

СОДЕРЖАНИЕ

<b>К юбилею Юджина Гарфилда</b>	3
<b>Сток В.Г.</b> Информетрические анализы систем организации знания (СОЗ)	4
<b>Торрес-Салинас Д., Робинсон-Гарсиа Н., Хименес-Контрерас Э., де ла Фуэнтэ Э.</b> Ранжирование издателей по библиометрическим показателям: основные результаты и методологические проблемы при проведении ранжирований академических издателей	14
<b>Хостейн С., Бауман Т.Д., Костас Р.</b> Когда статья фактически опубликована? Анализ дат онлайн доступности, публикации и индексации	20
<b>Селиванова И. В., Гуськов А.Е., Мазов Н.А.</b> Библиометрический анализ журнала «Вычислительные технологии» за 1996-2013 гг.	28

## РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Академик РАН **Ю.М. Арский** (Российская Федерация) — *главный редактор*,  
ВИНИТИ РАН, 125190, Москва, ул. Усневича, 20. Телекс 411249

Проф. д-р. **Р.С. Гиляревский** (Российская Федерация) — *заместитель главного редактора*,  
ВИНИТИ РАН, 125190, Москва, ул. Усневича, 20. Телекс 411249

**С. Дж. Паркер** (Канада) — *заместитель главного редактора*, IDCR, P.O. Box 8500,  
Ottawa, Ontario K1G 3H9, Canada

**А. Джикарайст** (Великобритания) — CURA Consortium and GAVEL g.e.i.e,  
38 Ship Street, Brighton BN1 1AB, UK

**М. Дрейк** (США) — Технологический институт шт. Джорджия, Библиотечный  
и информационный центр, 704 Cherry Street, Atlanta, Georgia 30332-0900, USA

**А. де Кемп** (Германия) — Издательство “Springer-Verlag”, Postfach 10 52 80,  
D-69042 Heidelberg, Germany

Д-р **Т. Кеннон** (Великобритания) — Отдел исследований и разработок  
Британской библиотеки, 2 Sheraton Street, London W1V 4BH, UK

**М. Миддлтон** (Австралия) — Школа информационных систем, QUT Gardens  
Point Campus, 2 George Street, Brisbane, 4000 QLD., Australia

**Т. Молвиг** (Норвегия) — Национальное управление по научной информации,  
вузовским и специальным библиотекам, P.O. Box 2439 Solli, N-0201, Oslo,  
Norway

**Х. Ринкон Феррейра** (Бразилия) — Бразильский институт информации по  
науке и технике (IBICT), SAS — Quadra 5, Lote 06, Bloco H, 700-70-000 Brasilia  
D.F., Brazil

**С. Феррейро** (Чили) — Чилийский университет, Системы информационных  
и библиотечных служб, Casilla de Correo 10D, Santiago, Chile

Проф. **Ю. Фуздивара** (Япония) — Университет Цукуба, Институт электроники  
и информатики, Tsukuba-shu, Ibaraki, 305 Japan

Д-р **М. Хименес** (Испания) — Испанское общество по научной документации  
и информации, Fuencarral, 123-6° dcha., 28010, Madrid, Spain

## К юбилею Юджина Гарфилда

16 сентября этого года выдающемуся представителю информационной профессии Юджину Гарфилду исполнилось 90 лет. Его заслуги перед информационной наукой и практикой неопределимы. Он разработал новый способ раскрытия накопленных человечеством знаний, получивший название *системы цитирования*; создал ряд уникальных информационных продуктов – указатели содержания текущих журналов (*Current Contents*), библиографических ссылок в научной литературе (*Science Citation Index*), ссылок на журналы (*Journal Citation Reports*). Для реализации своих идей основал в г. Филадельфия (США, 1956 г.) уникальный информационный центр *Institute for Scientific Information*, который теперь принадлежит корпорации *Thomson Reuter* и является ядром системы *Web of Knowledge*.

Ю. Гарфилд – неутомимый ученый-исследователь, которого глубоко интересуют современные проблемы информатики и научной коммуникации; в течение более 20 лет еженедельно он публиковал свои статьи в выпусках *Current Contents*, которые составили 17 томов сборников *Essays of an Information Scientist* (1977 – 1994) – своего рода энциклопедию информатики.

В 1948 г. Ю. Гарфилд получил степень бакалавра по химии, в 1954 г. – степень магистра в Колумбийском университете (Нью-Йорк), в 1962 г. стал доктором наук по структурной лингвистике в Пенсильванском университете (Филадельфия), защитив диссертацию «Алгоритм перевода химических наименований в молекулярные формулы». Его работы по индексированию медицинской литературы (1951 г.) в Университете Джонса Хопкинса (Балтимор, шт. Мэриленд) заложили теоретическую и практическую основу создания известной системы анализа и поиска этой литературы (*MEDLARS*). В 1956 г. он создал фирму, позднее названную Институтом научной информации (по его признанию, не без влияния прежнего названия ВИНТИ РАН), просуществовавшую до 1992 г., когда ее купил канадский газетный магнат Томсон.

С давних пор Ю. Гарфилд друг нашей страны, он любил приезжать к нам для чтения лекций, на книжные ярмарки.

Ю. Гарфилд является иностранным членом Российской академии естественных наук (1996 г.), ему присвоено звание почетного доктора наук Дальневосточного государственного университета (1997 г.).

Идеи Ю. Гарфилда, созданные им информационные системы и издания сегодня занимают одно из ведущих мест в информационном обеспечении науки, промышленности, экономики и культуры. Государственные органы и владельцы частных фирм оценивают эффективность своих инвестиций по библиографическим ссылкам на статьи, в которых излагаются результаты исследований, проводившихся на вложенные ими средства. Крупнейшие информационные системы *Web of Science (Thomson-Reuters)*, *Scopus (Elsevier)*, *Springer*, *EBSCO*, Российский индекс научного цитирования отслеживают потоки литературы при помощи инструментов Ю. Гарфилда.

Редакция и редколлегия нашего журнала, наши авторы и читатели, сотрудники Всероссийского института научной и технической информации Российской академии наук сердечно поздравляют Вас, дорогой Джин. Мы счастливы, что Вы в такой великолепной форме и продолжаете работать на благо нашего дела. Долгих Вам лет жизни, здоровья и успехов.

# Информетрические анализы систем организации знания (СОЗ)\*

Вольфганг Г. СТОК  
(Wolfgang G. STOCK)

Университет им. Генриха Гейне,  
факультет информатики, г. Дюссельдорф,  
Германия

*Система организации знания состоит из понятий и семантических отношений между понятиями, которые представляют область знаний терминологически. Различаются пять подходов к системам организации знания (СОЗ): номенклатуры, классификационные системы, тезаурусы, онтологии и как пограничный случай СОЗ, фолксономии. Вопрос исследования, поднятый в статье, следующий: как можно информетрически проанализировать эффективность СОЗ? Количественные информетрические измерения и показатели допускают описание, сравнительные анализы, а также оценку СОЗ и их качества. Описывается современное состояние оценки СОЗ. Большинство оценочных исследований, найденных в литературе, относятся к онтологиям. Вводятся измерения структуры СОЗ (например, приземленность, сложность, фактор рассеивания или структурированность) и показатели ее качества (полнота, последовательность, пересечение и использование).*

## Введение

Для большинства моделей оценки важным показателем качества информационной системы является качество ее информационного наполнения. В нашей модели оценки информационной системы [1] качеством информационного наполнения является подобласть основной сферы качества информационной системы (рис.). Качество информационного наполнения концентрируется на знаниях, хранящихся в системе [2-4]. Знание, непосредственно относящееся к информационным системам, состоит из двух аспектов, а именно знания документов (знания, вкладываемого авторами в их публикации) и знания суррогатов (знания, вкладываемого индексаторами в метаданные документа). В свою очередь знание суррогатов имеет две области: качество индексирования (применение правильных понятий для описания знания, содержащегося в документе, [5, с. 817-825]) и качество системы организации знания, которая используется для индексирования [5, с. 809-816]. СОЗ – система понятий, которая используется для представления (в большинстве случаев научных или других специализированных) документов. Общие типы СОЗ включают номенклатуры, классификационные системы, тезаурусы и онтологии. СОЗ применяются в профессиональных информационных системах, поддерживающих научную коммуникацию

предоставлением специализированной литературы. Несмотря на то, что имеется огромное количество исследований по качеству индексирования и его показателей (например, глубина индексирования, включая полноту индексирования суррогата и особенность индексирования атрибутивных понятий, эффективность индексирования понятий и последовательность индексирования суррогатов), в информатике есть только несколько работ, посвященных качеству СОЗ.

Информационные системы по науке и технологии (например, *Medline* по медицине, *Chemical Abstract Service* по химии и *Inspec* по физике) и информационные службы в контексте совместного управления знанием во многих случаях применяют так называемые «контролируемые словари» или «языки документации или (документационные языки)» в целях индексирования информации и ее поиска. Такие словари организуют понятия и семантические отношения между понятиями определенной области знания в «систему организации знания» (СОЗ).

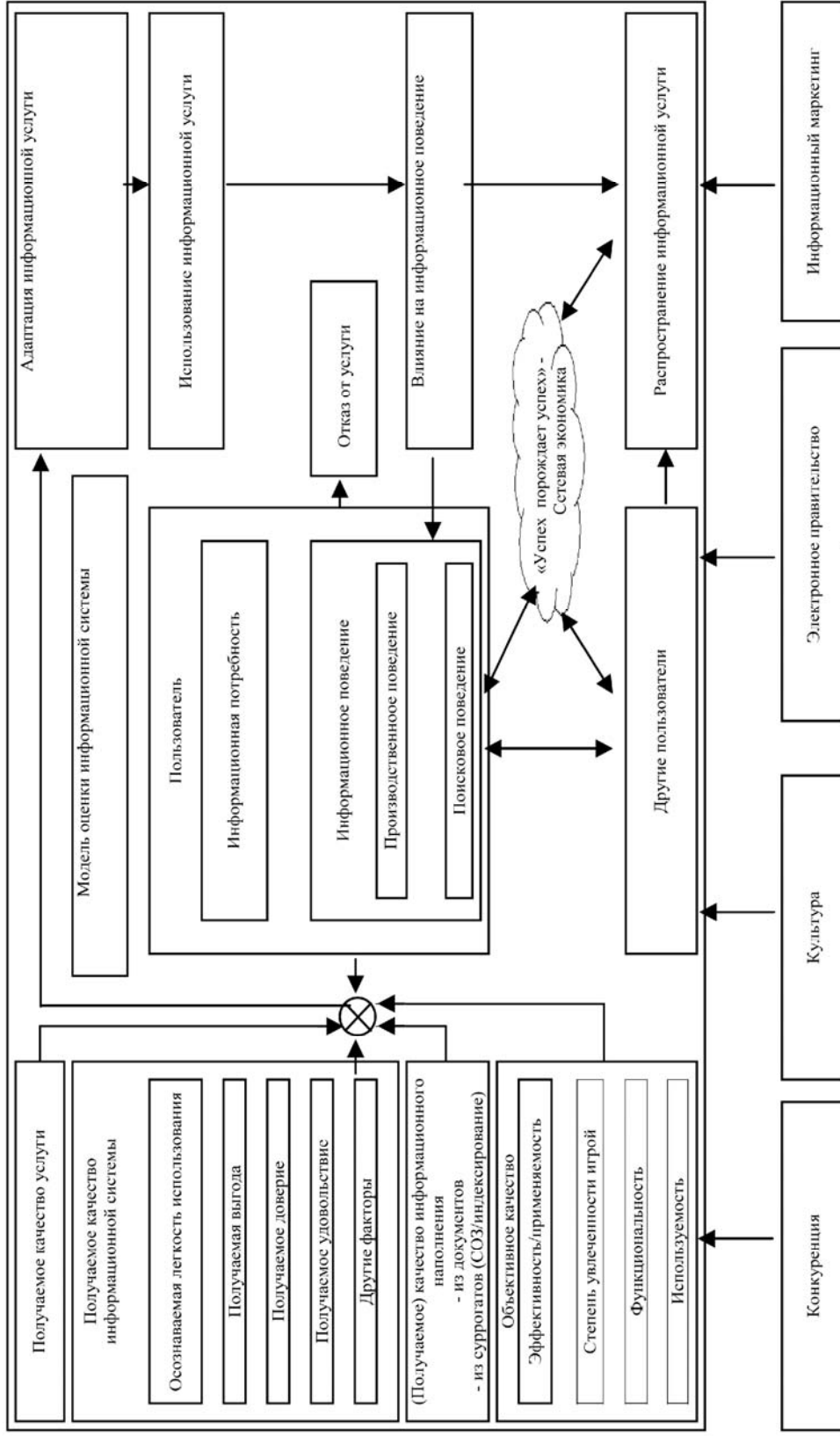
Цель данной статьи заключается в том, чтобы подчеркнуть важность оценки СОЗ как части эмпирической информатики, т.е. информетрии. Согласно Таге-Сатклиффу, информетрия является «изучением количественных аспектов информации в любой форме и в любой социальной группе» [6, с. 1]. Вольфрам делит информетрию на два аспекта, а именно «характеристики на основе системы, исходящие от документального содержания ИПС, и как они индексируются, и характеристики на основе использования, происходящие от способов, с помощью которых пользователи взаимодействуют с системными контентом и интерфейсами, предоставляющими доступ к этому контенту» [7, с. 6].

\*Перевод Stock W.G. Informetric analyses of knowledge organization systems (KOSs).— 2015.—  
<http://www.arxiv.org/ftp/arxiv/papers/1505/1505.03671.pdf>

**Величина 1:  
Информационная услуга**

**Величина 2:  
Пользователь информации**

**Величина 3:  
Принятие информации**



Источники: Schumann & Stock, 2014, р. 8 (модифицированный).

Рис. Измерение качества информационной наполнения в модели оценки информационной системы (ОИС).

Сток и Вебер [8] различают три предмета и соответственно три исследовательские области информетрии: 1) пользователи информации и использование информации (с областью исследования пользователя/использования); 2) сама информация, включающая специальную информацию (например, научную информацию) и сетевую информацию (с научными областями библиометрии, наукометрии и вебометрии); 3) информационные системы (с научной областью исследований принятия оценки и технологии). Информетрический анализ СОЗ является частью характеристик на основе системы, описанной Вольфрамом, и исследования Стока и Вебера по оценке информационных систем.

Исследования оценки способны дать ответ на два вопроса [9]: создаем ли мы нужные вещи (идя к оценке эффективности) и поступаем ли мы правильно (приходя на этот раз к оценке результативности)? Относительно оценки СОЗ эффективность означает создание подходящей СОЗ, а результативность – соответствующее строительство СОЗ (адекватно применяемые фонды, скорость внедрения, оптимальные средства программного обеспечения и т.д.) [10, с. 597]. Мы фокусируемся на эффективности и игнорируем результативность. Наш научный вопрос состоит в следующем: *как мы можем информетрически проанализировать эффективность СОЗ?* Количественные информетрические показатели позволяют как эмпирическое описание, сравнительные анализы, так и оценку СОЗ и их качества. С помощью эмпирического исследования СОЗ мы открываем новое поле деятельности в теориях информетрии.

В следующем параграфе будут кратко описаны СОЗ как системы понятий и семантических отношений. Далее будет дан обзор современного описания и оценки СОЗ. В основном разделе статьи представлены измерения и показатели для информетрической оценки СОЗ. Цель заключается не только в представлении синтеза большого числа подходов к оценке СОЗ, но и в предложении решения для исчерпывающего множества измерений основных структур СОЗ и критериев оценки

СОЗ. Для разработчиков СОЗ эти измерения и показатели должны предоставить полезные советы по созданию надежных номенклатур, тезаурусов, классификационных систем и онтологий.

### Понятия и семантические отношения

СОЗ состоит как из понятий, так и семантических отношений между понятиями в соответствующей области знаний (см. табл. 1, [11]). «Понятие» - это класс, содержащий определенные объекты в качестве элементов, где объекты имеют определенные свойства. Лингвистическим выражением понятия является «слово». Понятия не существуют независимо друг от друга, а входят во взаимосвязи. Мы будем называть отношения между понятиями «семантическими отношениями» [12,13]. В отличие от фольклонимии семантические отношения в СОЗ всегда являются «парадигматическими отношениями», т.е. отношениями, которые имеют силу независимо от документов (в противоположность синтагматическим отношениям, зависящим от совместной встречаемости понятий в документах). В СОЗ важны следующие семантические отношения:

- эквивалентность (синонимия, квазисинонимия или родовая идентичность между понятиями);
- иерархия (гипонимия, меронимия и пример);
- и как оставшийся класс дальнейшие отношения («см. также» как ассоциативное отношение или особые отношения, такие как полезность или имеет\_ дочернюю\_компанию в предприятии СОЗ).

Мы определяем системы организации знания через их кардинальность для выражения понятий и семантических отношений. Три «классических» метода в информетрике и практике – номенклатура, классификация, тезаурус – поддерживаются фольклонимиями и онтологиями. Фольклонимии представляют пограничный случай СОЗ, так как они не имеют ни одного парадигматического отношения [14].

Таблица 1

Системы организации знания (СОЗ) и используемые ими отношения

	Фольклонимия	Номенклатура	Классификация	Тезаурус	Онтология
	Тег	Ключевое слово	Индекс	Дескриптор	Понятие
<b>Эквивалентность</b>	-	да	да	да	да
-синонимия	-	да	да	да	да
-родовая идентичность	-	да	-	-	да
<b>Иерархия</b>	-	-	да	да	да
-гипонимия	-	-	-	да	да
-меронимия	-	-	-	да	да
-пример	-	-	-	по треб.	да
<b>Дальнейшие отношения</b>	-	-	-	да	да
-«См. также»	-	по треб.	по треб.	да	да
-видовые отношения	-	-	-	-	да
<b>Синтагматическое отношение</b>	да	да	да	да	нет

Источник: Stock, 2010, p. 1965 (модифицированный)

Номенклатуры (системы ключевых слов) различаются между собой главным образом использованием отношения эквивалентности и игнорированием всех форм иерархического отношения. В классификационных системах добавляется (не специально разработанное) иерархическое отношение. Тезаурусы также работают с иерархией; некоторые используют нехарактерное отношение иерархии, другие различаются по гипонимии (отношение «is-a») и по меронимии (отношение «часть целого»). В тезаурусах обязательно добавляется обычно не специально разработанное ассоциативное отношение («см. также»). Онтологии используют все описанные выше парадигматические отношения. Они моделируются в формальных языках, где терминологической логике также оказывается должное внимание. По сравнению с другими СОЗ онтологии категорически содержат примеры (индивидуальные понятия). Большинство онтологий работает с точно определенными дальнейшими отношениями. Тот факт, что онтологии напрямую представляют знание (а не просто документы, содержащие знание), допускает исчезновение синтагматических отношений в этом случае.

### Уровень развития оценки СОЗ

Большая часть исследований оценки, найденных в литературе, посвящена онтологиям (см. [15-22]). Первая статья по оценке СОЗ, написанная Гомес-Пересом в 1995 г. [23], также была по онтологиям. Сфера нашего исследования шире и охватывает все виды СОЗ. Существует только несколько оценочных исследований, затрагивающих другие виды СОЗ. Фогель [24] разработал ряд критериев качества для классификационных систем и тезаурусов, примененных в его поисковой системе *Convera*. Используя такие параметры как применимость, контекст, полнота и точность, Оунс и Кохрейн [25] разработали методы оценки тезауруса. Ванг, Кху и Шоуари [26] оценили эффективность навигации классификационной системы.

Гомес-Перес, Фернандес-Лопес и Корхо [27] различают проверку СОЗ и обоснование СОЗ. В то время как проверка основывается на правильном (формальном, а также неформальном) представлении понятий и семантических отношений (с такими аспектами, как последовательность, полнота и избыточность [28]), обоснование СОЗ относится к «реальному миру», т.е. сравнению между содержанием СОЗ и его «реальным» эквивалентом в соответствующей области знания [28].

Полагаясь на характеристику, данную Сабу и Фернандесом [29, с. 194], оценкой СОЗ является определение качества СОЗ на соответствие принципам построения. В этом определении имеются два ключевых понятия. Чем определяется «качество» и к чему относятся «принципы построения»? *Критерий качества* определяет «надежную» СОЗ. Врандечич [22, с. 295-296] предоставляет список таких критериев качества, среди которых точность (правильно ли СОЗ представляет свою область знаний?), адаптивность (предвидит ли СОЗ использование?), полнота (охватывается ли область знаний соответствующим образом?), последовательность (является ли СОЗ логично последовательной?) и коммерческая доступность (легко ли получить доступ и применять СОЗ?). Но все эти критерии качества являются «желательным аспектом, целями для управления созданием и оценкой онтологии. Ни один из них нельзя измерить непосредственно» [22, с. 296]. Важно иметь в виду, что мы не всегда можем работать с *измерениями качества*, а только с *показателями качества*. Есть несколько

интерпретаций в литературе, касающихся принципов построения [15, 29, с. 197 ff.], главным образом это примитивные метрики (как, например, число понятий), системы на основе данных (сравнения СОЗ с ее областью знаний), сравнения СОЗ между собой, подходы, касающиеся синтаксических и семантических структур СОЗ и, наконец, системы на основе пользователей (эксперименты с пользователями и опросами или интервью). Для всех принципов построения мы предоставляем иллюстрирующие примеры из литературы.

### Примитивные метрики

Простые метрики оценки – более известные как описательные метрики – основаны на подсчете понятий и отношений в СОЗ. Согласно Хуангу и Дьяо [30, с. 133], «оценка понятий количества заключается в подсчете числа понятий в онтологии», а «оценка вероятности наличия свойства предоставляет обзор избытка отношений между понятиями». Тартир и Арпинар [31] проводят различие между отношениями наследования (отношений, в которых свойства понятий становятся присущими более узким терминам понятий, таким как гипонимия) и другими отношениями (как, например, отношение ассоциации) и подсчетом обоих видов отношений. Помимо этого, Тартир и Арпинар [31, с. 187] принимают во внимание примеры, еще известные как понятия в СОЗ, которые представляют индивидуумов. Ян, Чжан и Йе [32, с. 165] работают со средним числом отношений на понятие. Более подходящим показателем служит «дерево баланса» СОЗ [30, с. 133]: «если дерево сбалансировано, то все его поддеревья имеют одинаковую структуру».

### СОЗ и их области знания

Представляет ли СОЗ свою область знания адекватно? Брюстер, Аланин, Дасмахпатра и Уилкс [33] переопределяют хорошо известные метрики полноты и точности применительно к СОЗ:

Желательно, чтобы *точность* отражала количество правильно установленного знания (в онтологии) по отношению к целому знанию, доступному в онтологии. Хотелось бы определить *полноту* как отражение количества правильно установленного знания по отношению ко всему знанию, которое должно быть установлено [33, с. 1].

Согласно авторам, «знание» относится к «понятиям», как представленное лингвистически словами. Они разработали совокупность типичных документов для области знаний и сравнили слова в текстах со словами в СОЗ. По мнению Брюстера и его коллег, онтология «может быть поставлена в невыгодное положение за термины, представленные в массиве и отсутствующие в онтологии, и за термины, представленные в онтологии, но отсутствующие в массиве» [33, с. 3]. Хотя нетрудно идентифицировать слова в СОЗ, нужна смелость, чтобы собрать типичные (или даже все) документы определенной области знаний.

### Данная и другие СОЗ

Чтобы показать уникальность некоей СОЗ, необходимо сравнить ее с другими СОЗ. Простой вопрос исследования здесь состоит в следующем: «итак, как мы можем измерить сходство онтологий или частей онтологий?» [34, с. 251]. Но ответ никоим образом не является столь же простым, как вопрос. В литературе имеется два подхода к изучению сходства между СОЗ: первый основан на общих словах и понятиях в слова-

ре [15,20,34], а другой работает с индексированными документами в случае полнорепрезентации (т.е. применяя различные СОЗ для индексации одинаковых документов) [35].

### **Синтаксическая и семантическая структура СОЗ**

Оценка синтаксической структуры нацелена на правильное использование формального языка. Для онтологий, предназначенных для применения в семантической сети, используются язык сетевой онтологии (OWL – Web Ontology Language) и подход к описанию ресурса (RDF – Resource Description Frame). Гораздо важнее оценка семантической структуры СОЗ. Фахад и Абдул Кадир [36] проводят различие между избыточностью, неполнотой (которая подобна подходу полноты и точности на основе данных) и непоследовательностью. Избыточность случается, когда определенная информация вводится больше одного раза в СОЗ, например, когда понятие дважды располагается в СОЗ в двух разных позициях. Непоследовательность является главным образом следствием ошибок цикличности (понятие определяется как более широкий термин или как сам по себе более узкий) и ошибок распределения (неправильные разбиения понятия на более узкие термины). Фахад, Абдул Кадир и Ношарван [37] определяют, что «главной причиной этих ошибок является то, что составители онтологий не классифицируют понятия правильно» [37, с. 286].

### **Подходы на основе пользователей**

Ной [38] называет показатели, такие как полнота, последовательность и правильность «объективными» критериями оценки: «хотя все эти виды оценки или сравнительные методы необходимы, ни один из них не является полезным для *потребителей онтологий*, которым нужно выяснять, какие существуют онтологии, и что важнее, какая из них подходит для их текущих задач» [38, с. 80]. Чтобы получить представление о качестве СОЗ на основе пользователя, некоторые авторы провели экспериментальные исследования тестируемых лиц или интервьюируемых пользователей с помощью опросов или инструкций.

Каселлас [10] оценил СОЗ через меры полезности. Он предложил два опроса, один с вопросами, касающимися понятий, определений, примеров и отношений СОЗ, а второй с более общим содержанием (как, например, «Я считал онтологию простой для понимания», или «Я полагаю, что существует больше противоречивости в этой онтологии»). Опрашиваемые были экспертами в области знаний СОЗ. Их просили выразить свое мнение по шкале от 0 до 5 (первый опрос) и от 1 до 10 (второй опрос).

Суомела и Кекалайнен [39] оценили онтологию как инструмент создания запроса. Ванг, Кху и Шоудри [25] оценили эффективность навигации классификационной системы. Оба исследования работали с экспериментами (метод тестирования на основе задачи), а также с помощью интервью [25] или опросов [39].

### **Меры и показатели оценки СОЗ**

В этом разделе мы вводим информетрические меры и показатели оценки СОЗ. Опираясь на обзор литературы и главу, касающуюся оценки СОЗ в *Handbook of Information Science* [5], представляем множество измерений структуры СОЗ и четыре показателя качества СОЗ (полнота, последовательность, пересечение и использование).

### **Основные измерения структуры**

Несколько простых параметров можно использовать для анализа структуры СОЗ [40]. Эти параметры относятся как к понятиям, так и к семантическим отношениям. Введем следующие структурные меры:

- Число понятий
- Семантическая выразительность (число и вид семантических отношений)
- Структурированность (среднее число семантических отношений на понятие)
- Число иерархических уровней
- Фактор рассеивания (число верхних терминов)
- Приземленность (число нижних терминов)
- Сложность (степень полииерархии)
- Степень перестроения (среднее число частных терминов на понятие).

Первоначальная базовая ценность – это число понятий в СОЗ. Здесь, в противоположность сказанному, применяется принцип «чем больше, тем лучше». Скорее цель состоит в получении оптимального значения числа терминов, адекватно представляющих область знания и документов, содержащихся в ней, соответственно. Если терминов слишком мало, то не все аспекты области знаний могут быть выборочно описаны. Если пользователь даже не находит «свой» поисковый термин, это будет иметь негативные последствия для полноты, а если он находит подходящий гипоним, будет страдать точность поисковых результатов. Если слишком много понятий допущено в СОЗ, то есть опасность, что пользователи утратят концентрацию внимания и очень мало документов будет найдено для каждого понятия. Когда документы индексируются через СОЗ (что является правилом, за исключением онтологий), среднее число документов на понятие служит хорошей оценкой для оптимального числа терминов в СОЗ. Дальнейший интерес представляет число обозначений (синонимов и квазисинонимов) на понятие. Среднее число обозначений (например, не дескрипторов) понятия (например, из дескриптора) является надежным показателем для использования обозначений в СОЗ.

По аналогии с понятиями число различных используемых семантических отношений обеспечивает показатель структуры СОЗ (семантическую выразительность). Общее число отношений в СОЗ представляет особый интерес. Рассматриваемые с точки зрения сети понятия СОЗ представляют собой узлы, тогда как их отношения – линии. Формат отношений является общим числом всех линий в СОЗ (без связей с обозначениями, так как они формируют свой собственный показатель). Полезным полученным параметром является среднее число семантических отношений на понятие, т.е. средняя степень терминов. Показатели количества понятий и размера связей могут быть сведены в «структурированность СОЗ».

Информация, касающаяся числа иерархических уровней, а также распределения терминов по всем этим индивидуальным уровням, представляет особый интерес. Также важны данные относительно числа верхних терминов (и таким образом различных фасетов) и нижних терминов (понятий на самом низком уровне иерархии), всякие данные относительно общего числа всех терминов в СОЗ. Отношение числа верхних терминов к числу всех терминов называется «фактором рассеивания», тогда как аналогичное отношение к нижним терминам может быть отнесено к «приземленности». «Сложность» в свою очередь измеряет степень полииерархии в СОЗ. Она относится к среднему числу гиперо-



нимов на каждое понятие. Подсчетом числа гипонимов ко всем понятиям, имеющим гипонимы (минус один), мы подбираем значение для среднего числа элементов одного уровня каждого понятия.

Сёргел [41] предлагает измерение степени перестроения термина. Степень перестроения СОЗ составляет среднее число частных терминов на понятие. Степень перестроения для *Сада* равняется единице, для *Вечеринки в саду* – 2, для *Обеда на вечеринке в саду* – 3 и т.д. Для СОЗ на английском языке более или менее легко подсчитать слова, образующие термин, для других языков, например, немецкого со многими составляющими (Garten: 1, Gartenfest: 2, Gartenfestessen: 3), мы должны прежде всего применить сложное разбиение, а затем подсчитать для каждого объекта СОЗ число ее частных терминов. Табл. 2 представляет обзор множества основных измерений структур СОЗ.

#### Показатель полноты

Полнота относится к степени терминологического охвата области знаний. Если область знания не слишком мала и не легко понятна, это значение будет очень трудно определить. Ю, Торн и Там [42, с. 775] определяют полноту посредством вопроса: «имеет ли СОЗ понятия, отсутствующие по отношению к релевантным принципам построения»? Порталуши [43] показывает, что полнота тематических областей СОЗ может оцениваться через выборки из индексированных документов. В случае исследования статьи по хронобиологии изучались в *Medline*. Были получены оригинальные документы, и размещенные понятия MeSH (Medical Subject Headings – «Медицинские предметные рубрики», которые являются тезаурусом по медицинской терминологии) анализировались в суррогатах. Порталуши [43] сообщает, что «с помощью прочтения каждой статьи было возможно идентифицировать общие хронобиологические понятия, еще не ассоциированные с определенными рубриками MeSH». Отсутствующие понятия, определенные таким образом, могли бы быть представлены в MeSH и могут быть ошибочно не замечены индексатором (в данном случае это будет ошибкой индексирования) или они просто не отражаются в СОЗ. В практике исследования некоторые общие хронобиоло-

гические понятия «не должны ассоциироваться с какой-либо определенной рубрикой MeSH» [43, с. 1213], таким образом, MeSH должна считаться незавершенной с точки зрения хронобиологии.

Если подсчитать понятия в тематическом подмножестве СОЗ и определить число терминов, отсутствующих с точки зрения тематики, отношение числа отсутствующих терминов и общего числа терминов (т.е. тех, которые отражены в СОЗ, плюс те, которые отсутствуют) приводит к оценочному значению полноты или полноты выдачи (по мнению авторов [33]) относительно соответствующей подобласти знания.

#### Семантические показатели

Последовательность СОЗ относится к пяти аспектам:

- Семантическая непоследовательность;
- Ошибка цикличности;
- Пропуск иерархических уровней;
- Избыточность;
- «Теннисная проблема».

Непоследовательности весьма вероятно возникают тогда, когда несколько СОЗ (последовательных по своей сути) объединяются в большую СОЗ. В случае семантической непоследовательности термины неправильно располагались в семантической сети всех понятий. Рассмотрим следующий дескрипторный объект:

Рыбы

BT: Морские животные

NT: Рыбы, обитающие в соленой воде

NT: Рыбы, обитающие в пресной воде

BT (более широкий термин) и NT (более узкий термин) расширяют семантическое отношение гипонимии в этом примере. В данном иерархическом отношении гипонимы наследуют все характеристики своих гиперонимов. Например, термин *Морские животные* содержит характеристику «живет в океане». Эта характеристика переходит к гипониму *Рыбы* и распространяется до его гипонимов *Рыбы, обитающие в соленой воде* и *Рыбы, обитающие в пресной воде*. Семантическая непоследовательность возникает в случае *Рыб, обитающих в пресной воде*, так как они не живут в океане.

Таблица 2

#### Основные измерения структуры СОЗ

Размерность	Информетрическое измерение	Подсчет
Структурированность	Количество понятий	Число понятий (узлы в сети)
	Формат отношений	Число отношений между понятиями (линии в сети)
	Семантическая экспрессивность	Число различных семантических отношений
	Документы на понятие	Среднее число документов на понятие (для определенной информационной системы)
Иерархия	Использование обозначений	Среднее число обозначений на понятие
	Глубина иерархии	Число уровней
	Иерархическое распределение понятий	Число понятий на различных уровнях
	Фактор рассеивания	Отношение числа верхних терминов к числу всех понятий
	Фактор приземленности	Отношение числа нижних терминов к числу всех понятий
	Фактор сложности	Среднее число гиперонимов на понятие
Перестроение	Фактор родства	Среднее число совместно встречающихся гипонимов на понятие
	Степень перестроения	Среднее число частных понятий на понятие

Источник: Stock & Stock, 2013, p. 815 (модифицированный)

Ошибки цикличности встречаются в иерархическом отношении, когда одно понятие появляется больше одного раза в понятийной лестнице [17]: «ошибки цикличности встречаются, когда класс сам по себе определен как специализация или обобщение» [17, с. 261]. Предположим появление двух СОЗ. Представим, что СОЗ 1 содержит следующее множество понятий:

Люди  
NT: Путешественники  
тогда как СОЗ 2 формулирует  
Путешественники  
NT: Люди.

Когда появляются обе СОЗ, результатом является логический цикл (пример, взятый из [44]).

Ошибки пропуска являются результатом не включения уровней иерархии. Эта ошибка хорошо описана Аристотелем в его работе *Topics* [45, кн. 6, гл. 5, с. 479-480]. Здесь мы также можем привести пример:

Коза  
NT: дикая коза  
NT: домашняя коза  
Дикая коза  
NT: Домашняя коза.

В биологической иерархии *Коза* является более широким термином по отношению к *Дикой козе* (*Sarpha aegagrus*). *Дикая коза*, в свою очередь, является более широким термином по отношению к *Домашней козе* (*Sarpha hirus*). Устанавливая прямое отношение между *Козой* и *Домашней козой*, наша СОЗ опускает иерархический уровень. Причиной ошибки пропуска является ошибочное предположение NT *Домашняя коза* внутри понятия *Коза*.

СОЗ избыточна, когда понятие появляется более одного раза в СОЗ. Такого рода ошибка может возникнуть, когда понятие интегрируется в несколько контекстов. В тезаурусе *Вишня* может быть гипонимом *Фруктового дерева* и гиперонимом *Гриота* (сорт вишни) и *Черешни*. В другом фасете того же тезауруса *Вишня* является более узким термином *Фруктового бренди* и более широким термином *Шерри бренди*. В этом примере второй вариант является ошибочным. Вишня должна быть убрана из фасета *Бренди*. Вместо этого дескриптора должна быть установлена ассоциативная связь между *Вишней* и *Шерри бренди*.

Что касается оценки онтологии, Хартманн и др. [18, с. 17] упоминают так называемую «теннисную проблему». Это явление, «в котором родственные слова могут встретиться в двух совершенно различных частях онтологии без соответствующей связи между ними, например, «мальчик, подбирающий мячи на корте» мог возникнуть как нисходящее слово от «мальчик», а «теннисный мяч» – от слов «спортивный инвентарь», несмотря на очевидное семантическое отношение». Действительно, если СОЗ состоит только из иерархии, теннисная проблема вызывает головную боль у разработчиков онтологии. Однако каждая СОЗ, допускающая использование ассоциативной связи, способна соотносить оба понятия:

Мальчик, подбирающий мячи на корте *СМ*.  
**ТАКЖЕ** Теннисный мяч  
(и наоборот). Задача того, кто осуществляет оценку, заключается в том, чтобы расположить понятия в СОЗ в соответствии с тесными семантическими отношениями, которые не связаны через кратчайшие траектории.

### **Пересечение с другими СОЗ**

Подход к изучению сходства между СОЗ заключается в подсчете общих слов и понятий в двух СОЗ. На уровне слов Медке и Стаб [34, с. 254] используют рас-

стояние Ливенштейна (т.е. число редакторских шагов между двумя строками). Слова с низким числом редакторских шагов считаются сходными. Если СОЗ 1 имеет объект «TopHotel», а СОЗ 2 – «Top\_Hotel», тогда расстояние Ливенштейна равняется единице (одна вставка) и слова поэтому являются одинаковыми. Но этот метод склонен к неудаче. Расстояние Ливенштейна между «Power» и «Tower» также равняется единице, несмотря на их различие. На уровне понятий сравнение даже более проблематично. Обрст и др. [20] так описывают эту проблему:

«Сказать, что два понятия имеют схожую семантику, подразумевает строго говоря то, что они занимают одинаковые места в их структурах. Однако вышеуказанная проблема ясна: систематизация онтологии определяется с точки зрения соответствия понятий (эквивалентность, тождество, сходство). Но как именно мы получаем доступ к понятиям, чтобы определить, находятся ли они в отношении соответствия?» [20, с. 146].

Обрст и др. [20] обнаружили, что большая часть исследований основана на словаре (т.е. словах – с учетом выше упомянутых проблем) или на структуре СОЗ (например, схожие более широкие термины и схожие более узкие термины). Подсчет общих слов и общих понятий является хорошей идеей на уровне теории, но когда дело доходит до практики, возникает проблема.

К счастью, есть альтернативный метод. В случае полирепрезентации [46] различные методы представления знания, также как различные СОЗ, используются для индексирования одних и тех же документов. Хостейн и Петерс [35] сравнивают теги (т.е. в известном смысле фолксономий, точки зрения читателей), тематические рубрики *Inspec* (перспектива составителей указателей), *Keywords Plus* (в качестве метода автоматического индексирования), а также авторские ключевые слова и слова из названий и рефератов (точка зрения авторов) более 700 журнальных статей. Авторы особенно заинтересованы в пересечении между тегами на основе фолксономии и другими методами представления знания. Безусловно, можно также сравнить несколько СОЗ друг с другом, поскольку они использовались для индексирования одних и тех же документов. Значение  $g$  представляет число идентичных понятий различных СОЗ на документ,  $a$  – является числом уникальных понятий из СОЗ 1 на документ,  $b$  служит числом уникальных понятий из СОЗ 2 на документ. Сходство между СОЗ 1 и СОЗ 2 можно подсчитать с помощью косинуса.

Метод Хостейна-Петерса можно использовать для сравнительной оценки различных СОЗ в контексте полирепрезентации. Когда меры сходства между двумя СОЗ относительно низки, это указывает на словари, дополняющие друг друга, что очень ценно для пользователей, так как предоставляет дополнительные точки доступа к документу. Если сходства, с другой стороны, являются высокими, одна из двух СОЗ будет вероятно становиться избыточной в такой практике.

### **Использование**

Из работы Ноя [38] мы узнали, что для оценки СОЗ необходимо рассматривать точку зрения пользователей. Соответственно оценка СОЗ должна включаться в более широкий подход, отражающий службу, пользователя, его принятие, окружение и время [1]. Аспекты методов на основе пользователей содержат показатели полученного качества услуги (охватывается, например, методом SERVQUAL), качества системы с подобластями реализованной легкости использования, выгоды, доверия,

удовольствия, других факторов (модель принятия технологии) и применимости.

Для оценки полученного качества услуги мы предлагаем использовать SERVQUAL [47]. SERVQUAL работает с двумя наборами утверждений: те, которые используются для измерения ожиданий относительно категории услуги вообще (EX), и те, которые измеряют ощущения (PE), касающиеся категории определенной услуги. Каждое утверждение сопровождается семибальной шкалой, изменяющейся от «категорически не согласен» (1) до «полностью согласен» (7). Для оценки ожидания можно отметить, что «в СОЗ в экономике полезно иметь отношение *имеет дочернюю компанию* при формулировке запросов», и затем обратиться к проверке предмета для его выражения в числовом исчислении по определенной шкале. Соответствующее утверждение для регистрации значения ощущения тогда будет таким: «в СОЗ X, отношение имеет дочернюю компанию, полезно при формулировке запросов». Здесь предмет также определяет численное значение. Для каждого элемента оценка различия определяется как  $Q = PE - EX$ . Если, например, тестируемый определяет значение 1 для ощущения после отметки *a*, равной 4, для ожидания, значение Q для системы X относительно этого атрибута в вопросе должно быть  $1 - 4 = -3$ .

При оценке полученного качества СОЗ используются опросы. Тестируемые должны быть знакомы с системой в целях получения правильных оценок. Для каждой подобласти набор утверждений формулируется так, чтобы пользователь мог произвести оценку по семибальной шкале (от «вполне вероятно» до «почти невероятно»). Дэвис [48, с. 380], например, утверждал: «использование системы X в моей работе способствовало моему гораздо быстрому завершению задач» (для измерения полученной выгоды) и «мое взаимодействие с системой X было ясным и понятным» (для аспекта осознаваемой легкости использования).

Полезными СОЗ являются те, которые не разочаровывают пользователей. Общая процедура по применимости тестов, по мнению Нилсона [49], является тестированием на основе задачи. Здесь исследователь определяет представительные задачи, которые могут

быть выполнены с использованием СОЗ и являются типичными для такого рода СОЗ. Такая задача по оценке применимости СОЗ в экономике может быть следующей: «искать понятия для подготовки запроса о пятом цикле Кондратьева!». Тестируемые должны быть «представительной выборкой конечных пользователей» [50, с. 25]. Тестируемым ставят задачи и они наблюдаются исследователем в ходе их выполнения. Полезно, чтобы тестируемые высказывали свои мысли при выполнении задач («мышление вслух»). Помимо тестов на основе задач исследователю полезно проводить интервью по СОЗ (например, относительно общего впечатления от СОЗ, полноты и семантической последовательности). В табл. 3 приведены все упомянутые показатели качества СОЗ.

### Заключение

Наши параметры в группе «Основная структура» являются простыми мерами, которые могут быть автоматически доступны системе. Действительно, это качественный аспект создания каждой СОЗ и поддержки программного обеспечения для предоставления такого рода данных основных структур [24]. Полнота, семантическая непоследовательность, пересечение с другими СОЗ и данные на основе пользователей являются показателями качества, которые «останутся задачей для интеллектуального уровня людей» [22, с.308].

Следующие шаги в исследовании по оценке СОЗ должны включать анализ различных методов оценки. Гомес-Перес [51, с. 74] упоминает такие научные вопросы – «насколько надежны методы оценки онтологий?» или «как платформы развития онтологии осуществляют оценку содержания?».

В нашей работе мы стремились сфокусировать внимание ученых в области информатики на широко отрицаемых аспектах информетрии: информетрическое описание и оценка СОЗ. В качестве основы дальнейшего обсуждения мы описали уровень развития оценки СОЗ и ввели предложения по измерениям, а также показателям качества СОЗ. Надеемся, что мы смогли расширить теорию информетрии путем введения методов оценки СОЗ.

Таблица 3

### Показатели качества СОЗ

Размерность	Информетрический показатель	Подсчет/метод
Полнота	Полнота подобласти знания	Отношение числа отсутствующих понятий и числа всех понятий (в определенной СОЗ и других) относительно подобласти
Семантика	Семантическая непоследовательность	Число ошибок семантической непоследовательности
	Цикличность	Число ошибок цикличности
	Пропуск иерархических уровней	Число ошибок пропуска
	Избыточность	Число ошибок избыточности
	Теннисная проблема	Число отсутствующих связей между ассоциативными понятиями
Множественные СОЗ	Степень полирепрезентации	Пересечение
	Полученное качество СОЗ	Опросы с помощью SERVQUAL
	Принятие СОЗ	Исследования принятия технологии
	Применимость	Тесты на основе задач

Источник: Stock & Stock, 2013, p. 815 (модифицированный)

## Литература

1. *Schumann L., Stock W. G.* The information service evaluation (ISE) model// *Webology*. — 2014.— Vol. 11, No.1, art. 115.
2. *DeLone W. H., McLean E. R.* Information systems success. The quest for the dependent variable// *Information Systems Research*.—1992.— Vol. 3, No.1.— P. 60-95.
3. *DeLone W. H., McLean E. R.* The DeLone and McLean model of information systems success. A ten-year update// *Journal of Management Information Systems*.— 2003.— Vol. 19, No. 4.—P. 9-30.
4. *Jennex M. E., Olfman L.* A model of knowledge management success// *International Journal of Knowledge Management*.— 2006.— Vol. 2, No. 3.— P. 51-68.
5. *Stock W. G., Stock M.* Handbook of Information Science.— Berlin, Germany, Boston, MA: De Gruyter Saur, 2013.
6. *Tague-Sutcliffe J.* An introduction to informetrics// *Information Processing & Management*.— 1992.—Vol. 28, No.1.—P. 1-4.
7. *Wolfram D.* Applied informetrics for information retrieval research.— Westport, CO, London, UK: Libraries Unlimited, 2003.
8. *Stock W. G., Weber S.* Facets of informetrics// *Information – Wissenschaft und Praxis*.—2006.— Vol. 57, No.8.— P. 385-389.
9. *Drucker P. F.* Managing for business effectiveness// *Harvard Business Review*.— 1963.— No. 41(May/June).— P. 53-60.
10. *Casellas N.* Ontology evaluation through usability measures. An experiment with the SUS scale in the legal domain// *Lecture Notes in Computer Science*.— 2009.— Vol. 5872.—P. 594-603.
11. *Stock W. G.* Concepts and semantic relations in information science// *Journal of the American Society for Information Science and Technology*.—2010.—Vol. 61, No. 10.—P. 1951-1969.
12. *Khoo C. S. G., Na J. C.* Semantic relations in information science// *Annual Review of Information Science and Technology*.— 2006.—Vol. 40.—P. 157-228.
13. *Storey V. C.* Understanding semantic relationships// *VLDB Journal*.—1993.—Vol. 2, No. 4.—P. 455-488.
14. *Peters I.* Folksonomies. Indexing and Retrieval in Web 2.0.— Berlin, Germany: De Gruyter Saur, 2009. (Knowledge & Information. Studies in Information Science.)
15. *Brank J., Grobelnik M., Mladenić, D.* A survey of ontology evaluation techniques// *Conference on Data Mining and Data Warehouses*. Ljubljana, Slovenia, October 17, 2005. — 2005.
16. *Gangemi A., Catenacci C., Ciaramita M., Lehmann L.* A theoretical framework for ontology evaluation and validation// *Proceedings of SWAP 2005, the 2nd Italian Semantic Web Workshop*. Trento, Italy, December 14-16, 2005.— 2005.
17. *Gómez-Pérez A.* Ontology evaluation/ S. Staab, R. Studer (Eds.), *Handbook on Ontologies* (pp. 251-273).—Berlin, Germany: Springer, 2004.
18. *Hartmann J., Spyns P., Giboin A., Maynard D., Cuel R., Suárez-Figueroa, Sure Y.* Methods for Ontology Evaluation / *Knowledge Web Consortium (EU-IST Network of Excellence IST-2004-507482 KWEB)*.—2005.
19. *Kebagias D. D., Papadimitriou I., Hois J., Tzocaras D., Bateman J.* A methodological approach for ontology evaluation and refinement// *ASK-IT International Conference*. Nürnberg, Germany, 26-27 June, 2008. — 2008.
20. *Obrst L., Ceusters W., Mani I., Ray S., Smith B.* The evaluation of ontologies/ C. J. O. Baker, K.-H. Cheung (Eds.), *Semantic Web. Revolutionizing Knowledge Discovery in the Life Sciences* (pp. 139-158). —Berlin, Germany: Springer, 2007.
21. *Pak J., Zhou L.* A framework for ontology evaluation// *Lecture Notes in Business Information Processing*.—2011.—Vol. 52.—P. 10-18.
22. *Vrandečić D.* Ontology evaluation/ S. Staab, & R. Studer (Eds.), *Handbook on Ontologies* (pp. 293-313). — Berlin, Germany: Springer, 2009.
23. *Gómez-Pérez A.* Some ideas and examples to evaluate ontologies// *Proceedings of the 11th Conference on Artificial Intelligence for Applications* (pp. 299-305).— Los Alamitos, CA: IEEE Computer Society, 1995.
24. *Vogel C.* Quality Metrics for Taxonomies. — Vienna, VA: Convera, 2002.
25. *Owens L. A., Cochrane P. A.* Thesaurus evaluation// *Cataloging & Classification Quarterly*.—2004.— Vol. 37, No.3-4.— P. 87-102.
26. *Wang Z., Khoo C. S. G., Chaudhry A. S.* Evaluation of the navigation effectiveness of an organizational taxonomy built on a general classification scheme and domain thesauri// *Journal of the Association for Information Science and Technology*.— 2014.—Vol. 65, No.5.— P. 948-963.
27. *Gómez-Pérez A., Fernández-López M., Corcho O.* *Ontological Engineering*. — London, UK: Springer, 2004.
28. *Lovrenčić S., Čubrillo M.* Ontology evaluation. Comprising verification and validation// *Central European Conference on Information and Intelligent Systems (CECIIS-2008)*. Zagreb, Croatia, September 24-26, 2008. — 2008.
29. *Sabou M., Fernandez M.* Ontology (network) evaluation/ M. C. Suárez-Figueroa et al. (Eds.), *Ontology Engineering in a Networked World* (pp. 193-212). —Berlin, Germany: Springer, 2012.
30. *Huang N., Diao, S.* Structure-based ontology evaluation// *Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE-2006)* (pp. 132- 137).— Los Alamitos, CA: IEEE Computer Society, 2006.
31. *Tartir S., Arpinar I. B.* Ontology evaluation and ranking using OntoQA// *International Conference on Semantic Computing (ICSC-2007)* (pp. 185-192).— Los Alamitos, CA: IEEE Computer Society, 2007.
32. *Yang Z., Zhang D., Ye C.* Evaluation metrics for ontology complexity and evolution analysis// *Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE-2006)* (pp. 162-169). — Los Alamitos, CA: IEEE Computer Society, 2006.
33. *Brewster C., Alani H., Dasmahapatra S., Wilks Y.* Data driven ontology evaluation// *International Conference on Language Resources and Evaluation (LREC-2004)*. Lisbon, Portugal, 24-30 May 2004. — 2004.
34. *Maedche A., Staab S.* Measuring similarity between ontologies// *Lecture Notes in Computer Science*.— 2002.—Vol. 2473.— P. 251-263.
35. *Haustein S., Peters I.* Using social bookmarks and tags as alternative indicators of journal content description// *First Monday*.— 2012.— Vol. 17, No.11.

36. *Fahad M., Abdul Qadir M.* A framework for ontology evaluation// 16th International Conference on Conceptual Structures (ICCS 2008) (pp. 149-158). Toulouse, France. — 2008.
37. *Fahad M., Abdul Qadir M., Nosbairwan M. W.* Semantic inconsistency errors in ontology// 2007 IEEE International Conference on Granular Computing (pp. 283-286). — Los Alamitos, CA: IEEE Computer Society, 2007.
38. *Noy N. F.* Evaluation by ontology consumers// IEEE Intelligent Systems.— 2004.— Vol. 19, No. 4.— P. 80- 81.
39. *Suomela S., Kekäläinen J.* User evaluation of ontology as query construction tool// Information Retrieval.— 2006.— Vol. 9, No. 4.— P. 455-475.
40. *Gangemi A., Catenacci C., Ciaramita M., Lehmann L.* Modelling ontology evaluation and validation// Lecture Notes in Computer Science.— 2006.— Vol. 4011.— P. 140-154.
41. *Soergel D.* Evaluation of knowledge organization systems (KOS). Characteristics for describing and evaluation KOS// Workshop „Classification Crosswalks: Bringing Communities Together” at the First ACM + IEEE Joint Conference on Digital Libraries. Roanoke, VA, USA, June 24-28, 2001. — 2001.
42. *Yu J., Thorn J. A., Tam A.* Requirements-oriented methodology for evaluating ontologies// Information Systems.— 2009.— Vol. 34, No. 8.— P. 766-791.
43. *Portaluppi E.* Consistency and accuracy of the Medical Subject Headings thesaurus for electronic indexing and retrieval of chronobiologic references// Chronobiology International.— 2007.— Vol. 24, No. 6.— P. 1213-1229.
44. *Cross V., Pal A.* An ontology analysis tool// International Journal of General Systems.— 2008.— Vol. 37, No. 1.— P. 17-44.
45. *Aristotle.* Topics.— Sioux Falls, SD: NuVision, 2005.
46. *Ingwersen P., Järvelin, K.* The Turn. Integration of information seeking and retrieval in context.— Dordrecht, NL: Springer, 2005.
47. *Parasuraman A., Zeithaml V. A., Berry L. L.* SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality// Journal of Retailing.— 1988.— Vol. 64, No. 1.— P. 12-40.
48. *Davis F. D.* Perceived usefulness, perceived ease of use, and user acceptance of information technology// MIS Quarterly. — 1989.— Vol. 13, No. 3.— P. 319-340.
49. *Nielsen J.* Usability Engineering.— Cambridge, MA, London, UK: Academic Press, 1993.
50. *Rubin J., Chisnell D.* Handbook of Usability Testing. 2nd Ed.— Indianapolis, IN: Wiley, 2008.
51. *Gómez-Pérez A.* Evaluating ontology evaluation// IEEE Intelligent Systems.— 2004.— Vol. 19, No. 4.— P. 74-76.

# Ранжирование издателей по библиометрическим показателям: основные результаты и методологические проблемы при проведении ранжирований академических издателей\*

**Даниэль ТОРРЕС-САЛИНАС**  
(Daniel TORRES-SALINAS)

Университет Наварры, Гранада, Испания

**Николас РОБИНСОН-ГАРСИА**  
(Nicolás ROBINSON-GARCIA)

**Эваристо ХИМЕНЕС-КОНТРЕРАС**

(Evaristo JIMÉNEZ-CONTRERAS)

**Энрике де ла ФУЭНТЕ**  
(Enrique de la FUENTE)

Университет Гранады, Гранада, Испания

*Приводятся результаты проекта под названием Библиометрические показатели для издателей (также известного как BiPublishers). Проект представляет первую попытку систематически развивать библиометрические ранжирования издателей. Данные для этого проекта были взяты из Book Citation Index, исследование охватило 2009-2013 гг. Получены 42 ранжирования: 4 по областям и 38 – по дисциплинам. Приведены шесть показателей, касающихся издателя, которые разделены на три типа: выход продукции, влияние и профиль издателя. Цель – охватить различные характеристики научной работы издателей. Обработаны и классифицированы 254 издателя в соответствии с типом издателя: коммерческие издатели и университетские издательства. Представлены основные издатели по областям. Проведено обсуждение ряда важных имеющих проблем при разработке этого типа инструментов. Ранжирование BiPublishers является действующим проектом, направленным на развитие и изучение новых источников данных и показателей для лучшего охвата и определения научного влияния издателей.*

## Введение

В последние годы сделано много продвижений в отношении развития библиометрических баз данных, включающих книги и главы книг. Исторически эти типы документов не входили в библиометрический анализ [1], однако появление таких продуктов как Google Scholar, Google Books, The Book Citation Index или включение их в базы данных, например, Scopus, открыли широкую сферу возможностей для их анализа [2, 3]. Подобно ранжированиям журналов, первый шаг по включению книг и книжных глав в библиометрический набор средств возможно позволит развить ранжирование издателей. Уже имеется ряд инициатив, следующих данным намерениям [4]. В предыдущей статье мы пред-

ложили разработку ранжирований академических издателей [5], основанную на Book Citation Index. Эта статья строится на идее разработки академических ранжирований на основе Book Citation Index [5]. Здесь мы представляем результаты проекта *BiPublishers – Bibliometric Indicators for Publishers* [6], доступного на <http://publishers.es>. Эта инициатива направлена на разработку новых методологий и показателей, которые могут лучше охватывать и определять научное влияние издателей академических и научных книг. Это действующая инициатива, в которой тестируются источники данных и показатели. Следовательно, приведенная информация не должна использоваться в целях научной оценки. Мы рассматриваем академических издателей как аналогию журналов, фокусируясь на них как на единице анализа; подход уже был применен в ряде других областей [7]. Мы включили шесть показателей для более 100 издателей в четырех широких областях и 38 различных дисциплинах. Данные основаны на Thomson Reuters' Book Citation Index.

\* Перевод Torres-Salinas D., Robinson-García N., Jiménez-Contreras E., de la Fuente E. The BiPublishers ranking: Main results and methodological problems when constructing rankings of academic publishers. – <http://www.arxiv.org/ftp/arxiv/1505/1505.01074.pdf>

## Материал и методы

### Общее описание используемой базы данных: The Book Citation Index

Book Citation Index (BKCI) был запущен в 2011 г. с целью пролить свет на научное воздействие монографий. Он заполнил пробел, на который уже обратил внимание Гарфила [8], создатель исходного указателя Science Citation Index. The Thomson Reuters' Book Citation Index (BKCI) был введен в действие в 2011 г. Он обеспечивает большое множество данных, касающихся цитирования и публикации, по монографиям и книжным главам и включен в основной массив Web of Science (Web of Science Core Collection) в рамках платформы Web of Science.

BKCI охватывает научную литературу с 1999 г. и так же, как это происходит в случае с Science Citation Index, Social Sciences Citation Index и Arts and Humanities Citation Index, следует строгому процессу отбора, используя следующие основные критерии отбора [9]: 1) распространение (обращение) публикаций, 2) полная библиографическая информация для всех процитированных ссылок и 3) внедрение процесса рецензирования коллегами. Как новый продукт BKCI имеет важные ограничения, которые должны учитываться при анализе приведенных результатов. Здесь мы суммируем основные [3]:

- *Языковое предпочтение.* BKCI сильно склоняется в сторону англоговорящих стран, в настоящее время (ноябрь 2014 г.) 97,7% записей приводятся на этом языке.

- *Большая концентрация издателей.* Только три издателя (Springer, Palgrave и Routledge) представляют половину этой базы данных.

- *Дисперсия (рассеяние) ссылок.* Из-за различия между книгами и главами книг ссылки на них также рассматриваются в качестве независимых.

## Обработка данных и нормализация

Все приведенные результаты основаны на сетевой версии BKCI (апрель 2014 г.). Временной период охватывал 2009-2013 гг. За этот период было найдено 482 470 записей, распространенных в 14 различных типах документов (см. рис. 2 в работе [6]). Что касается построения областей, то это было сделано с помощью агрегации предметных категорий Web of Science, как представлено в BKCI. В отличие от того, что происходит с журналами, книги индивидуально приписываются к одной или более категориям, это означает, что один издатель может иметь (и обычно имеет) свой выход продукции, распределенный среди различных категорий. Агрегация предметных категорий для областей и дисциплин доступна на <http://bipublishers.es/wp-content/uploads/2014/10/5.FieldsandDisciplinesConstruction.xlsx>.

Для каждой записи мы обработали библиографические области. Область Издатель (Publisher) обработана отдельно и нормализована вручную. Мы определили 342 различных издателя, хотя в конечном итоге было обработано 254. Чтобы обеспечить надежные результаты, издатели должны удовлетворять, по крайней мере, одному из следующих критериев для включения в ранжирование: а) иметь минимум пять книг, заиндексированных во временной период исследования или б) иметь минимум 50 книжных глав, заиндексированных в период исследования. В процессе нормализации мы приняли в качестве критерия тот факт, что если издатель был приобретен другим издателем, тогда вся его продукция будет приписана последнему. Кроме того, мы приписали типы издателей, делая различие между двумя типами: 1) коммерческие и академические издатели и 2) университетские издатели.

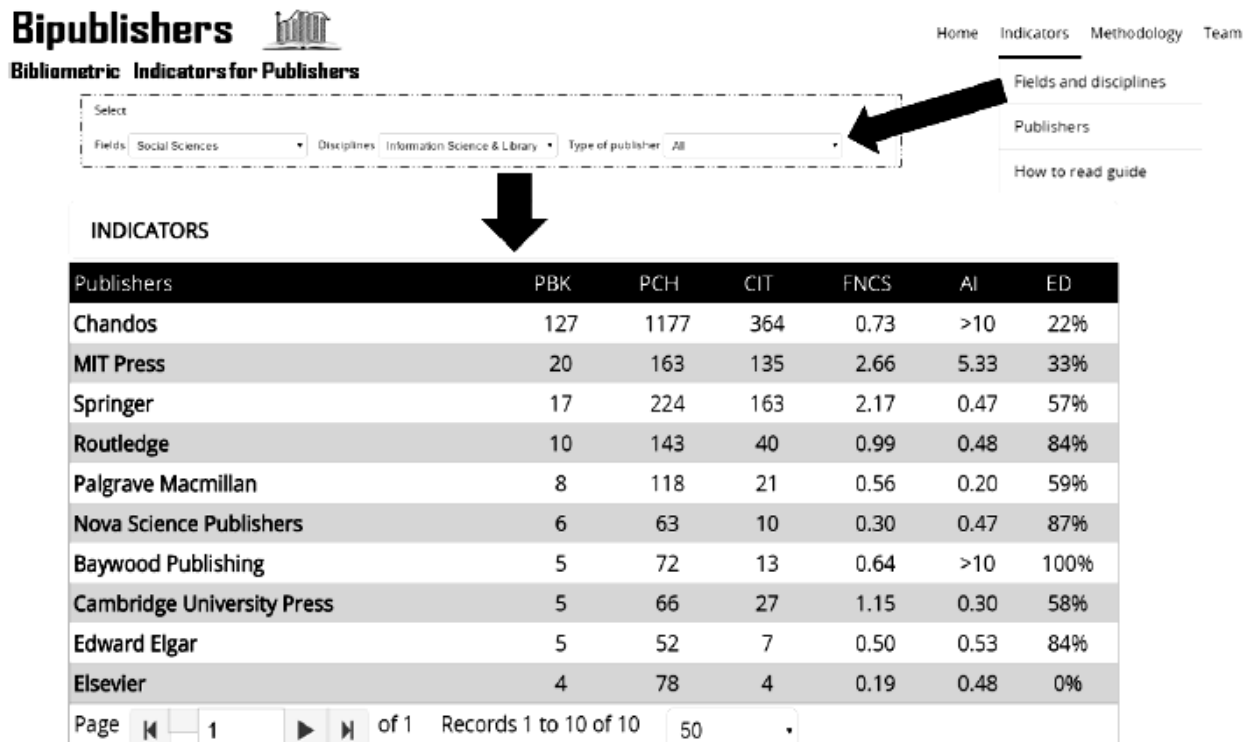


Рис. Моментальный снимок ранжирования для издателей в дисциплине «Информатика и библиотековедение»

## Результаты

### Краткое описание показателей и платформы web

Табл. 1 отражает шесть показателей, представленных для каждого издателя. Как видно, были отобраны три типа показателей, чтобы охватить различные аспекты научной деятельности издателей. Первый тип показателей отражает выход продукции издателя (РВК и РСН). Вторая группа концентрируется на показателях влияния, включая необработанное число полученных ссылок (СЦТ) и нормализованный показатель влияния (FNCS). И, наконец, третий тип показателей пытается охарактеризовать издателя. В этом случае мы включили индекс активности (АИ) и долю изданных глав из общего выхода продукции в определенной области (ЕD).

Эти шесть показателей приведены для каждого издателя, а также по областям и дисциплинам. Рисунок демонстрирует примеры того, как эти показатели представлены на сетевом сайте. Два основных входа доступны для обращения к ранжированиям. Первый – для просмотра по областям и дисциплинам. Здесь пользователи могут в первую очередь выбрать область, а затем дисциплину, в которой они хотят проконсультироваться, и отфильтровать результаты в зависимости от издателя. На рисунке мы показываем ранжирование для дисциплины «Информатика и библиотковедение», которая включена в область общественных наук. Как видно, ранжирования упорядочиваются по умолчанию по общему числу книг (РВК), однако они могут быть пересортированы пользователем с помощью клика на каждую рубрику.

Другая опция визуализации – непосредственно посмотреть на определенного издателя. Здесь пользователь может напрямую осуществлять поиск любого издателя, включенного в ранжирования. Страница профиля издателя показывает две метки (два контакта) на верху страницы. Первая метка (Данные) показывает основную

информацию об издателе (название и сетевой сайт). Метка нормализации показывает варианты названия (фамилии), обработанные и включенные под этим определенным издателем, вместе с городом и адресом, приписанными этому конкретному варианту. Под этими двумя метками все области и дисциплины, в которые включен издатель, показаны вместе со значениями шести показателей для каждой области или дисциплины. И снова результаты сортируются по умолчанию по числу книг.

### Основные характеристики ранжирований издателей по библиометрическим показателям

Всего 482 470 документов было обработано за временной период 2009-2013 гг. Мы определили 342 издателя. Из этого числа 254 издателя являются случаями для демонстрации. Мы создали 42 ранжирования: 4 ранжирования по широким областям и 38 ранжирований – по дисциплинам. Табл. 2 дает возможность посмотреть на то, как издатели, дисциплины, ссылки и документы распределяются среди этих четырех широких областей. Как видно, «Проектирование и технология» (Engineering and Technology) – это область, где располагается меньше дисциплин (приблизительно 4). Однако эта область обладает самым большим числом ссылок, показывая наивысшее среднее по числу цитирований на книгу (5,93).

Издатели с самым высоким числом книг, отраженных в ВЦИ, – это Springer (3 799 книг), Palgrave MacMillan (4 213 книг) и Routledge (2 176 книг). Из топ 20 наиболее продуктивных издателей в ВЦИ [6] только 7 представляют университетские издательства, тогда как остальные являются коммерческими издателями. Три самые продуктивные университетские издательства – это Cambridge University Press (1 775 книг), Princeton University Press (599 книг) и University of California Press (552 книги).

Таблица 1

### Определение показателей, приведенных по издателю

Показатели выхода продукции		
РВК	Общее число книг	Общее число книг, опубликованных данным издателем в определенной области или дисциплине за временной период исследования (2009-2013 гг.). Пороговый минимум
РСН	Общее число книжных глав	Общее число книжных глав, опубликованных данным издателем в определенной области или дисциплине за временной период исследования (2009-2013 гг.).
Показатели влияния		
СЦТ	Общее число цитирований	Общее число цитирований, полученных данным издателем в определенной области или дисциплине.
FNCS	Нормализованное значение цитирования по области	Нормализованное значение цитирований по области. Нормализованные цитирования, полученные в соответствии с нормализованным показателем, как определено в работе [10].
Профиль издателя		
АИ	Индекс активности	Распространение книг в определенной области или дисциплине в соответствии с общим выходом продукции данного издателя и относительно распространения всего ВКСИ
ЕD	Процент изданных документов	Доля книжных глав, принадлежащих к изданным книгам, из общего числа книжных глав, опубликованных данным издателем в определенной области или дисциплине за период исследования (2009-2013 гг.).



Общий обзор числа издателей, проанализированных по широким областям

Область	Количество дисциплин	Количество цитирований	Количество издателей	Количество документов	Среднее цитирование по типу документа
Гуманитарные науки и искусство	13	35918	Коммерческие –38	Книги: 8864	Книги 3,23
			Университетские – 41	Книжные главы: 87028	Книжные главы 0,08
			<b>Всего: 79</b>	<b>Всего:95892</b>	
Общественные науки	14	59609	Коммерческие –61	Книги: 10782	Книги 4,10
			Университетские – 23	Книжные главы: 114957	Книжные главы 0,13
			<b>Всего: 84</b>	<b>Всего:125739</b>	
Проектирование и технология	4	86324	Коммерческие –37	Книги: 2820	Книги 5,93
			Университетские – 38	Книжные главы: 33888	Книжные главы 0,35
			<b>Всего: 75</b>	<b>Всего:36708</b>	
Естественные науки			Коммерческие –35	Книги: 7757	Книги 5,44
	12	28591	Университетские – 2	Книжные главы: 109599	Книжные главы 0,40
			<b>Всего: 37</b>	<b>Всего:117316</b>	

Издатели равномерно распределены во всех областях, за исключением области «Естественные науки» (Science). Здесь меньшее число издателей (37), и все они, кроме двух, коммерческие. Также в области «Общественные науки» (Social Sciences) значительно больше коммерческих издателей (61), чем университетских (23). Что касается распределения типов документов, то книги в сфере «Искусство и гуманитарные науки» (Art and Humanities) имеют наименьшее среднее относительно книжных глав на книгу со значением 9,8, затем следуют «Общественные науки» (10,7). На другом конце располагаются «Естественные науки», показывающие среднее в 14,1 глав на книгу, а «Проектирование и технология» имеет среднее в 12,0.

#### **Релевантные издатели в ранжировании издателей по библиометрическим показателям**

В табл. 3 мы включили издателей топового (верхнего) ряда с самым большим числом книг (РВК) по области с их показателями работ. Как видно, имеются различия по большинству представленных издателей между областями «Естественные науки», «Проектирование и технология» и «Общественные и гуманитарные науки». В то время как издательства Palgrave Macmillan и Cambridge University Press присутствуют только в двух последних областях, издательства Elsevier и Nova Science Publishers присутствуют только в первой. С другой стороны, мы наблюдаем, что Springer присутствует во всех областях, однако индекс активности (AI) показывает низкие значения для областей «Гуманитарные науки и искусство» и «Общественные науки» (0,29 и 0,48 соответственно), в то время как он гораздо выше в области «Проектирование и технология» и «Естественные науки» (2,48 и 2,09 соответственно).

Что касается их влияния, то только два университетских издательства, включенные в топ 5 (Cambridge University Press - в гуманитарных науках и искусстве, общественных науках и естественных науках и Princeton University Press, представленное в гуманитарных науках и искусстве), всегда дают значения выше 1 в соответ-

ствии с их нормализованным влиянием цитирования (FNCS), освещая влияние своих изданий. Относительно коммерческих издателей, Springer и Elsevier являются единственными, которые демонстрируют значения выше 1, тогда как остальные показывают результаты ниже в соответствии с их FNCS.

#### **Методологические проблемы**

В этой статье мы описываем инициативу по созданию ранжирований для университетских и коммерческих издателей на основе данных цитирования. Выбранным источником данных был Book Citation Index. Книги и главы книг являются типами документов очень различного характера относительно тех документов, с которыми привыкли иметь дело специалисты по библиометрии [10]. Это выдвигает новые проблемы, отличающиеся от тех, которые были подняты при работе с журнальными публикациями. В этом разделе мы опишем основные проблемы, наблюдаемые в процессе развития ранжирования издателей.

##### *1. Варианты названий*

Thomson Reuters предоставляет основной список 499 издателей (<http://workinfo.com/mbl/publidhers>), после его анализа мы обнаружили много ошибок, заставивших нас выработать собственный процесс нормализации. Например, 15 вариантов названий были обнаружены в случае с Elsevier, также необходимо принять решения относительно того, как осуществляется такой процесс нормализации. В отличие от журналов издатели могут принадлежать к более крупным издательским корпорациям или могут иметь различные подразделения. Следует учитывать, что если ранжирование издателей должно включать все подразделения отдельного издателя, то надо отстаивать в качестве отдельных издателей тех, кто относится к той же самой корпорации, или нормализовать их до наивысшего найденного уровня. Здесь мы сделаем выбор в пользу этой последней опции; однако рациональность следования выбору одной или другой опции находится под вопросом, независимо от того, какая опция выбрана.

**Релевантные издатели и их показатели, основанные на четырех широких областях в ранжировании издателей по библиометрическим показателям издателей**

<b>Гуманитарные науки и искусство Издатель</b>	<b>PBK</b>	<b>PCH</b>	<b>CIT</b>	<b>FNCS</b>	<b>AI</b>	<b>ED</b>
Palgrave Macmillan	2108	19554	5772	0,68	1,42	38%
Cambridge University Press	1004	8167	4624	1,48	1,63	45%
Routledge	748	8303	3128	0,82	0,98	40%
Springer	383	4725	2418	1,12	0,29	59%
Princeton University Press	339	3022	3534	2,57	1,61	24%
<b>Общественные науки Издатель</b>	<b>PBK</b>	<b>PCH</b>	<b>CIT</b>	<b>FNCS</b>	<b>AI</b>	<b>ED</b>
Palgrave Macmillan	2680	26823	9249	0,68	1,49	42%
Routledge	1540	17427	9077	0,93	1,65	45%
Edward Elgar	814	10208	3434	0,84	1,91	62%
Springer	787	9779	6734	1,17	0,48	67%
Cambridge University Press	513	418	4373	2,05	0,68	49%
<b>Проектирование и технология Издатель</b>	<b>PBK</b>	<b>PCH</b>	<b>CIT</b>	<b>FNCS</b>	<b>AI</b>	<b>ED</b>
Springer	1054	12139	14831	129	2,48	53%
Elsevier	387	5238	3943	1,16	3,92	27%
Nova Science Publishers	267	2665	954	0,28	1,79	77%
Woodhead Publishing	192	2482	878	0,43	8,02	73%
Artech House	142	1759	676	0,55	8,08	30%
<b>Естественные науки Издатель</b>	<b>PBK</b>	<b>PCH</b>	<b>CIT</b>	<b>FNCS</b>	<b>AI</b>	<b>ED</b>
Springer	2446	40396	37013	1,17	2,09	73%
Nova Science Publishers	961	10711	3079	0,24	2,34	77%
Elsevier	538	10711	7787	1,28	1,98	56%
Cambridge University Press	417	5036	5728	1,61	0,77	44%
Routledge	361	4964	1948	0,47	0,54	43%

### 2. Кластеры издателей и корпорации

Следуя примеру Elsevier и придерживаясь описанных выше критериев, мы включили в рамки этой корпорации таких издателей как Pergamon, Academic Press или North Holland, все они принадлежат Elsevier. Поскольку рынок издателей очень нестабильный и подвергается постоянным изменениям, то такие изменения угрожают стабильности ранжирований и сравнений между современными данными. Последнее изменение непосредственно влияет на самых крупных издателей, включенных в Book Citation Index: MacMillan и Springer, которые недавно объединились [11].

Другим обнаруженным нами примером был случай издательства Willan Publ, приобретенного Taylor & Francis. Более сложно принять тип решений, когда продажа осуществлена в период проведения исследования. Так обстоит дело с издательством AK Peters, которое было приобретено CRC Press в 2010 г. Наконец, следует отметить, что этот вопрос представляет серьезные проблемы, так как не всегда ясна связь с зависимостью.

### 3. Конструкция областей и дисциплин

Как уже упоминалось ранее, конструкция областей и дисциплин осуществлялась путем агрегации предметных категорий из Book Citation Index. Это относительно общая практика в библиометрических исследованиях при работе с журнальными публикациями. В таком случае журналы приписываются к одной или более категориям. Следуя такому ходу мысли, можно предложить,

что издатели должны приписываться к категориям. Однако следуя более обоснованному (но также менее прозрачному) подходу, каждая книга приписывается к одной или более категориям. Было бы интересно подробнее узнать, в соответствии с какими критериями Book Citation Index классифицирует книги. Кроме того, в этой статье агрегация может быть сомнительной, поэтому мы придаем большое значение потребности в дальнейшем изучении альтернатив.

### 4. Типы публикаций: продолжающиеся издания в противоположность книгам

Серьезным ограничением Book Citation Index является включение продолжающихся изданий, таких как труды в базе данных [12]. Чтобы использовать эту базу данных для библиометрических целей, этот тип продукции должен быть удален до проведения анализа. С учетом этого все документы, помеченные как продолжающиеся издания, были удалены из нашего массива данных, т.е. документы, принадлежащие издательству Annual Reviews (как предложено в [12]).

### 5. Охват издателя

Важным ограничением при анализе выхода продукции издателей в Book Citation Index является то, что мы не знаем, какова степень ее охвата издателем. Включает ли она все книги, опубликованные издателем? Индексируют ли они только некоторые из них? После беглого просмотра кажется, что эта последняя опция является наиболее вероятной. Однако для подтверждения необходимо дальнейшее исследование.

## Заклучение (выводы и дальнейшее развитие)

В данной статье мы представляем первые результаты проекта Библиометрические показатели для издателей (также известного как ViPublishers). Проект предполагает проанализировать возможность развития библиометрических показателей для научных и академических издателей и является первым библиометрическим ранжированием таких характеристик. Это продолжающийся в настоящее время проект на основе данных из Book Citation Index. Это означает, что обнаруженные результаты наследуют все недостатки базы данных. Среди других ограничений мы придаем большее значение предпочтению в пользу английского языка и концентрации издателей. Мы обсуждаем основные проблемы, которые влечет за собой развитие библиометрического ранжирования для издателей, такие как нормализация названий издательств, имеющих место при объединении, конструкция областей и ранжирований, исключение определенных видов публикаций, включенных в Book Citation Index, а также неопределенности, касающиеся охвата издателем этой базы данных.

Чтобы проанализировать обоснованность наших результатов, а также исследовать другие массивы данных, мы надеемся в будущем включить другие источники данных (т.е. Scopus), а также разработать и включить новые библиометрические показатели, которые смогут лучше охватить другие характеристики издателей. Например, мы предлагаем проанализировать роль книжных серий в рамках издательств. В заключение мы полагаем, что появление в последние годы новых баз данных цитирования, включающих книги и книжные главы, должно вдохновить библиометрическое сообщество на углубление в новые дела, чтобы проанализировать научное влияние этих долгое время не принимаемых во внимание типов документов.

## Литература

1. *Nederhof A.J.* Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A review// *Scientometrics*. – 2006. – Vol. 66, No. 1. – P. 81-100.
2. *Kousha K., Thelwall M., Retzai S.* Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus// *Journal of the American Society for Information Science and Technology*. – 2011. – Vol. 62, No. 11. – P. 2147-2164.
3. *Torres-Salinas D., Robinson-Garcia N., Campanario J.M., Delgado López-Cózar E.* Coverage, field specialisation

and the impact of scientific publishers indexed in the Book Citation Index// *Online Information Review*. – 2014. – Vol. 38, No. 1. – P.24-42.

4. *Research School for Socio-Economic and Natural Sciences of the Environment, 2009.* SENSE Ranking of Academic Publishers. – [http://www.sence.nl/gfx\\_content/documents/ABCDE-indeling%20Scientific%20Publishers%20SENCE\\_approved\\_may\\_2009.pdf](http://www.sence.nl/gfx_content/documents/ABCDE-indeling%20Scientific%20Publishers%20SENCE_approved_may_2009.pdf) [March, 2015]

5. *Torres-Salinas D., Robinson-Garcia N., Jiménez-Contreras E., Delgado López-Cózar E.* Towards a Book Publishers Reports. First approach using the Book Citation Index// *Revista Espanola de Documentacion Cientifica*. – 2012. – Vol. 35, No. 4. – P. 615-620.

6. *Robinson-Garcia N., Jiménez-Contreras E., Fuente-Gutiérrez E., Torres-Salinas D.* Bibliometric indicators for publishers data processing, indicators and interpretation, 2014. – <http://arxiv.org/ftp/arxiv/papers/1411/1411/1361/pdf> [March 24, 2015].

7. *Torres-Salinas D., Moed H.F.* Library catalog analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics// *Journal of Informetrics*. – 2009. – Vol. 3, No. 1. – P. 9-26.

8. *Garfield E.* Citation indexes for retrieval and research evaluation/ *Consensus Conference on the Theory and Practice of research Assessment*, 7 October, Capri, Italy. – 1996. – <http://www.garfield.library.upenn.edu/papers/siretreseval-capri.html> [March 24, 2015].

9. *Testa J.* The book selection process for the Book Citation Index in Web of Science, 2010. – [http://workinfo.com/media/pdf/bkci-SelectionEssay\\_webpdf](http://workinfo.com/media/pdf/bkci-SelectionEssay_webpdf) [March 24, 2015].

10. *Moed H.F., de Bruin R.E., van Leenven T.N.* New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications.// *Scientometrics*. -1995. – Vol. 33, No. 3. – P. 381-422.

11. *Zuccala A., Guns R., Cornacchia R., Bod R.* Can we rank scholarly book publishers? A bibliometric experiment with the field of history// *Journal of the Association for Information Science and Technology*. – 2014. – doi:10.1002/asi.23267.

12. *Schweizer K.* Springer Science, MacMillan unit to merge into publishing giant .- Bloomberg, 2015. – <http://www.bloomberg.com/news/2015-01-15/springer-science-macmillan-unit-to-merge-into-publishing-giant.html> [March 24, 2015].

# Когда статья фактически опубликована? Анализ дат онлайн доступности, публикации и индексации\*

**Стефани ХОСТЕЙН**  
(Stefanie HAUSTEIN)

**Тимоти Д. БАУМАН**  
(Timothy D. BOWMAN)

Школа библиотековедения и информатики,  
Монреальский университет,  
г. Монреаль, Канада

**Родриго КОСТАС**  
(Rodrigo COSTAS)

Центр исследований науки и технологий,  
Лейденский университет, г. Лейден,  
Нидерланды

*С ускорением научной коммуникации в цифровую эру год публикации больше не является достаточным уровнем агрегации времени для библиометрических показателей и показателей социальных медиа. Статьи все больше цитируются до того, как они были официально опубликованы в выпуске журнала и упомянуты в Твиттере в течение дней онлайн доступности. В целях найти подходящую переменную для дня онлайн публикации, позволяющую вычисление более надежных критериев и точных интервалов цитирования и событий социальных медиа, сравниваются различные даты для массива из 58 896 статей, опубликованных Nature Publishing Group, PLOS, Springer и Wiley-Blackwell в 2012 г. Данные включают дату онлайн доступности, предоставленную издателями, месяц выпуска журнала, дату индексации в Web of Science, дату первого упоминания статьи в твите, а также публикацию в Altmetric.com и даты первых просмотров. Сравнивая эти даты, анализ показывает, что между издателями существуют большие различия, это приводит к заключению о необходимости большей прозрачности и стандартизации в представлении годов издания. Дата, по которой фиксированная журнальная статья (версия записи) впервые становится доступной на веб-сайте издателя, предлагается как согласованное определение даты онлайн доступности.*

## Введение

Процесс научной коммуникации, который обычно начинается с формулировки идеи исследования и гипотезы и заканчивается публикацией результатов для их распространения в научном сообществе [1], был ускорен посредством электронного издательства [2, 3]. Задержка публикации, которую Амет [4, с. 382] определил как “хронологическое расстояние между зафиксированной датой приема рукописи определенным журналом и ее появлением в любом печатном выпуске этого журнала”, усилилась с помощью электронной почты и систем обработки рукописи онлайн, а также онлайн публикацией [3]. Период задержки состоит из процесса рецензирования, который составляет главную задержку и завершается принятием рукописи, сопровождаемым техническими задержками производства журнала и портфелем заказов.

Различные исследования проанализировали задержки публикации и нашли различия между научными областями, журналами и издателями [например, 2, 4, 5-8].

Так как длительные задержки сталкиваются с проблемами приоритета и замедляют научный дискурс, скорость публикации играет важную роль для авторов и научной коммуникации [9-11]. Поэтому короткие задержки публикации можно рассматривать как качественный индикатор, отражающий уровень современных требований научных журналов [12]. Издатели начали сокращать задержки, делая так называемые версии принятых статей (первое представление, в печати, до печати или сначала онлайн) доступными до того, как они появляются в (печатном) выпуске. Для научных журналов о пищевых продуктах это показывает, что онлайн публикация до ее печатной версии сокращает задержку на 29 % [4], в то время как Дас и Дас [7] сообщили, что для 127 журналов в 2005 г. задержки в среднем составили три месяца между онлайн и печатным выпусками изданий со специфическими различиями между издателями. Торт, Тарджино и Амарал [13] показали, что со временем эта задержка значительно увеличилась для шести журналов по неврологии. Онлайн даты теперь записываются в таких библиометрических базах данных, как Scopus, которая влияет на библиометрические анализы [14-15]. Вместе с ростом популярности серверов препринтов (таких как arXiv и SSRN) и учрежденческих хранилищ, такие версии, как в печати, помогли ускорить цикл чтение-ссылка-чтение. В

\* Перевод Haustein S., Bowman T.D., Costas R. When is an article actually published? An analysis of online availability, publication, and indexation dates. — 2015.—  
<http://www.arxiv.org/ftp/arxiv/papers/1505/1505.00796.pdf>

результате рукописи все больше цитируют статьи, которые не были официально опубликованы в выпуске журнала. Хотя научная коммуникация всегда включала в себя обмен различными версиями рукописи между коллегами до, во время и после формальной публикации – например, обмен черновыми вариантами в целях обратной связи до представления или распространение препринтов после их принятия, – электронная эра делает эти версии “публичными”, доступными для поиска и (часто) постоянно находящимися в сети. Для определения и осуществления различия между множеством версий Национальная организация по стандартизации в области информации (NISO, США) согласовала следующие версии журнальной статьи [16]:

- Авторский оригинал (АО) - рукопись, готовая для представления.
- Представленная на рассмотрение версия (SMUR) – рукопись, находящаяся на формальном рецензировании.
- Принятая рукопись (AM) - версия журнальной статьи, принятой для публикации.
- Корректурa (P) - отредактированная версия принятой статьи.
- Версия записи (VoR) - зафиксированная версия журнальной статьи, формально изданная.
- Исправленная версия записи (CVoR) - VoR, в которой исправлены ошибки.
- Улучшенная версия записи (EVoR) - VoR, обновленная или улучшенная дополнительным материалом.

Важно отметить, что по определению NISO, VoR представляется как «зафиксированная версия журнальной статьи, ставшая доступной с помощью любой организации, которая действует как издатель, формально и эксклюзивно объявляя статью “изданной”» [16, с. 3]. Это определение включает «первые представления» и «статьи в печати» без информации относительно тома и номера или других идентификаторов, пока содержание и верстка статьи не устанавливаются окончательно.

Когда дело касается библиометрических показателей, то ускорение процесса публикации отражается как в моделях устаревания [17], так и полураспадов цитирования [18]. Эти увеличивающиеся задержки от онлайн до печати были показаны, чтобы искусственно увеличить темпы цитирования, включая индекс непосредственности и импакт фактор [13,15,19]. Скорость научной коммуникации становится особенно заметной в контексте метрики социальных медиа (так называемой альтметрии); например, упоминания научных документов в Твиттере происходят в течение часов (а иногда в течение минут) онлайн доступности [20].

Мы утверждаем, что в стремительно развивающуюся цифровую эру использование *года* публикации журнального выпуска как наименьшего уровня агрегации времени для библиометрических показателей становится недостаточным, особенно в контекстах оценки исследования, из-за следующих факторов:

- а) ускорение цикла чтение-ссылка–чтение благодаря электронной публикации,
- б) общепризнанность публикации онлайн до публикации выпуска журнала,
- в) увеличение задержек от версии онлайн до печати.

Следуя терминологии NISO, мы полагаем, что дата первого публичного появления онлайн VoR является наиболее релевантной и должна использоваться как основная единица времени для установления официальной даты публикации статьи. Это позволило бы создать более точные интервалы ссылок и событий социальных медиа, например, интервалы ссылок равной

длины (в днях или месяцах) для статей, опубликованных в январе или декабре, а также способствовало бы образованию более точных критериев с помощью соединения ссылок и событий социальных медиа за неделю (например, долей твитов и Фейсбука) или месяца (темпы ссылок) в зависимости от контекста оценки.

Хотя многие издатели теперь сообщают даты онлайн публикации, имеется много различных дат и предоставленная информация изменяется среди издателей, так как не существует никаких официальных стандартов относительно дат публикации. Эта статья исследует и нацелена на проверку различных дат “публикации”, чтобы найти хорошую переменную для фактической даты онлайн доступности. Таким образом, статья пытается ответить на следующие вопросы исследования:

1. Какие издатели определяют онлайн даты, и как они их предоставляют?
2. Насколько надежны даты, предоставленные издателями, и как они сравнивают их друг с другом?
3. Какие другие существующие даты могут использоваться в качестве переменной даты онлайн издания VoR?

## Методы и материалы

Массив данных этого исследования был извлечен из Web of Science (WoS) как главной базы данных цитирования, и определен 2012 годом издания, чтобы ограничить эффекты изменений со временем. Для подтверждения дат публикации, предоставленных издателями, даты первого упоминания в Твиттере конкретной статьи были получены от Altmetric.com. Мы утверждаем, что твит не может указывать на статью до ее существования, таким образом первый твит не может появиться до даты публикации онлайн. Твиты, охваченные Altmetric.com, указывают на документы с помощью DOI, выдавая в результате 313 301 статью WoS за 2012 г., по крайней мере, с одним событием, отмечаемым Altmetric.com [21]. Записи Altmetric, содержащие идентификаторы (ID) arXiv или системы данных по астрофизике (ADS), были удалены, чтобы исключить твиты к препринтам, которые, возможно, могли стать публичными до онлайн публикации VoR. Упоминания Твиттера таким образом ограничиваются упоминаниям или ссылками на веб-сайт издателя, DOI или PubMed ID.

10 ведущих издательств\* статей в массиве данных WoS-Altmetric можно найти в табл. 1 вместе с информацией о дате, предоставленной через программный интерфейс приложения, в метаданных, только в веб-странице или как динамическое содержание веб-страницы. Можно заметить (в заголовках табл.), что существуют многочисленные термины, чтобы описать дату публикации онлайн, и что многие типы дат стали доступными на веб-сайте, в метаданных или через программный интерфейс приложения; они включают полученную, пересмотренную, принятую, версию записи, онлайн, публикацию и дату. На основе выборочной проверки статей каждого издателя мы предполагаем, что даты, предоставленные как *версия записи, онлайн, публикация* и *дата* (табл. 1) относятся к (первым) появлениям онлайн VoR, требуемой для этого исследования. Издательства Wiley-Blackwell, Springer, PLOS и Nature Publishing Group (NPG) были выбраны из-за их охвата и технической выполнимости поиска информации по онлайн

\* Названия издательств из WoS были уточнены путем поиска вариантов, но слияния и приобретения не учитывались. Например, BMC рассматривается и является независимым издательством, хотя оно было приобретено Springer в 2008 г.

дате. Несмотря на то, что Elsevier – наиболее представительное издательство в этой выборке, было трудно получить необходимую информацию о дате для его статей, используя язык программирования PHP, потому что эта информация динамически вставлена в веб-страницу, использующую JavaScript; Elsevier предлагает программный интерфейс приложения, но при обращении к нему он обеспечивал доступ только к дате издания, а не к датам публикации онлайн, требуемым для этого исследования.

Используя DOI, мы обращались к сетевым платформам авторитетных издательств для поиска онлайн даты. PLOS, Springer и NPG – предлагают программный интерфейс приложения, но было обнаружено, что в некоторых случаях дополнительная информация о дате была доступна только при помощи поиска веб-страницы. Чтобы получить даты для Wiley, Springer и NPG, был написан скрипт PHP, который находил HTML страницы. Потом искался HTML для метаданных, содержащих информацию о дате (например <meta name = "prism.publicationDate" content = "2012-01-05"/>). Когда информация о дате была найдена, она сохранялась в реляционной базе данных для оценки. В случаях, когда веб-сайт статьи не имел (или пропускал) доступных метаданных, HTML анализировался синтаксически, и содержание определенных тегов HTML, по-видимому хранящих информацию о дате, извлекалось и сохранялось в реляционной базе данных; что касается статей Wiley, был написан второй скрипт, чтобы загрузить даты, не найденные в метаданных.

Чтобы сравнить различные доступные даты и проверить, насколько они могут использоваться как переменные для дат онлайн публикации, другая информация о дате была получена из WoS и Altmetric, чтобы вместе с информацией от издателей были доступны следующие даты:

- *Онлайн дата*: загружена с веб-сайтов издателей как часть метаданных статьи. Для NPG (“новая онлайн публикация”<sup>\*\*\*</sup>), Springer (“сначала онлайн”<sup>\*\*\*\*</sup>) и Wiley-Blackwell (“раннее представление”<sup>\*\*\*\*\*</sup>) эта дата отмечает, когда VoR стала публично доступной на веб-сайте издателя. Для PLOS онлайн дата соответствует дате публикации, поскольку нет никакого различия между датами онлайн и выпуска.

- *Дата выпуска журнала*: дата из выпуска журнала, как зарегистрировано WoS. Так как только меньшинство статей предоставили день месяца, дата выпуска журнала была преобразована в первое число каждого месяца. Из всех 1,3 млн. статей в WoS, изданных в 2012 г., 3,2% были опубликованы в выпусках, охватывающих несколько месяцев (таких как январь-февраль для двоекных выпусков). Они отражались как первый день первого месяца. Небольшая доля (0,5%) статей появлялась в сезонных выпусках (весна, лето, осень, зима). Так как дата указывает, что они изданы в начале, в середине, а также в конце определенного сезона, то эти даты игнорировались. Дополнительные 11,3% из всех статей 2012 г. не представили никакой даты выпуска. Рис. 1 дает обзор распределения 1,3 млн. статей WoS за 2012 г. согласно информации о дате выпуска журнала.

Таблица 1

**10 ведущих издателей по числу статей с типами доступных дат согласно данным, предоставленным издателем через программный интерфейс приложения (a), в метаданных (m) веб-страницы, только на веб-странице (w) или только как динамическое содержание (d). Издатели, отобранные для этого исследования, выделены серым цветом.**

Издатель	Статьи	Полученная	Пересмотренная	Принятая	Версия записи	Онлайн	Публикация	Дата	Выпуск журнала	Онлайн выпуск журнала
Elsevier	51 292	d	d	d		d	a		w	
Wiley-Blackwell	47 958	w		w		m, w <sup>i</sup>	m		w, m	w
Lippincott	21944							m	w, m	
Springer	19225					m	m, a	m	w, m, a	
PLOS	16 208	w		w			a, m		a, m	
BMC	11 930	w		w			w, m		w, m	
NPG	11 181	w, m		w, m		m, a	w, m, a		w, m, a	
ACS	11 024							m, w	w	
Oxford	10 368	w		w		w		m	w, m	
Sage	8 776				w	w		m	w, m	

<sup>i</sup>Wiley предоставляет две даты онлайн - “статьи, изданной онлайн”, а также “онлайн дату”. См. объяснения ниже.

\* Использование запроса программного интерфейса приложения <http://api.elsevier.com/content/abstract/doi/{doi}>

\*\* [http://www.nature.com/authors/author\\_resources/about\\_aop.html](http://www.nature.com/authors/author_resources/about_aop.html)

\*\*\* <http://www.springer.com/authors/journal+authors/helpdesk?SGWID=0-1723213-12-817311-0>

\*\*\*\* <http://olabout.wiley.com/WileyCDA/Section/id-404512.html#ev>

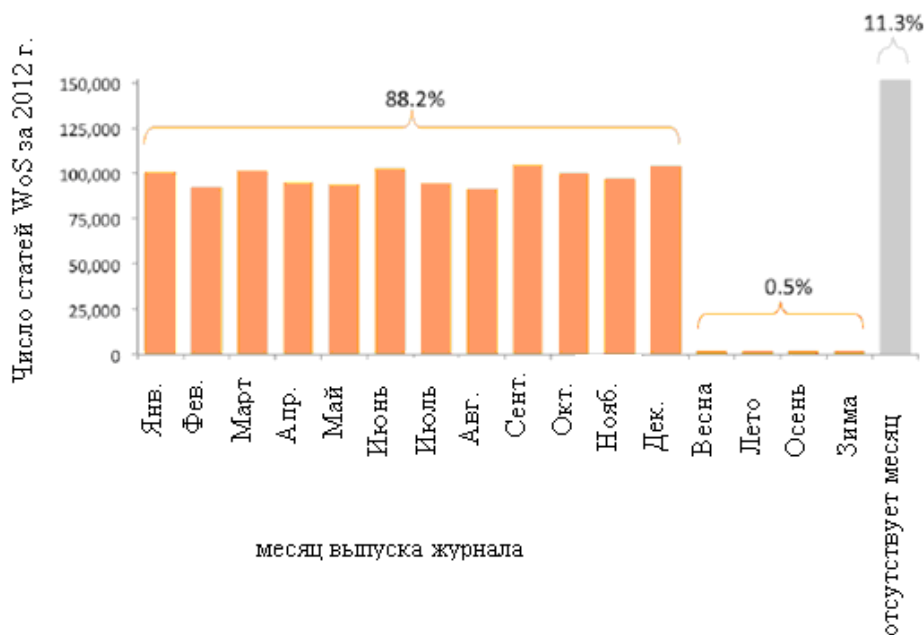


Рис.1. Число статей WoS за 2012 г. по месяцам выпуска журнала.

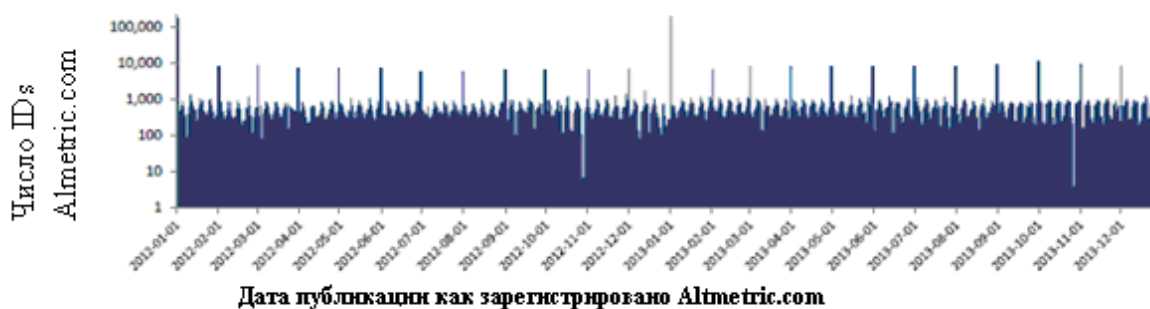


Рис. 2. Число Altmetric.com IDs на дату публикации в Altmetric.com с января 2013 г. по декабрь 2014 г.

- *Дата публикации в Altmetric*: дата публикации, как зарегистрировано Altmetric.com, которая является объединением даты выпуска журнала и онлайн даты, как найдено у издателя. Это также дата, которую использует Altmetric.com, чтобы вычислить оценку Altmetric и обеспечить критерии для статей того же самого возраста. Как показано на рис. 2, отдельные пики могут отмечаться как на 1 января каждого года, так и на первое или последнее число каждого месяца. Это может отражать общую практику издания, но также может быть вызвано соединением данных без фактической информации о дне (и месяце). Обнаружилось, что 15,1% записей\* Altmetric.com не имели никакой даты публикации или имели неправильные даты (например, даты вплоть до 2037 г.).

- *Первая замеченная Altmetric дата*: установление даты, когда Altmetric.com фиксировало первый случай для отдельного документа, который отсутствует для 4% всех записей\*\*.

- *Первая дата твита*: установление даты первого твита\*\*, зафиксированного Altmetric.com (за исключением всех статей со ссылками на arXiv ID или ADS ID, чтобы гарантировать, что твит не относится к препринту).

- *Дата индексации в WoS*: день, когда документ был внесен в WoS, который для статей за 2012 г. был по большей части в течение месяца (37,7%), месяцем ранее (11,5%) или месяцем позже (29,4%) выпуска журнала.

В дополнение к выше указанным датам мы также были в состоянии загрузить следующую информацию для статей, опубликованных Wiley-Blackwell:

- *Полученная рукопись*: дата, когда АО была представлена.

- *Принятая рукопись*: дата, когда АМ была принята.

- *Статья, сначала изданная онлайн*: мы не смогли определить точное значение этой даты; для 95,6% из всех 34 507 документов Wiley-Blackwell она была идентична *дате онлайн*, а для 1,6% отсутствовала. Для 2,3% статей *дата статьи, сначала изданной онлайн*, встречалась до *даты онлайн*, в среднем, за 35 дней, что предполагает, что она отмечает

\* На основе 2,1 млн. записей Altmetric.com, собранных в августе 2014 г.

\*\* На основе 2,1 млн. записей Altmetric.com, собранных в августе 2014 г.

\*\*\* Твиттер – самый общий источник, охваченный Altmetric.com [22], таким образом, это имеет смысл для того, чтобы работать с этой датой, а не взятой из других менее общих источников (например, Фейсбук или блоги).

публикацию AM. Однако в 137 случаях (0,4%), она следовала за *датой онлайн*, что в среднем составляло 52 дня.

Окончательный массив данных, т. е. соответствие WoS, Altmetric.com и статей датам онлайн, найденным у четырех издательств, включал 71 175 статей. Для лучшего сравнения он был ограничен статьями, для которых все пять дат, проверенных в качестве переменных для онлайн публикации (т. е. выпуск журнала, публикация Altmetric и первая замеченная дата, первый твит и дата индексации в WoS), были доступными. Это составляло в общей сложности 58 896 статей, 12,5% в NPG, 16,3% в PLOS, 24,6 % в Springer и 46,6% в Wiley-Blackwell.

## Результаты и обсуждение

Описательные статистические данные, сравнивающие дату онлайн с пятью потенциальными переменными, представлены в табл. 2, показывающей специфические различия для этих четырех издательств. Основываясь на предположении, что онлайн даты, предоставленные издательствами, были правильными, дата публикации в Altmetric, первая замеченная дата, а также дата первого твита, кажется, должны быть лучшими переменными для онлайн публикации, в то время как даты выпуска журнала и индексации в WoS показывают наибольшие отклонения от дат онлайн публикации.

Таблица 2

**Статистические данные для хронологического расстояния (в количестве дней) месяца выпуска журнала, публикации Altmetric и первой замеченной даты, даты первого твита и даты индексации в WoS с онлайн датами для NPG, PLOS, Springer и Wiley-Blackwell.**

Хронологическое расстояние до даты онлайн в количестве дней		NPG <i>n</i> =7391	PLOS <i>n</i> =9600	Springer <i>n</i> =14473	Wiley-Blackwell <i>n</i> =27432
Месяц выпуска журнала <sup>i</sup>	% до	20,04	n/a <sup>ii</sup>	3,47	9,09
	% идентичные	5,47		0,11	0,29
	% после	74,50		96,42	90,62
	среднее	61		146	84
	стандартное отклонение	78		111	93
	минимум	-330		-269	-423
	максимум	548		1850	1032
Дата публикации Altmetric	% до	55,38	66,49	43,37	63,83
	% идентичные	39,35	31,41	34,11	2,81
	% после	5,28	4,44	22,52	33,36
	среднее	12	27	9	121
	стандартное отклонение	68	79	48	322
	минимум	-3013	-697	-519	-16761
	максимум	411	526	1850	5016
Первая замеченная дата Altmetric	% до	3,48	0,00	0,08	14,59
	% идентичные	32,88	36,64	1,04	14,26
	% после	63,64	63,36	98,89	71,15
	среднее	35	12	90	63
	стандартное отклонение	87	49	164	122
	минимум	-459	0	-257	-533
	максимум	890	602	1843	1228
Дата первого твита	% до	3,52	0,00	0,08	14,52
	% идентичные	34,37	37,23	1,06	15,21
	% после	62,21	62,77	98,85	70,27
	среднее	37	15	92	65
	стандартное отклонение	92	59	169	127
	минимум	-459	0	-257	-533
	максимум	890	811	1843	1393
Дата индексации WoS	% до	2,72	0,00	0,10	0,05
	% идентичные	0,01	0,00	0,00	0,00
	% после	97,27	100,00	99,90	99,95
	среднее	83	39	163	97
	стандартное отклонение	81	20	113	94
	минимум	-302	9	-252	-359
	максимум	576	262	1866	1049

i –первый месяц выпуска журнала, как зарегистрировано WoS.

ii – PLOS не различает онлайн дату и дату выпуска, поэтому эти две даты фактически идентичны.



Эти различия отражают природу этих дат. Например, Altmetric собирает свои даты публикации с веб-сайтов издательств, в то время как первые твиты, как известно, случаются вскоре после публикации [20], обработка WoS отнимает больше времени, а именно, в среднем между 39 днями для PLOS или 163 днями для статей Springer. 61 день (NPG), 84 дня (Wiley-Blackwell) и 146 дней (Springer) между датами онлайн и датами журнальных выпусков главным образом отражают разницу между онлайн доступностью и публикацией выпуска журнала. Хотя (печатный) выпуск, как обычно предполагают, следует за публикацией онлайн хронологически, результаты табл. 2 показывают, что для проанализированных 3,47% статей Springer, 9,09% статей Wiley-Blackwell и 20,04% статей NPG онлайн дата идет после даты выпуска журнала, что считается негативной задержкой [7].

Хотя даты Altmetric и Твиттера работают лучше, чем выпуск журнала и индексация в WoS, ни одна из дат, кажется, недостаточно хорошо отражает дату онлайн, и могут отмечаться большие различия между издателями, в частности для Wiley-Blackwell, что ставит под сомнение обоснованность любых из пяти дат, как надежной переменной публикации VoR среди издателей. Дата публикации Altmetric, которая в целом показывает наименьшее различие по сравнению с датой онлайн, представленной издателями, – в среднем 9 дней для Springer, 12 дней для NPG, 27 для PLOS и 121 для Wiley-Blackwell, также проблематична, потому что она устанавливает

дату до публикации онлайн в 43,37% статей Springer, 55,38% статей NPG, 63,83% статей Wiley-Blackwell и 66,49% статей PLOS. Разница между издателями влияет на оценки Altmetric (а, возможно, также на оценки ссылок) при определении критериев оценок статей в противоположность критериям статей одного и того же сообщаемого возраста.

Основываясь на предположении, что твит не может упомянуть статью до того, как она появится в онлайн пространстве, на которое он ссылается, онлайн даты, представленные Wiley-Blackwell, кажутся наиболее проблематичными (рис. 3), поскольку 14,52%\* из 27 432 проанализированных статей имели твиты, указывающие на них до даты, которую издатель определяет как дату онлайн публикации. С другой стороны, ни одна из статей PLOS и несколько статей из Springer (0,08%) не были упомянуты в Твиттере до даты публикации онлайн. Хотя все проанализированные статьи получили упоминания в Твиттере, среднее число дней между датой онлайн и первым твитом было выше, чем ожидалось, варьируясь от 15 дней для PLOS до 92 дней для Springer. Кроме того, первое упоминание в Твиттере случилось в день онлайн публикации для 1,06% (Springer) и 34,47% (NPG) выборочных статей, что, в частности, с учетом приблизительно 80% последних статей, которые никогда не упоминаются в Твиттере [21], ограничивает полезность даты первого твита как переменной для онлайн публикации.

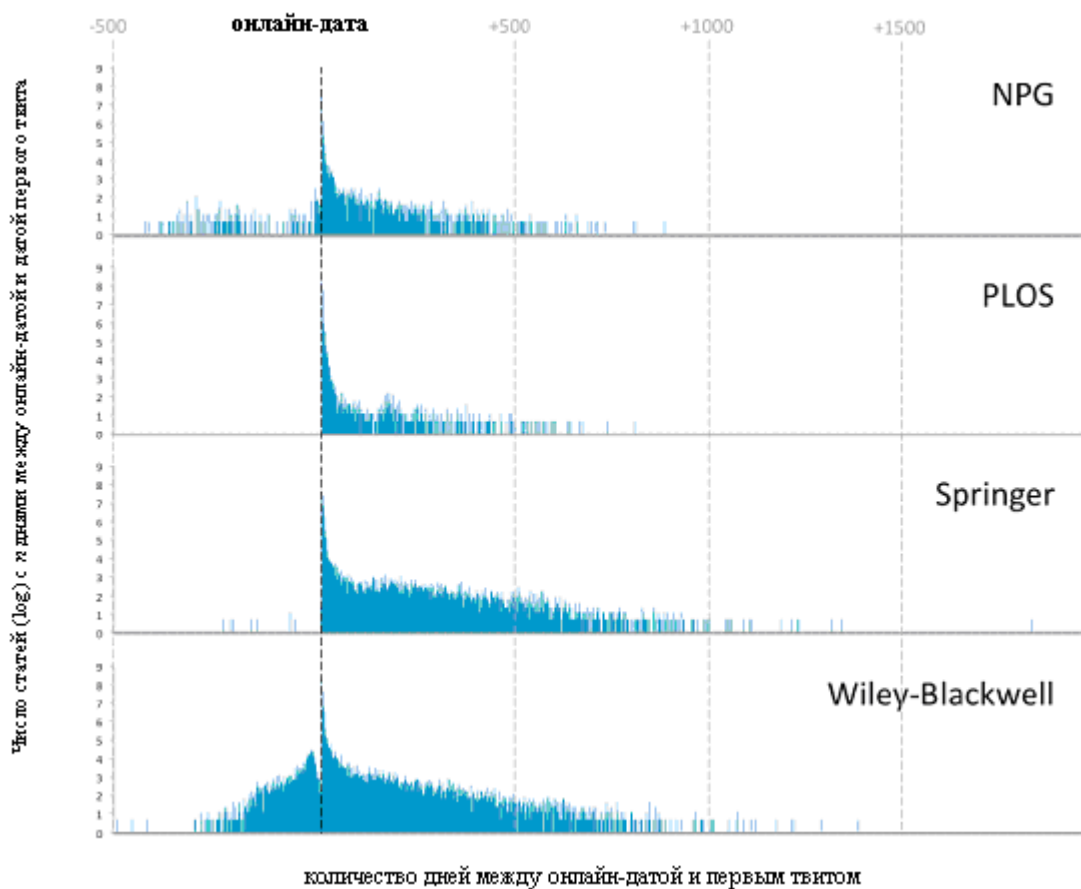


Рис. 3. Число статей (log) с  $n$  днями между онлайн-датой и датой первого твита по издательствам

\* Результаты только слегка меняются при использовании даты статьи, сначала изданной онлайн, т. е. 14,61% статей Wiley-Blackwell имели твит, появившийся до этой даты.

В настоящее время ни одна из исследованных дат не служит подходящей переменной для даты, когда журнальная статья была фактически доступна онлайн. В частности, найденные данные относительно того, что значительное количество статей Wiley-Blackwell были упомянуты в Твиттере до онлайн-даты, предполагают, что изменения имеют место по тому, как издатели сообщают об онлайн-датах. Это затрагивает как технические аспекты, так и фактический контент и используемый словарь. Таким образом, даже когда онлайн-даты могут быть загружены с веб-сайтов издателей или через программный интерфейс приложения, они, кажется, не всегда (и так у каждого издателя) отмечают фактическую точку времени, когда что-то стало доступно онлайн. Таким образом, существует настоятельная потребность в прозрачности и стандартизации различных дат, о которых сообщают издатели, чтобы обеспечить сравнимость онлайн-дат среди издателей. Принимая разработанный NISO словарь, определенные даты должны сообщаться для каждой версии журнальной статьи, а первое появление VoR, таким образом, будет отмечать дату фиксированной версии документа, представленного онлайн. Стандартизированный словарь и общее определение того, что значат различные даты публикации, не только улучшили бы эталонное тестирование в контексте оценки исследования, но также помогли бы точно определить начало периодов эмбарго открытого доступа, требуемых определенными финансирующими организациями, такими как Национальный институт здравоохранения в Соединенных Штатах Америки или Европейский исследовательский совет. В настоящее время эти периоды эмбарго, задерживающие зеленый открытый доступ на время от нескольких месяцев до нескольких лет, чтобы защитить доход издателей, начинают действовать с момента публикации статьи, который может относиться как к дате выпуска журнала, так и онлайн-дате\*. Установление даты начала эмбарго для даты онлайн публикации VoR устранило бы потенциальную лазейку, позволяющую издателям увеличивать период эмбарго, в течение которого они имеют исключительное право доступа.

Пока такой стандарт не внедрен исследование относительно метрики должно сосредоточиться на том, чтобы получать больше независимой от издателя информации о дате. Одной потенциальной переменной онлайн публикации могла бы стать дата успешного присвоения DOI в первый раз. Недавно CrossRef внедрил DOI Chronograph, инструмент, отслеживающий различные депозиты метаданных издателя, а также первый день успешного присвоения DOI [25]. Дальнейшая работа будет заключаться в исследовании того, насколько глубоко эти даты могут использоваться, чтобы создать точные критерии, необходимые в контексте метрик социальных медиа. Что касается цитат, где ежемесячные переменные достаточны, то индексация даты WoS должна быть исследована дальше.

*Благодарность.* Авторы хотели бы поблагодарить Юана Эйди и Altmetric.com за доступ к их данным и выразить признательность Фонду Альфреда П. Слоуна за финансирование, грант № 2014-3-25.

1. *Garvey W. D., Griffith B. C.* Scientific information exchange in psychology: The immediate dissemination of research findings is described for one science// *Science*. — 1964.— Vol. 146, No.3652.— P. 1655–1659. — doi:10.1126/science.146.3652.1655
2. *Dong P., Lob M., Mondry A.* Publication lag in biomedical journals varies due to the periodical's publishing model// *Scientometrics*. — 2006. — Vol. 69, No.2.— P. 271–286. — doi:10.1007/s11192-006-0148-3
3. *Wills M., Wills G.* The ins and the outs of electronic publishing// *Internet Research*. —1996.— Vol. 6, No.1.— P. 10–21. — doi:10.1108/10662249610123647
4. *Amat C. B.* Editorial and publication delay of papers submitted to 14 selected Food Research journals. Influence of online posting// *Scientometrics*. — 2008.— Vol. 74, No. 3. — P. 379–389. — doi:10.1007/s11192-007-1823-8
5. *Abt H. A.* Publication practices in various sciences// *Scientometrics*. — 1992. — Vol. 24, No.3.— P. 441–447. — doi:10.1007/BF02051040
6. *Björk B.-C., Solomon D.* The publishing delay in scholarly peer-reviewed journals// *Journal of Informetrics*. — 2013.— Vol. 7, No. 4. — P. 914–923. — doi:10.1016/j.joi.2013.09.001
7. *Das A., Das P.* Delay between online and offline issue of journals: A critical analysis// *Library & Information Science Research*. — 2006.—Vol. 28, No.3.— P. 453–459. —doi:10.1016/j.lisr.2006.03.019
8. *Diospatonyi I., Horvai G., Braun T.* Publication speed in analytical chemistry journals//*Journal of Chemical Information and Modeling*. — 2001.— Vol. 41, No.6.— P. 1452–1456. — doi:10.1021/ci010033d
9. *Rowlands I., Nicholas D.* The changing scholarly communication landscape: An international survey of senior researchers// *Learned Publishing*.— 2006.—Vol. 19.— P. 31–55.— doi:10.1087/095315106775122493
10. *Schauder D.* Electronic publishing of professional articles: Attitudes of academics and implications for the scholarly communication industry// *Journal of the American Society for Information Science*. — 1994.— Vol. 45, No. 2.— P. 73–100. — doi:10.1002/(SICI)1097-4571(199403)45:2<73::AIDASI2> 3.0.CO;2-5
11. *Tenopir C., King D. W.* Towards electronic journals: Realities for scientists, librarians, and publishers. — Washington, DC: Special Libraries Association, 2000.
12. *Haustein S.* Multidimensional journal evaluation. Analyzing scientific periodicals beyond the impact factor. — Berlin / Boston: De Gruyter Saur, 2012. — doi:10.1515/9783110255553
13. *Tort A. B. L., Targino Z. H., Amaral O. B.* Rising publication delays inflate journal impact factors// *PLOS ONE*.— 2012.— Vol. 7, No.12, e53374.— doi:10.1371/ journal.pone.0053374
14. *Gorraiz J., Gumpenberger C., Schlägl C.* Usage versus citation behaviours in four subject areas// *Scientometrics*. — 2014.— Vol. 101, No.2.— P. 1077–1095. — doi:10.1007/s11192-014-1271-1
15. *Heneberg P.* Effects of print publication lag in dual format journals on scientometric indicators // *PLOS ONE*.— 2013.— Vol. 8, No. 4, e59877. — doi:10.1371/journal.pone.0059877
16. *NISO/ALPSP Journal Article Versions (JAV) Technical Working Group.* Journal Article Versions (JAV):

\* [http://authorservices.wiley.com/bauthor/faqs\\_fundingbodyrequirements.asp](http://authorservices.wiley.com/bauthor/faqs_fundingbodyrequirements.asp)

Recommendations of the NISO/ALPSP JAV Technical Working Group. — Baltimore, 2008. — <http://www.niso.org/publications/rp/RP-8-2008.pdf>

17. *Egghé L., Rousseau R.* The influence of publication delays on the observed aging distribution of scientific literature// *Journal of the American Society for Information Science.*— 2000.— Vol. 51, No.2.— P. 158–165. — doi:10.1002/(SICI)1097-4571(2000)51:2<158::AID-ASI7>3.0.CO;2-X

18. *Luvvel M., Moed H. F.* Publication delays in the scientific field and the their relationship to the ageing of scientific literature// *Scientometrics.*— 1998.— Vol. 41, No.1-2.— P. 29–40.

19. *Seglen P. O.* Citations and journal impact factors: Questionable indicators of research quality// *Allergy.*— 1997.— Vol. 52, No.11.— P. 1050.

20. *Shuai X., Pepe A., Bollen J.* How the scientific community reacts to newly submitted preprints: Article

downloads, Twitter mentions, and citations// *PLOS ONE.*— 2012.— Vol. 7, No.11, e47523. — doi:10.1371/journal.pone.0047523

21. *Haustein S., Costas R., Larivière V.* Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns// *PLOS ONE.*—2015.— Vol. 10, No.3, e0120495. — doi:10.1371/journal.pone.0120495

22. *Robinson-García N., Torres-Salinas D., Zahedi Z., Costas R.* New data, new possibilities: Exploring the insides of Altmetric.com// *El Profesional de La Informacion.*— 2014.— Vol. 23, No.4.— P. 359–366. — doi:10.3145/epi.2014.jul.03

23. *Wass J.* Introducing the CrossRef Labs DOI Chronograph//Crosstech blog post. — 12 January 2015. — <http://crosstech.crossref.org/2015/01/introducingchronograph.html>

# Библиометрический анализ журнала “Вычислительные технологии” за 1996-2013 гг.\*

**И. В. СЕЛИВАНОВА**

Новосибирский государственный  
университет, г. Новосибирск, Россия

**А. Е. ГУСЬКОВ**

Государственная публичная научно-  
техническая библиотека СО РАН,  
г. Новосибирск, Россия  
Институт вычислительных технологий  
СО РАН, г. Новосибирск, Россия

**Н. А. МАЗОВ**

Институт нефтегазовой геологии  
и геофизики им. академика А.А. Трофимука  
СО РАН, г. Новосибирск, Россия  
Государственная публичная  
научно-техническая библиотека СО РАН,  
г. Новосибирск, Россия

*В работе приводятся результаты библиометрического анализа статей, опубликованных в журнале «Вычислительные технологии» за период 1996–2013 гг. Проведен количественный анализ публикаций по городам, организациям, странам; выявлены наиболее публикуемые и цитируемые авторы, построены графы соавторства. На примере данных журнала показано, что чем больше ссылок в пристатейном списке литературы имеет статья, тем больше раз она в среднем цитируется. Анализ цитирований журнала показал, что статьи в специальных выпусках цитируются в 1,5 раза меньше, чем в регулярных, а статьи на английском языке – в 4–5 раз меньше, чем на русском. Выполненные исследования будут полезны в первую очередь редколлегии журнала в качестве возможных рекомендаций по политике его дальнейшего развития; авторам – в качестве определяющего фактора для публикации их научных результатов; редакциям журналов подобной тематики для сравнения характеристик своих изданий с полученными результатами данной работы.*

## Введение

Журнал «Вычислительные технологии» (ЖВТ) издаётся Институтом вычислительных технологий СО РАН с 1996 г. С момента создания было выпущено 107 регулярных и 36 специальных выпусков, издано более полутора тысяч статей по математическому моделированию, интервальному анализу, вычислительной и прикладной математике, компьютерным технологиям, геоинформационным системам и электронным библиотекам. Библиометрический анализ журнала до настоящего времени не проводился.

Отметим, что библиометрические исследования обычно осуществляются с целью анализа текущей структуры рассматриваемого объекта (например, журнала, научной организации или целой области наук), что является необходимым для формирования принципов его дальнейшего развития.

Вопросам библиометрических исследований в последние годы уделяется большое внимание, о чем свидетельствует рост количества российских и зарубежных публикаций, посвященных данной тематике. Достаточно обратиться к мультидисциплинарным наукометрическим базам данных (например, Web of Science или Scopus), чтобы увидеть экспоненциальный рост количества таких работ. В последние годы большинство таких исследований наблюдается в области нанотехнологий, химии, медицины, биологии, наук о Земле. Например, в работах [1–3] проводится библиометрический анализ химических и биологических журналов. На основе по-

лученных результатов об авторском составе, организациях, цитированиях и тематиках статей и др. приведены некоторые советы для дальнейшего развития журналов. В наукометрических исследованиях по нанотехнологиям и наукам о Земле [4–6] выявлены наиболее значимые публикации, а также приведены наиболее продуктивные в этих областях страны, журналы, организации и ученые.

Однако в областях физико-математических и технических наук, к которым относятся статьи ЖВТ, таких исследований крайне мало. Среди них стоит выделить [7], где проведен библиометрический анализ математики как области наук в целом, а также [8], где изучено положение российских математических журналов в мировой науке. В данном исследовании проведен библиометрический анализ ЖВТ, который показывает сложившиеся авторские коллективы, географию публикаций. Выявлено, какие выпуски и статьи цитируются чаще всего, каков процент самоцитируемости, а также оценен вклад международного сотрудничества в развитие журнала.

Для проведения библиометрического анализа по выявлению авторских коллективов был изучен опыт исследований [9–10], где делается вывод о том, что ежегодно увеличивается число проектов, которые невозможно выполнить в одиночку, следствием чего является наблюдаемое быстрое уменьшение количества одиночных статей. Для выделения авторских коллективов в [11] предлагают использовать методику кластерного анализа. Для примера данный вид анализа

проводится для 50 наиболее продуктивных авторов в области геостатистики. В настоящей работе для исследования авторских коллективов построены графы соавторства, которые также позволяют выделять устойчивые коллективы.

Исследования международного научного сотрудничества часто проводятся на основе библиометрических баз данных Web of Science и/или Scopus. Например, в работах [12–14] построена карта международного сотрудничества стран Европейского союза с использованием данных Web of Science и выявлено, что оно проявляется больше всего в таких областях как физика, химия и клиническая медицина. Рассматривая национальную специфику, в работе [15] устанавливается факт низкой степени международного сотрудничества у индийских авторов. В России подобная проблема изучалась в работе [16], где выделены характерные тенденции для различных специализаций и стран сотрудничества. Поскольку ЖВТ не индексируется ни в Web of Science, ни в Scopus, для изучения международного сотрудничества используются данные Российского индекса научного цитирования (РИНЦ) о цитировании ЖВТ другими журналами, в том числе и международными, а также данные о географическом положении авторов журнала.

Для проведения анализа цитирования и самоцитирования был изучен опыт, отраженный в публикациях [17–21]. Так, в [17] проведено сравнение цитирования неанглоязычных статей по химии и физике с англоязычными, в результате чего выявлено, что неанглоязычные статьи цитируются меньше, что существенно отличается от нашего исследования, в котором англоязычные статьи цитируются значительно меньше, чем русскоязычные. Работы [18] и [19] посвящены анализу самоцитирования в журналах различного профиля: в результате двух независимых исследований было выяснено, что около половины статей содержат самоцитирование. Подобный анализ статей в ЖВТ показал, что самоцитирование в данном журнале находится на достаточно низком уровне. Вклад иностранных статей в журналы с точки зрения цитирования исследован в [20] и [21]. В настоящей статье проводится сравнительный анализ цитирования русскоязычных и англоязычных публикаций, а также определяется, в каких выпусках (специальных или регулярных) данный показатель является наибольшим.

Проведенный библиометрический анализ будет полезен для формирования и уточнения редакционной политики ЖВТ, а также для проведения других библиометрических исследований.

### Методология исследования

Объектом настоящего исследования являются статьи, изданные в ЖВТ за период с момента его образования в 1996 г. по 2013 г. включительно, пристатейные списки литературы, а также публикации и журналы, цитирующие вышеупомянутые статьи ЖВТ. В качестве источников данных для исследования были использованы:

- 1) База данных (БД) метаданных редакции журнала, содержащая:
  - а) информацию о статьях (название, авторы, год издания, номер, том и страницы журнала, язык публикации),
  - б) информацию об авторах (ФИО, год рождения, научная степень, организация),

с) информацию об организациях (название, город, страна);

- 2) База данных цитирований РИНЦ, содержащая:
  - а) списки цитируемых источников в ЖВТ,
  - б) списки источников, цитирующих ЖВТ.

В рамках подготовительного этапа данные об авторах и публикациях были проверены на дублетность. Затем была проведена нормализация названий организаций: названия, обозначающие одну организацию, были приведены к единому виду.

Выполнена декомпозиция пристатейных списков литературы на элементы, в зависимости от вида первоисточника (были использованы только приведенные ниже виды в связи с тем, что они составляют около 95% от общего числа источников):

- для вида “статья в журнале” выделялись следующие элементы: автор, название публикации, журнал, страницы, том и номер журнала, год издания;
- для вида “монография” - авторы, название, издательство, год.

Для проведения исследования были выбраны следующие элементы данных: автор, название публикации, журнал или издательство, год. Аналогичная работа проведена и с метаданными публикаций, цитирующих статьи в ЖВТ.

Затем информация из указанных источников была агрегирована в одной базе данных, структура которой соответствует схеме CERIF (Common European Research Information Format) [22]. Эта схема предназначена для структурированного хранения данных о научных исследованиях, включая информацию об организациях, ученых, их публикациях, журналах, наукометрических показателях и пр. Таким образом, была построена база данных публикаций журнала и всех связанных данных, необходимых для библиометрического анализа.

На основе сформированной БД, которая содержит всю необходимую информацию о связанных со статьями объектах, были сформулированы и решены следующие аналитические задачи:

- Анализ статей по типам и видам;
- Анализ географического распределения аффилиаций авторов;
- Исследование состава авторов журнала и построение графов соавторства;
- Исследование пристатейных списков литературы, а также списков источников, цитирующих ЖВТ;
- Исследование места ЖВТ среди журналов аналогичной тематики.

### Результаты исследования

Всего за указанный период было издано 145 выпусков журнала (включая 36 специальных выпусков), в которых были опубликованы 1596 статей. На рис.1 представлена динамика публикаций ЖВТ в регулярных и специальных выпусках. Ежегодно в шести регулярных выпусках журнала публикуются от 53 до 72 статей, причем преимущественно это статьи только российских авторов. В период 2002-2008 гг. активно издавались специальные выпуски, которые, как правило, содержали труды научных конференций. Однако позже эта практика была прекращена из-за наметившегося ухудшения библиометрических характеристик ЖВТ, которое рассматривается ниже.

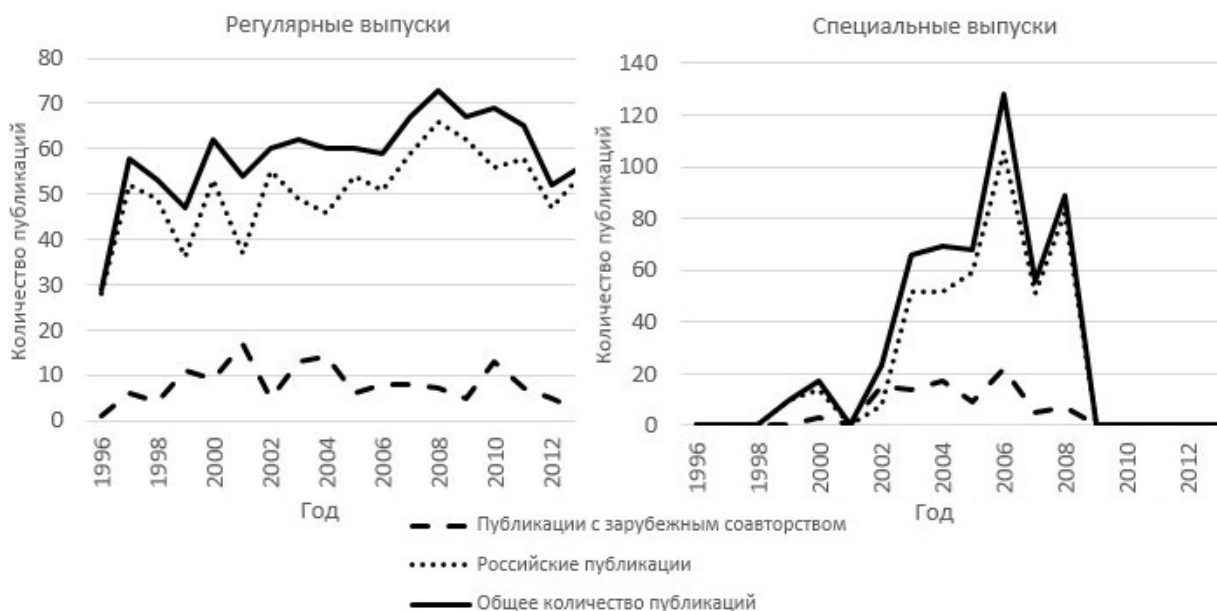


Рис.1. Динамика количества публикаций по годам

### Авторы и их аффилиации

Всего в журнале свои статьи опубликовали 1700 различных авторов, из них

- 43% - доктора физико-математических наук;
- 11% - доктора технических наук;
- 35% - кандидаты физико-математических наук;
- 8% - кандидаты технических наук;
- 3% - другие ученые степени.

Средний возраст авторов-женщин составляет 40 лет, мужчин - 48.

За рассматриваемый период в журнале были опубликованы статьи авторов более чем из 150 организаций. В табл. 1 представлены организации, сотрудники которых опубликовали в журнале больше всего статей, и их показатели, характеризующие степень внутрirosсийского и международного сотрудничества. Для каждой организации указано общее количество статей, количество организаций, имеющих с ней общие публикации, число таких публикаций и их доля в общем количестве; а также количество публикаций с зарубежными соавторами.

Наибольшее количество статей опубликовали авторы из ИВТ СО РАН, который является учредителем ЖВТ. Вторую и третью позиции занимают ИВМ СО РАН и ИВМиМГ СО РАН, которые являются наиболее близкими к ИВТ СО РАН по научной тематике, имеют с ним тесные научные связи и обладают большим числом результатов, подходящих для публикации в ЖВТ.

Наибольшую долю совместных с другими организациями публикаций имеют вузы - НГУ (79%) и НГТУ (71%), студенты и аспиранты которых проходят практику в институтах, что говорит об активном научном сотрудничестве.

Для примера рассмотрим состав организаций, имеющих общие публикации с ИДСТУ СО РАН. В их число входят 3 университета (4 статьи), 8 российских и зарубежных институтов (5 статей) и Президиум Монгольской академии наук (1 статья). В двух статьях представители ИДСТУ СО РАН публикуются совместно с представителями Одесского государственного экологического университета, еще в двух - с представителями Института систем энергетики им. Л. А. Мелентьева СО

РАН, Иркутского государственного технического университета и с представителями Института математики Вьетнамской академии наук и технологий, Ульяновского государственного университета.

В табл. 2-3 представлена география авторов журнала. Как видно из данных таблиц, большинство авторов являются представителями России (83%). Если же рассматривать распределение авторов по городам, то видно, что наибольшее количество статей в ЖВТ печатают авторы, работающие в Новосибирске (32%), далее идут ученые из Красноярска (10%), Томска (9%), Москвы (6%), Иркутска (4%), Алматы (3%), Кемерово (2%), Барнаула (2%), Екатеринбурга (2%) и Санкт-Петербурга (1%). Такое распределение сложилось исходя из, во-первых, интенсивности исследований по тематике ЖВТ в указанных городах, а во-вторых - наличия научных связей с "ядром" авторов журнала (табл. 6). Авторы из любого другого города опубликовали в ЖВТ менее 1% статей.

Рассмотрим изменение количества публикаций в соавторстве с зарубежными учеными во времени, а также количество публикаций по странам. Из рис. 2 видно, что ученые из Азии (кроме Казахстана) начинают активно публиковаться в ЖВТ начиная с 1999 г., из Европы (кроме Сербии) - в период 1999-2008 г. Это может объясняться тем, что за данный период было опубликовано несколько спецвыпусков журнала, посвященных докладам международных конференций и семинаров. Так, всплеск публикационной активности в 2002 г. обусловлен выходом специального выпуска, в котором были опубликованы труды конференции «Colloquy Optimization - Structure and Stability of Dynamical Systems», которая проходила в Германии в 2000 г., а в 2003 - выходом спецвыпуска, посвященного семинарам Российско-Германской рабочей группы.

Научные связи с учеными из Казахстана видны с первых выпусков журнала, где наблюдается явный скачок публикационной активности в 2004 г. Это связано с выходом спецвыпуска, посвященного трудам совещания Российско-Казахстанской группы по вычислительным и информационным технологиям. Еще одним примером положительного влияния международных конференций на привлечение иностранных авторов является актив-

ность сербских ученых. В 2009 и 2011 гг. проводились международные конференции из цикла «Математические и информационные технологии», организованные российскими и балканскими коллегами. Результатом стала публикация в журнале ряда статей при соавторстве сербских ученых.

Эти данные наглядно демонстрируют, что в редакционную политику журнал следует закладывать возможности использования международных конференций, как площадки для формирования корпуса зарубежных авторов и укрепления международного сотрудничества.

Таблица 1

**Количественный анализ публикаций по организациям**

№	Назв. орг-ции	Всего статей	Кол-во орг-ций, имеющих общие публ. с данной	Совместн. число публ. с другими орг-ми	Доля совместн. публ. в %	Кол-во публ. с зарубежн. соавторами
1	ИВТ СО РАН	313	57	125	40	12
2	ИВМ СО РАН	125	20	45	36	4
3	ИВМиМГ СО РАН	96	34	45	47	2
4	НГУ	66	25	52	79	3
5	ИДСТУ СО РАН	48	12	9	21	2
6	ИМ СО РАН	43	12	27	63	3
7	ИВЭП СО РАН	36	10	13	36	0
8	ИГиЛ СО РАН	36	6	10	28	1
9	ИЦиГ СО РАН	23	9	12	52	3
10	НГТУ	21	9	15	71	1
11	ИНГТ СО РАН	18	10	12	67	1
12	ВЦ ДВО РАН	13	3	5	38	0

Таблица 2

**Кол-во ученых по городам**

Город	Кол-во ученых
Новосибирск	554
Красноярск	168
Томск	155
Москва	113
Иркутск	71
Алматы	63
Кемерово	48
Барнаул	39
Екатеринбург	35
Санкт-Петербург	30

Таблица 3

**Кол-во ученых по странам**

Страна	Кол-во ученых
Россия	1421
Казахстан	54
Германия	42
Узбекистан	23
Турция	15
Сербия	13
США	13
Украина	11
Франция	10
Дания	8
Италия	8

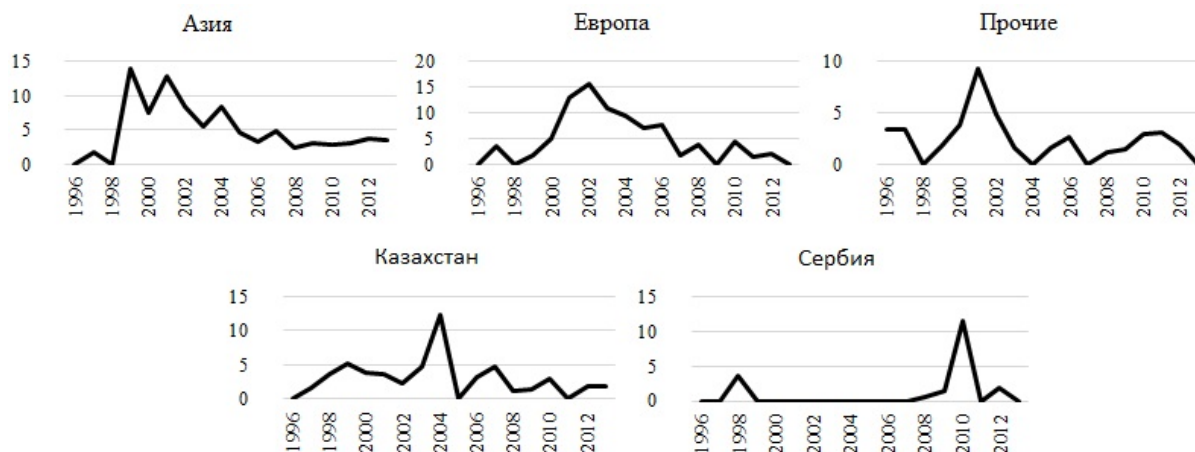


Рис. 2. Динамика количества зарубежных ученых во времени (Азия, Европа, Казахстан, Сербия, прочие: США, СНГ, Африка, Ближний и Дальний Восток, Океания) — % от общего кол-ва ученых

### Тематика публикаций

Для анализа публикаций по тематическим направлениям каждой организации была сопоставлена ее основная тематика из следующего списка\*:

- Математика и информатика (42%),
- Энергетика и механика (6%),
- Науки о Земле (6%),
- Физика (3%),
- Биология (1%),
- Космос (1%),
- Химия (<1%),
- Медицина (<1%),
- Не определено (40%)

Несмотря на то, что это предположение в некоторой степени условно, оно позволяет оценить количество публикаций по разным направлениям наук.

На рис. 3 представлена доля авторов ЖВТ обозначенных выше областей наук в общем числе публикаций. Как можно заметить из данного графика, с 2002 г. увеличивается доля авторов статей в области физики, практически не меняется доля в области математики и информатики, энергетике и механики, науках о Земле, нестабильно публикуются статьи по наукам о космосе, химии, биологии и медицине. Отметим, что резкий скачок в 2000 г. по биологии обусловлен выходом специального выпуска журнала, посвященного 10-летию лаборатории теоретической генетики ИЦиГ СО РАН.

Отметим высокий процент статей с неопределенной тематикой. Чаще всего это связано с тем, что в качестве места работы указан университет (НГУ, СФУ и др.), имеющий различные исследовательские профили.

### Анализ авторских коллективов

В табл. 4 показано распределение статей по количеству авторов.

В среднем на одну статью приходится 2 автора. 72% статей имеют не более двух авторов, остальные 28% - трех и более. Такое распределение характерно для журналов математических, физических и технических направлений: если в статье 1-2 автора, то в большинстве случаев это указывает на то, что статья имеет теоретический характер. Если же авторов больше, чем три, то данные исследова-

ния чаще направлены на практические результаты, которые трудно получить в одиночку [10, 23–24].

В табл. 5 указано распределение авторов по году рождения и году публикации. Из нее следует, что наибольшая доля авторов – это авторы 1947 – 1951 г. рождения. Пик их публикационной активности приходится на 2001 г., после чего замечен постепенный спад, однако, это поколение остается в числе доминирующих на протяжении всего периода. Замечен «провал» следующего поколения 1957 – 1971 г.р., которое ни в одном году не было в числе лидеров. Публикационная активность самого юного поколения 1977-1986 г.р. начала активно проявляться с 2005 г., а к 2011 г. вышла на лидирующие позиции.

Важно отметить, что такая динамика возрастного распределения является не нормальной, но вполне закономерной. При планомерном развитии научных исследований пик публикационной активности ученого приходится на возраст 40 – 50 лет (в ЖВТ такая ситуация имела место в конце 90-х гг.). Это значит, что в приведенной таблице с каждым годом должно наблюдаться постепенное смещение пика «вправо». Однако этого не происходит, поскольку социально-экономическая ситуация в России в период с 1990 по 2005 гг. привела к «выпадению» из науки поколения, рожденного в конце 50-х, 60-х и 70-х гг., что можно также наблюдать по кадровому составу многих научных организаций.

В табл. 6 показаны авторы, имеющие более десяти публикаций в ЖВТ, количество их публикаций в регулярных и специальных выпусках, а также количество цитирований (учитываются цитирования только статей в ЖВТ).

Лидером по количеству опубликованных в журнале статей, по числу цитирований, а также по количеству разных соавторов является академик Шокин Ю.И., основатель научной школы по развитию информационно-вычислительных технологий [25]. Также из таблицы видно, что авторы, имеющие более 10 публикаций в ЖВТ, публикуются преимущественно в регулярных выпусках. Большинство из этих авторов (18 из 26) - сотрудники ИВТ СО РАН. В специальных выпусках количество их статей равно 73 (11%), в регулярных - 240 (16%). Следует отметить, что большинство активно публикующихся соавторов имеют не более 3 «одиночных» публикаций (т.е. публикаций без соавторов), что свидетельствует о хорошем уровне сотрудничества.

\* Публикация относится к области наук, если хотя бы один из соавторов работает в организации, которая занимается этой областью



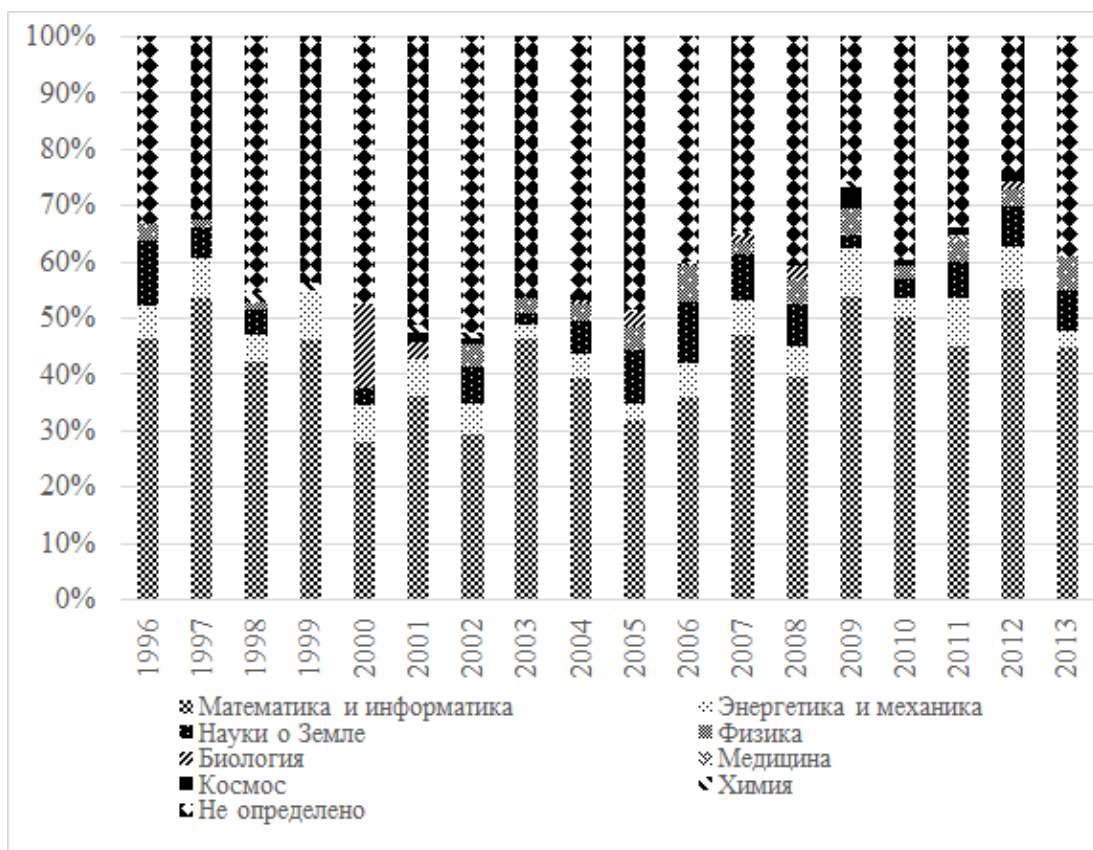


Рис. 3. Доля авторов соответствующих областей наук в общем числе публикаций ЖВТ

Таблица 4

Распределение статей по количеству авторов

Кол-во соавторов	1	2	3	4	5	6	7	8	9	10	12	14
Кол-во публикаций	610	546	282	82	43	19	8	2	1	1	1	1

Таблица 5

Распределение авторов по году рождения и году публикации (% от годового объема)

%	Год выпуска журнала	Год рождения автора													
		1925-1931	1932-1936	1937-1941	1942-1946	1947-1951	1952-1956	1957-1961	1962-1966	1967-1971	1972-1976	1977-1981	1982-1986	1987-1991	
	1996	0	0	4	13	26	13	4	4	13	26	0	0	0	
	1997	0	2	6	23	32	14	2	7	9	2	4	0	0	
	1998	0	2	5	14	28	18	5	7	9	12	5	0	0	
	1999	3	0	0	15	33	23	0	5	8	8	0	0	0	
	2000	3	3	0	21	30	15	12	0	9	0	0	0	0	
	2001	4	0	0	26	37	19	15	0	11	0	0	0	0	
	2002	0	3	10	9	26	15	8	3	2	6	5	0	0	
	2003	0	0	3	10	24	22	13	11	4	4	7	0	0	
	2004	2	2	8	15	25	14	8	6	4	6	8	0	0	
	2005	1	2	4	10	14	14	7	5	8	15	18	4	0	
	2006	1	2	8	9	18	11	5	4	7	8	16	11	0	
	2007	0	1	5	9	15	16	7	6	10	8	11	11	0	
	2008	0	1	1	4	14	12	7	6	8	10	17	18	0	
	2009	1	3	5	4	17	19	8	6	5	10	10	14	1	
	2010	1	1	6	9	19	12	9	7	5	10	8	11	3	
	2011	0	1	5	7	15	12	8	7	5	7	16	14	7	
	2012	0	1	3	8	15	17	7	9	2	9	12	10	6	
	2013	1	3	4	10	16	10	6	6	5	5	12	12	6	

Количество публикаций авторов в ЖВТ

Автор	Орг-ция	Всего публ.	Одиночн.публ.	Публ. в регу-ляр./спец. вы-пусках	Кол-во соавто-ров	Соавторы из ар. организ.	Заруб. соавто-ры	Цитирования
Шокин Ю. И.	ИВТ СО РАН	57	2	42/22	68	33	3	78
Хакимзянов Г.С.	ИВТ СО РАН	26	1	23/3	22	6	0	66
Федотов А.М.	ИВТ СО РАН	24	1	13/11	34	17	0	48
Федорук М.П.	ИВТ СО РАН	20	1	14/6	29	16	0	12
Ковеня В.М.	ИВТ СО РАН	19	5	16/3	7	3	0	19
Шокина Н.Ю.	ИВТ СО РАН	18	1	15/3	15	9	2	18
Бычков И.В.	ИДСТУ СО РАН	17	2	12/5	25	6	2	25
Вебер Г. В.	Институт прикладной математики (Турция)	15	3	13/2	19	13	11	1
Вшивков В.А.	ИВМиМГ СО РАН	15	0	15/0	24	20	0	66
Гордов Е.П.	ИМКЭС СО РАН	15	3	0/15	22	19	13	13
Рычков А.Д.	ИВТ СО РАН	15	2	11/4	10	7	5	4
Афанасьев К.Е.	КемГУ, ЦНИТ	14	1	8/6	12	0	0	19
Паасонен В.И.	ИВТ СО РАН	14	10	13/1	2	1	0	21
Черный С.Г.	ИВТ СО РАН	14	0	14/0	16	4	0	28
Барахнин В.Б.	ИВТ СО РАН	13	0	7/6	19	3	0	60
Баутин С.П.	УрГУПС	13	5	13/0	8	2	0	16
Пинчуков В.И.	ИВТ СО РАН	13	13	13/0	0	0	0	13
Черных Г.Г.	ИВТ СО РАН	13	1	8/5	14	11	0	9
Чубаров А.Б.	ИВТ СО РАН	13	0	8/5	21	12	0	26
Шарый С.П.	ИВТ СО РАН	13	8	12/1	8	5	5	38
Бочаров О.Б.	Бейкер Хьюз	11	0	11/0	4	3	0	25
Григорьев Ю.Н.	ИВТ СО РАН	11	2	9/2	5	4	0	8
Дудникова Г.И.	ИВТ СО РАН	11	0	11/0	19	12	2	37
Молородов Ю.И.	ИВТ СО РАН	11	1	3/8	17	8	0	11
Федотова З.И.	ИВТ СО РАН	11	4	8/3	10	4	0	40
Шлычков В.А.	ИВЭП СО РАН	11	8	9/2	5	4	0	14

Рассмотрим графы взаимодействия авторов, приведенных в табл. 6, изображенные на рис. 4-5.

В графе на рис. 4 можно выделить несколько пар стабильно сотрудничающих соавторов: Федотов А. М. – Барахнин В. Б. (период появления совместных публикаций 2005-2010 гг.), Федотов А. М. – Молородов Ю. И. (2004-2013), Шокин Ю. И. – Федотов А. М. (1997-2010), Шокин Ю. И. – Ковеня В. М. (2010-2013), Шокин Ю. И. – Чубаров А. Б. (1998-2008), Шокина Н. Ю. – Хакимзянов Г. С. (1996-2013), Шокина Н. Ю. – Рычков А. Д. (2002-2007). Отсутствие треугольников и четырехугольников, состоящих из жирных линий, свидетельствует об отсутствии устоявшихся коллективов более чем из двух человек. Следует отметить, что данные результаты получены лишь по публикациям одного журнала, полную

картину об устоявшихся коллективах соавторов необходимо составлять по базам данных публикаций сотрудников института.

Интересным примером развития научного направления является граф сотрудничества, построенный на основе статей д.ф.-м.н. Черного С. Г. (рис. 5). Он является заведующим лаборатории математического моделирования ИВТ СО РАН, а почти все его соавторы (8 из 13) были в разное время аспирантами или сотрудниками этой лаборатории: Чирков Д. В. (публиковался с Черным С. Г. в период 2000-2013 гг.), Лобарева И. Ф. (2005-2006), Астракова А. С., Банников Д. В. (оба - 2009-2013), Ешкунова И. Ф., Панов Л. В. (оба - 2011), Есипов Д. В., Лапин В. Н., Куранков Д. С. (все - 2011-2013), Авдюшенко А. Ю. (2012-2013).

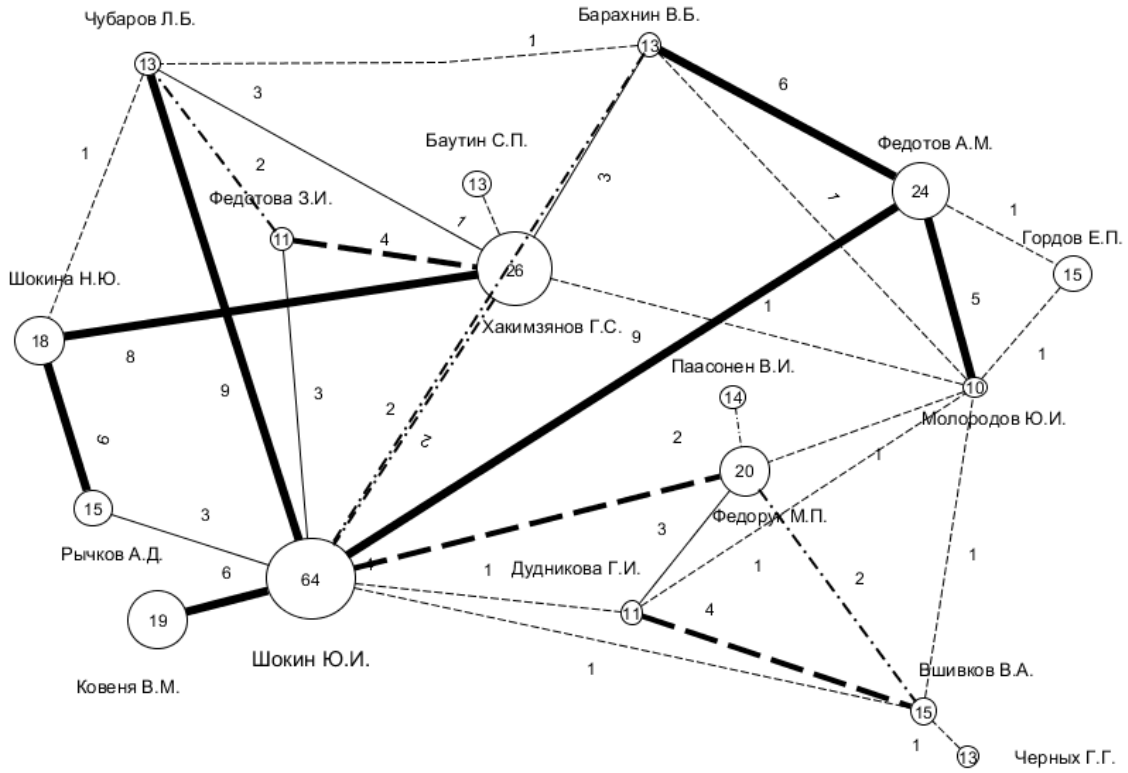


Рис. 4. Граф соавторства ЖВТ (Ребра: черный пунктир - 1 совместная публикация, черный штрих-пунктир - 2, черная линия - 3, жирный пунктир - 4, жирная линия - 5 и более)

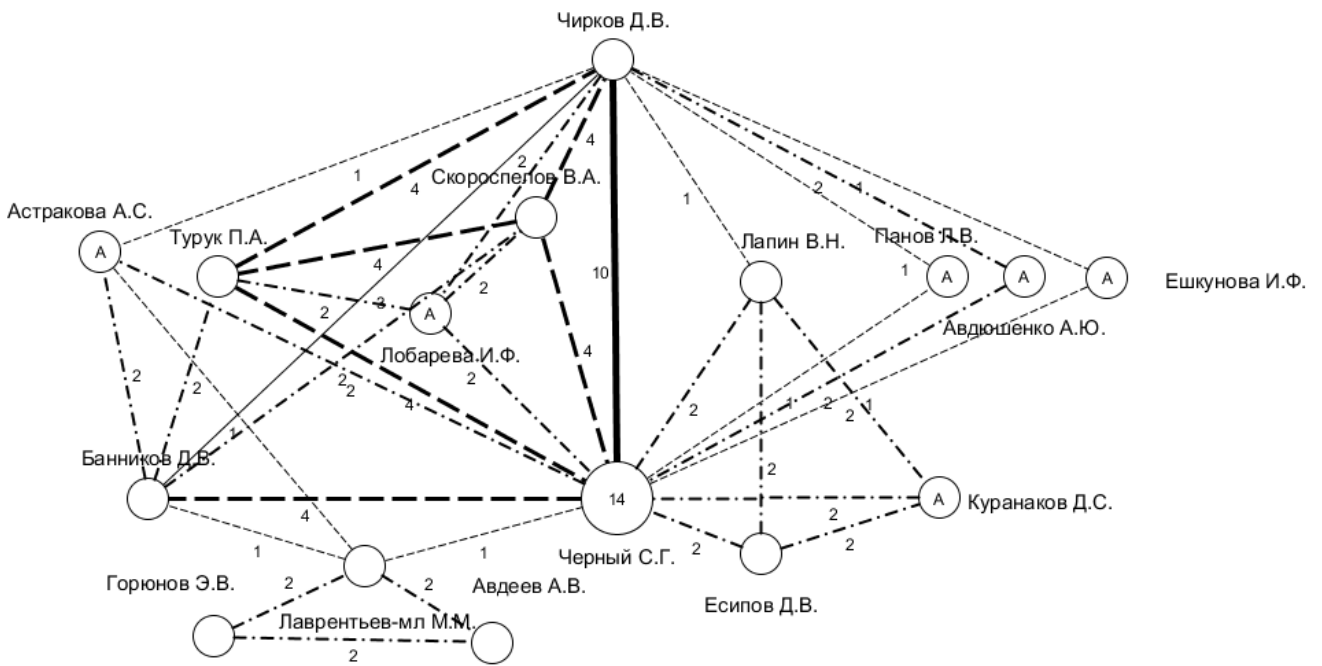


Рис. 5. Взаимодействие авторов ЖВТ. Математическое моделирование (Ребра: черный пунктир - 1 совместная публикация, черный штрих-пунктир - 2, черная линия - 3, жирный пунктир - 4, жирная линия - 5 и более; вершины: А-аспирант)

## Классификация статей

Рассмотрим следующую классификацию статей: англоязычные (статьи, опубликованные на английском языке), статьи с иностранным участием (статьи с иностранными соавторами), публикации совместно с вузами (хотя бы один из авторов работает в вузе), межинститутские публикации (соавторы работают только в институтах, российских и зарубежных). За все годы было опубликовано 890 межинститутских статей (55% от общего количества публикаций), 173 иностранных статьи (10%), 105 статей с иностранным участием (6%), 500 статей совместно с вузами (31%). Их динамика приведена на рис. 6.

За период существования журнала постепенно увеличивается доля совместных с вузами публикаций. Доля англоязычных статей и статей с иностранным участием после заметного всплеска в 2000-2004 гг. уменьшается, а доля межинститутских публикаций остается практически неизменной, о чем свидетельствуют линии тренда.

Наибольший процент иностранных статей и статей с иностранным участием был опубликован в период 2000-2004 гг., что связано с изданием ряда спецвыпусков журнала, посвященных семинарам, совещаниям Российско-Германской и Российско-Казахстанской рабочих групп, а также с материалами конференций с различными иностранными авторами. Следует отметить, что при этом снизилась доля межинститутских публикаций. Количество публикации, совместных с вузами, постепенно увеличивается. Вероятно, это связано с прохождением студентами и аспирантами вузов практики в институтах, где работают их научные руководители или коллеги по научному проекту.

## Цитирование

Анализ цитирования помогает определить, какие журналы, авторы и статьи являются самыми авторитетными и вызывают наибольший интерес у научного со-

общества; как быстро стареет то или иное направление исследования [26-27].

По данным РИНЦ о цитировании статей в первые десять наиболее цитируемых авторов ЖВТ входят: Шокин Ю. И., Вшивков В. А., Хакимзянов Г. С., Баракнин В. Б., Федотов А. М., Снытников В.Н., Федотова З. И., Шарый С. П., Дудникова Г. И., Шапеев В. П.

Среднее число цитирований на одного автора составляет 3,06; максимальное – 78. Среднее число самоцитирования – 1,6; максимальное – 19.

## Цитирующие журналы

Рассмотрим журналы, которые цитируют ЖВТ. Всего таких журналов 547 с общим количеством цитирований 1887. Упорядочим эти журналы в порядке убывания количества цитирований и разобьем данное множество на три группы, в каждой из которых будет одинаковое число цитирований. Количество элементов (журналов) в каждой из групп будет относиться друг к другу как  $1:q:q^2$ , при котором  $q > 1$ . Следуя закону Брэдфорда [28], журналы разбиваются на 3 группы, каждая из которых содержит в сумме  $2321/3=774$  цитирований. В результате, в первой, “ядерной”, зоне оказалось 10 журналов (табл. 7).

Самым цитирующим ЖВТ является сам ЖВТ, его доля от общего числа цитирований составляет 15%. Это соответствует мировой практике, поскольку в [29] отмечается, что “...среди всех журналов, представленных в Journal Citations Report Science Edition в 2010 г., 85% обладают уровнем самоцитирования 15% или менее. Таким образом, самоцитирование вполне нормально для большинства журналов, и его наличие следует предполагать. Однако в случае значительного отклонения данного показателя Thomson Reuters стремится определить, не влияет ли излишнее самоцитирование на увеличение импакт фактора журнала искусственным способом”. Заметим, что существенное увеличение доли самоцитирования уже может негативно сказаться на имидже журнала.

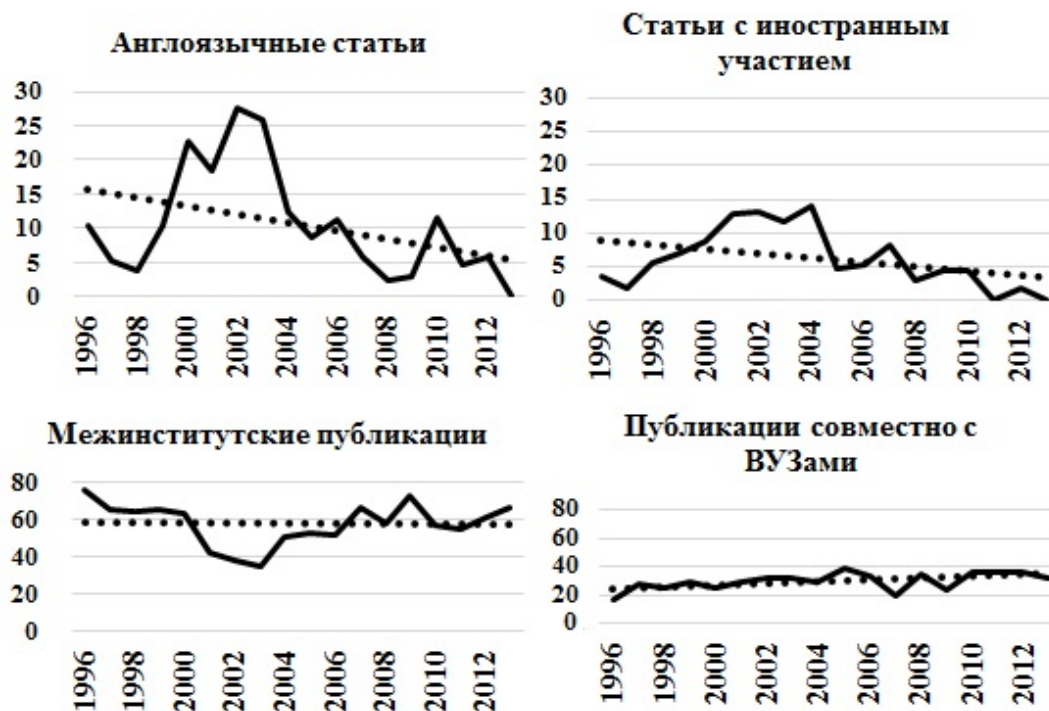


Рис. 6. Динамика различных видов публикаций в % от общего количества публикаций (пунктир - линия тренда)

## Цитирующие журналы

№	Журналы, цитирующие статьи ЖВТ по данным РИНЦ на 10 сентября 2014 г.	Кол-во цитир.	Доля от общ. числа цитир. в %	Накоп. сумма цитир.
1	Вычислительные технологии	337	15	337
2	Прикладная механика и техническая физика	122	5	459
3	Журнал вычислительной математики и математической физики	80	3	539
4	Вестник Новосибирского государственного университета. Серия: информационные технологии	55	2	594
5	Оптика атмосферы и океана	40	2	634
6	Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева	33	1	667
7	Вестник Томского государственного университета	32	1	699
8	Известия Российской академии наук	32	1	731
9	Вычислительные методы и программирование: новые вычислительные технологии	30	1	761
10	Математическое моделирование	28	1	789

Таблица 8

## Журналы, содержащие англоязычные статьи (не менее 5), которые цитируют ЖВТ (П–переводная версия журнала)

Название журнала	Число цитир. статей
(П) Journal of applied mechanics and technical physics	48
(П) Computational mathematics and mathematical physics	30
(П) Russian journal of numerical analysis and mathematical modelling	14
(П) Combustion, explosion, and shock waves	12
(П) Doklady physics	9
(П) High temperature	8
(П) Thermophysics and aeromechanics	7
(П) Bulletin of the novosibirsk computing center. Series: numerical analysis	6
(П) Automation and remote control	5
Journal of quantitative spectroscopy and radiative transfer	5
Lecture notes in computer science	5

Если рассматривать географию цитирующих журналов, попавших в “ядерную” зону, то можно заметить, что три из девяти оставшихся журналов издаются в Новосибирске, четыре – в Москве, два – в Томске. Большинство цитирующих статей являются русскоязычными (2258 из 2619, 86%), почти все остальные – англоязычными (336 из 2619, 13%). В табл. 8 представлены журналы, в которых содержатся цитирующие статьи на английском языке; большинство из них является переводными версиями русскоязычных журналов.

*Цитирование статей ЖВТ в хронологии*

Рассмотрим отдельно цитирование регулярных и специальных выпусков. В БД РИНЦ имеются данные лишь о спецвыпусках 2005–2008 гг., поэтому будем сравнивать данные только за этот период. Для одной статьи из регулярного выпуска в этот период число цитирований в среднем равно 2,05, а для одной статьи из спецвыпуска – 1,2. В табл. 9 показаны десять статей с наи-

большим количеством цитирований (№7,8 – статьи из специальных выпусков, остальные – из регулярных).

В табл. 10 представлены данные о статьях из регулярных и специальных выпусков, которые не цитировались ни разу. Доля статей с нулевым количеством цитирований из регулярных выпусков за период 2005–2008 гг. составляет 39% (101 из 259), из специальных – 58% (198 из 341). Более того, средняя цитируемость статей из специальных выпусков заметно ниже, чем из регулярных. На рис. 7 представлен график зависимости процитированных статей по отношению к не процитированным (обозначим его за  $k$ ) за весь исследуемый период.

Из графика видно, что после заметного скачка в 2005 г. данный коэффициент начинает уменьшаться, из чего можно сделать вывод, что число нецитируемых работ растет.

В мировой практике цитирования считается, что если работа не процитирована в первые четыре года после ее выхода, то вероятность того, что она будет про-

цитирована в дальнейшем, стремится к нулю [30 – 31]. Рассмотрим данную тенденцию на примере ЖВТ. Действительно, за период 2005–2008 гг. 220 публикаций (73%) из 301 были впервые процитированы в первые четыре года после публикации. Таким образом, вероятность того, что 299 (за период 2005–2008 гг.) статей из регулярных и специальных выпусков, имеющих в настоящее время нулевое цитирование, будут процитированы, крайне мала. Исключением из этого могут быть работы, про которые Ю. Гарфила говорил следующее: «Признание работы может запаздывать по нескольким причинам. Из-за засорения каналов информации в результате информационного взрыва идее может быть трудно проникнуть сквозь барьер устоявшихся шаблонов. Открытие может настолько опережать свое время, что его окажется невозможно связать с концептуальным строем

современной науки. Но оно может поначалу игнорироваться и просто потому, что его автор – молодой исследователь, работающий в малоизвестной организации» [32].

Существующие данные позволяют провести корректное сравнение для выпусков 2005–2008 г. и сделать вывод, что регулярные выпуски ЖВТ цитируются чаще, чем специальные (по данным РИНЦ).

Рассмотрим отдельно спецвыпуски, где среднее количество цитирований на статью больше, чем 1,2 (табл. 11). В 10 специальных выпусках ЖВТ количество ссылок на одну статью больше, чем среднее количество ссылок на одну статью во всех специальных выпусках журнала, из них статьи из семи выпусков в среднем цитируются больше, чем из регулярных. Большинство из этих выпусков посвящены обзорам докладов и трудов различных совещаний и конференций.

Таблица 9

Наиболее цитируемые статьи ЖВТ

№	Кол-во цитир.	Год	Название статьи, авторы
1	22	2007	М. П. Токарев, Д. М. Маркович, А. В. Бильский Адаптивные алгоритмы обработки изображений частиц для расчета мгновенных полей скорости
2	19	2003	А. А. Замышляева Фазовые пространства одного класса линейных уравнений соболевского типа второго порядка
3	18	2002	В. Н. Снытников, В. Н. Пармон, В. А. Вшивков, Г. И. Дудникова, С. А. Никитин, А. В. Снытников Численное моделирование гравитационных систем многих тел с газом
4	17	2004	Е. В. Никульчев Использование групп симметрий для идентификации сложных систем
5	16	2003	Н. В. Барановский, А. М. Гришин, Т. П. Лоскутникова Информационно-прогностическая система определения вероятности возникновения лесных пожаров
6	15	1998	С. П. Шарый Алгебраический подход во `внешней задаче` для интервальных линейных систем
7	14	2005	В. Б. Барахнин, Ю. В. Леонова Информационная модель отношений между документами в информационной системе
8	14	2007	О.Э. Якубайлик Геоинформационный Интернет-портал
9	13	1999	А. А. Литвиненко, Г. А. Хабахпапев Численное моделирование нелинейных достаточно длинных двумерных волн на воде в бассейнах с пологим дном
10	13	2001	В. А. Вшивков, К. В. Вшивков, Г. И. Дудникова Алгоритмы решения задачи взаимодействия лазерного импульса с плазмой

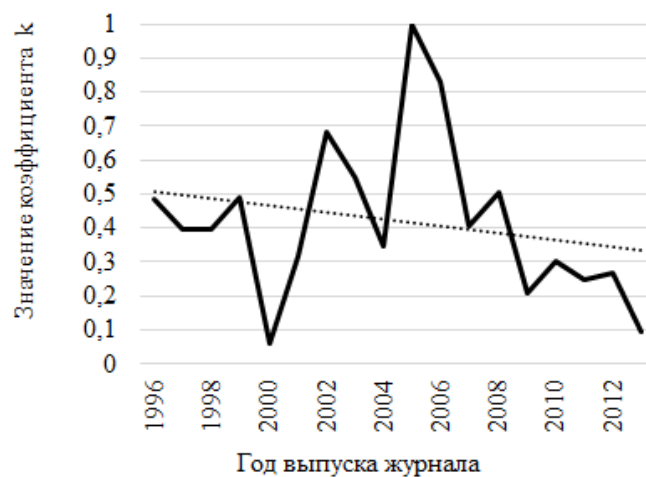


Рис. 7. График зависимости k от года выпуска журнала

Количество статей с нулевым числом цитирований

Год	Регулярный выпуск				Специальный выпуск			
	Всего статей	Кол-во статей с нулевым числом цитир.	Доля нецитир. статей (%)	Среднее число цитир. на статью	Всего статей	Кол-во статей с нулевым числом цитир.	Доля нецитир. статей (%)	Среднее число цитир. на статью
1996	29	13	45	1,7	-	-	-	
1997	58	29	50	1,3	-	-	-	
1998	54	27	50	1,7	-	-	-	
1999	47	21	45	1,8	10	Н/Δ	Н/Δ	
2000	62	54	87	0,5	17	Н/Δ	Н/Δ	
2001	54	30	56	1,4	-	-	-	
2002	60	22	37	2	23	Н/Δ	Н/Δ	
2003	62	26	42	2,1	66	Н/Δ	Н/Δ	
2004	60	32	53	1,8	69	Н/Δ	Н/Δ	
2005	60	17	28	2,8	68	42	62	1,3
2006	59	19	32	2,1	128	71	55	1,2
2007	67	33	49	1,7	56	23	41	2,03
2008	73	32	44	1,7	89	62	70	0,5
2009	67	44	66	1,3	-	-	-	
2010	69	39	57	1,3	-	-	-	
2011	65	40	62	0,6	-	-	-	
2012	52	31	60	0,9	-	-	-	
2013	56	45	80	0,4	-	-	-	
<b>Всего (2005-2008 гг.)</b>	<b>259</b>	<b>101</b>	<b>39</b>	<b>2,05</b>	<b>341</b>	<b>198</b>	<b>58</b>	<b>1,2</b>

Таблица 11

Наиболее успешные специальные выпуски (за 2005–2008 гг.)

№	Название спец. выпуска	Год	Среднее число цитирований на статью	Количество статей в выпуске
1	Труды IX рабочего совещания по электронным публикациям (EI-Pub2004). Новосибирск, 23-25 сентября 2004 г.	2005	4,6	16
2	ГИС- и веб-технологии в междисциплинарных исследованиях. Материалы Междисциплинарной программы СО РАН 4.5.2. Выпуск I	2007	3,15	13
3	Избранные доклады X Российской конференции "Распределенные информационно-вычислительные ресурсы" (DICR-2005), Новосибирск, 6-8 октября 2005 г.	2006	3,1	15
4	Труды Отдела проблем информатизации Томского научного центра СО РАН	2007	2,2	13
5	Специальный выпуск, посвященный 85-летию со дня рождения Н.Н. Яненко, часть 1	2006	1,8	12
6	Информационные технологии для эколого-биологических исследований. Междисциплинарный интеграционный проект СО РАН	2007	1,7	18
7	Избранные доклады семинара по численным методам и информационным технологиям Кемеровского государственного университета	2006	1,6	9
8	Специальный выпуск, посвященный 85-летию со дня рождения Н.Н. Яненко, часть 2	2006	1,5	12
9	Труды IX Всероссийской конференции "Современные методы математического моделирования природных и антропогенных катастроф", 18-21 сентября 2007 г., Барнаул	2008	1,4	12
10	Труды международной конференции и школы молодых ученых "Вычислительно-информационные технологии для наук об окружающей среде" (CITES 2005). Томск, 13-23 марта 2005 г. Часть 2.	2005	1,3	16

### Цитирование англоязычных статей из ЖВТ

За все время в журнале «Вычислительные технологии» опубликованы 173 статьи на английском языке, что составляет 11% от общего количества статей. В связи с тем, что 68 статей приходится на специальные выпуски, опубликованные до 2005 г., данных об их цитируемости в РИНЦ нет. 87 статей не были процитированы ни разу, а общее число цитирования остальных 18 статей по данным РИНЦ равно 35, что составляет около 2% от общего количества цитирования всех статей (1895).

В специальных выпусках была напечатана 91 англоязычная статья. Из них, по имеющимся данным, 19 статей имеют нулевое цитирование, две статьи получили 1 цитирование, одна – 2 и одна – 3. Получаем, что в среднем на одну англоязычную статью в специальном выпуске приходится 0,3 цитирования.

В регулярных выпусках было напечатано 82 англоязычные статьи. Из них 68 имеют нулевое число цитирований, а 13 статей в сумме – 28. Следовательно, на одну англоязычную статью в регулярном выпуске приходится также 0,3 цитирования.

Как было сказано в предыдущем разделе, на одну русскоязычную статью в регулярном выпуске приходится 1,5 цитирований, в специальном – 1,2. Таким образом, статьи на английском языке цитируются в 4 – 5 раз хуже, чем на русском. Скорее всего, это связано с тем, что журнал на текущий момент распространяется только в России и странах СНГ, соответственно его аудитория является преимущественно русскоязычной.

### Цитируемая литература

Данные о литературе, цитируемой в статьях ЖВТ, были также получены из РИНЦ. Всего было процитировано 6705 иностранных и 8638 русскоязычных источников. За все время авторами журнала было процитировано примерно одинаковое количество книг (7513 цитирований) и статей в журналах (7206), 346 интернет-ресурсов и 153 других источников – диссертаций, материалов конференций и пр.

По данным Thomson Reuters существует серьезное различие между российскими и зарубежными публикациями в подходах к оформлению пристатейных списков литературы. Публикации зарубежных соавторов в среднем содержат более 40 ссылок, и это число ежегодно растет. Российские публикации в последние 10 лет в среднем имеют 11–12 ссылок, и ЖВТ этой статистике полностью соответствует.

Разделим количество ссылок в списке литературы на несколько групп по размеру пристатейного списка литературы и рассчитаем среднее число цитирований в каждой группе. Результаты, представленные на рис. 8 и в табл. 12, позволяют сделать вывод, что чем больше ссылок на своих коллег делают авторы, тем больше интереса и/или доверия вызывает их статья, и тем больше она получает цитирований. Статьи без списка литературы имеют минимальные шансы быть процитированными.

Рассчитаем время полужизни («медианный возраст») статей, процитированных в ЖВТ [33]. В табл. 13 строки соответствуют годам выпуска ЖВТ, а колонки – годам выпуска цитируемых источников. Из нее можно увидеть, например, что 50% всех статей, процитированных в ЖВТ в 2011 г., приходится на возраст, равный примерно 12 годам (в колонке «>2000» находится 44% всех статей, значит, необходимо еще 6% статей из колонки

«1991 – 2000». В период 1991 – 2000 гг. в среднем на один год приходится примерно 3% процитированных статей. Следовательно, начиная отсчет статей с 1999 г., 50% статей наберется к 2011 г. примерно за 11 – 12 лет).



Рис. 8. Зависимость среднего числа цитирований от количества ссылок в списке литературы

Таблица 12

### Количество ссылок в списке литературы

№	Группа	Среднее число цитир. на статью
1	0 ссылок	0,1
2	1-2 ссылок	1,3
3	3-4 ссылок	1
4	5-6 ссылок	1,5
5	7-8 ссылок	1,5
6	9-10 ссылок	1,6
7	11-20 ссылок	1,9
8	21-30 ссылок	2
9	>30 ссылок	2,4

По данным за 2012 г. в рейтинге РИНЦ по убыванию показателя «Время полужизни статей, процитированных в журнале за год» ЖВТ занимает первое место в разделе «Информатика», второе – в разделе «Автоматика и вычислительная техника» и лишь 38 в разделе «Математика».

### Место ЖВТ среди журналов с аналогичным профилем

В Российском индексе научного цитирования журнал «Вычислительные технологии» находится в трех тематических рубриках: информатика, математика, автоматика и вычислительная техника. В табл. 14 показаны места ЖВТ по показателям РИНЦ среди журналов аналогичных тематик. В целом позиции ЖВТ среди российских журналов можно назвать очень уверенными. Если рассматривать наиболее характерные показатели – «Пятилетний импакт фактор РИНЦ» и «Среднее число ссылок у статьи в журнале», то по ним в рубриках «Автоматика и ВТ» и «Информатика» журнал находится близко к лидерам.



Доля цитируемых статей по годам (в %)

	<1951	1951-1960	1961-1970	1971-1980	1981-1990	1991-2000	>2000	Время полужизни ( $\pm 1.5$ года)
1996	1	2	8	24	33	33	0	8
1997	2	2	8	19	38	30	0	11
1998	1	1	6	18	34	40	0	10
1999	2	2	9	22	33	32	0	14
2000	2	2	9	19	30	38	0	13
2001	1	2	7	13	26	52	0	9
2002	1	1	6	13	24	56	0	9
2003	2	2	9	15	22	51	0	10
2004	1	1	5	9	18	41	24	8
2005	1	1	5	10	16	36	31	8
2006	1	1	5	7	14	31	41	9
2007	2	2	6	7	12	29	43	8
2008	1	2	4	9	15	25	45	8
2009	1	2	4	7	13	23	50	9
2010	1	2	4	8	14	20	53	8
2011	1	2	7	7	13	26	44	12
2012	1	3	7	8	16	21	44	11

Таблица 14

Место ЖВТ в рейтинге журналов (по данным РИНЦ на 20.10.2014)

Показатель РИНЦ (2012г.)	Значение данного показателя в ЖВТ	Место ЖВТ в рейтинге РИНЦ		
		Математика (всего 159 журналов)	Автоматика и ВТ (всего 141 журнал)	Информатика (всего 42 журнала)
Рейтинг Science Index	4793	23	24	4
Число цитирования журнала за год	333	19	7	3
Пятилетний импакт фактор РИНЦ(с учетом переводной версии)	0,342	34	13	8
Пятилетний импакт фактор РИНЦ без самоцитирования	0,318	24	12	4
Число статей в журнале за год	50	43	58	16
Среднее число ссылок у статьи в журнале	16,4	18	7	7
Число цитирований журнала за год без учета самоцитирований	300	19	7	3

В рубрике «Математика» ЖВТ находится на 34-м месте по показателю «Пятилетний импакт фактор РИНЦ», по показателю «Среднее число ссылок у статьи в журнале» – на 18-м, что объясняется тем, что больше половины журналов, опережающих его, индексируются в БД Web of Science и Scopus (17 из 33 и 10 из 17 соответственно).

#### Заключение

За 18 лет с момента основания в журнале «Вычислительные технологии» опубликовано 1580 статей 1700 ав-

торов из более чем 150 организаций по следующим тематикам: математическое моделирование, интервальный анализ, вычислительная и прикладная математика, компьютерные технологии, геоинформационные системы, электронные библиотеки. Наибольшее количество статей принадлежит ученым из Новосибирска, среди зарубежных ученых лидерами по количеству публикаций являются ученые из Казахстана.

Доля межинститутских публикаций составляет 55% от общего количества публикаций, англоязычных статей – 10%, статей с иностранным участием – 6%, статей совме-

стно с вузами – 31%. Зафиксировано увеличение публикаций, подготовленных совместно с вузами, количество иностранных статей и статей с иностранным участием после заметного всплеска в 2000-2004 гг. уменьшается.

В рамках заявляемой тематики ЖВТ занимает лидирующие позиции среди других российских журналов, уступая, в основном, переводным журналам, входящим в международные базы цитирования.

При этом авторы этой статьи считают необходимым рекомендовать редакционной коллегии журнала внести в редакционную политику определенные изменения, направленные на повышение качества принимаемых статей:

1. Определить минимально допустимый размер приростов списка литературы. Рекомендуемый размер – не менее 20 ссылок.

2. Ограничить «самоцитирование» авторов: не более 10-20% процентов ссылок на свои статьи.

3. Отслеживать коэффициент самоцитирования журнала и не допускать его превышение 15% -го порога.

4. Крайне избирательно относиться к изданию специальных выпусков в виду их низкой цитируемости.

5. Ввести годовую квоту на количество статей от сотрудников базовой организации ИВТ СО РАН. Рекомендуемый размер квоты – 10-20% от общего числа статей.

6. До начала распространения журнала за рубежом избегать публикации статей на английском языке, которые в настоящее время цитируются в 4-5 раз хуже, чем на русском.

7. Разработать меры дополнительного стимулирования подготовки обзорных статей, которые пользуются большей популярностью, по сравнению с традиционными статьями.

8. Рассмотреть возможность уменьшения времени «задержки» с момента выхода статьи до ее публикации в открытом доступе до одного года, либо полностью перейти на распространение по принципу Open Access.

9. Использовать международные конференции, как площадки для формирования корпуса зарубежных авторов и укрепления международного сотрудничества.

Уверенные позиции журнала «Вычислительные технологии» позволяют говорить о том, что дальнейший рост наукометрических показателей (в первую очередь, количества цитирований) и выход на ведущие позиции должны быть связаны с переходом в новое качество – выпуска переводной версии журнала (на английском языке), с последующим включением журнала в международные наукометрические базы данных Web of Science и Scopus. В случае принятия такого решения, будет необходимо проводить регулярную работу по привлечению ведущих зарубежных специалистов к публикации в журнале и работе в составе редакционной коллегии.

*Благодарность.* Авторы благодарят ООО «Научная электронная библиотека» (elibrary.ru) и лично Григория Зацмана за предоставленные данные о цитируемости журнала «Вычислительные технологии».

## Литература

1. *Зибарева И.В., Писляков В.В., Теплова Т.Н., Нефедов О.М.* Библиометрический анализ журнала «Успехи химии» // Вестник РАН. – 2008. – Т. 78. – № 6. – С. 490–499.

2. *Pudovkin A.I., Fuseler E.A.* Indices of journal citation relatedness and citation relationships among aquatic biology journals // *Scientometrics*. – 1995. – Vol. 32, No. 3. – P. 227–236.

3. *He T., Zhang J., Teng L.* Basic research in biochemistry and molecular biology in China: A bibliometric analysis // *Scientometrics*. – 2005. – Vol. 62, No. 2. – P. 249–259.

4. *Biglu M., Eskandari F., Asgharzadeh A.* Scientometric Analysis of Nanotechnology in MEDLINE // *BioImpacts*. – 2011. – Vol. 1, No. 3. – P. 193–198.

5. *Hullmann A.* Measuring and assessing the development of nanotechnology // *Scientometrics*. – 2007. – Vol. 70, No. 3. – P. 739–758.

6. *Мазов Н.А.* Библиометрический анализ журнала «Геология и геофизика» как основа для оценки исследований в области наук о Земле // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса. Крым, 2011. – С. 1–3.

7. *Bebrens H., Luksch P.* Mathematics 1868–2008: A bibliometric analysis // *Scientometrics*. – 2010. – Vol. 86, No. 1. – P. 179–194.

8. *Marshakova-Shaikovich I.* Russian mathematical journals in world and national corpora of scientific journals: Bibliometric analysis // *Sociol. Sci. Technol.* – 2012. – Vol. 3, No. 2. – P. 79–101.

9. *Per O.S., Aksnes D.W.* Scientific productivity and group size: A bibliometric analysis of Norwegian microbiological research // *Scientometrics*. – 2000. – Vol. 49, No. 1. – P. 125–143.

10. *Abt H. A.* The future of single-authored papers // *Scientometrics*. – 2007. – Vol. 73, No. 3. – P. 353–358.

11. *Zhou F., Guo H. C., Ho Y. S., Wu C. Z.* Scientometric analysis of geostatistics using multivariate methods // *Scientometrics*. – 2007. – Vol. 73, No. 3. – P. 265–279.

12. *Marshakova-Shaikovich I.* Bibliometrik maps of scientific collaboration of EU countries in science and social Science // *Sociol. Sci. Technol.* – 2010. – Vol. 1, No. 2. – P. 13–16.

13. *Anuradha K.T., Urs S.R.* Bibliometric indicators of Indian research collaboration patterns: A correspondence analysis // *Scientometrics*. – 2007. – Vol. 71, No. 2. – P. 179–189.

14. *Chuang K.-Y., Huang Y.-L., Ho Y.-S.* A bibliometric and citation analysis of stroke-related research in Taiwan // *Scientometrics*. – 2007. – Vol. 72, No. 2. – P. 201–212.

15. *Rajendran P., Jeysankar R., Elango B.* Scientometric analysis of contributions to journal of scientific and industrial research dr. P. Rajendran // *Int. J. Digit. Libr. Serv.* – 2011. – Vol. 1, No. 2. – P. 78–161.

16. *Клязева С.Ю., Слащева Н.А.* Научно-техническое сотрудничество России и ЕС: библиометрический анализ // Форсайт. – 2008. – Т. 1. – № 5. – С. 30–41.

17. *Tiew W.S.* Characteristics of self-citations in Journal of Natural Rubber Research 1988–1997: A ten-year bibliometric study // *Malaysian J. Libr. Inf. Sci.* – 2000. – Vol. 5, No. 1. – P. 95–104.

18. *Liang L., Ronssean R., Zhong Z.* Non-English journals and papers in physics and chemistry: Bias in citations? // *Scientometrics*. – 2012. – Vol. 95, No. 1. – P. 333–350.

19. *Rattan G.K.* Self-citations in Annals of Library and Information Studies // *Libr. Philos. Pract.* – 2013. – Vol. 6, No. 1.

20. *Wang S., Wang H., Weldon P.R.* Bibliometric analysis of English-language academic journals of China and their internationalization // *Scientometrics*. – 2007. – Vol. 73, No. 3. – P. 331–343.

21. *Abt H. A.* The frequencies of multinational papers in various sciences // *Scientometrics*.— 2007.— Vol. 72, No. 1. — P. 105–115.
22. *Main features of CERIF*. — EuroCRIS. — URL: <http://www.eurocris.org/cerif/main-features-cerif>
23. *Liu Z.* Paper to digital: Documents in the information age.— USA: Libraries Unlimited, 2008.
24. *Harzing A.-W.* The publish or perish book, part 3.— USA: Tarma Software Research, 2011.
25. Юрий Иванович Шокин (к 70-летию со дня рождения) // *Вычислительные технологии*. — 2013. — Т. 18. — № 3. — С. 110–113.
26. *Налимов В.В., Мульченко З.М.* Наукометрия. Изучение развития науки как информационного процесса. — М.: Наука, 1969.
27. *Хайтун С.А.* Наукометрия. Состояние и перспективы.— М.: Наука, 1983.
28. *Bradford S.C.* Sources of information on specific subjects // *Lect. notes Comput. Sci.* — 1985. — Vol. 10, No 4. — P. 173–180.
29. *Тесла А.* Процесс отбора журналов. — Thomson Reuters. — URL: [http://wokinfo.com/media/essay/journal\\_selection\\_essay-ru.pdf](http://wokinfo.com/media/essay/journal_selection_essay-ru.pdf)
30. *Schubert A., Glanzel W.* Mean response time — A new indicator of journal citation speed with application to physics journals // *Czechoslovak Journal of Physics B*. — 1986. — Vol. 36, No. 1. — P. 121–125.
31. *Egghe L.* A heuristic study of the first-citation distribution // *Scientometrics*. — 2000. — Vol. 48, No. 3. — P. 345–359.
32. *Гарфилад Ю.* Можно ли выявлять и оценивать научные достижения и научную продуктивность? // *Вестник АН СССР*. — 1982. — № 7. — С. 42–50.
33. *Писляков В.В.* Методы оценки научного знания по показателям цитирования // *Социологический журнал*. — 2007. — № 1. — С. 128–140.

Приглашаем российских и зарубежных авторов к сотрудничеству  
в журнале «Международный форум по информации».  
Оригинальные статьи и другие материалы (рецензии, письма)  
можно присылать на русском или английском языке  
по почтовому адресу, указанному в «Памятке для авторов»  
или по электронной почте: [mfi@viniti.ru](mailto:mfi@viniti.ru).

Ответственный за выпуск *Л. В. Кобзева*

Компьютерная верстка *М. А. Филимонова*

ИД № 04689 от 28.04.2001 г.

Подписано в печать 11.09.2015 г.

Бумага офсетная. Формат 60x841/8. Гарн. литер. Печать цифровая

Усл. печ. л. 5,50 Уч.-изд. л. 5,66 Тираж 34 экз.

Адрес редакции: 125190, Россия, г. Москва, ул. Усиевича, д. 20

Тел. (499) 155-44-95