

Анализ качественных и количественных характеристик журнала «Искусственный интеллект и принятие решений»¹

Аннотация. В статье представлены результаты исследования публикаций журнала «Искусственный интеллект и принятие решений». Приводятся различные количественные и качественные характеристики журнала, получаемые с помощью инструментов eLibrary и с использованием современных средств интеллектуального анализа текстов Exactus Expert. С помощью методов качественного анализа научных публикаций установлено, к каким научным направлениям относятся статьи журнала, какая часть статей принадлежит лидерам научных коллективов, кем финансируются исследования, результаты которых представлены в статьях журнала. Выполнена оценка качества текстов, осуществлен поиск заимствований в текстах публикаций. В заключении приводится общая оценка журнала на основе выявленных характеристик.

Ключевые слова: анализ научных журналов, выявление направлений научных исследований, полнотекстовая кластеризация научных публикаций, Exactus Expert.

Введение

В современных условиях анализ возросшего количества научных публикаций в российских и зарубежных журналах немыслим без использования наукометрических инструментов. Однако эти инструменты ориентированы в основном на количественную, а не качественную оценку содержания конкретных статей или журналов. В последнее время широко обсуждается недостаточность количественных оценок. Так, в [1] отмечается: «Количественная оценка должна дополнять качественную, экспертную оценку ... Специалисты, проводящие оценку научной деятельности, не должны перекладывать принятие решений на числа. Индикаторы – не замена информированному суждению. Каждый сохраняет ответственность за свою оценку». Российские ученые [2] также указывают на значимость качественных показателей: «Наиболее объективными являются рейтинги,

выстроенные на основании синтеза количественных и качественных показателей и характеристик эффективности научно-исследовательской деятельности ученого, уровня его разработок и значимости их для научного мира и широкой общественности».

В настоящем исследовании проводится анализ публикаций журнала «Искусственный интеллект и принятие решений» (ИИПР) с целью рассмотрения качественных показателей, полученных с применением методов интеллектуального анализа полуструктурированных данных. В том числе анализа полных текстов, и той дополнительной информации, которую они дают по сравнению со стандартными инструментами количественной оценки научных журналов. Для получения количественных оценок используется основное средство анализа российских журналов – электронная библиотека eLibrary [3]. Качественные показатели формируются посредством методов и алгоритмов, реализованных в системе интеллектуального по-

¹ Работа выполнена при финансовой поддержке РФФИ (гранты №14-29-05008-офи_м, №14-29-05075-офим_м).

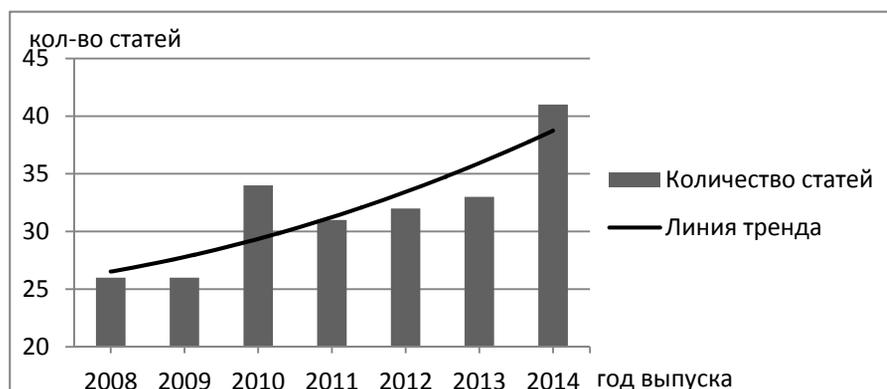


Рис. 1. Распределение статей в журнале по годам

иска и анализа научных публикаций Exactus Expert [4]. Они позволяют определить научные направления, представленные в журнале, долю статей, принадлежащих лидерам научных коллективов, выявить текстовые заимствования и определить оригинальность текстов, оценить их качество и т.д. Результаты исследования могут быть полезны не только редколлегии самого журнала и авторам, которые публикуются или хотят опубликоваться в данном журнале, но и всем, кто интересуется инструментами оценки научных журналов.

Статья оформлена следующим образом: в первом разделе приводится статистика и сведения о журнале, полученные стандартными методами. На основе этой статистики сделаны отдельные выводы. Второй раздел посвящен методам интеллектуального анализа полуструктурированных данных, представлены результаты, которые могут быть получены только с применением автоматического анализа полных текстов статей. В заключении приведены основные выводы и направления дальнейших исследований.

1. Количественные показатели журнала

Журнал «Искусственный интеллект и принятие решений» выходит 6 раз в год, в том числе 4 номера на русском языке и 2 на иностранном. Один номер журнала содержит от 8 до 12 статей, в среднем в русскоязычных номерах в год выходит 32 статьи, при этом в 2014 году число публикаций увеличилось до 41, что говорит об увеличении интереса авторов к журналу как к изданию для опубликования (Рис. 1).

Количественные показатели журнала могут быть получены с использованием электронной

библиотеки научных публикаций eLibrary. Библиотека интегрирована с Российским индексом научного цитирования (РИНЦ), который позволяет получить количественную оценку уровня журналов. Эта оценка формируется на основе анализа структурированных данных, добавляемых в систему вручную и включающих годы опубликования статей, авторство, списки цитируемых источников, ключевые слова. В результате вычисляются различные библиометрические показатели журнала. К таким показателям относятся как классический импакт-фактор, который широко используется во всем мире для оценки уровня научных журналов, так и другие показатели, учитывающие ряд дополнительных факторов, влияющих на величину импакт-фактора, и позволяющие скорректировать это влияние. В частности, учитывается тематическое направление исследований, объем, состав и хронологическое распределение журналов в базе данных, самоцитирование и цитирование соавторами, возраст публикации, число соавторов, авторитетность ссылок (кто процитировал).

Стоит отметить, что характеристики журнала, приведенные в eLibrary, в целом соответствуют действительности, но не являются точными. Это связано с ошибками, допускаемыми при ручном вводе данных. Например, согласно данным eLibrary в журнале опубликовались 258 авторов, но при этом фамилии 39 авторов в системе eLibrary либо указаны с опечатками, либо добавлены случайно, либо наоборот продублированы, т.е. имеется информация лишь о 219 авторах из 282. Это говорит о низкой эффективности добавления имен авторов вручную: теряется до 22% информации, кроме того большой процент данных оказывается зашумлен.

Табл. 1. Количественные характеристики журнала ИИПР, представленные в eLibrary²

Название показателя		2008	2009	2010	2011	2012	2013	2014
Число статей, опубликованных за предыдущие два года		-	26	52	60	65	63	65
Двухлетний импакт-фактор РИНЦ с учетом цитирования из всех источников		-	0,92	0,92	0,75	0,72	1,02	0,68
Пятилетний импакт-фактор РИНЦ		-	-	0,885	0,953	0,735	0,906	0,577
Число цитирований статей предыдущих двух лет	Из журнала ИИПР	-	1	7	6	6	3	7
	Всего	-	24	48	45	47	64	44
Число статей, опубликованных за предыдущие 5 лет		-	-	52	86	117	149	156
Число цитирований статей предыдущих 5 лет	Из журнала ИИПР	0	1	7	9	13	12	18
	Всего	0	21	46	82	86	135	90
Пятилетний индекс Херфиндаля по цитирующим журналам		-	884	643	1071	608	290	570
Индекс Херфиндаля по организациям авторов		903	1224	841	593	759	1159	1212
Десятилетний индекс Хирша		-	-	4	4	5	7	8
Время полужизни статей из журнала, процитированных в текущем году		-	-	2,2	2,8	2,9	3,6	4,1
Время полужизни статей, процитированных в журнале в текущем году		-	-	3,9	9,4	7,5	8,4	7,0

² Показатели журнала за 2014 год являются предварительными, поскольку в базу данных РИНЦ еще не загружен полный массив данных за 2014 год

Приведем в Табл. 1 некоторые показатели, рассчитанные средствами электронной библиотеки eLibrary для журнала «Искусственный интеллект и принятие решений».

Приведем определения некоторых показателей, указанных в Табл. 1, как они даются в системе eLibrary, и проанализируем, о чем могут говорить значения этих показателей по отношению к журналу ИИПР. Методики для корректного подсчета значений указанных показателей рассмотрены в работе [5].

Импакт-фактор журнала вычисляется как число цитирований, которые были получены в течение текущего года статьями, опубликованными в журнале за несколько предыдущих лет, поделенное на число всех статей данного журнала за этот период. Видно, что в 2011 и 2012 годах двухлетний импакт-фактор снизился по сравнению с 2009 и 2010 годами, но к 2013 году снова вырос: в среднем на каждую статью приходится одно цитирование. Значение показателя в 2014 году является заниженным в связи с неполнотой данных РИНЦ за этот год. По пятилетнему импакт-фактору журнал занимает 66 место в рейтинге российских журналов. Стоит отметить, что по специальному комплексному

показателю, рассчитываемому в eLibrary с использованием многих показателей из Табл. 1, а также с учетом различий в практике цитирования для различных научных направлений, журнал занимает 29 место.

Индекс Херфиндаля-Хиршмана по цитирующим журналам рассчитывается как сумма квадратов процентных долей журналов, цитирующих данный, по отношению к общему количеству цитирований. При расчете учитываются ссылки из текущего года на предыдущие 5 лет, в том числе самоцитирования (в eLibrary этот показатель рассчитывается только на основании цитирований из журналов, включенных в РИНЦ). Чем больше количество цитирующих журналов и чем равномернее распределены по ним ссылки на данный журнал, тем меньше этот показатель. Максимальное значение равно 10000 и достигается, когда все ссылки сделаны из одного журнала. Аналогично, индекс Херфиндаля-Хиршмана по организациям авторов рассчитывается как сумма квадратов процентных долей количества статей, опубликованных различными организациями, по отношению к общему количеству статей в журнале в текущем году, в которых организа-

ция идентифицирована. Согласно Табл. 1, значение первого индекса Херфиндаля от года к году снижается, минимальное значение достигалось в 2013 году. Это означает, что все больше сторонних журналов цитируют журнал ИИПР. Одной из причин увеличения разнообразия цитирующих журналов может являться расширение читающей аудитории, однако сделать однозначные выводы по одному индексу не представляется возможным. Значение второго индекса, наоборот, в последние годы увеличивается, что говорит либо об уменьшении числа организаций, которые представляют авторы журнала, либо о перевесе принадлежности публикаций в сторону нескольких отдельных организаций.

Индекс Хирша вычисляется на основе распределения цитирований статей и имеет значение N , если среди наиболее цитируемых статей журнала есть N статей, на каждую из которых сослались как минимум N раз, а остальные статьи имеют число цитирований не более N . Учитываются все статьи, опубликованные в журнале за 10 лет, и цитирования за этот же период. К 2014 году индекс Хирша журнала ИИПР вырос до 8, рост числа ссылок на наиболее цитируемые статьи может говорить о том, что опубликованные однажды результаты исследований затем долгое время остаются актуальными.

Время полужизни статей из журнала, процитированных в текущем году – медианный возраст этих статей, такой, что половина ссылок на журнал, сделанных в текущем году, идет на статьи моложе этого возраста, другая половина – на статьи старше. Аналогично, время полужизни процитированных в журнале статей в текущем году – медианный возраст этих статей, такой, что половина ссылок из журнала, сделанных в этом году, идет на статьи моложе этого возраста, другая половина – старше. Согласно Табл. 1, сторонние издания от года к году цитируют статьи журнала ИИПР 3-4 летней давности, т.е. больше половины ссылок в каждом году приходится на работы, опубликованные в первых номерах журнала. Значения в последней строке Табл. 1 показывают, что авторы рассматриваемого журнала цитируют сторонние издания 7-9 летней давности. Это может говорить о том, что в основу работ кладутся не только современные методы и последние достижения науки, но и более ранние, возможно, фундаментальные работы.



Рис. 2. Распределение публикаций журнала ИИПР по тематике в eLibrary

Кроме различных количественных показателей для каждого журнала в электронной библиотеке eLibrary построены различные распределения:

- распределение публикаций по тематике (по заданным вручную тематическим рубрикам);
- распределение публикаций по ключевым словам (указанным авторами этих публикаций);
- распределение публикаций по организациям;
- распределение публикаций по числу соавторов;
- хронологическое распределение статей из журнала, процитированных в 2013 году;
- распределение цитирований по цитирующим журналам;
- распределение цитирований по типу цитирующих публикаций.

Рассмотрим два первых распределения. Распределение по тематикам позволяет определить общее направление журнала – относится он к математике, физике или, например, к гуманитарным наукам. Распределение строится на основе тематических рубрик, указанных при добавлении статей в базу данных РИНЦ.

На Рис. 2 видно, что основная часть работ в журнале ИИПР относится к рубрикам «Кибернетика», «Организация и управление», «Математика» и «Автоматика. Вычислительная техника». Разделение статей на такие крупные рубрики является малоинформативным для исследователей – оно не дает возможности определить, чему они посвящены, какие научные направления представлены в журнале.



Рис. 3. Распределение публикаций журнала ИИПР по ключевым словам в eLibrary

Распределение ключевых слов, выделенных авторами статей, также не всегда позволяет идентифицировать направления исследований (Рис. 3). Так, из первых 11 ключевых слов, лишь эксперт в предметной области сможет выделить два ключевых слова, характеризующих научные направления: «ДСМ-метод» и «интегрированные экспертные системы». Остальные ключевые слова без контекста можно отнести ко многим областям исследований. Например, статьи с термином «нейронные сети» могут относиться к совершенно разным темам.

Из сказанного выше следует, что электронная библиотека eLibrary является полезным ресурсом для получения некоторой общей статистики по журналу. Однако ни количественные показатели, которые могут быть интерпретированы по-разному, ни распределения публикаций, которые сами по себе являются малоинформативными, не позволяют получить качественные показатели, определяемые только путем обращения к полным текстам статей.

Рассмотрим возможности применения современных методов качественного анализа научных публикаций, реализованных в системе Exactus Expert.

2. Методы качественного анализа научных публикаций

В связи с ограниченными возможностями методов количественной оценки, для определения содержательных и интерпретируемых качественных характеристик журнала предлагается использовать современные методы анализа полуструктурированных данных, реализованных в системе Exactus Expert [6]: методы кластеризации

текстов для выявления научных направлений и анализа графов для определения научных коллективов, синтактико-семантический анализ текстов для формирования ключевых слов научных направлений и построения поискового индекса, методы анализа качества текстов научных статей. Показано, что различные методы анализа полных текстов статей могут быть эффективно использованы для поддержки научно-технической деятельности [7].

В качестве данных используются полные тексты 224 статей журнала ИИПР, выпущенных в 2008-2014 годы включительно. Статьи находятся в открытом доступе на сайте журнала.

2.1. Тематическая кластеризация статей журнала

Кластеризация статей выполняется согласно методу, предложенному авторами настоящей работы [8]. Особенностью метода является учет лингвистической информации, содержащейся в текстах статей, при расчете меры близости документов. Отличием от других методов, позволяющих выделять научные направления по текстам [9, 10], является использование инвертированного поискового индекса. Он уменьшает время обнаружения схожих по содержанию текстов, выполняет кластеризацию параллельно на разных узлах распределенной вычислительной системы, а также позволяет работать без предварительного задания количества кластеров.

Кластеризация реализуется в два этапа. Первый заключается в выделении групп тематически близких статей на разных узлах и построении их дескрипторов с использованием синтактико-семантического анализа. Дескрипторы представляют собой наборы слов и словосочетаний, содержат термины предметных областей и, в отличие от наборов отдельных слов, позволяют получить интерпретируемое описание статей, объединенных в один кластер. Второй этап состоит в классификации статей по этим дескрипторам.

В рамках исследования выполнена кластеризация коллекции для выявления терминологии различных направлений, представленных в журнале ИИПР. Получено 14 кластеров, в каждый из которых входят тексты близкие по теме. Для каждого кластера автоматически с помощью методов обработки естественного языка выделена характерная терминология. В Табл. 2

представлены характеристики наиболее крупных кластеров. Большому размеру шрифта во втором столбце соответствуют более значимые словосочетания. Число статей, образующих кластер, указано в последнем столбце.

Перечислим названия выявленных кластеров, которые даны авторами настоящей статьи в результате анализа содержания полных текстов из этих кластеров:

- многокритериальная оптимизация;
- семантический поиск;
- ДСМ-метод автоматического порождения гипотез;

- многокритериальная порядковая классификация;
- интегрированные экспертные системы;
- кластерный анализ и распознавание текстовой и графической информации;
- автоматизация полета вертолета.

Видно, что извлеченная терминология кластеров, указанная в Табл. 2, соответствует названиям, данным вручную. Это означает, что применение метода кластеризации позволяет определить основные темы исследований и получить представление о том, чему посвящены работы авторов журнала, еще до анализа полных текстов.

Табл. 2. Результаты кластеризации коллекции статей журнала ИИПР в Exactus Expert

№	Терминология кластера	Авторы	N
1	"аксиоматический подход" "аппроксимация оболочки" "бинарные отношения" "важность критериев" "векторные критерии" "выбираемые варианты" "выбираемый вектор" "готовность лпр" "граница парето" "диалоговые карты" "задача выбора" "иррефлексивный" "искусственный интеллект" "лпр" "многокритериальная оптимизация" "многокритериальная среда" "многокритериальные задачи" "многокритериальный выбор" "мно-гомерная граница" "множество векторов" "множество парето" "недоминировать" "неизвестное множество" "отношение предпочтения" "парето-оптимальные векторы" "парето-оптимальный" "предпочтение лпр" "принцип эджворта-парето" "проблемы сужения" "сужение множества" "транзитивный"	В.Д. Ногин В.В. Подиновский А.В. Лотов А.И. Рябиков А.Л. Бубер А.В. Холмов О.Н. Климова Н.Б. Брусникина О.В. Басков А.О. Захаров А.П. Нелюбин	14
2	"именная группа" "интеллектуальный поиск" "информационный поиск" "коллекция документов" "компьютерная лингвистика" "морфологический анализ" "поисковые запросы" "семантический поиск" "синтаксический анализ" "слова запроса"	В.Ф. Хорошевский Г.С. Осипов И.В. Смирнов А.О. Шелманов Е.С. Кузнецова И.В. Храмоин И.А. Тихомиров Р.Е. Суворов И.В. Соченков	8
3	"абдуктивный" "автоматическое порождение" "аксиома полноты" "база фактов" "дсм-метод" "дсм-метод порождения" "дсм-рассуждение" "изучаемые эффекты" "индуктивный метод" "истинностное значение" "каузальная полнота" "конструирование систем" "логические средства" "методы милля" "множество гипотез" "операция сходства" "подформула" "познавательные процедуры" "порождение гипотез" "порожденные гипотезы" "правдоподобный вывод" "правила вывода" "применение правил" "синтез процедур" "система интеллекта" "степень противоречивости" "условия применимости" "фактическое противоречие" "эмпирическая зависимость"	В.К. Финн О.М. Аншаков М.А. Михеенкова А.В. Жожикашвили В.Л. Стефанюк	8

4	"вербальный" "вербальный анализ" "вычислительный" "выявление правил" "гипотетический" "диагностические решения" "задача классификации" "знания экспертов" "искусственный интеллект" "классифицировать" "ларичев" "метод классификации" "многокритериальная классификация" "многокритериальный" "мошкович" "непротиворечивый" "номинальная классификация" "отношение доминирования" "порядковая классификация" "построение систем" "решаемые правила" "шкала критериев" "экспертная классификация" "экспертные знания"	А.Б. Петровский Г.В. Ройзензон Е.М. Фуремс В.Н. Лобанов Д.Ю. Кочин	7
5	"ат-технология" "базы знаний" "диалоговый" "задачно-ориентированная методология" "задачно-ориентированный" "имитационный" "инструментальная база" "инструментальные средства" "инструментальный комплекс" "информационный-измерительный" "искусственный интеллект" "иэс" "кибернетика" "компоненты комплекса" "масштабировать" "многоагентный" "модель представления" "новый компонент" "поддержка построения" "поддержка разработки" "прикладные иэс" "программный инструментальный" "производственный" "процессы функционирования" "средства поддержки" "технология построения" "формализм" "экспертные системы"	Г.В. Рыбина М.Г. Иващенко Ю.М. Блохин А.О. Дейнеко А.В. Мозгачев	6
6	"аппараты сети" "вектор" "выбор метрики" "евклид-махаланобис" "задача кластеризации" "задача распознавания" "значения признака" "инс" "искусственный интеллект" "кластеризация" "ковариация" "матрица ковариации" "машинная графика" "набор" "первоначальное размещение" "признаковое пространство" "пространство признаков" "размещение кластеров" "сеть кохонена" "число кластеров" "экспертный"	В.М. Хачумов М.В. Хачумов А.А. Талалаев В.П. Фраленко И.Л. Толмачев И.П. Тищенко	6
7	"автоматизация" "вертолет" "искусственный интеллект" "летательный" "летательный аппарат" "полет"	С.В. Емельянов Л.Н. Никифорова В.М. Хачумов К.С. Яковлев Д.А. Макаров	5

2.2. Научные коллективы, публикующиеся в журнале

Определение коллективов выполняется с помощью метода, разработанного авторами настоящей работы [11]. Метод заключается в обходе взвешенного графа авторов (весом ребра является количество совместных публикаций авторов, соответствующих примыкающим вершинам) и последовательном выделении связанных компонент графа, каждая из которых образует основу или ядро отдельного коллектива. При этом учитываются только ребра, имеющие вес выше заданного порога, который

уменьшается каждый раз после извлечения всех ядер коллективов при данном пороге. На каждом шаге извлеченные коллективы пополняются авторами, имеющими меньшее число публикаций с авторами ядра. Они образуют множество нерегулярных авторов коллектива. Автоматически определяется также лидер как автор, входящий в ядро коллектива, на которого ссылаются чаще, чем на других. Так, лидером коллектива считается автор, имеющий больший вес в научном сообществе, а не формальный руководитель по отдельным проектам, выполняемым коллективом.

На основе коллекции журналов из Перечня ВАК и патентов на полезные модели, изобре-

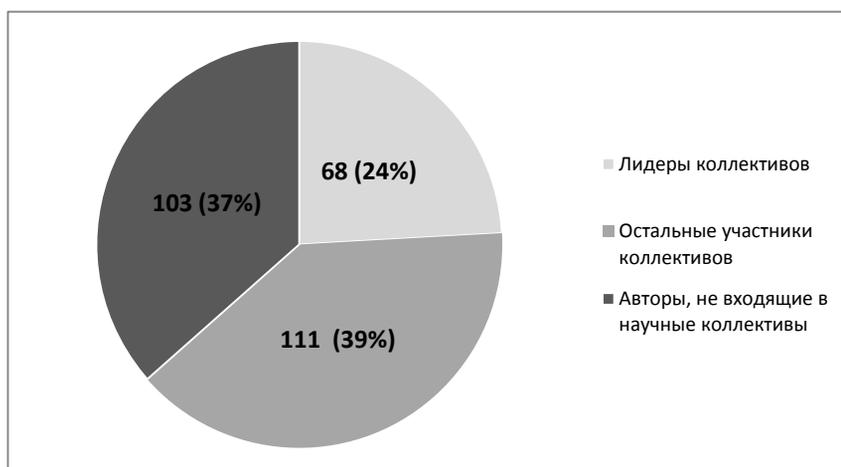


Рис. 4. Соотношение авторов журнала по их роли в коллективе



Рис. 5. Соотношение статей в журнале по типу авторства

ния и промышленные образцы выявлено более 50 тыс. российских научных коллективов. Данные о коллективах доступны в системе Eхastus Expert.

Путем сопоставления авторов статей журнала ИИПР с выявленными коллективами, установлено, что 75 статей принадлежат научным коллективам. Статья рассматривается как написанная коллективом, если соавторами являются хотя бы два автора из одного коллектива. На Рис. 4 приведено соотношение авторов журнала по их роли в коллективе, на Рис. 5 - соотношение статей по типу авторства.

Приведенные данные в определенной степени свидетельствуют о высоком качестве опубликованных в журнале работ. Согласно Рис. 5, почти половина статей (47%) написана лидерами коллективов или с их участием и более 60% авторов (сумма значений двух правых секторов на Рис. 4) являются участниками различных

научных коллективов, т.е. уже имеют опыт совместного проведения исследований и написания научных публикаций. Однако это не означает, что статьи, написанные без участия лидеров и иных членов научных коллективов, имеют низкое качество – для них всего лишь отсутствует один из признаков, который в определенной мере характеризует высокое качество статьи.

2.3. Поддержка исследований грантами

Благодаря поисковому индексу Eхastus Expert [12, 13], построенному по полным текстам, становится возможным построение дополнительных распределений при отсутствии структурированных данных. Так, с применением поиска по текстам журналов с запросами, соответствующими названиям различных программ и грантов, получено распределение статей, выполненных с финансовой поддержкой

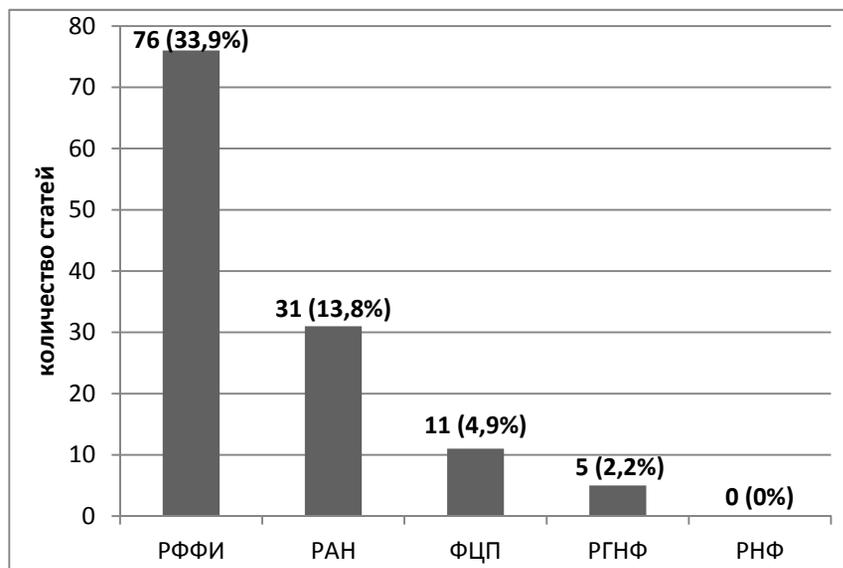


Рис. 6. Количество статей, выполненных при поддержке программ и грантов

(Рис. 6). В скобках указана доля от общего количества всех статей журнала за все годы.

Согласно Рис. 6, многие статьи журнала описывают результаты работы, проведенной при поддержке различных программ и грантов, большая часть из которых приходится на гранты РФФИ и программы РАН. Можно сказать, что публикуемые в журнале работы, поддержанные грантами, чаще посвящены фундаментальным исследованиям, чем прикладным.

2.4. Определение качества текстов статей журнала

В системе Exactus Expert реализованы методы, позволяющие оценивать качество текстов научных статей. Общая оценка выставляется в интервале от 1 до 5 и зависит от различных нарушений, допущенных в тексте. Во-первых, проверяется соответствие словосочетаний общенаучному словарю. Текст с низким употреблением общенаучных словосочетаний может ввести в заблуждение в силу неоднозначности толкования ненаучных слов и словосочетаний, имеющих различную интерпретацию, и поэтому получает оценку ниже. Во-вторых, проверяется соблюдение структуры научного текста. Научный текст, как правило, имеет структуру, соответствующую формату IMRAD [14], т.е. включает фрагменты текста, содержание которых соответствует разделам «Введение и постановка проблемы», «Методы», «Результаты» и «Выводы и обсуждение» (Introduction,

Methods, Results and Discussion). Наконец, выявляются нарушения правил согласования, синтаксической и семантической связности текста, последовательности изложения и лексической избыточности [15]. Также выявляются текстовые заимствования из ранее опубликованных работ [16] с целью определения оригинальности текста. При этом Exactus Expert позволяет выявлять как дословные заимствования, так и частично переписанные фрагменты текста, обладающие схожей синтаксической структурой.

Экспериментальное исследование качества текстов журнала показало, что почти все статьи журнала удовлетворяют требованиям к структуре научного текста, имеют достаточное количество общенаучных словосочетаний и не содержат значимых нарушений. Лишь для восьми статей автоматически не удалось выявить фрагменты, соответствующие разделу «Выводы», что связано с размытыми формулировками и отсутствием обсуждения проведенного исследования. В связи с этим 216 статей журнала получило в системе Exactus Expert оценку «5», свидетельствующую о высоком качестве текстов и 8 статей – оценку «4» из-за отсутствия четко сформулированных выводов в работе (Рис. 7).

Дополнительно средствами Exactus Expert проведен анализ списков использованных источников и соответствующих им ссылок внутри текстов. В результате анализа выявлены следующие недостатки нескольких статей журнала:

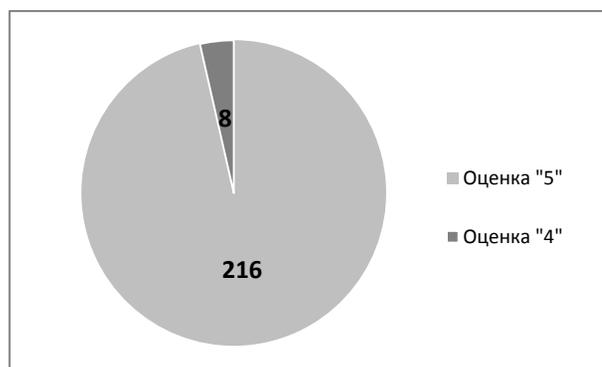


Рис. 7. Оценка качества текстов журнала ИИПР

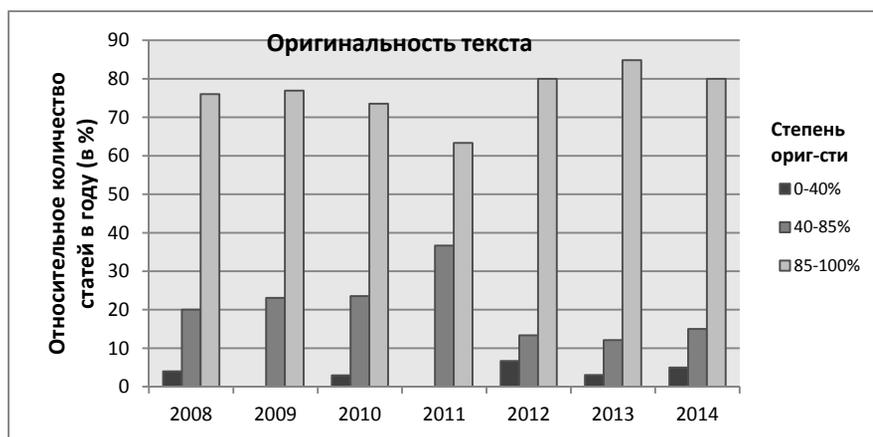


Рис. 8. Процент оригинальности текстов журнала ИИПР

1) в четырех работах все или большинство использованных источников являются неавторитетными (Интернет-сайты);

2) имеется 3 статьи, в которых в списке литературы указано больше источников, чем ссылок, расставленных автором в тексте, что может говорить о невнимательности автора или о намеренном расширении списка литературы дополнительными источниками для придания большей значимости работе;

3) журнал содержит 5 статей, в которых в одной из ссылок, указанных в тексте в квадратных скобках, приведено одновременно более 10 источников. Наибольшее число источников в одной ссылке равно 16. В некоторых случаях это может говорить о неоправданном расширении списка использованных источников;

4) имеется 4 статьи, в которых ссылки ведут преимущественно лишь на одного автора (3 статьи со ссылками автора на себя и 1 статья с большинством ссылок на другого автора). Такое распределение ссылок свидетельствует о низком

знакомстве автора с работами других исследователей по этой теме.

Тексты журнала проанализированы на предмет заимствований. Результаты представлены на Рис. 8. Видно, что в течение пяти лет распределение заимствованных фрагментов в текстах журнала почти не изменялось, при этом большинство текстов имеют высокую степень оригинальности.

Авторами настоящей статьи вручную проведен анализ статей, содержащих заимствованные фрагменты. Он показал, что авторы этих статей, как правило, используют фрагменты из своих предыдущих работ, во многих случаях ссылаясь на эти работы.

Заключение

В настоящей работе выполнен анализ публикаций журнала «Искусственный интеллект и принятие решений» как с использованием стандартных статистических методов, так и с применением современных средств, позволяю-

щих получать качественные характеристики, опирающиеся на полные тексты статей.

Количественные показатели, полученные с помощью системы eLibrary, позволяют сделать вывод о том, что журнал становится все более продуктивным в области искусственного интеллекта и принятия решений. Количество статей в номере от года к году растет, увеличивается число авторов и среднее количество ссылок на статьи, расширяется аудитория, использующая результаты опубликованных исследовательских работ. При этом результаты долгое время остаются актуальными.

Качественные характеристики, полученные с помощью системы Exactus Expert, позволяют проанализировать содержательную часть журнала:

– авторы публикуют работы по обширному кругу тем, при этом как минимум по семи темам регулярно появляются новые статьи, в результате чего образовались крупные кластеры текстов, соответствующие следующим научным направлениям: «многокритериальная оптимизация», «семантический поиск», «ДСМ-метод автоматического порождения гипотез», «многокритериальная порядковая классификация», «интегрированные экспертные системы», «кластерный анализ и распознавание текстовой и графической информации», «автоматизация полета вертолета»;

– результаты работ публикуют как отдельные исследователи, так и целые научные коллективы, при этом активное участие принимают лидеры коллективов, имеющие большой вес в научном сообществе по сравнению с другими членами, который, в том числе, выражается наиболее высокими показателями цитируемости;

– многие работы поддерживаются различными фондами и министерствами, большая часть финансирования приходится на фундаментальные исследования;

– авторы статей публикуют тексты высокого качества: работы написаны с использованием большого числа общенаучных словосочетаний, которые обеспечивают ясность текстов, тексты хорошо структурированы и содержат все разделы, необходимые для возможности дальнейшего анализа и использования результатов работы. Лишь в нескольких публикациях не представлены в явном виде выводы, несколько статей имеют ссылки, почти все из которых являются самоцитированиями, и две публикации содержат большое число заимство-

ванных фрагментов. Незначительное количество текстов невысокого качества не снижает общей положительной оценки журнала, однако использование методов автоматического анализа текстов позволило бы редакции обратить внимание авторов на недостатки их работ и полностью избежать досадного попадания в журнал текстов с заимствованиями, которые трудно определить вручную.

Таким образом, в настоящей статье показано, что методы интеллектуального анализа текстов позволяют получить важную дополнительную информацию, которую не отражают стандартные статистические методы, и тем самым помогают преодолеть ограниченность количественной библиометрии. В дальнейшей работе планируется расширить область исследований и перейти от анализа отдельных журналов к исследованию масштабных коллекций научных публикаций, предоставляемых крупными научными электронными библиотеками, в том числе научной электронной библиотекой открытого доступа «КиберЛенинка» [17] и научной электронной библиотекой eLibrary [3].

Литература

1. Hicks D., Wouters P., Waltman L., de Rijcke S., Rafols I. Bibliometrics: The Leiden Manifesto for research metrics // Nature. – 2015. – Vol. 520. P. 429-431.
2. Ефимова Г.З. Анализ эффективности наукометрических показателей при оценке научной деятельности // Вестник Тюменского государственного университета. – 2012. – Вып. 8. С. 101-108.
3. eLIBRARY.RU – научная электронная библиотека [Электронный ресурс]: URL: <http://elibrary.ru/> (дата обращения: 01.09.2015).
4. Exactus Expert – система интеллектуального поиска и анализа научных публикаций [Электронный ресурс]: URL: <http://expert.exactus.ru/> (дата обращения: 01.09.2015).
5. Писляков В.В. Методы оценки научного знания по показателям цитирования // Социологический журнал, 2007, № 1. С. 128-140.
6. 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: Поиск-аналитическая система поддержки научно-технической деятельности // Труды тринадцатой национальной конференции по искусственному интеллекту с международным участием КИИ-2012. Белгород: Изд-во БГТУ, 2012. Т. 4. С. 100-108.
7. Osipov Gennady, Smirnov Ivan, Tikhomirov Ilya, Sochenkov Ilya, Shelmanov Artem, Shvets Alexander. Information Retrieval for R&D Support // Professional Search in the Modern World. Lecture Notes in Computer Science, Vol. 8830, 2014. P. 45-69.

8. Shvets A., Devyatkin D., Sochenkov I., Tikhomirov I., Popov K., Yarygin K. Detection of Current Research Directions Based on Full-Text Clustering // Proceedings of Science and Information Conference 2015. London, 2015. P. 483-488.
9. Lee W. H. How to identify emerging research fields using scientometrics: An example in the field of information security // Scientometrics. 2008. Vol. 76(3). P. 503-525.
10. Glanzel W. Bibliometric methods for detecting and analysing emerging research topics // El profesional de la informacion. – 2012. – Vol. 21(2). P. 194-201.
11. Девяткин Д.А., Швец А.В. Экспериментальный метод автоматического выделения тем исследований и научных коллективов // Труды тринадцатой национальной конференции по искусственному интеллекту с международным участием КИИ-2012. Белгород: Изд-во БГТУ, 2012. Т. 2. С. 90-99.
12. Соченков И.В., Суворов Р.Е. Сервисы полнотекстового поиска в информационно-аналитической системе (Часть 1) // Информационные технологии и вычислительные системы. 2013. №2. С. 69-78.
13. Соченков И.В., Суворов Р.Е. Сервисы полнотекстового поиска в информационно-аналитической системе (Часть 2) // Информационные технологии и вычислительные системы. 2013. №3. С. 71-87.
14. Sollaci L.B., Pereira M.G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey // J. Med. Libr. Assoc. 2004. V. 92. № 3. P. 364-371.
15. 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.
16. Zubarev D., Sochenkov I.: Using Sentence Similarity Measure for Plagiarism Source Retrieval — Notebook for PAN at CLEF 2014. In: CEUR Workshop Proceedings, CEUR-WS.org, Eds. L. Cappellato, N. Ferro, M. Halvey and W. Kraaij. 2014. P.p. 1027-1034, / [Электронный ресурс] URL: <http://ceur-ws.org/Vol-1180/CLEF2014wn-Pan-ZubarevEt2014.pdf>, (дата обращения 01.09.2015)
17. КиберЛенинка – научная электронная библиотека открытого доступа [Электронный ресурс]: URL: <http://cyberleninka.ru/> (дата обращения: 01.09.2015).

Швец Александр Валерьевич. Младший научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН. Окончил Сибирский федеральный университет в 2011 году. Автор 25 печатных работ. Область научных интересов: компьютерная лингвистика, искусственный интеллект, машинное обучение, математическое моделирование, методы оптимизации. E-mail: shvets@isa.ru

Девяткин Дмитрий Алексеевич. Младший научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН. Окончил Рыбинскую государственную авиационную технологическую академию в 2011 году. Автор 14 печатных работ. Область научных интересов: адаптивные методы дистанционного обучения, классификация и кластеризация текстов, интеллектуальный веб-краулинг, искусственный интеллект. E-mail: devyatkin@isa.ru

Зубарев Денис Владимирович. Инженер-исследователь Федерального исследовательского центра «Информатика и управление» РАН. Окончил Российский университет дружбы народов в 2014 году. Область научных интересов: искусственный интеллект, поисковые системы, распределенные системы. E-mail: zubarev@isa.ru

Тихомиров Илья Александрович. Доцент. Заведующий лабораторией «Интеллектуальные технологии и системы» Федерального исследовательского центра «Информатика и управление» РАН. Окончил Рыбинскую государственную авиационную технологическую академию в 2002 году. Кандидат технических наук. Автор более 50 печатных работ. Область научных интересов: искусственный интеллект, компьютерная лингвистика, поисковые системы, информационная безопасность, Интернет-системы. E-mail: tih@isa.ru

Григорьев Олег Георгиевич. Главный научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН. Окончил Московский институт электронного машиностроения в 1980 году. Доктор технических наук. Автор 57 печатных работ. Область научных интересов: искусственный интеллект, компьютерная лингвистика, поисковые системы, информационная безопасность, Интернет-системы, СУБД. E-mail: ogrigoriev@gmail.com