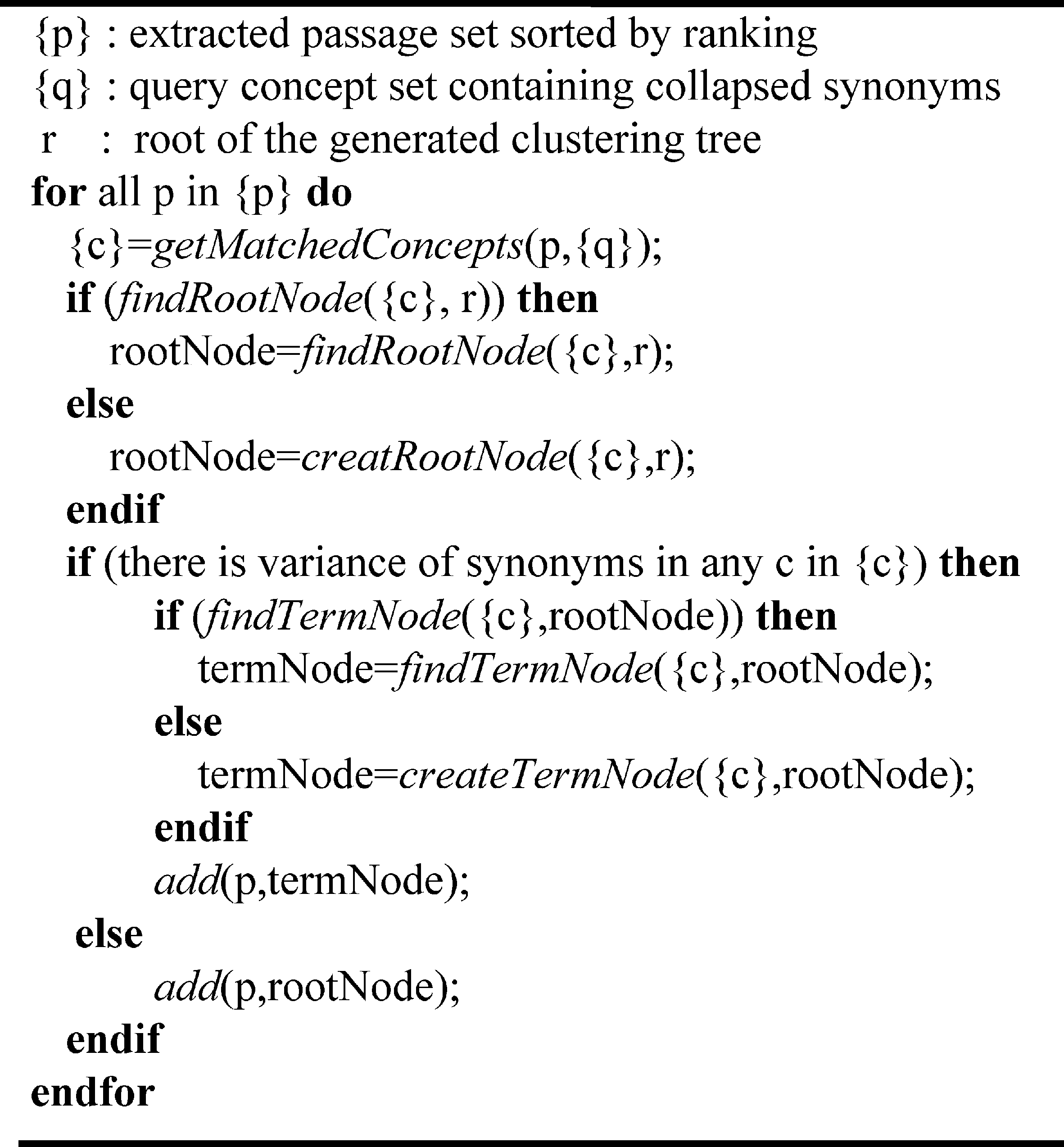


Рис. 4. Алгоритм кластеризации на основе терминов запроса.

Уступая q, мы проводим второй слой кластеризации на основе содержания пассажей по более утонченным семантическим измерениям. В нашей нынешней системе и p, и q эмпирически присваиваются 5. Для этой задачи мы использовали Lingo [98], который использует подход к декомпозиции с одним вектором для поиска общих меток для кластеров и извлекает соответствующее содержимое (отрывки из ответов кандидатов в нашей системе) для каждого кластера. Это может генерировать читаемые метки для кластеров и позволяет помещать проход в несколько кластеров вместо жесткого разделения.

### 3.5.2. Удаление избыточности на основе самой длинной общей подстроки

Поскольку ответы кандидатов извлекаются из нескольких источников, неизбежно, что они будут содержать некоторую избыточную информацию. Чтобы решить эту проблему, мы исследовали самую длинную общую подстроку (LCSubstring) [99], которая представляет собой алгоритм для идентификации самой длинной строки (или строк), которая является подстрокой (или являются подстроками) двух или более строк. Мы применили LCSubstring в нашей системе, чтобы устранить избыточность между предложениями, а также отрывками, которые были извлечены в качестве ответов-кандидатов. Разница между самой длинной общей подстрокой и самой длинной общей подпоследовательностью (LCS) заключается в том, что первая должна быть непрерывной подстрокой из исходных строк, в то время как последняя состоит из всех общих подпоследовательностей, которые имеют один и тот же порядок, но включают интервалы между исходными строками. Таким образом, самая длинная общая подстрока имеет больше ограничений, чем LCS, и мы используем ее для удаления избыточности в

наша система. Чтобы гарантировать, что две единицы достаточно похожи, чтобы считаться дублирующими, мы установили порог эмпирически. Хотя самая длинная общая подстрока использовалась для задач автоматической оценки суммирования, таких как ROUGE-L [100],, и использовалась для перефразирования [101,102], о ней еще не сообщалось об устранении избыточности как части суммирования.

# 4. Внедрение системы

Система AskHERMES построена на фреймворке J2EE, в котором JBoss используется для сервера приложений, а JBoss Seam для построения пользовательского интерфейса. JBoss имеет встроенное кэширование EJB (Enterprise JavaBeans) и механизм повторного использования, который обеспечивает доступ к большой нагрузке. Мы также создали циклический балансировщик нагрузки на фронтальном веб-сервере для распределения нагрузки доступа между шестью внутренними серверами. Шесть серверов работают под управлением операционных систем linux/solaris, в которых развернут AskHERMES. В настоящее время время отклика системы AskHERMES составляет в среднем 20 с.

# 5. Результаты

В этом разделе мы сообщаем о нашей пилотной оценке системы AskHERMES, которая сравнивается с двумя часто используемыми современными системами: коммерческой поисковой системой Google и системой клинических баз данных UpToDate.

## 5.1. Расчетный проект

Чтобы оценить нашу систему AskHERMES, мы случайным образом выбрали 60 вопросов из коллекции ClinicalQuestions [40] и попросили трех врачей (AB, JJC и JE) провести ручную оценку результата. Для сравнения, мы также оценили поисковую систему Google (используя как Google, так и Google Scholar) и систему баз данных UpToDate по одному и тому же набору вопросов. Наша цель состояла в том, чтобы изучить, насколько хорошо каждая из трех систем отвечает на вопросы, и определить четыре метрики в оценке следующим образом:

1. Простота использования (шкала от 1 до 5).

Таблица 3

Медианный оценочный балл (с межквартильным диапазоном, показанным в скобках) Google, UpToDate и AskHERMES.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Гугл | UpToDate | АскГЕРМЕС |
| Простота использования | 4 (3, 5) | 4 (4, 5) | 4 (3.75, 5) |
| Качество ответа | 3 (1, 4.25) | 4 (1, 5) | 2.5 (1, 4) |
| Затраченное время (ы) | 2.5 (2, 5) | 3 (2, 5) | 4 (2, 5) |
| Общая производительность | 3 (1, 4) | 4 (1, 5) | 3 (1, 4) |

Таблица 4

Тест с подписью Уилкоксона (основанный на отрицательных рангах) для общего сравнения производительности трех систем (Z — нормальное аппроксимационное значение; значение p указывает на уровень значимости).

|  |  |  |
| --- | --- | --- |
|  | С | значение p (2-хвостовое) |
| В целом: Google - Общий: AskHERMES | 0.800 | .423 |
| В целом: Актуально – В целом: AskHERMES | 1.604 | .109 |
| В целом: UptoDate - В целом: Google | 1.175 | .240 |

1. Качество ответа (шкала от 1 до 5).
2. Затраченное время (в с).
3. Общая производительность (шкала от 1 до 5).

Для этой пилотной оценки каждому врачу-субъекту был представлен взаимоисключающий набор из 20 вопросов, случайным образом выбранных из нашей коллекции вопросов. Для каждого вопроса каждому субъекту было предложено определить ответы от каждой из трех систем: AskHERMES, Google и UpToDate, а затем присвоить оценку для каждой системы по каждой метрике оценки, определенной выше.

5.2. Производительность AskHERMES по сравнению с Google и

# UpToDate

В таблице 3 приведены результаты работы трех систем. Эти результаты показывают, что оценка простоты использования AskHERMES была очень конкурентоспособной с тем же средним баллом оценки 4, что и обе другие системы, достигнув среднего балла 4,079 по сравнению с лучшим баллом 4,132 для UpToDate, что говорит о том, что наш интерфейс представления на основе кластеризации довольно эффективен и выгоден с точки зрения удобства для пользователя. Это также отражается в том, что

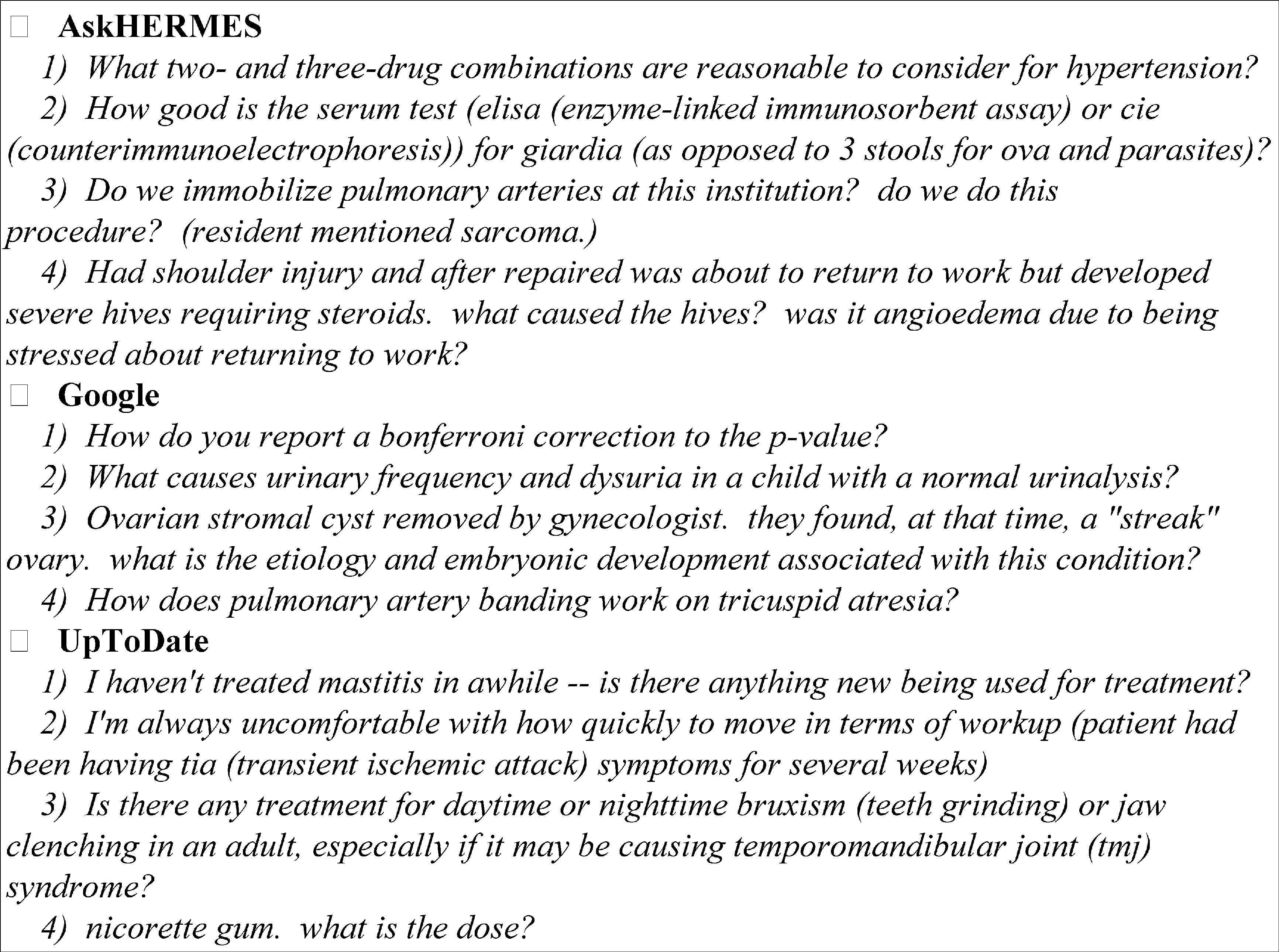


Рис. 6. Примеры вопросов, по которым лучше всего работают разные системы.

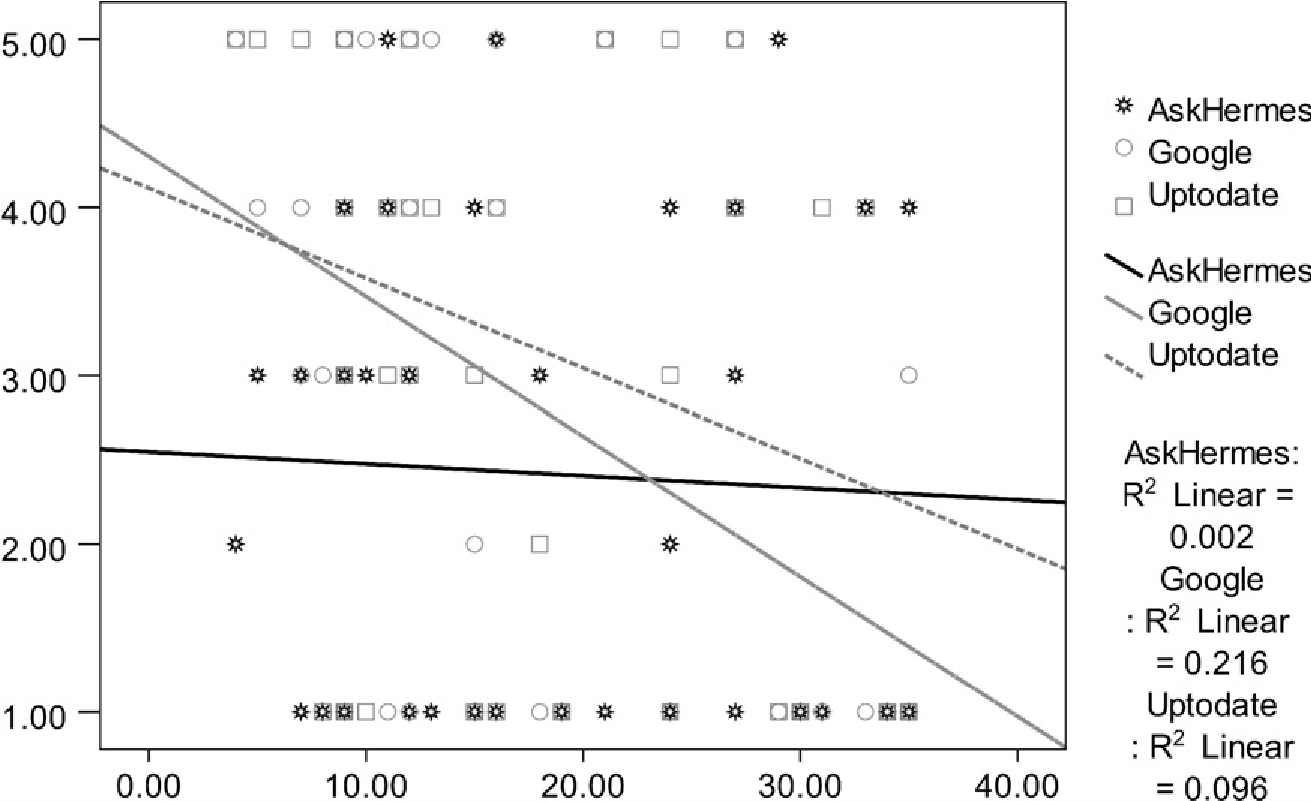


Рис. 7. Точечный график качества ответа и количества слов, о которых идет речь.

AskHERMES имеет более медленное время отклика, чем другие системы (задержка в несколько секунд), простота использования системы компенсирует это потерянное время, так что оценки Time Spent сопоставимы.

Что касается качества ответа, UpToDate получил лучший балл 4, что следует ожидать, поскольку он включает в себя богатый ресурс знаний для конкретной области, который более клинически ориентирован. Хотя AskHERMES получил самый низкий балл в этой категории, его отличие от Google в этом отношении не является статистически значимым (тест Wilcoxon signed, p > 0,1).

Аналогичным образом, оценка общей производительности AskHERMES была немного ниже, чем в системе UpToDate, но, как показывает таблица 4 , сравнение производительности трех систем попарно не показывает статистически значимых различий (p > 0,1) на основе двустороннего теста ранга с подписью Wilcoxon. Мы также обнаружили, что AskHERMES дал наименьшее стандартное отклонение в показателях качества ответа (1,445) и общей производительности (1,427) по сравнению с Google (1,706 / 1,603) и UpToDate (1,653 / 1,636), продемонстрировав свою надежность и способность адаптироваться.

## 5.3. Влияние длины вопроса на качество ответов

Чтобы получить более глубокое понимание качества работы AskHERMES, мы рассмотрели вопросы, по которым каждая система работала лучше всего, как показано на рисунке 6. Мы обнаружили, что каждая система имеет свои сильные стороны для различных видов вопросов; Как показано на рисунке 6, AskHERMES лучше всего отвечает на сложные вопросы в конкретных контекстах, таких как отношения, сравнения и ограничения; Google лучше всего справляется с ответами на короткие вопросы, на которые можно ответить в открытом домене; и UpToDate лучше всего справляется с ответами на короткие клинические вопросы по конкретным темам.

7 показана связь между качеством ответа и количеством слов в данном вопросе, демонстрируя, что производительность Google и UpToDate колеблется для разных вопросов, в то время как AskHERMES работает последовательно по разным вопросам. Мы заметили, что когда вопросы содержат более 25-35 слов, Google и UpToDate дают особенно плохие ответы по сравнению с нашей системой. Кроме того, мы обнаружили, что только тогда, когда количество слов в вопросе меньше 25, UpToDate работает статистически лучше (p < 0,04), чем AskHERMES для качества ответа. Эти результаты демонстрируют, что AskHERMES имеет большой потенциал для ответа на клинические вопросы, которые обычно длинные и сложные, несмотря на то, что он является полностью автоматической системой, в отличие от UpToDate, которая опирается на большое количество ручных усилий.

# 6. Обсуждение

В настоящее время разработка AskHERMES находится на ранней стадии, и ресурсы данных, которые он проиндексировал, очень ограничены. Кроме того, AskHERMES является автоматической системой контроля качества, в то время как UpToDate использует экспертов предметной области для ручного выбора только клинически связанных знаний. Кроме того, AskHERMES в настоящее время не имеет доступа ко всем полнотекстовым статьям по теме. По мере того, как полнотекстовые биомедицинские статьи с открытым доступом становятся все более доступными, мы предполагаем, что производительность AskHERMES будет значительно улучшена.

Наша пилотная оценка показала, что AskHERMES конкурентоспособен с другими клиническими информационными ресурсами. Общий оценочный балл AskHERMES не так высок, как у UpToDate, но статистически значимых различий не было, согласно тестам, подписанным Уилкоксоном. Обратите внимание, что это все еще предварительная оценка нашей системы, и статистический тест может быть недостаточно мощным в зависимости от текущего размера выборки, что, следовательно, требует большего исследования и проверки в более масштабной оценке с более независимыми выборками в будущей работе. Примечательно, что мы обнаружили, что AskHERMES может на самом деле работать лучше с более длинными и сложными клиническими вопросами, что говорит о том, что его интеграция компонента анализа вопросов и предметно-ориентированной онтологии дает AskHERMES возможность понимать и правильно распознавать информационные потребности врачей.

UpToDate компилируется экспертами, в то время как AskHERMES автоматически собирает информацию из текстов на естественном языке. Часто тексты, из которых AskHERMES извлекал информацию, не имеют клинической значимости, и поэтому нерелевантность приводит к снижению производительности. Кроме того, AskHERMES построен на ограниченных текстовых ресурсах, в то время как UpToDate и Google Scholar имеют гораздо более богатый ресурс, включая полнотекстовые статьи и электронные книги. Все вышеупомянутые факторы способствуют производительности AskHERMES. Мы подчеркиваем, что, несмотря на преимущества UpToDate и Google Scholar в их ресурсах, различия между AskHERMES и Google Scholar или UpToDate, как показано в таблице 4 , не были статистически значимыми. Кроме того, AskHERMES достиг того же общего показателя производительности (как показано в таблице 3), предполагая, что AskHERMES может превзойти UpToDate и Google Scholar, если доступны богатые ресурсы.

|  |
| --- |
| Рис. 8. AskHERMES' отвечает на вопрос ''В чем причина и лечение стоматита этого старика?" ' (оба фокуса освещены в сжатой форме).    Рис. 9. AskHERMES' ответы на вопрос ''В чем разница между Denver II и обычным Denver Developmental Screening Test?" ". |

С другой стороны, AskHERMES имеет ряд преимуществ перед Google и UpToDate. Во-первых, новое обобщение и презентация на основе кластеризации, которые он предлагает, имеют явное преимущество перед длинными списками документов, полученными из Google, с потенциалом сэкономить время занятых врачей при извлечении потенциально нерелевантных документов. Во-вторых, AskHERMES является полностью автоматической системой, обеспечивающей непревзойденное преимущество перед UpToDate в том, что она не полагается на трудоемкие, трудоемкие человеческие усилия по обслуживанию и обновлению системы. В-третьих, чтобы использовать UpToDate, врачи должны четко и лаконично формулировать свои информационные потребности и должны знать работу системы UpToDate, в то время как AskHERMES требует небольшой подготовки, поскольку система была разработана для выполнения этой работы самостоятельно.

Кроме того, есть некоторые специфические особенности нашей системы, выходящие за рамки самой оценки, которые можно резюмировать следующим образом:

1. Используя структурную кластеризацию, основанную на содержательных терминах запроса, AskHERMES может обрабатывать вопросы, охватывающие несколько направлений или тем. Например, в вопросе «Какова причина и лечение стоматита этого старика?» есть два очага: «причина» и «лечение», и очень трудно найти одно предложение или лаконичный отрывок, который мог бы охватить и то, и другое. AskHERMES может автоматически разделять ''причину'' и ''лечение'' на основе кластеризации на основе терминов запроса, как показано на рисунке 8.
2. Ранжирование и сопоставление на основе LCS позволяет AskHERMES немедленно определить лучший ответ, когда ответ достаточно похож на вопрос, независимо от того, насколько сложным является вопрос. 9 показан ответ на вопрос «В чем разница между Denver II и обычным Denver Developmental Screening Test?», где предложение с самым высоким рейтингом в выходных данных нашей системы является правильным ответом.
3. Чтобы помочь врачам легко получать информацию с разных точек зрения, интерфейс представления ответов в AskHERMES предоставляет как кластеризованные ответы (проиллюстрированные на рис. 8), так и ранжированные ответы (проиллюстрированные на рис. 10). Если вопрос пользователя достаточно прост или очень специфичен, пользователь может быстрее найти ответы из ранжированных ответов. Кроме того, связанные вопросы (рис. 10), полученные из нашей коллекции вопросов, также предоставляются нашим «интерактивным» интерфейсом, чтобы помочь врачам, которые могут захотеть просмотреть ответы на связанные вопросы.

Таким образом, ответы на клинические вопросы являются очень сложной задачей, и ни одна современная система не всегда может хорошо работать с множеством вопросов, которые могут быть заданы ей. AskHERMES предоставляет практическую и конкурентоспособную альтернативу, чтобы помочь врачам найти ответы.

# 7. Выводы и будущая работа

Мы представляем нашу онлайн-систему ответов на клинические вопросы AskHERMES, которая призвана помочь врачам быстро удовлетворить свои информационные потребности. Система опирается на использование контролируемых и неконтролируемых методов обучения в различных компонентах для изучения различных лингвистических особенностей. AskHERMES в настоящее время может

|  |
| --- |
| Рис. 10. Иллюстрация интерфейса AskHERMES для ранжированных ответов и связанных с ними вопросов на вопрос «Какова причина и лечение стоматита этого старика?» ". |

анализировать и понимать сложные клинические вопросы различных типов, на которые нельзя ответить фактоидами или отдельными предложениями.

Наша пилотная оценка показывает, что AskHERMES работает сопоставимо с такими современными системами, как Google и UpToDate. В частности, наша система демонстрирует лучшую способность отвечать на длинные и сложные клинические вопросы, чем другие системы, демонстрируя надежность по вопросам с различным количеством слов. В целом, согласно нашим предварительным результатам, статистически значимых различий между AskHERMES и двумя другими системами не было.

Поскольку AskHERMES в настоящее время не интегрирует компонент идентификации клинических доказательств, который вручную вводится UpToDate, мы планируем разработать автоматическую систему распознавания клинической информации для дальнейшего повышения качества ответа AskHERMES. Вместо поиска на основе документов мы будем исследовать извлечение клинической информации непосредственно на основе единиц прохождения, для которых мы также проанализируем более систематические способы интеграции информации о ключевых словах. Кроме того, мы ищем более эффективные способы повышения точности расширения запросов, как описано в [103–105]. Наконец, будущее направление исследований заключается в проведении более обширных оценок системы AskHERMES, включая внутреннюю оценку кластерного подхода для обобщения, внешнюю оценку всей системы с использованием большего набора данных, а также сравнение с существующими системами, такими как система Essie ([http://essie.nlm.nih. gov/](http://essie.nlm.nih.gov/)) и Semantic Medline (<http://skr3.nlm.nih.gov/SemMedDemo/>).

# Подтверждения

Авторы признают поддержку со стороны Национального института здравоохранения (NIH), номер гранта 2R01LM009836. Любые мнения, выводы или рекомендации принадлежат авторам и не обязательно отражают взгляды NIH.

# Ссылки

1. Тимпка Т., Арборелиус Е. Дилеммы врача общей практики: изучение необходимости и использования знаний во время консультаций по вопросам здравоохранения. Methods Inform Med 1990;29:23–9.
2. Bergus GR, Randall CS, Sinift SD, Rosenthal DM. Влияет ли структура клинических вопросов на исход консультаций со специалистами по специальности? Arch Family Med 2000;9:541–7.
3. Эли Джей Ви, Берч РЖ, Винсон Д.К. Информационные потребности семейных врачей: клинические вопросы для конкретных случаев. J Family Pract 1992;35:265–9.
4. Ошерофф ДЖА, Форсайт ДЕ, Бьюкенен БГ, Банковиц РА, Блюменфельд БХ, Миллер РА. Информационные потребности врачей: анализ вопросов, поставленных в ходе клинического обучения. Ann Int Med 1991;114:576–81.
5. Covell DG, Uman GC, Manning PR. Информационные потребности в офисной практике: удовлетворяются ли они? Ann Int Med 1985;103:596–9.
6. Смит Р. Какая клиническая информация нужна врачам? BMJ 1996;313:1062–8.
7. Эли JW, Ошерофф JA, Чамблисс ML, Ebell MH, Розенбаум ME. Ответы на клинические вопросы врачей: препятствия и потенциальные решения. J Am Med Inform Assoc 2005;12:217–24.
8. Cimino JJ, Li J, Graham M, Currie LM, Allen M, Bakken S, et al. Использование онлайн-ресурсов при использовании клинической информационной системы. В: AMIA ann symp proc; 2003. С. 175–9.
9. Флетчер Р.Х., Флетчер С.У. Доказательный подход к медицинской литературе . J Gen Int Med 1997;12(Suppl. 2):S5–S14.
10. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Friedman CP, Tidmarsh P и др. Факторы, связанные с успехом в поиске MEDLINE и применении доказательств для ответа на клинические вопросы. J Am Med Inform Assoc

2002;9:283–93.

1. Каллен РЖ. В поисках доказательств: использование семейными врачами Интернета для получения клинической информации. J Med Lib Assoc 2002;90:370–9.
2. Стивенс МБ, фон Тун АМ. Военно-медицинская информатика: доступ к информации в развернутой среде. Военная медицина 2009;174:259–64.
3. Китчин Д.Р., Эпплгейт К.Е. Изучение радиологии обследование, исследующее радиологию, использование учебников, журналов и Интернета. Акад Радиол 2007;14:1113–20.
4. Тан Х, Нг Дж. Поиск в Google для постановки диагноза - использование Google в качестве диагностического пособия: интернет-исследование. BMJ 2006;333:1143–5.
5. Перселл ГП, Уилсон., Деламот Т. Качество медицинской информации в Интернете. BMJ 2002;324:557–8.
6. Джадад А.Р., Гальярди А. Рейтинг медицинской информации в интернете: навигация к знаниям или к Вавилону? ДЖАМА 1998;279:611–4.
7. Зильберг ВМ, Лундберг ГД, Мусаккьо РА. Оценка, контроль и обеспечение качества медицинской информации в Интернете: Caveant lector et viewor – пусть читатель и зритель остерегаются. ДЖАМА 1997;277:1244–5.
8. Гленни Э., Кирби А. Карьера рентгенолога: информация в Сети. J Diag Radiogr Imaging 2006;6:25–33.
9. Чайлдс С. Оценка качества медицинской информации в Интернете. Perform Meas Metrics 2005;6:80–96.
10. Griffiths KM, Tang TT, Hawking D, Christensen H. Автоматизированная оценка качества сайтов депрессии. J Med Int Res 2005; 7:е59.
11. Griffiths KM, Christensen H. Качество веб-информации о лечении депрессии: поперечное обследование. BMJ 2000;321:1511–5.
12. Уайетт Джей Си Комментарий: измерение качества и влияния Всемирной паутины. BMJ 1997;314:1879–81.
13. Бенигери М., Плуйе. Недостатки медицинской информации в Интернете. Укрепление здоровья Int 2003;18:381–6.
14. Cline RJ, Haynes KM. Поиск информации о здоровье потребителей в Интернете: состояние техники. Здоровье Educ Res 2001;16:671–92.
15. Freeman MK, Lauderdale SA, Kendrach MG, Woolley TW. Google Scholar против PubMed в поиске первичной литературы, чтобы ответить на вопросы, связанные с наркотиками. Энн Фармакотер 2009;43:478–84.
16. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML. Evans ER: анализ вопросов, задаваемых семейными врачами относительно ухода за пациентами. BMJ 1999;319:358–61.
17. De Leo G, LeRouge C, Ceriani C, Niederman F. Веб-сайты, наиболее часто используемые врачом для сбора медицинской информации. AMIA Анну Симп Прок 2006:902.
18. McCord G, Smucker WD, Selius BA, Hannan S, Davidson E, Schrop SL и др. Ответы на вопросы в месте оказания помощи: практикуют ли жители EBM или управляют источниками информации? Акад Мед 2007;82:298–303.
19. Goodyear-Smith F, Kerse N, Warren J, Arroll B. Оценка электронных учебников. DynaMed, MD Consult и UpToDate. Aust Family Phys 2008;37:878–82.
20. Фуа Дж, Лим Т.К. Как резиденты и стажеры используют и воспринимают персонального цифрового помощника и UpToDate. BMC Med Educ 2008;8:39.
21. Чимино Джей Джей, Бороцов Д.В. Приведение лошади к воде: использование автоматических напоминаний для увеличения использования онлайн-поддержки принятия решений. AMIA Анну Симп Прок

2008:116–20.

1. Hoogendam A, Stalenhoef AF, Robbe PF, Overbeke AJ. Ответы на вопросы, задаваемые во время ежедневного ухода за пациентами, с большей вероятностью будут даны UpToDate, чем PubMed. J Med Int Res 2008; 10:е29.
2. Yu H, Sable C, Zhu HR. Классификация медицинских вопросов на основе таксономии электронных доказательств. В: Материалы семинара AAAI 2005 года по ответам на вопросы в ограниченных областях; 2005 год.
3. Ю Х, Соболь С. Быть Эрланг Шен. Определение ответных вопросов. В: Материалы девятнадцатой международной совместной конференции по искусственному интеллекту по знаниям и рассуждениям для ответов на вопросы; 2005.
4. Ю. Х. К ответам на биологические вопросы с экспериментальными доказательствами. Автоматическая идентификация текста, который суммирует содержимое изображения в полнотекстовых статьях. AMIA Анну Симп Прок 2006:834–8.
5. Lee M, Cimino J, Zhu HR, Sable C, Shanker V, Ely J, et al. Помимо поиска информации - медицинский вопрос. AMIA Анну Симп Прок 2006:469–73.
6. Lee M, Wang W, Yu H. Изучение контролируемых и неконтролируемых методов обнаружения тем в биомедицинском тексте. BMC Биоинформатика 2006;7:140.
7. Ю Х., Кауфман К. Когнитивная оценка четырех онлайн-поисковых систем для ответа на вопросы определения, поставленные врачами. Pacific Symp Biocomput 2007;12:328–39.
8. Yu H, Lee M, Kaufman D, Ely J, Osheroff J, Hripcsak G, et al. Разработка, реализация и когнитивная оценка дефиниционной системы ответов на вопросы для физицийцев. J Biomed Inform 2007;40:236–51.
9. Ю Х, Цао Ю. Автоматическое извлечение информационных потребностей из специальных клинических вопросов. AMIA Анну Симп Прок 2008:96–100.
10. Ю Х, Цао Ю. Использование взвешенных моделей ключевых слов для улучшения поиска биомедицинской информации. В: Саммит AMIA по трансляционной биоинформатике, Сан-Франциско, США; 2009 год.
11. Cao YG, Ely J, Antieau L, Yu H. Оценка презентации клинического ответа на вопрос. БиоНЛП 2009.
12. Ю Х, Цао Ю. Использование взвешенных моделей ключевых слов для улучшения ответов на клинические вопросы. В: IEEE международная конференция по биоинформатике и биомедицине семинар НЛП подходы для неудовлетворенных информационных потребностей в автомобилях здравоохранения; 2009.
13. Hersh W, Cohen A, Ruslen L, Roberts P. TREC 2007 genomics track overview. В: Конференция трека геномики TREC; 2007 год.
14. Hersh W, Cohen A, Roberts P, Rekapalli H. TREC 2006 genomics track overview. В: Конференция по геномике TREC; 2006 год.
15. Вурхиз ЭМ. Отчет о треке ответов на вопросы TREC-8. В: Труды ТРЕК; 1999 год.
16. Херш В. Качество информации во Всемирной паутине. J Am Coll Dent 1999;66:43–5.
17. Херст М., Педерсен Дж. Пересмотр кластерной гипотезы: рассеяние/сбор по результатам поиска. В: 19-я ежегодная международная конференция ACM по исследованиям и разработкам в области информационного поиска (SIGIR-96); 1996.
18. Пратт В. Динамическая организация результатов поиска с помощью UMLS. Proc AMIA Annu Fall Symp 1997:480–4.
19. Cimino JJ, Aguirre A, Johnson SB, Peng P. Общие запросы для удовлетворения потребностей в клинической информации. Bull Med Lib Assoc 1993;81:195–206.
20. Цвайгенбаум. Ответы на вопросы в биомедицине. В: Семинар EACL по обработке естественного языка для ответов на вопросы, Будапешт; 2003. С. 1–4.
21. Цвайгенбаум. Вопросы-ответы по биомедицине: методы и состояние техники. В: семинар MIE 2005; 2005 год.
22. Niu Y, Hirst G. Анализ семантических классов в медицинском тексте для ответов на вопросы. В: Семинар ACL 2004 по ответам на вопросы в ограниченных областях; 2004 год.
23. Niu Y, Hirst G, McArthur G, Rodriguez-Gianolli P. Ответы на клинические вопросы с идентификацией ролей. В: Семинар ACL по обработке естественного языка в биомедицине; 2003 год.
24. Дель Фиоль Г, Хауг Пи Джей, Чимино Джей Джей, Нарус СП, Норлин С, Митчелл Дж.А. Эффективность тематических информационных кнопок: рандомизированное контролируемое исследование. J Am Med Inform Assoc 2008;15:752–9.
25. Collins SA, Currie LM, Bakken S, Cimino JJ. Информационные потребности, использование менеджера информационной кнопки и удовлетворенность по типу клинициста: тематическое исследование. J Am Med Inform Assoc 2008.
26. Cimino J. Infobuttons: упреждающая поддержка пассивного принятия решений. AMIA Анну Симп Прок 2008:1203–4.
27. Cimino JJ. Использование, удобство использования, полезность и влияние менеджера информационных кнопок. AMIA Анну Симп Прок 2006:151–5.
28. Lei J, Chen ES, Stetson PD, McKnight LK, Mendonca EA, Cimino JJ. Разработка инфобутонов в беспроводной среде. AMIA Анну Симп Прок 2003:906.
29. Cimino JJ, Li J. Обмен информационными кнопками для удовлетворения информационных потребностей клиницистов. AMIA Анну Симп Прок 2003:815.
30. Чимино Джей Джей, Ли Джей, Баккен С, Патель ВЛ. Теоретические, эмпирические и практические подходы к решению неудовлетворенных информационных потребностей пользователей клинических информационных систем. Proc AMIA Symp 2002:170–4.
31. Cimino JJ, Elhanan G, Zeng Q. Поддержка информационных кнопок с терминологическими знаниями. Proc AMIA Annu Fall Symp 1997:528–32.
32. Mendonça EA, Kaufman D, Johnson SB. Ответы на информационные потребности в рабочем процессе. В: Материалы 9-го Всемирного конгресса по медицинской информации и библиотекам; 2005 год.
33. Chase HS, Kaufman DR, Johnson SB, Mendonca EA. Голосовой захват потребностей медицинских резидентов в клинической информации во время стационарной ротации. J Am Med Inform Assoc 2009;16:387–94.
34. Elhadad N, Kan M, Klavans JL, McKeown KR. Кастомизация в единой структуре для обобщения медицинской литературы. Artif Intell Med 2005;33:179–98.
35. Эльхадад Н., МакКаун К., Кауфман Д., Джордан Д. Облегчение доступа врачей к информации посредством индивидуального текстового резюме. AMIA Анну Симп Прок 2005:226–30.
36. McKeown K, Chang SF, Cimino JJ, Feiner SK, Friedman C, Gravano L и др. PERSIVAL, система персонализированного поиска и обобщения мультимедийной медицинской информации. В: Материалы 1-й совместной конференции ACM/IEEE-CS по электронным библиотекам; 2001 год.
37. Ely JW, Osheroff JA, Ferguson KJ, Chambliss ML, Vinson DC, Moore JL. Пожизненное самостоятельное обучение с использованием компьютерной базы данных клинических вопросов. J Family Pract 1997;45:382–8.
38. Д'Алессандро ДМ, Крейтер CD, Петерсон МВ. Оценка информационно-искательного поведения педиатров общего профиля. Педиатрия 2004;113:64–9.
39. Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA и др. Таксономия общих клинических вопросов: классификационное исследование. BMJ 2000;321:429–32.
40. Seol YH, Kaufman DR, Mendonca EA, Cimino JJ, Johnson SB. Сценарная оценка информационных потребностей врачей. Мединфо 2004;11: 306–10.
41. Демнер-Фушман Д., Лин Дж. Ответы на клинические вопросы с помощью основанных на знаниях и статистических методов. Компьютерный лингвист 2007;33:63–103.
42. Huang X, Lin J, Demner-Fushman D. Оценка PICO как представления знаний по клиническим вопросам. AMIA Анну Симп Прок 2006:359–63.
43. Demner-Fushman D, Lin J. Извлечение ответа, семантическая кластеризация и экстрактивное обобщение для ответа на клинические вопросы. В: Материалы 21-й международной конференции по компьютерной лингвистике и 44-го ежегодного собрания ассоциации компьютерной лингвистики (COLING/ACL); 2006. С. 945–52.
44. Линь Дж., Демнер-Фушман Д. Роль знаний в концептуальном поиске: исследование в области клинической медицины. В: Материалы 29-й ежегодной международной конференции ACM SIGIR по исследованиям и разработкам в области информационного поиска (SIGIR), Сиэтл, Вашингтон; 2006. С. 99–106.
45. Frakes WB, Baeza-Yates R. Поиск информации: структура данных и алгоритмы. Прентис-Холл; 1992 год.
46. SaltonG.Avectorspacemodelforinformationretrieval. ЦАКМ

1975;18:613–20.

1. Мюллер HM, Кенни ЭЭ, Штернберг PW. Textpresso: онтологическая система поиска и извлечения информации для биологической литературы. ПЛос Биол 2004; 2:e309.
2. Дирвестер СК, Дюме СТ, Ландауэр Т.К., Фурнас Г.В., Харшман РА. Индексация латентным семантическим анализом. J Am Soc Inform Sci 1990;41(6):391–407.
3. Гофманн Т. Вероятностная латентная семантическая индексация. В: Материалы 22-й ежегодной международной конференции ACM SIGIR по исследованиям и разработкам в области информационного поиска, Калифорния, США; 1999. С. 50–7.
4. Понте Дж., Крофт У. Подход языкового моделирования к поиску информации. В: Материалы 21-й ежегодной международной конференции ACM SIGIR по исследованиям и разработкам в области информационного поиска, Мельбурн, Австралия; 1998. С. 275–81.
5. Снайдерман К.А., Демнер-Фушман Д., Фишман М., Иде НК, Риндфлеш ТЦ. Основанные на знаниях методы, помогающие клиницистам найти ответы в MEDLINE. J Am Med Inform Assoc 2007;14:772–80.
6. Шринивасан., Риндфлеш Т. Изучение интеллектуального анализа текста из MEDLINE. Proc AMIA Symp 2002:722–6.
7. Хамфрис БЛ, Линдберг Д.А. Проект UMLS: создание концептуальной связи между пользователями и необходимой им информацией. Bull Med Lib Assoc 1993;81:170–7.
8. Ide NC, Loane RF, Demner-Fushman D. Essie: концептуальная поисковая система для структурированного биомедицинского текста. J Am Med Inform Assoc 2007;14:253–63.
9. Хови Э., Хермякоб У., Лин С.Ю. Использование внешних знаний в фактоидном QA. В: TREC 2001; 2001. С. 644–52.
10. Иоахимс Т. Категоризация текста с помощью машин опорных векторов: обучение со многими соответствующими функциями. Конспекты лекций по информатике. Берлин/Гейдельберг: Шпрингер; 1997. С. 137–42.
11. Lafferty J, McCallum A, Pereira F. Условные случайные поля: вероятностные модели сегментации и маркировки данных последовательности. В: Материалы восемнадцатой международной конференции по машинному обучению (ICML); 2001. С. 282–9.
12. Аронсон АР. Эффективное сопоставление биомедицинского текста с метатезавром UMLS: программа MetaMap. Proc AMIA Symp 2001:17–21.
13. Robertson S, Zaragoza H, Taylor M. Простое расширение BM25 на несколько взвешенных полей. В: АКМ ЦИКМ; 2004 год.
14. Сарагоса H, Craswell N, Taylor M, Saria S, Robertson S. Microsoft Cambridge at TREC-13: web и HARD треки. В: Проект TREC 2004; 2004 год.
15. Tellex S, Katz B, Lin J. Количественная оценка алгоритмов поиска пассажей для ответов на вопросы. В: Материалы 26-й ежегодной международной конференции ACM SIGIR по исследованиям и разработкам в области информационного поиска; 2003. С. 41–7.
16. Gobeill J, Ehrler F, Tbahriti I, Ruch P. Словарный поиск пассажа для вопросов-ответов в геномике. В: Шестнадцатая конференция по поиску текста, TREC, Гейтерсбург, штат Мэриленд; 2007 год.
17. Лю Икс, Крофт ВБ. Извлечение отрывков на основе языковых моделей. В: Проц ЦИКМ; 2002. С. 375–82.
18. Херст МА. TextTiling: сегментация текста на многоабзацные подтемы. Компьютерный лингвист 1997;23:33–64.
19. Патерсон М., Данчик В. Самые длинные общие подпоследовательности. В: Прок 19-го МФКС, No 841 в LNCS, т. 841; 1994. С. 127–42.
20. Лю Фэйфань, Лю Ян. Изучение корреляции между ROUGE и человеческой оценкой в резюме встреч. Обработка звука, речи и языка. IEEE Trans 2010;18:187–96.
21. Osinski S, Stefanowski J, Weiss D. Lingo: алгоритм кластеризации результатов поиска на основе декомпозиции сингулярных значений. Intell Inform Syst 2004: 359–68.
22. Хиршберг Д.С. Алгоритмы для самой длинной общей задачи подпоследовательности. J ACM 1977;24:664–75.
23. Lin C. ROUGE: пакет для автоматической оценки резюме. В: Материалы рабочего совещания ACL: резюме текста за 2004 год; 2004. С. 74–81.
24. Ng RT, Zhou X. Масштабируемое обнаружение скрытых писем из больших папок. В: ACM SIGKDD'05; 2005. С. 544–9.
25. Чжан Ю. Патрик Дж. Перефразирование идентификации путем канонизации текста. В: Труды семинара по австралазийским языковым технологиям 2005 года; 2005 год.
26. Abdou S, Savoy J. Поиск в Medline: расширение запросов и ручная оценка индексации. Inform Process Manage 2008;44:781–9.
27. Зигельник Л., Курланд О. Предотвращение дрейфа запросов для надежного расширения запросов. В: Материалы 31-й ежегодной международной конференции ACM SIGIR по исследованиям и разработкам в области информационного поиска; 2008. С. 825–6.
28. Stokes N, Li Y, Cavedon L, Zobel J. Исследование критериев успешного расширения запросов в геномной области. Inform Retrieval 2009;12:17–50.
29. Rinaldi F, Dowdall J, Schneider G, Persidis A. Ответы на вопросы в области геномики. рабочее совещание ACL-2004 по ответам на вопросы в ограниченных доменах, Барселона, Испания; 2007.