

Сервисы поддержки научно-образовательной деятельности в системе Exactus Expert

Services of system Exactus Expert to support research and education

*Г. С. Осипов, И. В. Смирнов, И. А. Тихомиров, И. В. Соченков, А. В. Швец
Институт системного анализа РАН,
Москва, Россия*

*Gennady Osipov, Ivan Smirnov, Ilya Tikhomirov, Ilya Sochenkov and Alexander Shvets
Institute for Systems Analysis of the Russian Academy of Sciences,
Moscow, Russia*

В докладе представлены сервисы системы интеллектуального поиска и анализа научных публикаций Exactus Expert, предназначенные для поддержки научной и образовательной деятельности. Рассмотрены сервисы исследования предметной области и анализа качества научных публикаций.

Exactus Expert system for intelligent search and scientific publication analysis is presented. The services of domain search and analysis of research paper quality are described.

Система Exactus Expert [1] предназначена для информационно-аналитической поддержки научно-технической и образовательной деятельности. Система является инструментом, помогающим начинающим и опытным исследователям, студентам и аспирантам, в выполнении курсовых, дипломных, кандидатских и других научных работ. Система позволяет получить информацию о состоянии дел в заданной научной области, оценить перспективность темы научных исследований, оценить качество и оригинальности научных текстов. Система выполняет семантический анализ [2] научных публикаций из различных источников для решения поставленных задач.

Исследование предметной области

Задача изучения предметной области возникает на первых этапах научного исследования. Эта задача состоит в ознакомлении с публикациями по заданной теме, с авторскими коллективами, работающими в заданном научном направлении, в поиске научных тем, вызывающих последнее время повышенный интерес у исследователей.

В настоящее время система Exactus Expert предоставляет пользователю следующие возможности по исследованию предметной области:

- семантический поиск научных публикаций по запросу, выражающему научную тему или направление, с учётом метаданных;
- поиск содержательно близких документов;
- аннотирование научно-технических документов;
- предоставление ключевых слов, словосочетаний и терминов, характеризующих тематику отдельных документов и коллекций.
- анализ публикационной активности по заданной тематике во временном разрезе;
- выявление и анализ научных направлений;
- выявление и анализ научных коллективов.

Семантический поиск применяется для получения списка публикаций по интересующей пользователя тематике. Для каждой найденной публикации имеется возможность просмотреть ключевые слова, характеризующие содержание публикации, и аннотацию – сжатое изложение содержания документа. Ключевые слова и аннотация формируются автоматически и позволяют быстро ознакомиться с содержанием научной публикации. Для каждой найденной публикации можно найти похожие на неё другие публикации, проанализировать сходство и различие двух публикаций. Таким образом пользователь быстро получает представление о научных исследованиях, выполняемых по заданной теме. В процессе работы с публикациями, пользователь может отбирать заинтересовавшие его работы и формировать свои персональные коллекции публикаций.

Оценка перспективности научных направлений и тем основана на анализе публикационной активности, отражающей интенсивность исследований, ведущихся по интересующей пользователя тематике. Выявляются тенденции к росту или потере интереса к данной тематике. В графическом виде отображается интенсивность исследований по заданной теме с трендами, что позволяет пользователю сделать вывод о развитии или угасании научной темы.

Exactus Expert позволяет проанализировать коллективы, работающие в одном или нескольких научных направлениях. Для каждого коллектива выделяются лидеры, регулярные авторы, нерегулярные авторы, приглашенные авторы. Имеется возможность просмотреть динамику публикационной активности и других показателей научной деятельности коллектива. Функции работы с коллективами помогают сориентироваться в научном ландшафте анализируемой предметной области.

Анализ качества научных публикаций

Сервис анализа качества научных текстов – рефератов, курсовых и дипломных работ, диссертаций и авторефератов, научных статей, монографий и научно-методических материалов полезен как исследователям, так и преподавателям. Сервис позволяет автоматически оценить соответствие текста требованиям, предъявляемым к научным публикациям, и проанализировать содержание публикации.

Анализ качества научных публикаций состоит в следующем:

- определение соответствия научных публикаций формальным требованиям к структуре (наличие необходимых разделов, например, введение, постановка проблемы, список литературы и т.д.);
- определение содержания научной и ненаучной лексики;
- выявление наличия логических и семантических дефектов в тексте научной публикации;
- выделение авторских терминов – новых понятий, определяемых авторами публикаций;
- выделение результатов, представленных в публикациях;
- вычисление интегральной оценки качества научной публикации.

Структура текста считается полной, если в нем присутствуют «постановка проблемы», «методы», «результаты», «выводы», «список источников» [3].

Под логическими и семантическими дефектами [4] подразумеваются главным образом грамматические ошибки в структуре языковой единицы (слова, словосочетания или предложения), нарушения грамматической нормы – словообразовательного, морфологического или синтаксического оформления мысли в соответствии с требованиями сочетаемости, согласования и управления. К грамматическим ошибкам также относятся: ошибочное словообразование; ошибки в образовании форм слов; ошибки в согласовании и управлении; ошибки в употреблении причастных и деепричастных оборотов; ошибки в построении сложных предложений и др.

Выделение в тексте авторских терминов и результатов состоит в поиске формулировок, которыми обычно описываются полученные научные результаты или вводятся дефиниции. Наличие таких формулировок является ещё одним критерием качества научного текста [5].

Пример отчёта об анализе качества научного текста представлен на рисунке 1.

Уровень качества данного научно-технического текста равен 5 (0..5).

Индекс соответствия текста формальным требованиям: 5 (0..5).

- Текст содержит 51% общенаучной лексики. Рекомендуемо более 16%, допустимо от 7%.
- Текст содержит 1% ненаучной лексики. Рекомендуемо менее 2%, допустимо до 4%.
- Список литературы присутствует.
- Постановка проблемы присутствует. Коэффициент достоверности: 0.88 (0..1).
- Описание методов присутствует. Коэффициент достоверности: 0.85 (0..1).
- Выводы присутствуют. Коэффициент достоверности: 0.64 (0..1).

Степень логической и семантической правильности (отсутствие дефектов): 4 (0..5).

- Количество нарушений падежного согласования: 0 (0..5).
- Низкая степень нарушения синтаксической связности.
- Количество нарушений согласования однородных существительных и управляющего слова: 2.
- Содержание плеоназмов - умеренное.

Результаты:

1. Описанный в статье методологический подход к моделированию катастрофоустойчивости информационной системы был разработан для оценки больших территориально-распределенных информационных систем и применялся при выполнении работ по повышению катастрофоустойчивости Государственной автоматизированной системы Российской Федерации Выборы .

Термины:

1. Под *катастрофами* подразумеваются не только пожар, наводнение или землетрясение, но и возможные непредвиденные сбои в работе служб, разрушение данных или повреждение ЦОД, например, в случае разрыва телекоммуникационных линий, возникшего в результате проведения ремонтных работ, умышленной диверсии или саботажа или просто по халатности и ошибочных действий обслуживающего персонала.
2. Процесс создания такой модели будем называть *моделированием катастрофоустойчивости информационной системы*.

Рис. 1 – Результаты анализа качества научной публикации

При анализе качества научной публикации выполняется также поиск похожих публикаций.

Использование системы

Система Exactus Expert в настоящее время работает в демонстрационном режиме и доступна по адресу <http://expert.exactus.ru> после несложной регистрации. На данный момент информационная база системы включает публикации научных журналов, авторефераты, материалы конференций и патенты. В дальнейшем планируется пополнение информационной базы новыми документами, добавление сервисов оценки оригинальности научных документов и других сервисов для поддержки научно-образовательной деятельности.

Литература

1. Gennady Osipov, Ivan Smirnov, Ilya Tikhomirov, Olga Vybornova. Technologies for Semantic Analysis of Scientific Publications. //Proceedings of 2012 IEEE 6th International Conference Intelligent Systems (Yager, R.R., V. Sgurev and M. Hadjiski, Eds.), Vol. II, pp. 58–62.
2. Осипов Г.С., Смирнов И.В., Тихомиров И.А. Реляционно-ситуационный метод поиска и анализа текстов и его приложения //Журнал «Искусственный интеллект и принятие решений». Номер 2-2008. – С. 3–10.
3. Швец А.В., Кузнецова Ю.М., Осипов Г.С., Латышев А.В. Метод и алгоритм обнаружения признаков лингвистических дефектов в научно-технических текстах // Информационные технологии и вычислительные системы. – 2013. – № 2. – С. 79–87.
4. Кузнецова Ю.М., Осипов Г.С., Чудова Н.В., Швец А.В. Автоматическое установление соответствия статей требованиям к научным публикациям // Труды ИСА РАН. – 2012. – Т. 62. – Вып. 3. – С. 132–138.
5. Тихомиров И.А., Смирнов И.В., Соченков И.В., Девяткин Д.А., Шелманов А.О., Зубарев Д.В., Швец А.В., Лешкин А.В., Суворов Р.Е. Exactus Expert: Поисково-аналитическая система поддержки научно-технической деятельности // Труды тринадцатой национальной конференции по искусственному интеллекту с международным участием КИИ-2012. Б.: БГТУ, 2012. т. 4. – С. 100–108.