

УДК 004.934

*А. Г. Татарина, Д. Е. Прозоров*

**АЛГОРИТМ СОПОСТАВЛЕНИЯ БУКВ И ФОНЕМ  
НА ОСНОВЕ ПРАВИЛ  
В ЗАДАЧЕ ПОИСКА РЕЧЕВЫХ ДОКУМЕНТОВ  
ПО ТЕКСТОВОМУ ЗАПРОСУ\***

Поиск речевой информации по текстовому запросу является актуальной, но трудной задачей. Один из подходов к решению этой задачи основан на преобразовании речи в текст. Преимущество данного подхода заключается в возможности использования лингвистической информации. К его недостаткам можно отнести ошибки, возникающие при работе систем автоматического распознавания речи и искажающие текст, по которому впоследствии выполняется поиск. Одним из способов уменьшения влияния ошибок распознавания на эффективность поиска речевой информации является сочетание алгоритмов фонемного транскрибирования и алгоритмов приближённого сравнения строк. В некоторых алгоритмах фонемного транскрибирования и, в том числе, при определении меры сходства фонем требуются оценки условных распределений фонем, сопоставляемым буквам транскрибируемого слова. Задача сопоставления известна как графемно-фонемное выравнивание. В статье представлен алгоритм графемно-фонемного выравнивания на основе правил. Алгоритм может быть использован для оценки условных распределений фонем при наличии обучающего словаря с соответствующими транскрипциями.

*Ключевые слова:* графемно-фонемное выравнивание, фонемное транскрибирование, поиск речевых документов.

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 с речевым содержанием. Примерами мультимедийных документов с речевым содержанием (МДРС) являются радио- и видеонОВОСТИ, аудиокниги, доклады конференций и, в последнее время, образовательные ресурсы в виде аудио- и видео-лекций, интерактивные учебные фильмы и мультимедийные методические разработки [1]. Также развитие речевых технологий привело к появлению человеко-машинных интерфейсов с голосовым управлением.

Задача поиска МДРС в широком смысле относится к т.н. области «Spoken Content Retrieval» (SCR – поиск речевой информации) [2, 3]. Анализу проблем в области SCR и поиску путей их решения посвящено большое количество как зарубежных [2–9], так и отечественных [10–12] научных исследований. В общем случае SCR-задача заключается в определении соответствия содержания речевого сигнала текстовому или устному запросу пользователя. В данной работе соответствие содержания подразумевает совпадение слов запроса со словами, произносимыми в речевом сигнале. Такая задача имеет название «обнаружение речевых терминов» (Spoken Term Detection – STD). Следует заметить, что известная задача «обнаружение ключевых слов» (Keyword Spotting) отличается от задачи STD. Обнаружение ключевых слов подразумевает поиск слов из заранее определенного множества [12].

Один из подходов к поиску речевой информации заключается в восстановлении произнесенных слов в речевых документах на основе лингвистической информации, и последующее сравнение полученных слов со словами запроса. Методы данного подхода основаны на использовании автоматического распознавания речи (АРР). Основным недостатком использования АРР является наличие ошибок распознавания, которые искажают содержание речевой информации, что снижает эффективность поиска в целом. Качество распознавания зависит от разнообразных акустических параметров речевого сигнала

(например, шум, темп речи, интонация, акцент), а также содержания произносимой речи (разнообразии произносимых слов).

Объединение распознавания речи в текст и фонемного транскрибирования распознанного содержимого позволяет повысить эффективность поиска по сравнению с обычным текстовым поиском в сочетании с АРР [5]. В этом случае поиск речевых документов выполняется на фонемном уровне, что позволяет частично снизить влияние ошибок распознавания. Ключевая идея использования фонемного транскрибирования совместно с АРР заключается в представлении фонем в виде векторов, компоненты которых равны значениям условных вероятностей появления фонемы в зависимости от букв в слове. Это позволяет сравнивать распознанные слова документов со словами запроса посредством вещественной меры сходства, настраиваемой по обучающему множеству в автоматическом режиме. Данное обучающее множество состоит из буквенного и фонемного представления слов. Обучающее множество может быть сформировано на основе лексикона.

Важную роль в некоторых методах фонемного транскрибирования имеет оценка условных вероятностей  $P(\varphi_j | c_i)$ , где  $\varphi_j$  – фонема алфавита  $\Phi$ , а  $c_i$  – буква алфавита  $C$ . Данная подзадача может быть решена путём подсчета количества соответствий фонемы  $\varphi_j$  букве  $c_i$ , которые встречаются в транскрипции и буквенном представлении слов обучающего словаря  $\Psi$ .

Сопоставление букв слова с фонемами его транскрипции выполняется посредством алгоритма графемно-фонемного выравнивания (grapheme-to-phoneme alignment, G2P alignment) [13, 14].

Существует большое разнообразие алгоритмов графемно-фонемного выравнивания, в которых используются разные способы сопоставления букв с фонемами. Например, можно выделить алгоритмы на основе [13, 14]:

- составленных вручную экспертом набора правил,
- динамического программирования,
- деревьев решений,

- нейронных сетей,
- EM-алгоритма.

Формальная постановка задачи графемно-фонемного выравнивания может быть сформулирована следующим образом.

Пусть даны буквенное  $w^c = c_1c_2...c_m$  и фонемное  $w^\phi = \phi_1\phi_2...\phi_n$  представления слова  $w$ . Необходимо взаимно-однозначно сопоставить элементы последовательности  $w^c$  с элементами последовательности  $w^\phi$ .

Возможны следующие три вида сопоставлений букв и фонем [13]:

- один к одному,
- один ко многим,
- многие ко многим.

Примеры видов сопоставлений представлены в таблице 1.

Таблица 1

### Виды сопоставлений букв и фонем

Вид сопоставления	Пример
один к одному	с а м о л ё т
	[s] [ay] [m] [a] [ll] [jo] [t]
один ко многим	ф а к у ль т е т
	[f] [ay] [k] [u] [ll] [tt] [je] [t]
многие ко многим	а ль я н с
	[a] [ll] [j_ja] [n] [s]

Рассмотрим алгоритм графемно-фонемного выравнивания на основе правил.

Имеется обучающий словарь  $\Psi$ , содержащий слова с их фонемными транскрипциями. В данной работе для экспериментов использовался словарь, сформированный из 4845 слов «Частотного словаря публицистики» [15], являющегося частью «Нового частотного словаря русской лексики» [16], который создан на основе «Национального корпуса русского языка» [17]. Фонемные транскрипции слов взяты из словаря [18], при формировании которого исполь-

зовался транскриптор [19] на основе правил. Примеры транскрипций приведены в таблице 2.

Таблица 2

### Примеры слов обучающего словаря

Буквенное представление слова	Фонемная транскрипция слова
библиотека	[bb] [i] [b] [ll] [i] [a] [tt] [je] [k] [ay]
благодарить	[b] [l] [ay] [g] [ay] [d] [a] [rr] [ii] [tt]
богатство	[b] [a] [g] [aa] [c] [t] [v] [ay]
масштаб	[m] [a] [sh] [t] [aa] [p]
явиться	[j] [i] [vv] [ii] [c] [ay]

Для записи транскрипций слов обучающего словаря  $\Psi$  используется фонемный словарь, состоящий из 16 гласных и 36 согласных фонем. Данные фонемы приведены в таблице 3.

Таблица 3

### Начальный фонемный словарь

Тип	Фонемы
гласные	[a] [aa] [ay] [ee] [i] [ii] [ja] [je] [jo] [ju] [oo] [u] [uj] [uu] [y] [yy]
согласные	[p] [pp] [b] [bb] [t] [tt] [d] [dd] [k] [kk] [g] [gg] [c] [ch] [f] [ff] [v] [vv] [s] [ss] [z] [zz] [sh] [sch] [zh] [h] [hh] [m] [mm] [n] [nn] [l] [ll] [r] [rr] [j]

Буквы русского алфавита также можно разделить на гласные и согласные. Дополнительно существуют буквы, являющиеся нейтральными по звучанию (см. Таблицу 4).

Таблица 4

### Типы букв русского алфавита

Тип	Буквы
гласные	а, у, о, ы, и, э, я, ю, ё, е
согласные	б, в, г, д, ж, з, й, к, л, м, н,

## Технические науки

	п, р, с, т, ф, х, ц, ч, ш, щ
нейтральные	ь, ь

Произношение гласных букв отражается гласными фонемами, согласных букв – согласными фонемами. Окружение буквы в слове влияет на её произношение. Произношение гласных звуков может уточняться. Например: ударный, безударный, смягчающий звуки. Произношение согласного звука может быть твёрдыми, мягким, шипящим, свистящим, звонким, глухим и т. д. Формируя правила графемно-фонемного выравнивания, основанные на чтении букв в зависимости от окружения в слове, необходимо учитывать и правила получения транскрипции, по которой выполняется сопоставление. Иначе необходимо предусматривать исключения в правилах сопоставления букв слова с фонемами его транскрипции. Правильных вариантов фонемной транскрипции слова может быть несколько, тогда как буквенное представление слова – одно. А поскольку формирование обучающего словаря  $\Psi$  не ограничивается определенным транскриптором, необходимы правила сопоставления, использующие сами слова и их транскрипции. Например, более общие правила, которые учитывают только тип буквы и фонемы, а также позиции буквы в слове и фонемы в его транскрипции. Данные правила основаны на том, что порядок следования согласных и гласных букв в слове сохраняется и в его фонемной транскрипции.

Буквенное представление слова  $w$  является последовательностью букв  $w^c = c_0, \dots, c_m$ , где  $c_i \in C$ . Фонемная транскрипция слова  $w$  представляет последовательность фонем  $w^\phi = \varphi_0, \dots, \varphi_n$ , где  $\varphi_j \in \Phi$ . Введем индексы  $k$  и  $t$ , указывающие позиции элементов в соответствующих последовательностях. Обозначим  $k$ -й элемент последовательности  $w^c$  как  $c[k]$ , а  $t$ -ый элемент последовательности  $w^\phi$  как  $\varphi[t]$ .

В процессе одновременного поэлементного просмотра  $w^c$  и  $w^\phi$ , в зависимости от типов текущих буквы  $c[k]$  и фонемы  $\varphi[t]$  будет выполняться изменение индексов  $k$  и  $t$ , а также вставка символа "|" для разделения соседних букв

и фонем разных типов. Полученные в результате последовательности обозначим как «выровненные». Примеры таких выровненных последовательностей представлены в таблице 5.

Таблица 5

### Примеры графемно-фонемного выравнивания на основе правил

б   и   б   л   и   о   т   е   к   а
bb   i   b   ll   i   a   tt   je   k   ay
б   л   а   г   о   д   а   р   и   ть
b   l   ay   g   ay   d   a   rr   ii   tt
б   о   г   а   т   с   тв   о
b   a   g   aa   c   t   v   ay
м   а   с   шт   а   б
m   a   sh   t   aa   p
я   в   и   тьс   я
j i   vv   ii   c   ay

Рассмотрим правила построения выровненных последовательностей, определяемые возможными значениями типов текущих буквы  $c[k]$  и фонемы  $\varphi[t]$ , а также следующих буквы  $c[k+1]$  и фонемы  $\varphi[t+1]$ :

Псевдокод алгоритма выравнивания:

–  $c[k]$  – согласная,  $\varphi[t]$  – согласная,  $\varphi[t+1]$  – согласная (кроме "j")

**ЕСЛИ**  $c[k+1]$  – согласная **ТО** {

$\varphi[t]\varphi[t+1] \rightarrow \varphi[t]|\varphi[t+1]$

$c[k]c[k+1] \rightarrow c[k]|c[k+1]$

$t = t + 1; k = k + 1; \}$

**ЕСЛИ**  $c[k+1]$  – гласная **ТО** {  $t = t + 1; \}$

**ЕСЛИ**  $c[k+1]$  – нейтральная **ТО** {

$\varphi[t]\varphi[t+1] \rightarrow \varphi[t]|\varphi[t+1];$

$c[k]c[k+1] \rightarrow c[k]c[k+1]|c[k+2];$

## Технические науки

$t = t + 1; k = k + 2; \}$

–  $c[k]$  – согласная,  $\varphi[t]$  – согласная,  $\varphi[t + 1]$  – гласная

**ЕСЛИ**  $c[k + 1]$  – согласная **ТО**  $\{ k = k + 1; \}$

**ЕСЛИ**  $c[k + 1]$  – гласная **ТО**  $\{$

$\varphi[t]\varphi[t + 1] \rightarrow \varphi[t] | \varphi[t + 1]$

$c[k]c[k + 1] \rightarrow c[k] | c[k + 1]$

$t = t + 1; k = k + 1; \}$

**ЕСЛИ**  $c[k + 1]$  – нейтральная **ТО**  $\{ k = k + 2; \}$

–  $c[k]$  – согласная,  $\varphi[t]$  – согласная,  $\varphi[t + 1] = [\langle\langle j \rangle\rangle]$

**ЕСЛИ**  $c[k + 1]$  – согласная **ТО**  $\{ k = k + 1; \}$

**ЕСЛИ**  $c[k + 1]$  – гласная **ТО**  $\{$

$\varphi[t]\varphi[t + 1] \rightarrow \varphi[t] | \varphi[t + 1]$

$c[k]c[k + 1] \rightarrow c[k] | c[k + 1]$

$t = t + 1; k = k + 1; \}$

**ЕСЛИ**  $c[k + 1]$  – нейтральная **ТО**  $\{$

$\varphi[t]\varphi[t + 1] \rightarrow \varphi[t] | \varphi[t + 1];$

$c[k]c[k + 1] \rightarrow c[k]c[k + 1] | c[k + 2];$

$t = t + 1; k = k + 2; \}$

–  $c[k]$  – гласная,  $\varphi[t]$  – гласная,  $\varphi[t + 1]$  – согласная (кроме  $\langle\langle j \rangle\rangle$ )

**ЕСЛИ**  $c[k + 1]$  – согласная **ТО**  $\{$

$\varphi[t]\varphi[t + 1] \rightarrow \varphi[t] | \varphi[t + 1];$

$c[k]c[k + 1] \rightarrow c[k] | c[k + 1];$

$t = t + 1; k = k + 1; \}$

**ЕСЛИ**  $c[k + 1]$  – гласная **ТО**  $\{ k = k + 1; \}$

–  $c[k]$  – гласная,  $\varphi[t]$  – гласная,  $\varphi[t + 1]$  – гласная

**ЕСЛИ**  $c[k + 1]$  – согласная **ТО**  $\{ t = t + 1; \}$

**ЕСЛИ**  $c[k+1]$  – гласная **ТО** {

$\varphi[t]\varphi[t+1] \rightarrow \varphi[t] | \varphi[t+1];$

$c[k]c[k+1] \rightarrow c[k] | c[k+1];$

$t = t + 1; k = k + 1; \}$

–  $c[k]$  – гласная,  $\varphi[t]$  – гласная,  $\varphi[t+1] = [«j»]$

**ЕСЛИ**  $c[k+1]$  – согласная (кроме «Й») **ИЛИ**  $c[k+1]$  – гласная **ТО** {

$\varphi[t]\varphi[t+1]\varphi[t+2] \rightarrow \varphi[t]\varphi[t+1] | \varphi[t+2];$

$c[k]c[k+1] \rightarrow c[k] | c[k+1];$

$t = t + 2; k = k + 1; \}$

**ЕСЛИ**  $c[k+1] = «Й»$  **ТО** {

$\varphi[t]\varphi[t+1] \rightarrow \varphi[t] | \varphi[t+1];$

$c[k]c[k+1] \rightarrow c[k] | c[k+1];$

$t = t + 1; k = k + 1; \}$

–  $c[k]$  – гласная,  $\varphi[t] = [«j»]$

$t = t + 1;$

Символы "|" выровненных последовательностей  ${}^1w^C$  и  ${}^1w^\Phi$  разделяют буквы и фонемы на группы внутри данных последовательностей. Таким образом, буквы и фонемы, принадлежащие группам на одной позиции в обеих последовательностях, сопоставляются между собой.

В результате работы описанного алгоритма графемно-фонемного выравнивания для каждого слова словаря  $\Psi$  получена пара выровненных последовательностей  $\langle {}^1w^C; {}^1w^\Phi \rangle$ . Используя данные пары, можно составить множество  $G = \{g_v\}$  графемно-фонемных совместных мультиграмм (grapheme-phoneme joint multigram), которые известны как графоны (graphone) [13]. Графоном  $g_v = \langle x^{s_v}; y^{s_v} \rangle \in C \times \Phi$  называется пара, состоящая из сопоставленных между собой последовательности букв  $x^{s_v} = c_r \dots c_q$  и последовательности фонем

$y^{s_v} = \varphi_1 \dots \varphi_p$ . В таблице 6 показаны примеры графонов, полученные на основе выровненных последовательностей из таблицы 5.

Таблица 6

### Примеры графонов

Слово	Графоны
библиотека	<б-bb> <и-i> <б-b> <л-ll> <и-i> <о-a> <т-tt> <е-je> <к-k> <а-ay>
благодарить	<б-b> <л-l> <а-ay> <г-g> <о-ay> <д-d> <а-a> <р-rr> <и-ii> <ть-tt>
богатство	<б-b> <о-a> <г-g> <а-aa> <т-c> <с-t> <тв-v> <о-ay>
масштаб	<м-m> <а-a> <с-sh> <шт-t> <а-aa> <б-p>
явиться	<я-i j> <в-vv> <и-ii> <тьс-c> <я-ay>

Используя множество графонов  $G$  можно оценить условные вероятности  $P(\varphi_j | c_i)$  для всех  $\varphi_j \in \Phi$  и  $c_i \in C$ :

$$P(\varphi_j | c_i) = \frac{|G^{c_i, \varphi_j}|}{|G^{C, \varphi_j}|}, \quad (1)$$

где  $G^{c_i, \varphi_j} = \{g_v : c_i \in x^{s_v}, \varphi_j \in y^{s_v}\}$  – множество графем, содержащих букву  $c_i$  и фонему  $\varphi_j$ ,  $G^{C, \varphi_j} = \{g_v : \varphi_j \in y^{s_v}\}$  – множество графем, содержащих фонему  $\varphi_j$ .

Таким образом, представленный алгоритм графемно-фонемного выравнивания позволяет формировать множество графонов. Множество графонов используется для оценки условных вероятностей  $P(\varphi_j | c_i)$ , которые необходимы при поиске речевых документов по текстовому запросу на основе фонемного транскрибирования.

### Список литературы

1. Шемончук Д. С. Разработка и исследование методов улучшения функционала сетевых мультимедийных порталов в сфере управления образовательными процессами: дис. ... канд. техн. наук. М.: МИЭМ, 2009.

2. *Larson M., Jones G. J. F.* Spoken Content Retrieval: A Survey of Techniques and Technologies // Foundations and Trends in Information Retrieval. Vol. 5. Nos. 4.5 (2011). P. 235–422.
3. *Lee L.-s., Glass J., Lee H.-y., Chan C.-a.* Spoken content retrieval – beyond cascading speech recognition with text retrieval // IEEE ACM Transactions on Audio, Speech and Language Processing. 2015. Vol. 23. № 23. P. 1389–1420.
4. *Brown M. G.* Video Mail Retrieval by Voice: An Overview of the Cambridge/Olivetti Retrieval System / M. G. Brown, J. T. Foote, G.J.F. Jones, K. Sparck Jones, S.J. Young. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.18.1832&rep=rep1&type=pdf>
5. *Wechsler M., Munteanu E., Schauble P.* New Approaches to Spoken Document Retrieval // Information Retrieval. 2000. V. 3. P. 173–188.
6. *Glavitsch U., Schauble P.* A system for retrieving speech documents // In Proceedings of ACM. SIGIR. 1992. P. 168–176.
7. Large Vocabulary Continuous Speech Recognition of an Inflected Language using Stems and Endings // SPECOM. 2007. Vol. 49. Iss. 6. P. 437–452.
8. *Hauptmann A.* Automatic spoken document retrieval. URL: <http://repository.cmu.edu/compsci/960/>
9. *Ng K., Zue V.* Subword Unit Representations for Spoken Document Retrieval // Proceedings of Eurospeech 97. 1997. P. 1607–1610.
10. *Кипяткова И. С., Ронжин А. Л., Карнов А. А.* Автоматическое распознавание русской речи // СПИИРАН. СПб.: ГУАП. 2013. 314 с.
11. *Янь Цзинбинь, Хейдоров И. Э., Ткачя А. А.* Исследование характеристик системы поиска ключевых слов на основе минимального интервала редактирования и мер достоверности // Речевые технологии. 2009. С. 5–14.
12. *Гусев М. Н.* Методы и модели распознавания русской речи в информационных системах: дис. ... д-ра техн. наук. СПб.: СПб ГУТ, 2014.
13. *Bisani M., Ney H.* Joint-sequence models for grapheme-to-phoneme conversion // SPECOM. 2007.
14. *Широкова А. М.* Буквенно-звуковое преобразование в системах автоматической обработки речи // Структурная и прикладная лингвистика. СПб.: межвуз. сб. / под ред. А. С. Герда. Вып. 10. 2014. 384 с.
15. Частотный словарь публицистики. URL: <http://dict.ruslang.ru/>
16. *Ляшевская О., Шаров С.* Новый частотный словарь русской лексики // Азбуковник. 2009. С. 1112.
17. Национальный корпус русского языка. URL: <http://ruscorpora.ru/>

18. Акустические и языковые модели. URL: <https://sourceforge.net/projects/cmuspinx/files/Acoustic%20and%20Language%20Models/Russian/>

19. ru4sphinx. URL: <https://github.com/zamiron/ru4sphinx>

**ТАТАРИНОВА Александра Геннадьевна** – ассистент кафедры прикладной информатики, Вятский государственный университет. 610000, г. Киров, ул. Московская, 36.

E-mail: [tatarinova.alexg@gmail.com](mailto:tatarinova.alexg@gmail.com)

**ПРОЗОРОВ Дмитрий Евгеньевич** – доктор технических наук, профессор кафедры радиоэлектронных средств, Вятский государственный университет. 610000, г. Киров, ул. Московская, 36.

E-mail: [prozorov.de@gmail.com](mailto:prozorov.de@gmail.com)