

На правах рукописи



Татарина Александра Геннадьевна

**ИНФОРМАЦИОННЫЙ ПОИСК РЕЧЕВЫХ ДОКУМЕНТОВ НА  
ОСНОВЕ МОДЕЛИ С ФОНЕМНЫМ ПРЕДСТАВЛЕНИЕМ  
СОДЕРЖАНИЯ**

Специальность: 05.13.17 – «Теоретические основы информатики»  
(Технические науки)

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата технических наук

Нижний Новгород – 2019 г.

Работа выполнена на кафедре прикладной математики и информатики Федерального государственного бюджетного образовательного учреждения высшего образования (ФГБОУ ВО) «Вятский государственный университет» (ВятГУ)

Научный руководитель: **Прозоров Дмитрий Евгеньевич**, доктор технических наук, доцент, профессор кафедры радиоэлектронных средств, ВятГУ, г. Киров

Официальные оппоненты: **Приоров Андрей Леонидович**, доктор технических наук, доцент, доцент кафедры инфокоммуникаций и радиофизики, Федеральное государственное бюджетное образовательное учреждение высшего образования «Ярославский государственный университет им. П.Г. Демидова», г. Ярославль

**Акатьев Дмитрий Юрьевич**, кандидат технических наук, доцент, доцент кафедры экономики, управления и информатики, Федеральное государственное бюджетное образовательное учреждение высшего образования «Нижегородский государственный лингвистический университет им. Н.А. Добролюбова», г. Нижний Новгород

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ИСП РАН), г. Москва

Защита диссертации состоится «14» марта 2019 года в 12 часов в ауд. 1315 на заседании диссертационного совета Д 212.165.05 при Нижегородском государственном техническом университете им. Р.Е. Алексеева по адресу: 603950, г. Нижний Новгород, ул. Минина, 24.

С диссертацией можно ознакомиться в библиотеке Нижегородского государственного технического университета им. Р.Е. Алексеева и на сайте <https://www.nntu.ru/structure/view/podrazdeleniya/fpsvk/obyavleniya-o-zashhitah>.

Автореферат разослан «    » \_\_\_\_\_ 2019 года.

Учетный секретарь  
диссертационного совета



Суркова Анна Сергеевна

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность работы.** Исторически, первые информационно-поисковые системы (ИПС) были ориентированы, в основном, на обработку документов текстового формата. Однако, широкое распространение профессиональных и бытовых технических устройств, таких как: видео- и фотокамеры, смартфоны и т.п., позволяющих создавать, хранить и распространять контент медиа-формата в совокупности с быстрым развитием информационно-коммуникационных систем привели, начиная с 90-х годов XX века, к взрывному росту количества документов с мультимедийным, в том числе и речевым содержанием. Примерами мультимедийных документов с речевым содержанием (МДРС) являются радио- и видео-новости, аудиокниги, записи докладов конференций и, в последнее время, образовательные ресурсы в виде аудио- и видео-лекций, интерактивные учебные фильмы и мультимедийные методические разработки.

Поиск речевой информации находится на пересечении таких научных областей как обработка естественного языка (Natural Language Processing, NLP) и цифровая обработка сигналов (Digital Signal Processing, DSP). Анализу проблем в области поиска речевой информации и разработке методов их решения посвящено большое количество как отечественных (Галунов В.И., Карпов А.А., Кипяткова И.С., Потапова Р.К., Савченко В.В., Утробин В.А., Фархадов М.П., Хейдоров И.Э.) так и зарубежных (Glass J., Hauptmann A., Jones G.J.F., Rose R.C., Spark-Jones K., Wechsler M., Zue V.W.) научных исследований. Часть методов поиска речевой информации построена на основе анализа акустических признаков речевого сигнала. К недостаткам такого подхода можно отнести высокую зависимость от состояния и параметров речи диктора. Кроме того, обработка только акустических признаков не позволяет учитывать лингвистическую информацию, содержащуюся в речевом сигнале. Другая часть методов использует системы автоматического распознавания речи (АРР) на предварительных этапах обработки речевых сигналов.

В процессе АРР речевой сигнал обрабатывается на акустическом, фонемном и языковом уровнях. Использование лингвистической информации снижает влияние вариабельности параметров речи дикторов, таких как интонация, темп, на качество распознавания речи, а значит повышает эффективность поиска речевой информации при условии произношения речи различными дикторами. Основным недостатком использования АРР является наличие ошибок распознавания, которые искажают полученное содержание речевой информации. В качестве причин появления ошибок распознавания выступают различные факторы, например, такие как зашумлённость сигнала, наличие акцента у диктора. Кроме того, традиционно распознавание речи происходит по ограниченному словарю слов. Поэтому отдельно стоит выделить ошибки, связанные с распознаванием несловарных слов – т.н. Out-Of-Vocabulary (OOV) проблема. Способы решения OOV проблемы связаны с изменением единиц распознавания речи на подслова (CVC-/VCV-признаки, морфемы, N-граммы) или фонемы, что приводит к усложнению моделей,

используемых при распознавании, значительному увеличению времени распознавания, а также снижению точности распознавания слов, входящих в словарь распознавания. Поэтому актуальной является научно-практическая задача разработки и исследования моделей и методов информационного поиска речевых документов по содержанию, позволяющих снизить влияние ошибок распознавания речи, с целью повышения эффективности поиска.

**Объектом исследования** являются тексты, полученные в результате автоматического распознавания содержания речевых документов.

**Предметом исследования** являются модели и алгоритмы поиска речевых документов по содержанию.

**Целью работы** является повышение эффективности поиска речевых документов путём разработки модели информационного поиска на основе фонемного представления содержимого речевых документов и запросов пользователя.

Для достижения указанной цели поставлены и решены следующие задачи:

- анализ и классификация моделей информационного поиска речевых документов;
- разработка векторной модели информационного поиска речевых документов, в том числе:
  - разработка функции релевантности основанной на нечетком сравнении слов при текстовом или фонемном представлении слов,
  - разработка механизма ранжирования речевых документов по релевантности, оценивание которой выполняется по их текстовому и фонемному представлениям;
- разработка метода фонемного поиска речевых документов, в том числе:
  - разработка алгоритмов фонемного транскрибирования, использующих вероятностные характеристики букв и фонем, полученные на основе обучающего словаря,
  - разработка меры близости фонем, используемой при сравнении фонемных транскрипций слов,
  - разработка алгоритма графемно-фонемного выравнивания;
- составление тестовой коллекции, содержащей речевые документы на русском языке;
- разработка программной модели системы информационного поиска речевых документов на основе разработанных алгоритмов;
- экспериментальные исследования влияния разработанных моделей и алгоритмов фонемного транскрибирования, графемно-фонемного выравнивания на эффективность информационного поиска речевых документов по запросу.

**Методы исследования.** Для решения поставленных в работе задач используются методы алгебры, математической статистики, теории вероятностей, теории информации, теории графов и динамического программирования.

**Научная новизна работы** состоит в следующем:

1. Предложен метод информационного поиска речевых документов по содержанию, отличающийся алгоритмом сравнения слов по фонемным транскрипциям и использованием вероятностной меры сходства фонем.

2. Предложена вероятностная мера сходства фонем, используемая при сравнении фонемных транскрипций слов и отличающаяся: описанием фонем посредством условных распределений графем, ассоциированных со сравниваемыми фонемами (тогда как известные подходы основаны на описании фонем в виде векторов акустических признаков); способом количественной оценки указанной меры сходства на основе расстояния Кульбака-Лейблера.

3. Разработана модель поиска речевых документов, использующая функцию релевантности на основе приближенного сравнения слов при их текстовом и/или фонемном представлении, в отличие от существующих моделей, основанных на точном сравнении слов. Модель позволяет частично учитывать ошибки распознавания слов в речевых документах, что повышает эффективность поиска при низком качестве распознавания речи.

4. Разработан рекуррентный алгоритм фонемного транскрибирования текстовых последовательностей на основе математического аппарата конечных цепей Маркова.

**Соответствие диссертации паспорту специальности.** Работа соответствует паспорту специальности 05.13.17 – «Теоретические основы информатики» по техническим наукам в пунктах: п.2 – «Исследование информационных структур, разработка и анализ моделей информационных процессов и структур» и п.5 – «Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений».

**Обоснованность и достоверность результатов** диссертационной работы подтверждается полученными в результате эксперимента значениями эффективности поиска по коллекции речевых документов, составленной из реальных данных, и обеспечиваются корректным использованием математического аппарата.

**Практическая значимость работы**

Полученные в диссертационной работе теоретические и практические результаты показывают:

- повышение средней точности поиска при использовании приближённого сравнения слов по фонемным транскрипциям на основе 2-связной цепи

Маркова относительно приближенного поиска, реализованного в сторонней библиотеке Lucene, в среднем на 4% и системы поиска ключевых слов речевых документов VoiceDigger на 0,5%;

- повышение средней точности поиска речевых документов на [1-15]% при использовании разработанного алгоритма фонемного транскрибирования на основе многосвязной цепи Маркова относительно алгоритма фонемного транскрибирования на основе скрытых марковских моделей при искажении и замене слов в результате распознавания;

- целесообразность использования приближенного сравнения слов на основе фонемных транскрипций для поиска речевых документов по содержанию при низком качестве распознавания речи;

- эффективность применения математического аппарата многосвязных дискретных цепей Маркова для построения фонемных транскрипций при информационном поиске речевых документов по содержанию.

**Реализация и внедрение результатов.** Теоретические и практические результаты, полученные при выполнении диссертационной работы, использованы в НИР по темам «Разработка и исследование интеллектуальной системы информационного поиска и анализа тональности текстовых и речевых документов» (проект РФФИ № 16-07-00342а), «Разработка и исследование словарей оценочной лексики для анализа тональности текстов» (государственное задание Минобрнауки России, проект № 34.2092.2017/4.6); внедрены в АО «НИИ СВТ» (г. Киров) в виде технических предложений при доработке автоматизированного комплекса информационного анализа материалов на базе перспективных технологий обработки неструктурированной информации. Разработанная система поиска речевых документов зарегистрирована в Реестре программ для ЭВМ, свидетельство № 2015617364 от 08.07.2015 г. Разработанный программный модуль фонемного транскрибирования на основе нелинейного рекуррентного алгоритма зарегистрирован в Реестре программ для ЭВМ, свидетельство № 2018660458 от 23.08.2018 г.

**Апробация результатов.** Основные результаты докладывались и обсуждались на следующих конференциях: международной конференции «15th IEEE East-West Design & Test Symposium» (Сербия, Нови-Сад, 2017); международной конференции «9th IEEE International Conference on Application of Information and Communication Technologies» (Ростов-на-Дону, 2015); международной конференции «Цифровая обработка сигналов и ее применение» (Москва, 2013); всероссийской научно-практической конференции «Общество, наука, инновации» (Киров, 2013-2017).

#### ***Основные положения, выносимые на защиту***

1. Модель информационного поиска на основе текстового и фонемного представления содержания речевых документов позволяет повысить

эффективность поиска по коллекции документов при низком качестве распознавания речи.

2. Алгоритмы приближенного сравнения слов, отличающиеся от традиционных использованием фонемных транскрипций, обеспечивают снижение влияния ошибок распознавания речи на эффективность поиска речевых документов.

3. Рекуррентный алгоритм фонемного транскрибирования на основе оценки апостериорной вероятности состояний многосвязной цепи Маркова.

4. Результаты экспериментальных исследований.

### ***Публикация результатов***

По результатам исследования опубликовано 19 работ, в том числе: 4 статьи в рецензируемых научных изданиях, рекомендуемых ВАК, 3 статьи в издательствах, индексируемых в базах Web of Science и Scopus. Получено 2 свидетельства официальной регистрации программы для ЭВМ.

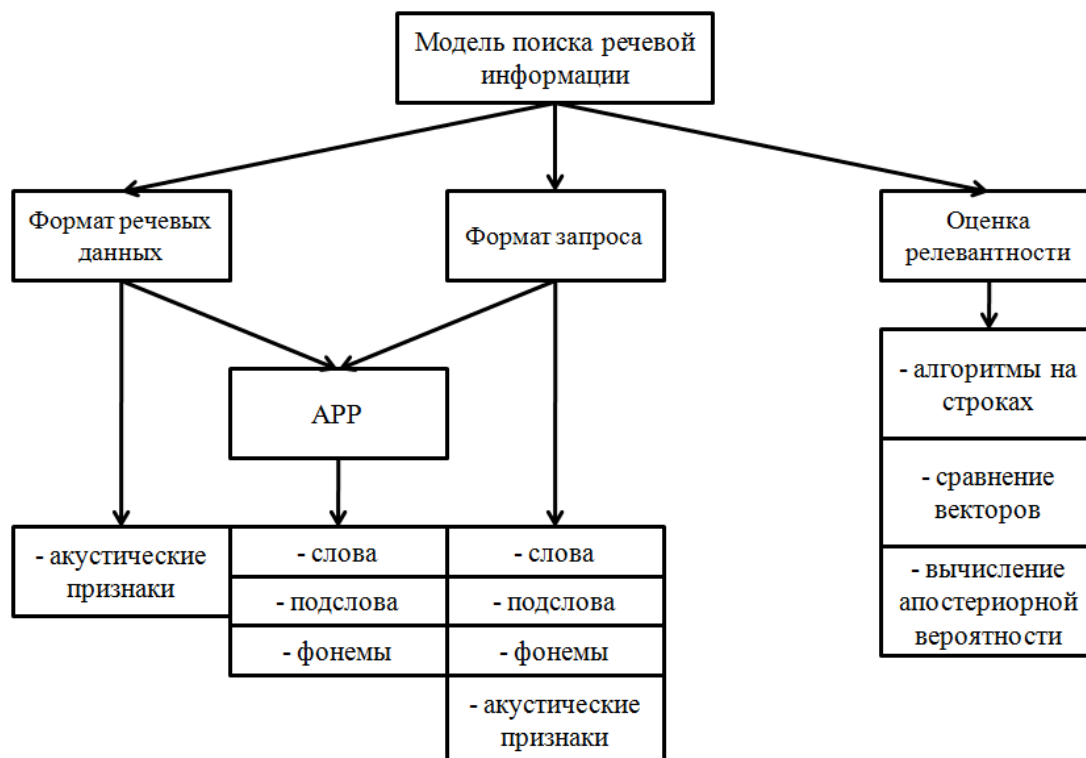
### ***Структура и объем работы***

Диссертация состоит из введения, четырёх глав, списка использованной литературы из 119 наименований. Общий объем работы составляет 145 страниц текста, 43 рисунка и 35 таблиц.

## **СОДЕРЖАНИЕ РАБОТЫ**

Во **введении** приведена общая характеристика диссертационной работы, обоснована актуальность исследования, сформулирована цель и задачи исследования, раскрыта научная новизна и практическая значимость полученных результатов.

В **первой главе** рассмотрены и проанализированы основные методы поиска речевой информации. Выделены две группы методов. Первая группа методов выполняет поиск речевых документов на акустическом уровне. Существенным недостатком данных методов является недоступность лингвистической информации на акустическом уровне обработки речевой информации. Методы второй группы выполняют поиск по содержанию, полученному в результате автоматического распознавания спонтанной слитной речи. Поиск по распознанному содержанию речевых документов затрудняется при наличии ошибок распознавания. Представлены основные модели информационного поиска, определяемые форматом представления документов, запроса и функцией релевантности. Показано, что традиционные для текстового поиска модели не учитывают возможные искажения содержания документов, возникающие в результате ошибок, например, опечаток. Предложена классификация моделей поиска в зависимости от формата представления речевой информации и способа оценки релевантности (рисунок 1).



**Рисунок 1** – Классификация моделей поиска речевой информации.

Выполнен обзор существующих систем поиска речевых документов, приведены количественные показатели эффективности поиска, такие как: полнота, точность, F1-мера, 11-точечный интерполированный график «полнота/точность» и макроусредненная средняя точность ( $map@K$ ).

**Вторая глава** посвящена разработке и исследованию векторной модели поиска на основе текстового представления содержания речевых документов и запроса пользователя, а также функции релевантности на основе приближенного сравнения строк. Рассмотрены этапы автоматического распознавания слитной спонтанной речи. Проанализированы алгоритмы нечеткого сравнения строк.

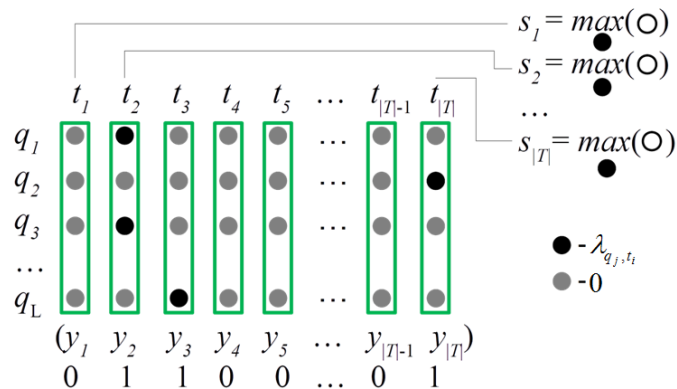
Речевой документ  $d_k$  коллекции  $D$  посредством APP преобразуется в текст, представляющий набор независимых слов  $\{w_i\}^{d_k}$ . Текстовый запрос  $Q$  задается последовательностью слов  $\{q_j\}$ . Функция релевантности  $F(d_k, Q)$  вычисляет значение  $r_k$ , характеризующее степень соответствия содержания речевого документа запросу.

Пусть  $T = \{t_i\}$  является множеством слов, встречающихся в  $D$ . Каждому документу  $d_k \in D$  сопоставляется вектор  $\vec{d}_k = (x_0, \dots, x_n)$   $n$ -мерного пространства ( $n = |T|$ ), где  $x_i$  – вес слова  $t_i \in T$  в документе  $d_k$ , вычисленный по методу *tf-idf*.

Определим матрицу  $\mathbf{A} = \|a_{ji}\|$ , такую что  $a_{j,i} = \lambda_{q_j, t_i}$ , если  $\lambda_{q_j, t_i} \geq \tau$ , и  $a_{j,i} = 0$ , если  $\lambda_{q_j, t_i} < \tau$ ;  $\lambda_{q_j, t_i}$  – значение меры сходства слов  $q_j$  и  $t_i$ , а  $\tau$  –



пороговое значение. Текстовому запросу  $Q$  сопоставляется вектор  $\vec{q} = (y_0, \dots, y_n)$ , где  $y_i = 1$ , если  $\sum_j a_{j,i} > 0$ , иначе  $y_i = 0$ . Процесс формирования вектора запроса  $\vec{q}$  и вектора весовых коэффициентов  $\vec{s} = (s_0, \dots, s_n)$ , где  $s_i = \max_j(a_{j,i})$ , иллюстрирует рисунок 2.



**Рисунок 2** – Формирование векторов запроса и весовых коэффициентов

Значение функции релевантности  $F(d_k, Q)$  документа  $d_k \in D$  запросу  $Q$  вычисляется как взвешенная косинусная мера близости векторов  $\vec{d}_k$  и  $\vec{q}$

$$F(d_k, Q) = \cos'(\vec{d}_k; \vec{q}) = \frac{\sum_{i=1}^n s_i x_i y_i}{\sqrt{\sum_{i=1}^n s_i x_i^2} \cdot \sqrt{\sum_{i=1}^n s_i y_i^2}}. \quad (1)$$

Результаты сравнения эффективности предложенной модели относительно других представлены в таблице 1. Использована тестовая коллекция comb\_msu\_ru\_nsh.

**Таблица 1.** Максимальные значения  $F_1$ -меры.

Модель	Оценка релевантности	Сравнение слов	$\max F_1$
1	косинусная мера	точное	52,6
2	взвешенная косинусная мера	наибольшая общая подстрока	71,2
3		расстояние Левенштейна	70,8
4	точный поиск Lucene		55,5
5	приближённый поиск Lucene		62,2

Использование предложенной взвешенной косинусной меры (1) повышает показатель  $F_1$ -меры до 14% относительно модели приближённого поиска, реализованной в библиотеке Lucene.

В **третьей главе** рассматривается векторная модель информационного поиска речевых документов на основе фонемных транскрипций слов и функции релевантности (1). Фонемная транскрипция передает общее произношение слова.

Разработаны алгоритмы графемно-фонемного выравнивания на основе правил и алгоритма DTW, используемые для расширения фонемного алфавита и построения условных распределений фонем и графем по обучающему словарю  $\Psi$ .

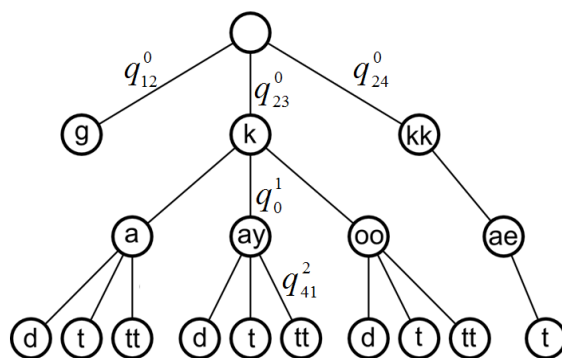
Пусть задан алфавит графем (букв)  $C = \{c_i\}$  некоторого языка и алфавит фонем  $\Phi = \{\varphi_j\}$ , составленный на основе обучающего словаря  $\Psi$ . Каждую фонему  $\varphi_j$  будем описывать распределением  $P_{C|\varphi_j} \equiv P(C|\varphi_j)$ , где  $P(C|\varphi_j)$  – условная плотность вероятности графем  $\{c_i\}$ , соответствующих фонеме  $\varphi_j$  в транскрипциях различных слов. Фонемное транскрибирование слова  $w = \langle c^0 c^1 \dots c^m \rangle$  заключается в построении фонемной транскрипции  $\langle \varphi^0 \varphi^1 \dots \varphi^n \rangle$ .

Разработано несколько алгоритмов фонемного транскрибирования.

*Алгоритм транскрибирования на основе дерева альтернатив.* Статистическая зависимость между графемами  $c_i$  и фонемами  $\varphi_j$ , при наличии обучающего словаря  $\Psi$ , позволяет оценить вид условных дискретных распределений  $F(\varphi_j) \equiv P(c_i|\varphi_j)$ . В соответствии с теоремой Байеса и критерием максимума апостериорной вероятности, результатом транскрибирования является последовательность фонем  $\langle \varphi^0 \dots \varphi^n \rangle$ , для которой выполняется условие

$$P^{ac}(\varphi^0 \dots \varphi^n) = \frac{P(\varphi^0) F(\varphi^0) \prod_{k=1}^m P(\varphi^k | \varphi^{k-1}) \cdot \prod_{k=0}^m F(\varphi^k)}{P(c^0) \prod_{k=1}^m P(c^k | c^{k-1})} \rightarrow \max. \quad (2)$$

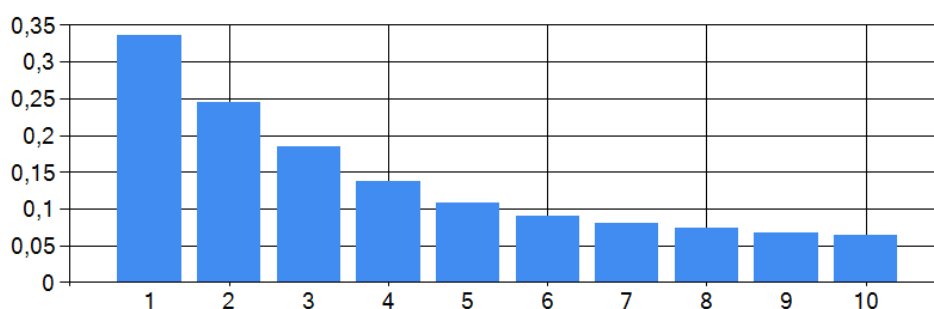
Фонемную транскрипцию слова в соответствии с выражением (2) удобно находить путём построения дерева  $T$ , пример которого приведён на рисунке 3.



**Рисунок 3** – Пример дерева альтернатив  $T$  слова «кот».

Пути от корня до листьев нижнего уровня являются вариантами фонемных транскрипций слова  $\langle c^0 \dots c^n \rangle$ . В качестве «наилучшей» фонемной транскрипции слова выбирается последовательность узлов, образующая путь от корня до листа с наибольшим значением метки  $q_j^n$ .

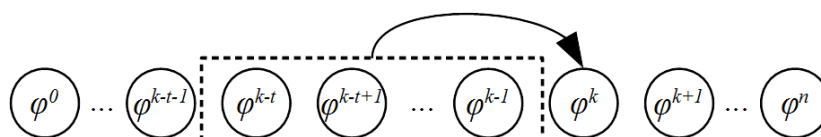
*Рекуррентный алгоритм фонемного транскрибирования.* Исследована зависимость появления фонемы  $\varphi_i^l$ , находящейся на  $l$ -й позиции в транскрипции слова, относительно предшествующей  $(l-k)$ -й фонемы  $\varphi_j^{l-k}$  при  $l = k..|w^\Phi|$ , где  $w^\Phi$  - транскрипция слова обучающего словаря  $\Psi$  (рисунок 4).



**Рисунок 4** – Условная вероятность появления наиболее вероятной фонемы на  $l$ -й позиции в транскрипции слова в зависимости от  $(l-k)$ -й предшествующей фонемы при  $l = k..|w|$

Результаты показали, что статистическая зависимость между соседними фонемами убывает по экспоненциальному или близкому к нему закону. С учётом этого факта, опираясь на теорему Дуба, для формирования наиболее вероятной последовательности фонем транскрибируемого слова предложено использовать математический аппарат теории многосвязных (сложных) конечных цепей Маркова.

В общем случае фонема транскрипции зависит от конечного числа  $t$ -предыдущих фонем как показано на рисунке 5.

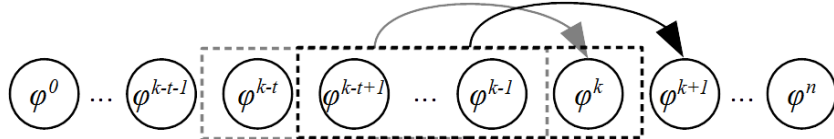


**Рисунок 5** – Зависимость фонемы от  $t$ -предыдущих фонем транскрипции слова

Тогда

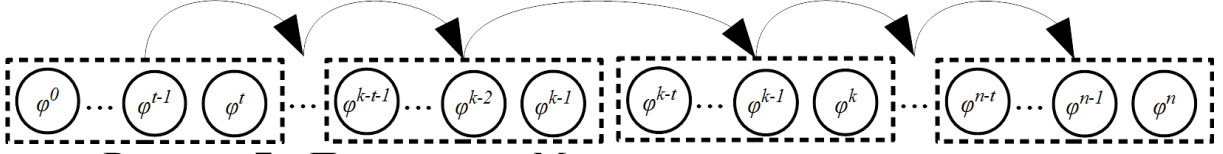
$$P(\varphi_i^k | \varphi_j^{k-1} \varphi_r^{k-2} \dots \varphi_q^{k-t} \varphi_l^{k-t-1} \dots) = P(\varphi_i^k | \varphi_j^{k-1} \varphi_r^{k-2} \dots \varphi_q^{k-t}). \quad (3)$$

Многосвязная конечная цепь Маркова, определяемая условными вероятностями (3), представлена на рисунке 6.



**Рисунок 6** – Многосвязная (сложная) конечная цепь Маркова

Многосвязную цепь Маркова с  $N$  скалярными состояниями можно преобразовать в простую цепь Маркова с  $N^t$  векторными состояниями, показанную на рисунке 7.



**Рисунок 7** – Простая цепь Маркова с векторными состояниями

Векторные состояния простой цепи Маркова (рисунок 7):

$$\vec{\varphi}_{ij..r}^k = (\varphi_i^k \varphi_j^{k-1} \dots \varphi_r^{k-t+1}). \quad (4)$$

В результате формируется простая векторная цепь Маркова  $\{\vec{\varphi}_{ij..r}^1, \vec{\varphi}_{j..rq}^2, \dots, \vec{\varphi}_{vx..z}^n\}$  с матрицей переходов размера  $N^t \times N^t$ .

На основе предложенной модели получена система уравнений:

$$u_{ij..r}^{k+1} = f\left(\vec{\varphi}_{ij..r}^{\rightarrow k+1}\right) + \ln\left(\sum_{t=1}^N \exp(u_{j..rq_t}^k) \cdot P\left(\vec{\varphi}_{ij..r}^{\rightarrow k+1} \mid \vec{\varphi}_{j..rq_t}^{\rightarrow k}\right)\right) + c, \quad (6)$$

где  $u_{ij..r}^{k+1}$  - логарифм апостериорной вероятности векторного состояния  $\vec{\varphi}_{ij..r}^{\rightarrow k+1}$ ,  $f\left(\vec{\varphi}_{ij..r}^{\rightarrow k+1}\right)$  - логарифм функции правдоподобия векторного состояния  $\vec{\varphi}_{ij..r}^{\rightarrow k+1}$  и  $c$  - коэффициент нормировки.

Критерием распознавания  $k$ -ой фонемы транскрипции слова является критерий максимума апостериорной вероятности, в соответствии с которым

$$\varphi_i : (\varphi_i^k \varphi_j^{k-1} \dots \varphi_r^{k-t+1}) = \vec{\varphi}_{ij..r}^{\rightarrow k} \Rightarrow i = i_t = \arg \max_{t=1..N} (u_{i_t j_t \dots r_t}^k). \quad (7)$$

В предлагаемой векторной модели поиска на основе приближённого сравнения слов в функции релевантности (1) используются весовые коэффициенты  $s_i$ . Определяемые сходством фонемных транскрипций слов запроса и речевых документов значения  $s_i$  вычисляются посредством расширенного алгоритма наибольшей общей подстроки, в котором вместо равенства букв применяется мера сходства фонем. Предложена мера сходства фонем  $\varphi_j$  и  $\varphi_k$  на основе расстояния Кульбака-Лейблера  $D_{KL}(P_{C|\varphi_j} \parallel P_{C|\varphi_k})$ :

$$\rho_{KL}(\varphi_j; \varphi_k) = 1 - \frac{\max(D_{KL}(P_{C|\varphi_j} \parallel P_{C|\varphi_k}), D_{KL}(P_{C|\varphi_k} \parallel P_{C|\varphi_j}))}{\max_{t_1, t_2=1..N} D_{KL}(P_{C|\varphi_{t_1}} \parallel P_{C|\varphi_{t_2}})}. \quad (8)$$

Исследована производительность разработанных алгоритмов фонемного транскрибирования и алгоритма на основе скрытой марковской модели (СММ). В таблице 2 приведено среднее время предобработки обучающего словаря  $\Psi$ , среднее время обучения СММ и среднее время транскрибирования слов в зависимости от их длины. Используемые обозначения алгоритмов транскрибирования:

**TREE** - алгоритм на основе построения дерева альтернатив,

**Mark1** - рекуррентный алгоритм на основе простой цепи Маркова  $t = 1$ ,

**Mark2** - рекуррентный алгоритм на основе сложной цепи Маркова  $t = 2$ ,

**HMM** - алгоритм на основе СММ.

**Таблица 2.** Средние временные затраты фонемного транскрибирования

Метод		TREE	Mark1	Mark2	HMM
Время обучения, мс		3 606		8 488	600 620
Время транскрибирования слова, мс	5 букв (189)	198,06	56,24	128,16	1,47
	6 букв (331)	282,35	70,24	167,88	4,55
	7 букв (403)	286,38	83,17	211,36	0,51
	8 букв (399)	322,66	97,23	251,58	0,72
	среднее время	272,36	76,72	189,75	1,81

Таким образом, предлагаемые алгоритмы TREE, Mark1 и Mark2 являются менее затратными на этапе обучения по сравнению с алгоритмом на основе СММ, но проигрывают по времени транскрибирования.

Проанализирована эффективность предлагаемой модели поиска на основе разработанных алгоритмов фонемного транскрибирования. Исследуемые в экспериментах модели поиска можно разделить на три группы: модель, используемая в системе VoiceDigger от компании «Центр речевых технологий», модели на базе поискового движка Lucene, векторная модель на основе взвешенной косинусной меры (1).

Названия моделей, методов поиска и их краткое описание представлены в таблице 3.

**Таблица 3.** Основные характеристики моделей поиска

Модель	Поиск	Характеристики
VoiceDigger		Поиск ключевых слов по речевым документам на основе СММ, построенных по транскрипциям слов на акустическом уровне
Lucene	Base	Точный поиск, реализованный в Lucene
	Fuzzy	Поиск на основе расстояния Левенштейна
	Ngram	N-граммный поиск при $N=3, 4, 5$
VSM*	Text Substr	Векторная модель с взвешенной косинусной мерой (1), где $s_i$ вычисляются на основе длины наибольшей общей подстроки

Модель	Поиск		Характеристики
		Lev	... $s_i$ вычисляются посредством расстояния Левенштейна
		MetaphoneRu	... $s_i$ вычисляются посредством фонетического сравнения строк по алгоритму MetaphoneRu
VSM*	Phoneme	Tree	... $s_i$ вычисляются посредством сравнения слов по фонемным транскрипциям, построенных по алгоритму на основе дерева альтернатив
		HMM	... построенных по алгоритму на основе СММ
		Mark1	... построенных по алгоритму на основе простой цепи Маркова
		Mark2	... построенных по алгоритму на основе сложной цепи Маркова связности равной 2

В таблице 4 приведены значения макроусреднённой средней точности MAP, полученные при поиске по коллекции радио-новостей.

Таблица 4. Значения MAP при поиске по радио новостям

Модель	Поиск		MAP	MAP
VoiceDigger			0,6786	
Lucene	Base		0,4442	0,7874
	Fuzzy		0,6514	0,8672
	Ngram		0,6429	0,8339
VSM*	Text	Substr	0,6712	0,8929
		Lev	0,6809	<b>0,9015</b>
		MetaphoneRu	0,4787	0,8008
	Phoneme	Tree	0,6770	0,8905
		HMM	0,6817	0,8968
		Mark1	0,6808	0,8907
		Mark2	<b>0,6821</b>	0,8948

Поиск на основе векторной модели с взвешенной косинусной мерой (1) и сервиса распознавания речи Yandex SpeechKit показывает наилучшие результаты.

Yandex SpeechKit является веб-сервисом, предоставляющим доступ к распознаванию аудио файлов на серверах компании Яндекс по протоколу HTTP. Используемая в экспериментах, языковая модель распознавания в Yandex SpeechKit определяется параметром topic равным notes, что соответствует модели для распознавания текстов при свободной диктовке. Точность распознавания русской речи сервисом Yandex SpeechKit в среднем составляет 80-90%, в то время как CMU Pocketsphinx при использовании моделей cmusphinx\_ru-5.2 - 40-80%.

Однако сервис Yandex SpeechKit в условиях сильных акустических помех может вернуть распознанный текст с пропущенными словами. В результате из содержания речевого документа, по которому выполняется поиск, исключаются длинные последовательности слов, что приводит к снижению эффективности поиска. В таблице 5 приведены значения точности распознавания фрагментов «трудных» аудио-файлов.

**Таблица 5.** Точность распознавания новостных аудио-файлов, записанных в условиях акустических помех

Документ	Точность распознавания посредством CMU Pocketsphinx	Точность распознавания посредством сервиса Yandex SpeechKit
0	54,70%	88,89%
1	35,14%	68,92%
2	20,74%	72,59%
3	28,50%	47,66%

Точность распознавания CMU Pocketsphinx ниже относительно Yandex SpeechKit, но в первом случае ошибки распознавания преимущественно являются заменой слов, а во втором - удалением слов.

В таблице 6 представлены значения эффективности поиска по аудио-файлам, записанных в условиях акустических помех.

**Таблица 6.** Значения MAP при поиске по новостным аудио файлам, записанных в различных условиях акустических помех

Модель	Поиск		MAP		MAP	
	VoiceDigger		0,4596			
Lucene	Base		CMU Pocketsphinx	0,0588	Yandex SpeechKit	0
	Fuzzy			0,2059		0,0441
	Ngram			0,4216		0,3284
VSM*	Text	Substr	0,5699	0,3922		
		Lev	0,5294	0,4179		
		MetaphoneRu	0,0882	0		
	Phoneme	Tree	<b>0,6299</b>	0,4498		
		HMM	0,5588	0,3897		
		Mark1	0,5907	<b>0,4841</b>		
		Mark2	0,5650	0,4179		

Результаты показывают достижение наилучшей средней точности поиска при использовании библиотеки CMU Pocketsphinx и разработанной модели фонемного поиска по сравнению с APP на основе сервиса YandexSpeechKit и исследуемыми моделями поиска.

Проанализирована точность поиска относительно ошибок распознавания речевых документов. В тестовой коллекции радио-новостей выделены ошибки

замены, искажения, разбиения на части и удаления слов. В таблице 7 приведены значения эффективности поиска в случае ошибок распознавания речи, заключающиеся в замене слов на близкие по произношению.

**Таблица 7.** Значения MAP при поиске по коллекции радио-новостей относительно ошибок замены слов в результате APP

Набор запросов		qs	
		err1	
Lucene	Base	0,0119	
	Fuzzy	0,4528	
	Ngram	0,4351	
VSM*	Text	Substr	0,4053
		Lev	0,4702
		MetaphoneRu	0,0060
	Phoneme	Tree	0,4876
		HMM	0,4543
		Mark1	<b>0,5264</b>
		Mark2	0,5252

Таким образом результаты экспериментов показывают, что: предлагаемая векторная модель с приближенным сравнением слов повышает точность поиска документов при низком качестве распознавания слов, предлагаемые алгоритмы транскрибирования на основе цепи Маркова позволяют повысить точность поиска в случае замены слов в результате APP.

В **четвёртой главе** разработана архитектура системы информационного поиска речевых документов на основе предложенной модели, оценивающей релевантность посредством приближенного сравнения слов при текстовом и/или фонемном их представлении.

Описано построение тестовой коллекции речевых документов, представляющих записи радио новостей на русском языке и наборов запросов с учётом типов ошибок, полученных в результате APP.

В **заключении** изложены основные научные и практические результаты диссертационной работы.

### **ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ**

1. Выполнен анализ и предложена классификация моделей информационного поиска речевых документов, выявлены достоинства и недостатки алгоритмов приближённого сравнения строк.

2. Разработана векторная модель информационного поиска речевых документов, основанная на использовании взвешенной косинусной меры и фонемного представления содержания, позволяющая частично учитывать ошибки автоматического распознавания речи в отличии от традиционных моделей поиска, основанных на точном сравнении слов.



3. Разработан метод фонемного поиска речевых документов на основе алгоритмов фонемного транскрибирования, использующих вероятностные характеристики букв и фонем обучающего словаря, вместо акустических параметров фонем, и позволяющий в отличие от нечёткого поиска выполнять сравнение слов, учитывая их произношение.

4. Составлена тестовая коллекция, содержащая речевые документы на русском языке, а также множества запросов с указанием экспертной оценки релевантности.

5. Разработана программная модель системы информационного поиска речевых документов на основе разработанных алгоритмов.

### **ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

#### ***Публикации в рецензируемых изданиях, рекомендованных ВАК РФ***

1. Прозоров Д.Е., Яшина А.Г. (Татарина А.Г.) Анализ алгоритмов фонемного транскрибирования в задачах контекстного поиска речевых документов // Инфокоммуникационные технологии – Самара, 2013 – Том 12. – № 4 – С. 62-65.

2. Прозоров Д.Е., Яшина А.Г. (Татарина А.Г.) Взвешенная косинусная мера векторной модели информационного поиска речевых документов // Информационные технологии. – №9, т. 21, 2015. – С. 715-720.

3. Прозоров Д.Е., Яшина А.Г. (Татарина А.Г.) Меры сходства слов, используемые в методах текстового и фонетического поиска речевых документов // Информатизация образования и науки. – 1(29), 2016. – С. 130-137.

4. Прозоров Д.Е., Плетнёв К.В., Яшина А.Г. (Татарина А.Г.) Апостериорная оценка состояний многосвязной цепи Маркова // Информация и космос. – 1(6), 2016. – URL: <http://openbooks.ifmo.ru/read/15422/15422.pdf>

#### ***Публикации, индексируемые международными базами***

5. Prozorov D., Yashina A. (Tatarinova A.) The extended longest common substring algorithm for spoken document retrieval // IEEE, 9th International Conference, Application of Information and Communication Technologies (AICT). – 2015. – pp. 88-90.

6. Tatarinova A.G., Prozorov D.E. Spoken Document Retrieval System based on Phonemic Transcribing // IEEE East-West Design & Test Symposium (EWDTS) - 2017. - pp.337-340. - URL: <http://ieeexplore.ieee.org/document/8110136/>

7. Prozorov D.E., Tatarinova A.G. Grapheme-to-phoneme conversion based on high-order Markov chain for spoken term detection by text query // IEEE East-West Design & Test Symposium (EWDTS) - 2017. - pp.646-650. - URL: <http://ieeexplore.ieee.org/document/8110058/>

#### ***Публикации в других научных изданиях***

8. Яшина А.Г. (Татарина А.Г.) Алгоритм контекстного поиска речевых аудио-файлов на основе фонемного сравнения слов // Advanced Science, 2012. - №1. - С. 73-85. - URL: [http://www.vyatsu.ru/uploads/file/1210/1\\_\(2\).pdf](http://www.vyatsu.ru/uploads/file/1210/1_(2).pdf)

9. Яшина А.Г. (Татарина А.Г.) Анализ алгоритма контекстного поиска речевых документов при использовании буквенного или фонемного представления слов // Материалы XV Международной конференции «Цифровая обработка сигналов и её применение DSPA-2013» – Москва, 2013. – том I – С. 213-216.

10. Яшина А.Г. (Татарина А.Г.) Поиск речевых документов на основе фонемного транскрибирования слов с использованием скрытой марковской модели // Материалы Всероссийской ежегодной научно-технической конференции «Общество, наука, инновации» (НТК–2013) – Киров, 2013. – 1 электрон. опт. диск (CD-ROM).

11. Яшина А.Г. (Татарина А.Г.) Поиск речевых документов с использованием различных мер сравнения фонем // Материалы Международной научно-практической конференции «Актуальные направления фундаментальных и прикладных исследований» – М. : spc Academic, 2013. – Т. 2. – С. 73–76.

12. Яшина А.Г. (Татарина А.Г.) Контекстный поиск речевых документов по текстовому запросу на основе *tf-idf* меры и фонемного транскрибирования // Всероссийская ежегодная научно-техническая конференция «Общество, наука, инновации» (НТК-2014)– Киров, 2014. – 1 электрон. опт. диск (CD-ROM).

13. Прозоров Д.Е., Яшина А.Г. (Татарина А.Г.) Модель информационного поиска речевых документов по текстовому запросу на основе фонемного транскрибирования и *tf-idf* меры // Алгоритмы, методы и системы обработки данных – 1 (26), 2014 – URL: [http://amisod.ru/index.php?option=com\\_content&view=article&id=242:amisod-2014-1-26-yashina-prozorov&catid=26:amisod-2014-1-26](http://amisod.ru/index.php?option=com_content&view=article&id=242:amisod-2014-1-26-yashina-prozorov&catid=26:amisod-2014-1-26)

14. Яшина А.Г. (Татарина А.Г.) Поиск речевых документов на основе алгоритма MetaphoneRu // Всероссийская научно-практическая конференция «Общество, наука, инновации» (НПК-2015) – Киров, 2015. – 1 электрон. опт. диск (CD-ROM).

15. Татарина А.Г., Прозоров Д.Е. Функции оценки релевантности речевых документов текстовому запросу векторной и вероятностной моделей поиска // Всероссийская научно-практическая конференция «Общество, наука, инновации» (НПК-2016) – Киров, 2016. – с. 2743-2751. – 1 электрон. опт. диск (CD-ROM).

16. Татарина А.Г., Прозоров Д.Е. Анализ моделей информационного поиска документов с речевым содержанием // Информационные технологии моделирования и управления. – 5(101), 2016

17. Татарина А.Г., Прозоров Д.Е. Алгоритм фонемного транскрибирования на основе дерева альтернатив // Всероссийская научно-практическая конференция «Общество, наука, инновации» (НПК-2017) – Киров, 2017. - С. 2236-2244.

18. Татарина А.Г., Прозоров Д.Е. Алгоритм сопоставления букв и фонем на основе правил в задаче поиска речевых документов по текстовому запросу //

ВятГУ: Advanced Science. - №4. - 2017. - URL: [http://advanced-science.ru/assets/mgr/docs/4\(2017\)/Технические/tatarinova-prozorov.pdf](http://advanced-science.ru/assets/mgr/docs/4(2017)/Технические/tatarinova-prozorov.pdf)

19. Прозоров Д.Е., Татаринова А.Г. Фонемное транскрибирование в задаче поиска документов с речевым содержанием // деп. в ВИНТИ РАН № 77-В2017. - 2017. - 34 с.

***Свидетельства о государственной регистрации программ для ЭВМ и баз данных***

20. Яшина А.Г. (Татаринова А.Г.), Прозоров Д.Е. Система поиска речевых документов «WISEARCH». Свидетельство № 2015617364 от 08.07.2015 г.

21 Татаринова А.Г., Прозоров Д.Е. Программный модуль фонемного транскрибирования на основе нелинейного рекуррентного алгоритма. Свидетельство № 2018660458 от 23.08.2018 г.