

ИССЛЕДОВАТЕЛЬСКИЙ ПОРТАЛ ДЛЯ АНАЛИЗА И ОЦЕНКИ СТИЛЯ НАУЧНЫХ ПУБЛИКАЦИЙ

Шучалова Юлия Сергеевна, Ланин Вячеслав Владимирович

Национальный исследовательский университет «Высшая школа экономики», 614070, Россия,
г. Пермь, ул. Студенческая, 38, shuchalova@gmail.com

Необходимость соблюдения норм академического стиля при создании научных публикаций на английском языке – актуальная проблема для многих авторов, не являющихся носителями языка. Рекомендации для улучшения академического письма чаще всего носят теоретический характер, но существует возможность исследовать особенности их применения на практике с помощью анализа текстовых корпусов. В данной работе предложена концепция и функциональные требования для Интернет-портала, предоставляющего доступ к сервису анализа и оценки стиля научных публикаций. Портал предназначен для нескольких категорий пользователей и позволит проводить исследования с использованием различных корпусов и наборов лингвистических данных, а также просто получать оценку характеристик загружаемой работы относительно других статей по данной тематике и соответствующие рекомендации. На текущий момент исследовательский прототип портала находится в стадии реализации.

Ключевые слова: академический стиль английского языка, стилистический анализ, корпусная лингвистика, оценка качества публикаций

Качественный перевод статьи на английский язык – одна из задач, с которой сталкиваются многие авторы научных публикаций, не являющиеся носителями этого языка. Особую трудность представляет соблюдение норм академического стиля, необходимое для грамотного представления результатов своей работы в научном сообществе. Учебные материалы по академическому письму (*EnglishforAcademicPurposes*) в основном направлены на изучение часто встречающихся фраз и основных теоретических приёмов академизации текста, но, как правило, не уточняют частоту и интенсивность их использования на практике. Таким образом, для полноценного развития навыка необходимо выполнение большого количества практических заданий под контролем эксперта, что не всегда возможно в условиях ограниченного времени и трудозатрат.

В работе [1] представлен метод оценки качества стиля научной статьи, основанный на её сравнении с корпусом (коллекцией текстов), состоящим из «эталонных» статей по подобной тематике, т.е. прошедших профессиональное рецензирование. Сравнение проводится на основе маркеров стиля – слов, выражений, конструкций, характерных для научных текстов. Примеры критериев академического стиля: использование терминологии,

глаголы в пассивном залоге, существительные с абстрактными суффиксами. Всего экспертом было выделено более 15 признаков научного текста.

Статистика, полученная в результате разметки корпуса в соответствии с маркерами стиля, позволит сделать вывод о встречаемости тех или иных критериев «академичности» текста и их значимости при оценке стиля. Эти данные в дальнейшем позволят определять уровень качества научной публикации с точки зрения стиля и формировать рекомендации по его улучшению.

Для решения задач компьютерной лингвистики существуют программные средства универсального характера, такие как продукты семейств GATE и Ant, а также создаются специализированные программы, например, система Fapas (Full Automatic Paper Analysis System) для анализа стиля академических текстов, описанная в работе [2]. Для реализации метода, который является базовым для создаваемого портала, адаптируется универсальная система GATE. На данный момент в разработке находится исследовательский прототип приложения на базе среды GATEDeveloper и библиотек GATEEmbedded[3], предназначенных для работы с текстовыми корпусами. Прототип предоставит возможность начать исследование статей и корпусов, определить значимость различных критериев академического стиля и делать на основе этих данных предположение о качестве текста. После реализации прототипа планируется создать web-портал с расширенным функционалом.

Формат портала предназначен увеличить спектр пользователей, которые смогут пользоваться программой. GATEDeveloper – удобная среда для программистов и лингвистов, однако работа в ней требует определённых знаний о среде в целом и структуре приложения в частности. В web-портале технические особенности будут скрыты от обычного пользователя.

Предполагается несколько основных сценариев использования услуг сервиса:

1. Режим «Исследование»: пользователь загружает или выбирает корпус или документ, создаёт или редактирует необходимый набор маркеров академического стиля, запускает сервис и получает размеченный корпус или документ, а также отчёт о его статистических характеристиках. Этот режим предназначен для лингвистов, филологов, преподавателей академического письма.
2. Режим «Проверка»: пользователь загружает документ, выбирает наиболее подходящий по тематике корпус и получает отчёт по результатам сравнения характеристик документа с корпусом, оценку качества стиля и рекомендации по улучшению. Режим, предназначенный для авторов научных публикаций, студентов.

Таким образом, к базовым функциям портала относятся загрузка и хранение корпусов и отдельных документов, их разметка по маркерам стиля, создание аналитических отчётов,

оценка и вывод рекомендаций по отдельным работам. Кроме базового сервиса, составляющего ядро портала, планируется реализация дополнительных компонентов – хранилища данных и визуального редактора лексико-синтаксических шаблонов, которые используются для создания разметки по маркерам стиля. Социальная часть портала включает в себя возможность совместной работы, размещение обучающих материалов, организацию форума для пользователей портала. Примерная архитектура портала представлена на рис. 1.

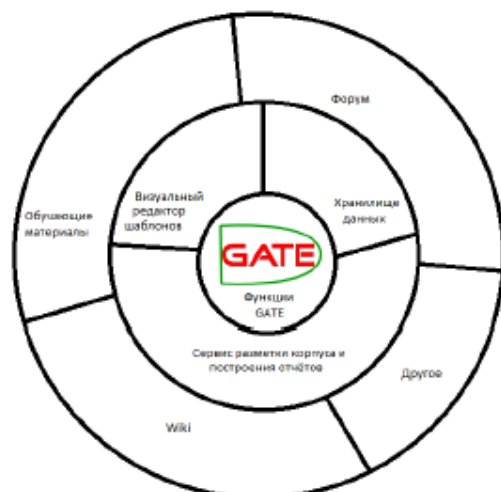


Рис. 12. Архитектура портала

Внедрение таких компонентов как хранилище данных и визуальный редактор шаблонов и расширение функционала базового сервиса относительно исследовательского прототипа необходимо для поддержания одного из главных свойств реализуемого метода анализа стиля – гибкости. Обеспечить гибкость метода можно в следующих аспектах:

- использование корпусов по различным тематикам, от разных издателей;
- изменение набора искомых маркеров стиля и изменение их базовых лингвистических ресурсов;
- добавление пользовательских типов разметки текста для выделения новых маркеров стиля;
- использование различных метрик для сравнения документов и корпусов;
- выделение различной статистики (относительно документа, относительно корпуса; в разрезе слов, в разрезе предложений; и т.д.);

Необходимо также перенести на портал одну из удобных возможностей среды GATE-ручное редактирование разметки. Таким образом, продвинутые пользователи смогут

загружать собственные корпуса текстов (и хранить их, например, в специальном облаке), исследовать их и получать индивидуальные настраиваемые отчёты.

Визуальный редактор лексико-синтаксических шаблонов [4] позволит изменять параметры разметки текста и добавлять новые маркеры стиля. Кроме того, в перспективе возможно реализовать нанесение пользовательской разметки с помощью кода на языке Java, например, в виде механизма плагинов, как в среде GATEDeveloper.

Стоит отметить, что концепция реализуемого в рамках портала метода достаточно универсальна, и в теории он может быть применён для других языков и стилей при наличии размеченных корпусов (в первом случае) и разработанных лексико-синтаксических шаблонов для выделения нужных маркеров стиля. В любом случае, успешная реализация даже части вышеописанных функций позволит представить пользователям сервис, благодаря которому авторы научных публикаций и студенты получат возможность проверить свои работы перед отправкой на рецензирование, а филологи, лингвисты и преподаватели английского языка смогут проводить исследования, делиться их результатами с коллегами, а также внедрять данный сервис в учебный процесс.

Библиографический список

1. Lanin V., Strinyuk S., Shuchalova Y. Academic Papers Evaluation Software // 2015 9th International Conference on Application of Information and Communication Technologies (AICT), Conference Proceedings. IEEE, 2015, Rostov-on-Don, Russia, October 14-16, 2015, –С. 506-510.
2. Scholz T., Conrad S. Style Analysis of Academic Writing // Natural Language Processing and Information Systems: 16th International Conference on Applications of Natural Language to Information Systems, Proceedings. NLDB 2011, Alicante, Spain, June 28-30, 2011. – С. 246-249.
3. GATE.ac.uk [Офф. сайт] // URL: <http://gate.ac.uk/> (датаобращения: 31.03.2016).
4. Ланин В.В., Шучалова Ю.С. Визуальный предметно-ориентированный язык описания лексико-семантических шаблонов // «Информатизация и связь» №1, 2016г. –С.44-48

RESEARCH PORTAL FOR ACADEMIC PAPER ANALYSIS AND EVALUATION

Shuchalova Y., Lanin V.

National Research University Higher School of Economics, st. Studencheskaya, 38, Perm, 614070,
shuchalova@gmail.com

The problem of mastering the academic style of English language is relevant to many non-English speaking authors. Recommendations on style improvement are often of a theoretical nature, but there is a possibility to research their implementations using text corpora. This paper presents a concept and functional requirements for a Web-portal, which provides services of academic style analysis and evaluation. The portal is suitable for various types of users and will give them the opportunity to perform research using different corpora and linguistic data as well as get a simple summary of the document's characteristics compared with other papers on relevant topics and recommendations on style improvement. To this day the research prototype of the portal is in development.

Keywords: English for Academic Purposes, style analysis, corpus linguistics, academic paper evaluation