

УДК 004.912

## ПРИМЕНЕНИЕ СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ПРОГРАММЕ ФОНОСЕМАНТИЧЕСКОГО АНАЛИЗА

*Звягинцева А.А., Гранковский В.А.*

### **Введение**

Фоносемантика — это наука, изучающая связь между звучанием и значением слова. Она появилась сравнительно недавно, в XX веке. В 1957 году Чарльз Осгуд в своей работе [1] изложил основные принципы этой науки (и уточнил их в следующей книге [2]), а впоследствии эти методы были адаптированы для русского языка А. П. Журавлевым [3]. На основе его материалов в дальнейшем было разработано большое множество программ [4-7], выполняющих фоносемантический анализ русских слов и текстов.

Однако всё рассмотренное нами программное обеспечение не лишено недостатков. Нашей целью стала разработка программного продукта технически более совершенного и более удобного с точки зрения интерфейса. Мы попытались избежать существующие изъяны.

Мы разработали программу и интернет-сервис для выполнения фоносемантического анализ слов, прозаических и поэтических текстов, предоставляемые возможности которых совершенно бесплатны для пользователей.

### **Технология анализа**

В рассмотренных нами программах при вводе текста для анализа необходимо вручную расставлять ударения, у нас же этот процесс автоматизирован. Для этого в отдельном файле содержится орфоэпический словарь русского языка, необходимое слово в котором мы ищем с помощью алгоритма бинарного поиска за короткое время.

В существующих аналогах учитывается исключительно буквенное написание, что в фоносемантическом анализе не корректно. Мы же разработали алгоритм фонетического транскрибирования текста для получения более правильных результатов. Таким образом, наша программа анализирует звучание слова, а не его написание.

В других программах значения результатов анализа текста усреднены, а это неверно, поскольку настройки отрывка могут меняться с течением времени. У нас эти изменения можно наблюдать.

Для более наглядного представления результатов анализа они визуализируются. Данные отображаются в виде графика поверх текста, а также изменяющихся цвета, размера и отступов от левого края. Пример результата анализа текста, полученный в нашей программе приведен на рис. 1.

В нашей программе анализ выполняется по двадцати пяти различным шкалам. Для выведения результатов выбираются три наиболее ярко выраженные для текста (сумма отклонений от средних значений у которых максимальна). Расшифровка условных обозначений показана над результатами.

Использование технологий .NET Framework позволило применить широкий спектр средств для визуализации. Результирующее изображение имеет векторный формат, что положительно влияет на качество. Также в программе имеется возможность векторной печати, кроме того результаты можно сохранить в формате \*.jpg, \*.gif и \*.png.

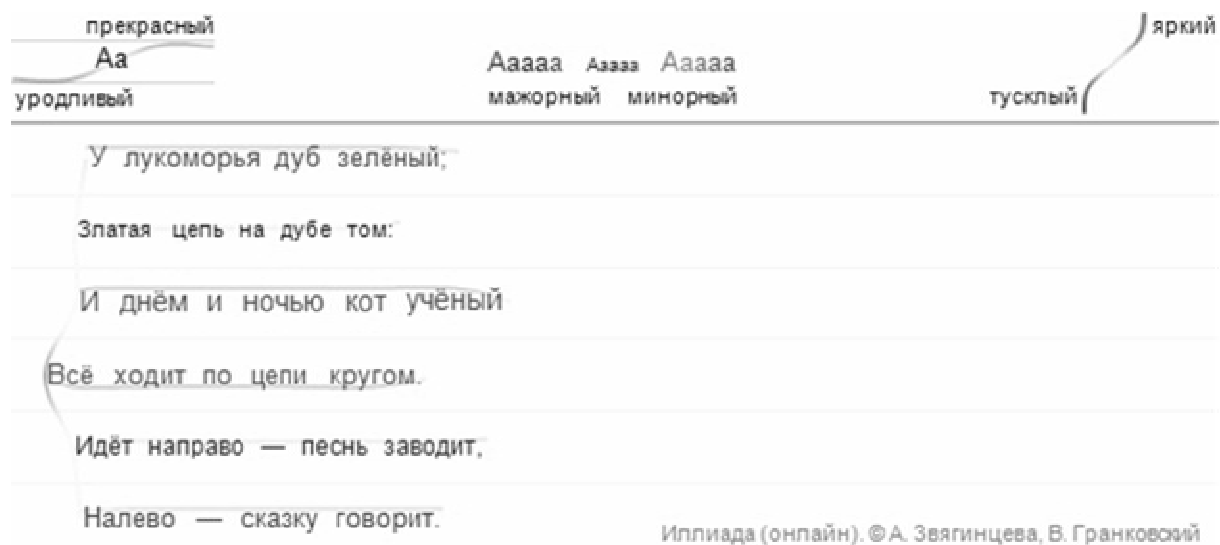


Рисунок 1 — Пример анализа, выполненного в нашей программе

### Особенности офлайн-версии разработки

На рис. 2 представлен интерфейс программы.

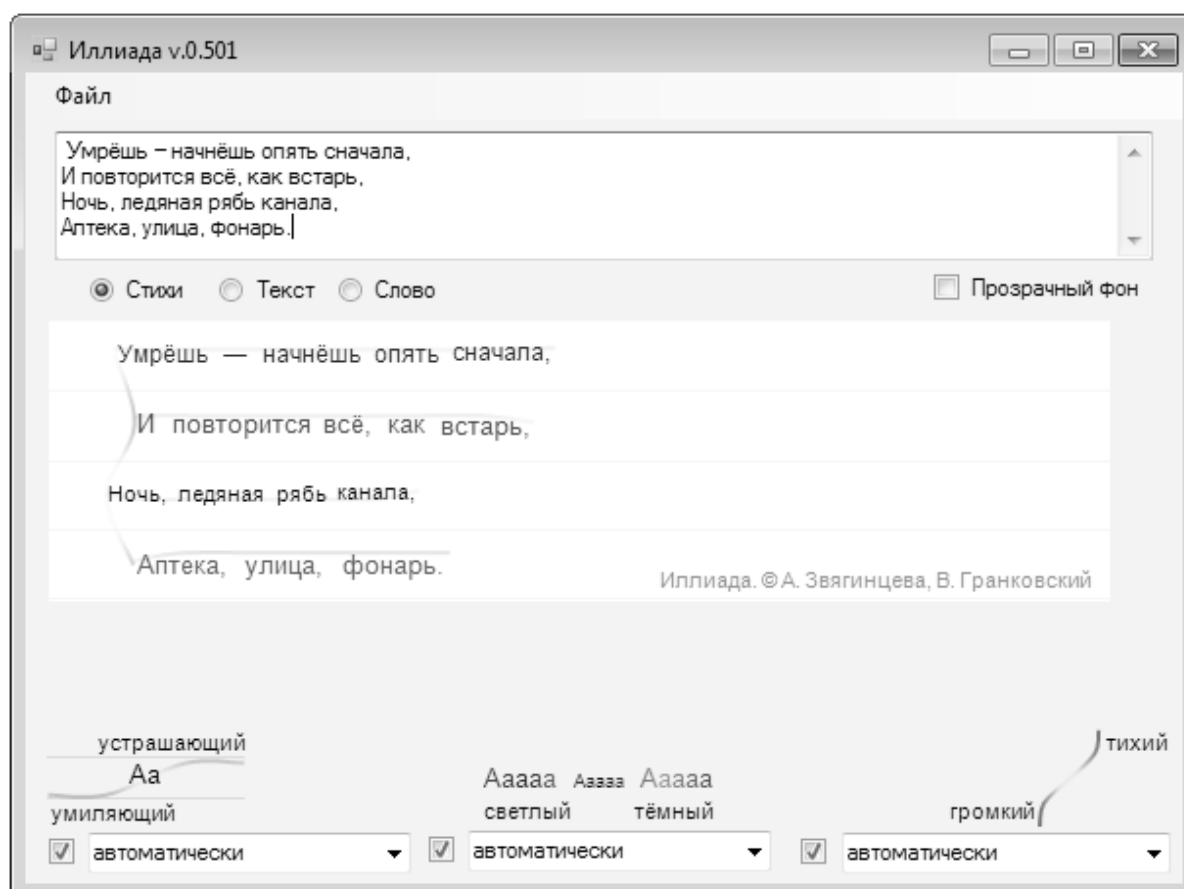


Рисунок 2 — Интерфейс программы

В текстовом поле, находящемся в верхней части формы, пользователю предлагается ввести данные для анализа. Ниже необходимо выбрать один из трёх режимов: «стихи», «текст» или «слово». Результаты практически мгновенно отобразятся в поле ниже. Результаты рассчитываются в реальном времени, таким образом при изменении входных данных они сразу же обновляются. По умолчанию программа сама выбирает наиболее подходящие для данного отрывка шкалы, однако при желании их можно задать вручную, выбрав нужные в размещенных внизу полях выбора. Также можно отключить визуализацию по некоторым шкалам.

На рис. 3 изображена работа программы в режиме «слово». В текстовое поле вводится одно слово. В поле вывода результатов отображается транскрипция в знаках международного фонетического алфавита. Ниже в порядке убывания значимости показаны значения по двадцати пяти шкалам. Для большей наглядности более сильные шкалы изображены более ярким цветом.

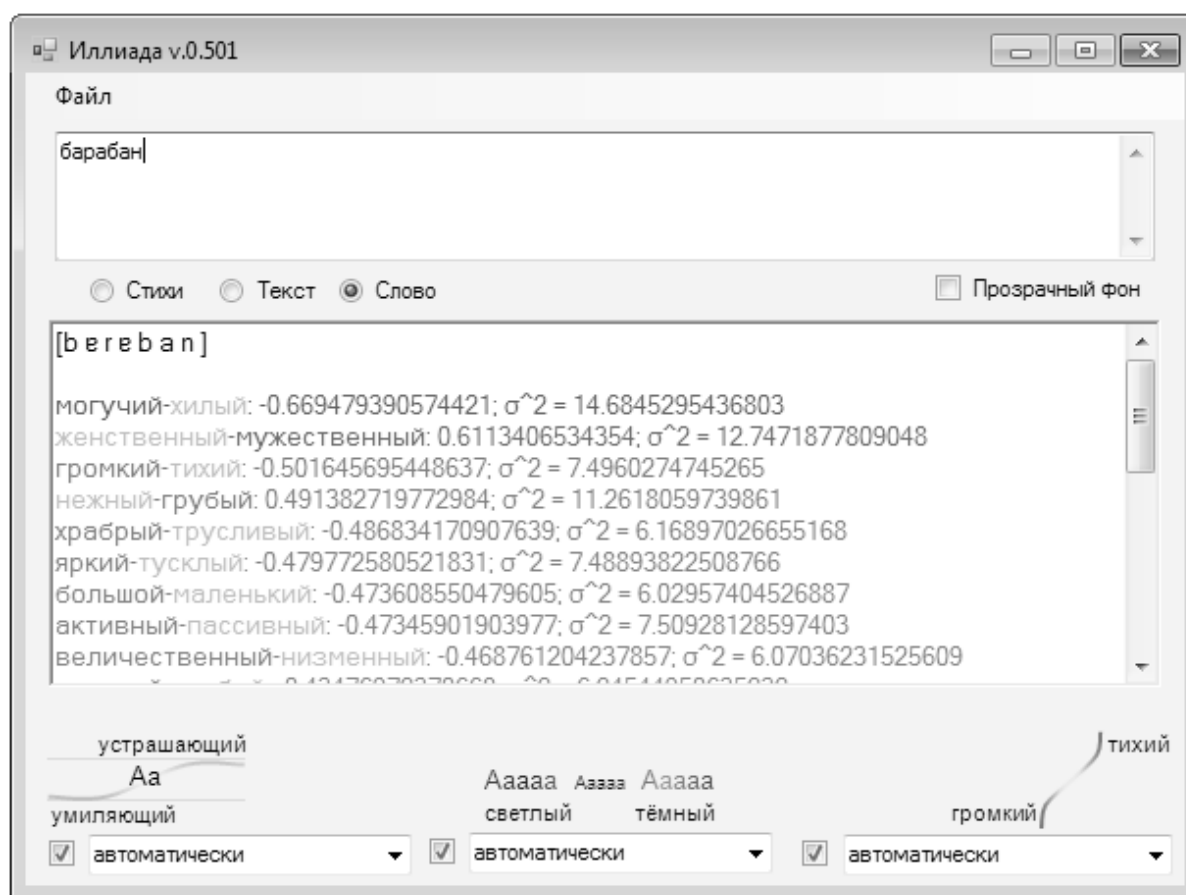


Рисунок 3 — Анализ в режиме «слово»

Для повышения скорости работы программы была использована многопоточность. Задачи распараллеливаются на несколько одновременно выполняющихся потоков. Интерфейс и вычисления работают в разных потоках, благодаря чему программа не зависает во время анализа больших объемов текста. При изменении входных данных программа практически мгновенно пересчитывает результаты анализа. При большом объеме текста, когда вычисления занимают более длительный промежуток времени, появляется полоса заполнения, а поле с результатом

становится серым (рис. 4). В том случае, если входные данные были изменены, а анализ к тому времени еще не закончился, то он перезапускается.

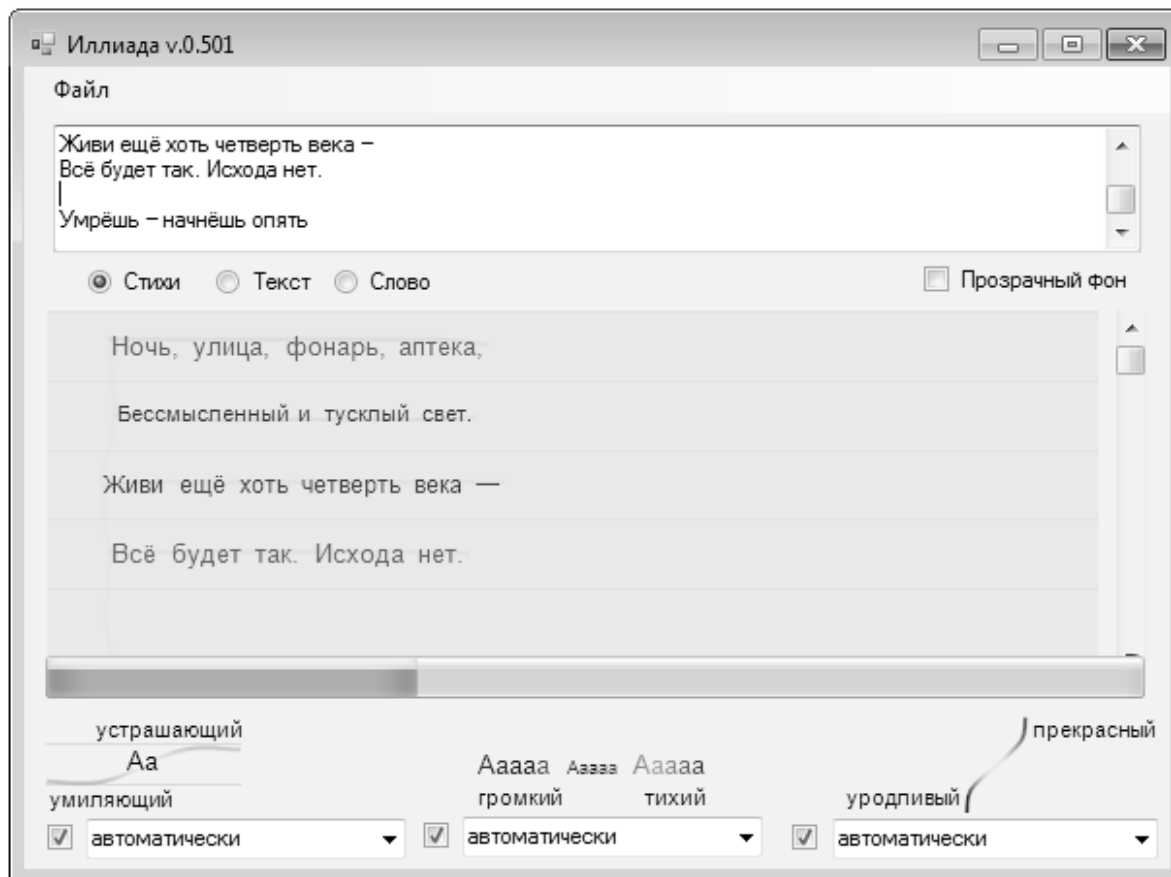


Рисунок 4 — Полоса заполнения при больших объемах текста

### Особенности онлайн-версии разработки

Кроме офлайн-версии программы, мы разработали сайт проекта «Иллиада», который доступен по адресу <http://illiada.net>. Анализ текста в онлайн-версии не уступает по качеству программе. Однако, в первой имеются ограничения на количество вводимого текста и отсутствует возможность ручного выбора шкал. Пользователь может скачать полную версию программы после бесплатной регистрации.

На отдельной странице сайта предоставлена возможность бесплатного анализа. Там в поле следует ввести текст, выбрать необходимый режим и нажать кнопку «Выполнить анализ». На сайте анализ работает не в реальном времени, а только по нажатию кнопки. Для меньшего расхода трафика благодаря средствам AJAX при нажатии на кнопку веб-браузер перезагружает не всю страницу, а только её часть.

Для начинающих работу пользователей предусмотрена справка. Для повышения удобства мы использовали различные скрипты. Например, при нажатии на поле для ввода текста оно автоматически растягивается вниз. Кроме того справка отображается в отдельном плавно всплывающем окне. На рисунке 5 изображен интерфейс онлайн-анализатора.

В дальнейшем планируется развивать наш сайт. В частности, мы собираем базу примеров проанализированных текстов, снабженную новостной RSS-лентой.

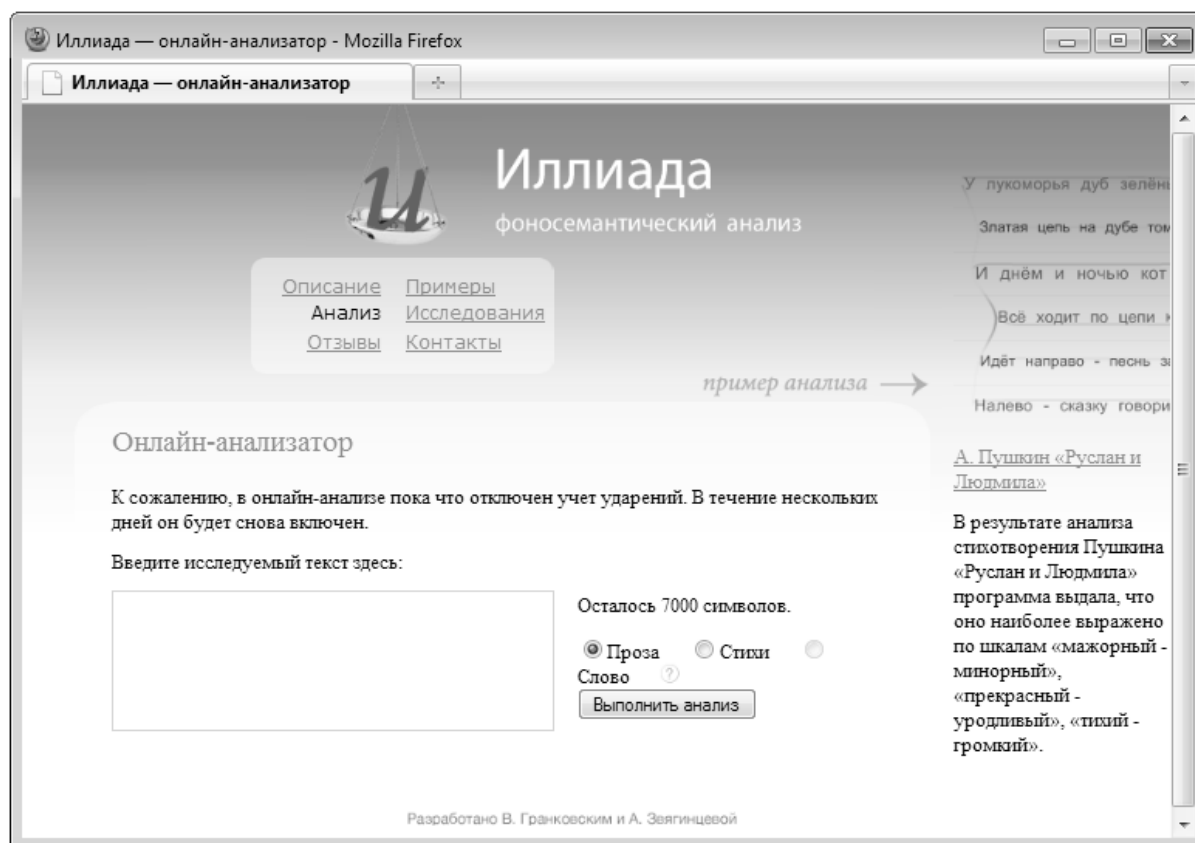


Рисунок 5 — Страница анлиза на сайте

### Заключение

На основе теории фоносемантики, в частности, разработок А. П. Журавлева, мы построили программный продукт, обладающих гораздо большей функциональностью из рассмотренных нами существующих аналогов. Однако на этом работа не останавливается. Мы надеемся, что наша программа будет интересна как профессионалам, так и обычным пользователям.

### Литература

- [1] Charles E. Osgood. The Measurement of Meaning/ Charles E. Osgood, George Suci, & Percy Tannenbaum. University of Illinois Press, 1957.
- [2] Charles E. Osgood. Method and Theory in Experimental Psychology. Oxford, 1956.
- [3] Журавлёв А.П. Фонетическое значение. Л.: ЛГУ, 1974. 159 с.
- [4] Интернет-ресурс системы фоносемантического анализа «ВААЛ». <http://www.vaal.ru/>
- [5] Интернет-ресурс по анализу имен и фамилий. <http://psevdonim.ru/>
- [6] Интернет-ресурс по анализу имен и фамилий. <http://www.analizfamilii.ru/>
- [7] Интернет-ресурс по анализу писем. <http://www.analizpisem.ru/>