

**Keywords:** *programmable logical device (PLD), parameters, objective functional, decision making, optimal choice.*

Литвинская Ольга Сергеевна, к.т.н., доцент Кафедры вычислительных машин и систем (ВМиС) Пензенского государственного технологического университета (ПензГТУ). Тел. (8-412) 49-61-56; 8-927-366-59-45. E-mail: oslit@ya.ru

Турьгин Игорь Геннадьевич, аспирант Кафедры ВМиС ПензГТУ. Тел. (8-412) 49-61-56; 8-937-413-72-21. E-mail: tigseir@ya.ru

## НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 025.4.03

### АНАЛИЗ АЛГОРИТМОВ ФОНЕМНОГО ТРАНСКРИБИРОВАНИЯ В ЗАДАЧАХ КОНТЕКСТНОГО ПОИСКА РЕЧЕВЫХ ДОКУМЕНТОВ

*Прозоров Д.Е., Яшина А.Г.*

В статье описаны и проанализированы методы фонемного транскрибирования: метод, использующий скрытую марковскую модель, и метод, основанный на теории графов. Указанные методы используются в разработанной системе контекстного поиска речевых документов. Произведен анализ времени фонемного транскрибирования, приведены значения эффективности поиска речевых документов, которые достигаются при использовании данных методов.

**Ключевые слова:** информационный поиск, распознавание речи, фонемное транскрибирование, поиск речевых документов.

#### Введение

Качество работы современных информационных систем существенно зависит от применяемых методов контекстного поиска. Существующие методы поиска ориентированы в основном на обработку текстовых документов. В то же время все большее распространение получают коллекции мультимедийных, в том числе речевых, документов. Примерами таких документов являются радио- и видеонews, аудиокниги, записи лекций, доклады конференций и т.п. Задача контекстного поиска речевых документов по текстовому или устному запросу относится к области Spoken Document Retrieval (SDR).

Основными этапами работы SDR-системы являются: распознавание речи, индексирование и поиск документов, релевантных запросу [1]. При этом ошибки, возникающие на этапе распознавания, снижают эффективность SDR-системы в целом.

Качество распознавания речи в основном зависит от используемого метода, количества и качества обучающих акустических данных, качества записи речевых документов, а также числа

дикторов [3]. Улучшить качество распознавания речи можно переходом к распознаванию подслов [4] или фонов/фонем [5], что позволяет не зависеть от объема словаря, используемого при распознавании слов.

Другой подход, позволяющий повысить эффективность SDR-системы, основан на использовании методов поиска, учитывающих ошибки распознавания речевых единиц [2; 6].

Так, в [2] результат фонемного распознавания корректируется на основе статистики ошибок, связанных с пропуском, заменой и добавлением фонем. Метод, изложенный в [6], использует фонемное транскрибирование текстового представления распознанных речевых документов и запросов пользователя, что позволяет учитывать сходство произношения речевых единиц.

В данной работе анализируются два метода получения фонемных транскрипций слов, которые позволяют повысить эффективность контекстного поиска речевых документов. Данные методы частично учитывают сходство произношения речевых единиц и статистику ошибок распознавания слов речевых документов. Первый метод основан на теории графов, во втором методе используется скрытая марковская модель.

#### Постановка задачи

Транскрибирование представляет собой запись слов посредством некоторого алфавита, которая осуществляется для передачи особенностей произношения. Пусть задан алфавит фонем  $\{\varphi_k\}$  и слово  $w_i$  представлено в виде последовательности букв  $v_1 v_2 \dots v_m$ . Тогда задача фонемного транскрибирования заключается в построении соответствующей последовательности фонем для слова

$$f(v_1 v_2 \dots v_m) = \varphi_1 \varphi_2 \dots \varphi_n. \quad (1)$$

### Алгоритм поиска

Речевой документ  $d_k$  коллекции  $D$  будем рассматривать как фразу  $\psi^{d_k}$ , состоящую из распознанных слов  $\{w_i\}^{d_k}$  записанных слитно (без пробелов и знаков препинания). Текстовый запрос  $Q$  является набором ключевых слов  $\{w_j\}^Q$ . Задача SDR сводится к поиску документов  $d_k$  таких, что

$$\operatorname{argmax}_D F(d_k, Q), \quad (2)$$

где функция  $F$  вычисляет оценку релевантности документа  $d_k \in D$  запросу  $Q$ .

Определим функцию  $F$  как

$$F(d_k, Q) = \begin{cases} 1, & L_0(\psi^{d_k}; w_j^Q) \geq h, j = 1 \dots t; \\ 0, & L_0(\psi^{d_k}; w_j^Q) < h, j = 1 \dots t, \end{cases} \quad (3)$$

где  $h$  – пороговое значение,  $t$  – количество слов запроса,  $L_0(\psi^{d_k}; w_j^Q)$  – нормированная длина наибольшей общей подпоследовательности фразы  $\psi^{d_k}$  и слова  $w_j^Q$

$$L_0 = \frac{l(\psi^{d_k}; w_j^Q)}{l(w_j^Q)}, \quad (4)$$

$l(\psi^{d_k}; w_j^Q)$  – длина наибольшей общей подпоследовательности фразы  $\psi^{d_k}$  и слова  $w_j^Q$ , а  $l(w_j^Q)$  – длина слова  $w_j^Q$ .

Значения  $l(\psi^{d_k}; w_j^Q)$  вычисляются посредством алгоритма поиска наибольшей общей подстроки [7], модифицированного для обработки фонемных последовательностей. В разработанном алгоритме используется вспомогательная матрица  $A$  размерности  $(R+1)(K+1)$ , где  $R$  – длина фразы  $\psi^{d_k}$ , а  $K$  – длина слова  $w_j^Q$ . Элементы матрицы  $A$  определяются рекуррентным соотношением

$$\begin{aligned} a_{0,k} &= 0, \quad k = 0, \dots, K; \\ a_{r,0} &= 0, \quad r = 0, \dots, R; \\ a_{r,k} &= a_{r-1,k-1} + \lambda, \end{aligned} \quad (5)$$

где параметр  $\lambda$  соответствует мере близости между  $(r+1)$ -элементом фразы  $\psi^{d_k}$  и  $(k+1)$ -элементом слова  $w_j^Q$ . Тогда

$$l(\psi^{d_k}; w_j^Q) = \max_A(a_{r,k}). \quad (6)$$

Каждое слово фразы или запроса представлено кортежем  $(c_1^{w_i}, c_2^{w_i}, \dots, c_N^{w_i})$ , где  $c_j^{w_i}$  –  $j$ -ая

фонема слова  $w_i$ . Соответственно, фонемное транскрибирование слова  $w_i$  заключается в получении кортежа  $(c_1^{w_i}, c_2^{w_i}, \dots, c_N^{w_i})$ .

### Фонемное транскрибирование с использованием дерева вариантов фонем (метод TREE)

В данном методе фонемного транскрибирования слово рассматривается как последовательность комбинаций букв, по которой строится последовательность фонем. Каждая  $j$ -ая фонема определяется на основе  $j$ -ой и  $(j-1)$ -ой комбинации букв и  $(j-1)$ -ой фонемы слова.

Для получения фонемного представления слова строится дерево возможных вариантов  $T$ . Вершины  $i$ -го уровня дерева  $T$  соответствуют возможным вариантам  $i$ -ой фонемы слова. Каждой дуге приписывается метка

$$q_i = P(\varphi_{i-1} | s_{i-1}) \cdot P(\varphi_i | \varphi_{i-1}) \cdot P(\varphi_i | s_i), \quad (7)$$

где  $P$  – условная вероятность,  $\varphi_i$  –  $i$ -ая фонема, а  $s_i$  –  $i$ -ая комбинация букв слова.

На рис. 1 приведен пример дерева возможных вариантов фонем для слова «кот».

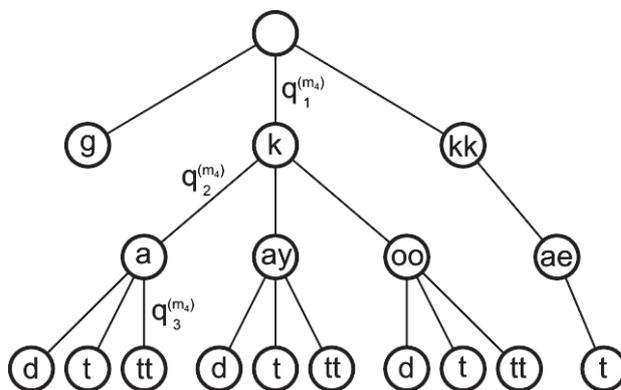


Рис. 1. Пример дерева возможных вариантов фонемного представления слова «кот»

Фонемное транскрибирование заключается в нахождении пути от корневой вершины до одного из листьев дерева  $T$ , который содержит дуги с максимальным значением произведения меток.

Пусть  $M = \{m_j\}$ ,  $j = 1 \dots z$  – множество путей от корневой вершины до листьев дерева  $T$ , где  $z$  – число листьев. Тогда вершины, принадлежащие пути  $m_j$  с максимальным произведением меток  $q_i$ , представляют фонемную транскрипцию слова

$$\operatorname{argmax}_j \prod_{i=1}^{N^{(m_j)}} q_i^{(m_j)}, \quad (8)$$

где  $N^{(m_j)}$  – число дуг пути  $m_j$ . Подробное описание алгоритма построения дерева приведено в [6].

### Фонемное транскрибирование на основе скрытой марковской модели (метод НММ)

Для решения задачи фонемного транскрибирования можно использовать скрытую марковскую модель (Hidden Markov Model, НММ) [8], в которой наблюдаемые последовательности являются буквенными значениями слова, а скрытые состояния – фонемами.

Для описания НММ требуется определение значений  $N$  – число состояний модели,  $M$  – число различных наблюдаемых реализаций, вероятностных мер  $A$ ,  $B$  и  $\pi$ . Величины  $A$  и  $B$  соответствуют распределениям вероятностей переходов между состояниями и появления наблюдаемых реализаций в конкретном состоянии, соответственно. Мера  $\pi$  задает начальное распределение вероятностей состояний.

Начальные параметры НММ определяются на основе статистических данных полученных при обработке словаря, который содержит слова в буквенном представлении и соответствующие им фонемные последовательности. Состояния НММ определяются алфавитом фонем, а наблюдения – возможными буквенными значениями, которые были выделены при сопоставлении значений фонемного и буквенного представлений слов из словаря. Начальное распределение вероятностей состояний вычисляется как частота появления фонем в словаре. Матрица  $A$  определяется вероятностями переходов между фонемами в слове. Элементы матрицы соответствуют значениям вероятностей  $P(\varphi_i | \varphi_j)$  и вычисляются как частоты появления последовательностей фонем  $\varphi_i \varphi_j$  в словах словаря. Матрица  $B$  определяется условными вероятностями  $P(\varphi_k | s_j)$  соответствия фонем возможным буквенным значениям, которые вычисляются на основе используемого словаря.

Для обучения НММ применяется алгоритм Баума-Уэлча [8].

### Эксперимент

Для проведения эксперимента разработана SDR-система на языке C#, использующая библиотеку rocketsphinx 0.7 [9] для распознавания речи и реализацию НММ [10]. Эксперимент заключался в выполнении поиска в коллекции [11], состоящей из 620 речевых документов. Эффективность поиска оценивалась посредством значений показателей полноты  $R$  (recall) и точности

$P$  (precision), которые усреднялись по 250 запросам, составленным на основе содержания речевых документов.

Таблица 1. Показатели эффективности поиска

Буквенное представление		Фонемное представление			
$R$	$P$	метод TREE		метод НММ	
		$R$	$P$	$R$	$P$
46,9	48,2	57,09	50,51	58,2	49,3

В таблице 1 приведены результаты поиска, полученные на основе фонемного представления слов (методы TREE и НММ), а также без использования фонемного транскрибирования (буквенное представление). Результаты (см. таблицу 1) иллюстрируют возможность повышения эффективности контекстного поиска речевых документов при использовании методов фонемного транскрибирования.

Оба метода фонемного транскрибирования содержат этап предварительной обработки. Так, в методе TREE требуется вычислить вероятности  $P(\varphi_{i-1} | s_{i-1})$ ,  $P(\varphi_i | \varphi_{i-1})$  и  $P(\varphi_i | s_i)$  на основе обучающего множества. Метод НММ, в свою очередь, включает этап обучения модели. В эксперименте использовалось обучающее множество, включающее речевые документы и соответствующие им фонемные транскрипции.

Для сравнения эффективности рассмотренных алгоритмов фонемного транскрибирования использовалась коллекция из 620 речевых документов, содержащая 9422 слова. В таблице 2 приведено время, затрачиваемое на этапы предварительной обработки и транскрибирования коллекции документов рассмотренными методами.

Таблица 2. Временные затраты при обработке коллекции речевых документов

Метод	TREE	НММ
Время предварительной обработки, с	1,43	22,78
Время транскрибирования, с	38,04	9,46
Общее время, с	39,47	32,24

### Выводы

Фонемное транскрибирование текстов, полученных в результате распознавания речевых документов, позволяет повысить эффективность контекстного поиска SDR-систем примерно на

23% по показателю полноты и на ~3,5% по показателю точности (см. таблицу 1). При использовании методов TREE и HMM достигаются сравнимые значения полноты и точности контекстного поиска (см. таблицу 1).

Метод TREE проигрывает методу HMM в 4 раза по времени фонемного транскрибирования речевых документов и выигрывает в 15,9 раза на этапе предварительной обработки. Дальнейшее направление работы связано с разработкой параллельной реализации метода TREE и повышением эффективности контекстного поиска посредством учета ошибок распознавания при поиске речевых документов.

### Литература

1. Kompatsiaris Y., Hobson P. *Semantic Multimedia and Ontologies: Theory and Applications*. Springer, 2008. – 290 p.
2. Wechsler M. *New Approaches to Spoken Document Retrieval // Information Retrieval*. Vol. 3, 2000. – P. 173-188.
3. Jones G., Foote J., Jones K.S., Young S. *Video mail retrieval using voice: An overview of the stage-2 system // Electronic Workshops in Computing*, Glasgow. Springer, 1995. – P. 7-7.
4. Glavitsch U. *The First Approach to Speech Retrieval // Technical Reports 238*, ETH Zürich, Institute of Information Systems, 1995. – 51 p.
5. Ng K., Zue V.W. *Phonetic Recognition for spoken document retrieval // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Vol. 1, 1998. – P. 325-328.
6. Яшина А.Г. Алгоритм контекстного поиска речевых аудиофайлов на основе фонемного сравнения слов // *Advanced Science*. №1, 2012. – С. 73-85. [http://www.vyatsu.ru/uploads/file/1210/1\\_\(2\).pdf](http://www.vyatsu.ru/uploads/file/1210/1_(2).pdf)
7. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. *Алгоритмы. Построение и анализ*. М.: ИД «Вильямс», 2005. – 1290 с.
8. Рабинер Л.Р. Скрытые марковские модели и их применение в избранных приложениях при распознавании речи. // *ТИИЭР*. Т. 77, № 2, 1989. – С. 86-120.
9. CMU Sphinx. *Open Source Toolkit For Speech Recognition // http://cmusphinx.sourceforge.net*
10. Accord.Net Framework // <http://code.google.com/p/accord/>
11. FestLang // <http://sourceforge.net/projects/festlang.berlios>

## PHONEMIC TRANSCRIPTION METHODS FOR THE SDR SYSTEM

Prozorov D.E., Yashina A.G.

**Two phonemic transcription methods are described and compared in this paper. The first method is based on graph theory and the second method uses hidden Markov model (HMM). These methods are used in the spoken document retrieval (SDR) system. These two methods are compared on work time and effectiveness of search of spoken documents by text query. Values of precision and recall which are obtained by these methods are shown.**

**Keywords:** *information retrieval, speech recognition, phonemic transcription, spoken document retrieval.*

Прозоров Дмитрий Евгеньевич, д.т.н., профессор Кафедры радиоэлектронных средств Вятского государственного университета (ВятГУ). Тел. 8-912-711-21-11. E-mail: prozo-rov.de@gmail.com

Яшина Александра Геннадьевна, ассистент Кафедры прикладной математики и информатики ВятГУ. Тел. 8-922-909-60-45. E-mail: ayashina.vshu@gmail.com

УДК 519.712

## ТЕХНОЛОГИЯ P2P-АУТСОРСИНГА ДЛЯ ОРГАНИЗАЦИИ РАСПРЕДЕЛЕНИЯ РАБОТ В ЕДИНОМ ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ ПРЕДПРИЯТИЯ

Иващенко А.В., Леднев А.М.

В статье описывается технология P2P-аутсорсинга для организации взаимодействия персонала предприятия в едином информационном пространстве по распределению задач. Для реализации P2P-аутсорсинга предлагается организация проведения виртуальных аукционов. Приводится пример реализации такой системы на предприятии нефтегазовой отрасли.

**Ключевые слова:** информационные технологии, управление проектами, P2P-сети, аутсорсинг, аукцион.

### Введение

Одной из актуальных проблем управления современным предприятием является организация выполнения сложных задач в условиях не-