ФАКТОЛОГИЯ РЕЗУЛЬТАТОВ АВТОМАТИЧЕСКОГО АНАЛИЗА ПОТОКА НАУЧНЫХ СТАТЕЙ НА ОСНОВЕ ELIBRARY И РИНЦ

Н.А. Резник, доктор педагогических наук, профессор, ведущий научный сотрудник
Лаборатория науковедческих и наукометрических исследований,
Федеральное государственное научное учреждение
«Институт научной и педагогической информации» Российской академии образования (Москва), Россия

Аннотация. Статья продолжает представление материалов НИР «Информационный поток публикаций в российских научных и научно-методических журналах за 2006–2011 годы по методике использования информационных ресурсов в преподавании математики».

Ключевые слова: тематические запросы, индекс Хирша, информация, математика, методика обучения, научные статьи, подборки, поток, списки публикаций.

В данной статье рассматривается «недокументированная» и практически неизученная возможность проведения в РИНЦ анализа потоков научных публикаций по определённому тематическому запросу. В ней излагаются факты, обнаруженные в РИНЦ в период с 22.09 по 22.11 2013 года по позициям: формирование тематического массива данных, автоматизированная оценка результатов.

За основу значения слова ПОТОК автором было принято его толкование в словаре русского языка: «Поток — непрерывное движение массы, большого количества ... чего-л.». В соответствии с этим в данное исследование были положены следующие позиции: число «опускаемых» в ходе статистической обработки результатов экспериментов в педагогических исследованиях.

1. Для создания потока научных статей необходима генеральная совокупность, пригодная для извлечения из неё «хороших» выборок. Такая совокупность должна быть достаточно велика, а выборки, подвергаемые испытанию, — действительно случайны.

2. При репрезентативности (сохранении основных признаков генеральной совокупности) потоки главным в его анализе должны являться отделение в нём массива статей методического содержания по разным предметам (по горизонтали) в разные периоды (по вертикали).

Позиция 1. Поиск генеральной совокупности

Одной из попыток построить искомый поток был вызов статей по запросу Математика. Начальная страница вывела статьи, с содержанием недоступным во всем НЭЗ (рис. 1.1). После Расширенного поиска, выдавшего 12552 статьи, их осталось немного больше третьи. При этом на первых страницах подборок все статьи оказались закрыты (рис. 1.2). Релевантность в представлении статей также не была соблюдена.

В Поиск и Расширенный поиск вводились и вариации наименования специальности 13.00.02 «Теория и методика преподавания и обучения (математика)» в целях обнаружения наличия значительных различий или совпадений в них. Решили в последнем случае (даже при открытых текстах), согласно одному из феноменов информационных потоков [6] явились:

- «паразитная (ненасыщированная, невостребованная)» информация;
- «несоответствие формально релевантной (...относящейся к делу) информации действительною потребностям» вводимого в Поиск РИНЦ тематического запроса.
На рисунке 2 данные образцы просмотра результатов Рассширенного поиска в пяти подобных подборах по:
- применению словоформы математи в интервале 2006–2011 годы (рис. 2.1, Поиск) и демонстрации примеров того, что уделялось вручную (рис. 2.1, Просмотр);
- удаленности из-за очевидной недоступности (рис. 2.2 и 2.5), явной ненаучности (рис. 2.3), отказе программы в просмотре (рис. 2.4),
- при непосредственном знакомстве с их содержанием (рис. 2.1, Чтение).

Рис. 2. Демонстрация результатов построения пяти подборок научных статей по запросу «Теория и методика обучения и воспитания математике»

При переходе к построению генеральной совокупности в основу были положены предметные словоформы из наименований разделов этих дисциплин (например, Площадь фигур, Тела вращения, Точки и прямые, Уравнение и неравенство) и важнейших понятий (например, Модуль и радиус, Степени и корни, Цифры и Числа, Многочлен, Дроб, Иррациональные числа и комплексные числа, Функция и рекуррентные зависимости).

Максимальная генеральная совокупность была составлена объединением 23-х подборок разного вида, полученных путем введения запросов для них в Поиск РИНЦ (рис. 3.1, А).
Рис. 3. Пример преобразования Генеральной совокупности в Популяции научных статей

Здесь использовались словоформы типа Анализ, Арифмет, Вероятн, Геометр, Комбинац, Логик, Модел, Прикл, Пропед. Статист, Тополог. Физик. Цифр и Числ (рис. 3.1.В).

По Расширенному поиску полученные результаты преобразовывались в новые подборки по «маршрутах»:

- общая подборка «!_ГЕНЕРАЛЬНАЯ_ все запросы» (рис. 3.1.А);
- в подборку «_ГЕНЕРАЛЬНАЯ_ статьи_06-11_математик» (рис. 3.2);
- в подборку списка Д (рис. 3.1.D) каждая из 23-х тематических подборок из списка В (рис. 3.1.B)

Далее по повторному Расширенному поиску было формирование подборок по годам (рис. 3.3). Разница в количествах статей, переданных в подборки 2006_статьи, ..., 2011_статьи, оказалась столь незначительной, что ею стало возможным пренебречь.

Результатирующий анализ полученных подборок шёл вручную (рис. 3.4).

Сначала удалялись статьи с недоступным содержанием и с наименованиями, не согласующимися с тематикой основного запроса.

Затем шёл просмотр содержания статей. Здесь отсев осуществлялся только случаи, когда статьи ни по каким параметрам не подходила в разряд «методического». В завершении «запускался» Анализ публикаций в подборке (рис. 3.5).

Позиция 3. Описываемый репрезентативный выборок

Была исследована реакция Поиска РИНД и на более узкие вопросы типа: Алгебр, Пропедевт, Таблицы умнож, Арифмет задач, Аксиом и постановок, Анализ, Арифм, Вариан интегр, Высш алгебр, Дифференц, Интерг, Логик, Операция, Ряды, Множества и другие. Эти же словоформы входили в Поиск с добавками типа метод, препод, обуч, в различных сочетаниях, соединениях и порядках, а также с добавлением к ним уровня образования (нуд, колледж, школа) и уточнениями (начал, сред, старши, высш, подготов). В качестве примеров даны фрагменты подборок по целевым установкам (Таблице умножение) и прикладной направленности (Методика обучения алгебре).
А. СОДЕРЖАТЕЛЬНЫЙ АСПЕКТ
Смысловое значение словоформ, вводимых в Пояск, при выдаче результата должно быть приоритетным, обеспечивая его релевантность. Поэтому ожидается увидеть в большом количестве слова таблица и умножение уже на первой странице подборки Таблицное умножение. На деле же присутствовали они в наименованиях только двух статей с закрытым содержанием: №2 в №3 (рис. 4.1).

Рис. 4. Фрагменты результатов поиска публикаций по теме Таблицное умножение

При просмотре других страниц (номера: 5, 10, 15, 20, 25, 30, 40) этой же подборки также было замечено большое количество публикаций, недоступных непосредственному прочтению. Точно такая же картина наблюдалась и в подборках с большим количеством слов: Решение задач на табличное умножение (рис. 4.2) и Методика решения задач на табличное умножение в начальной школе (рис. 4.3).

В. ПРИКЛАДНАЯ НАПРАВЛЕННОСТЬ
Математика в учебные планы высших учебных заведений введена и для «технарей» и для «гуманитария». Приложения математики «обслуживают ВСЕ», и каждый случай – это отдельная «ептаня». Составить же представление о потоках в 2006–2011 годы с помощью РИИЦ оказалось слишком сложно. Доступ к содержанию многих публикаций (составляемых автором многочисленных потоков) отсутствует (в примере на рис. 5 отмечено красным цветом и подчеркнуто). Для открытых же статей (отмечены синим цветом) из этого же списка по их наименованиям (вполне возможно отнести к обучению по специальностям):
№1 и №3 – к 31.00.00 – Химия.
№7 – к 10.00.00 – Государство и право. Юридические науки.
№9, №20 и №27 – к 14.00.00 – Народное образование. Педагогика (14.25.09 – Методика преподавания учебных дисциплин в общобразовательной школе).
№10 – к 73.00.00 – Транспорт.
№11 и №16 – к 06.00.00 – Экономика. Экономические науки.
№14 – к 87.00.00 – Охрана окружающей среды. Экология человека.
Рис. 5. Фрагмент результатов поиска публикаций по теме «Методика обучения алгебре в вузе»

По наименованию же их можно предполагать: по Рубрикатору в РИНЦ всё это касается пункта 14.35.09 – Методика преподавания учебных дисциплин в высшей профессиональной школе (где почему-то лишь 6 статей) раздела 14.35.00 – Высшее профессиональное образование. Педагогика высшей профессиональной школы.

Для статей же №13 и №21 (обучение инженеров) и №23 №29 (обучение иностранных граждан) аналогичное определить не удалось.

Недочёты – повторы одних и тех же публикаций (рис. 6.1) и неряшливость в оформлении (рис. 6.2) – на экранах РИНЦ разнообразны. Их много и в заголовках статей, и в их содержании (рис. 6.3).

Рис. 6. Примеры недочётов на экранах РИНЦ
Содержательных ошибок в открытых доступу материалах также предостаточно.
Например, в последней подборке (рис. 6) только за один сутки (с 12 по 13 ноября сего года) обнаружилась публикация, содержание которой вновь не открылось (рис. 7.1-2), а ночью была выведена статья того же автора, но совершенно иного содержания (рис. 7.3).

Рис. 7. Примеры ошибок в представлении содержания статей в РИНЦ

В статье другого автора (рис. 7.4) содержание её было выведено на экран без наименования в начале, колоннитуле и в конце (рис. 7.5). Есть и пустые схемы (рис. 7.6). Довольно часто РИНЦ отказывается открывать содержание статьи, которая обозначена как открытая (рис. 7.7).

С. ИНВОЛЮЦИЯ ПОТОКА

Ниже приведены таблицы, полученные в результате вызова и просмотра доступных статей из генеральной совокупности !!_генеральная статьи_2006-2011_математик, составленной из совокупности по 23 различным запросам (рис. 3.5).

Всего в этой совокупности присутствуют: статей – 2350, журналов с учётом которых и томов – 1795, наименований журналов – 635. Показан этот процесс для открытых статей на рис. 8.

Здесь:
— слева – данные потока при очерёдном (по имеющимся неизменным по параметрам) Расширенном поиске в РИНЦ;
— справа – идут примеры журналов, открытые статьи из которых попали в эти потоки.

Лучшего слова, чем инволюция, для происходящего в этих подборках, найдено не было.
Например: 2006_статьи_открыто (рис. 3.5).
Далее (рис. 8.1-3, верхние строки):
2006_статьи_открыто – всего статей 87 – отбор автоматический (рис. 8.1);
2006_статьи_матем – всего статей 8 – отбор вручную (рис. 8.2);
2006_статьи_матем_метод – всего статей 3 – результат просмотра (рис. 8.3).

Предварительный погр:
Содержание публикаций, попадающих в искомый поток благодаря усилиям РИНЦ, соответствует смыслу запроса в его Поиск далеко не всегда.
ЗАКЛЮЧЕНИЕ

Методика преподавания математики (как и любого другого учебного предмета) как специальная отрасль соответствующего Научного Знания есть наука. Но связана она не столько с 'передачей' её открытий и достижений человеческому сообществу, сколько с переработкой научных знаний в учебное, с процессом адаптации идей и открытых для постижения темы, кого они интересуют и кому они нужны. То же, что было описано выше, относится и к построению генеральной совокупности и возможности извлечения из неё валидных выборок. То есть, предлагаемый автором подход подразумевал поиск ответа на вопрос: «возможно ли в настоящее время их построение в РИНЦ?»

В этом плане самым важным моментом в подведении итогов была «степень, в которой результаты исследования, системы мерений или статистики являются точными или представляют то, что были предназначены представлять». Иначе говоря, «показатель качества метода, его способность давать результаты, адекватно отражающие изучаемое явление» [1].

В данной работе за основу была взята «учебная знаковая информация, которую есть система знаков (слов, формул, иллюстраций), призвана помочь учащимся воспринимать материал, побудить их к поиску решений возникающих проблем». В методике обучения и воспитания математике «это особенно важно — мы должны ориентироваться на значительные усилия самого ученика и не столько строить его запас сведений, сколько организовывать, представлять возможности для их получения [7, с. 329]. Поэтому при окончательном составлении потоков по школе и вузу не принимались во внимание статьи, содержание которых явно и полностью сосредоточено на методическом аспекте преподавания предмета. (Например, психологический, культурологический, психологический и другие).

Предполагалось, что в силу релевантности, РИНЦ в первую очередь будет представлять публикации типа научные статьи, диссертации и монографии, содержание которых непосредственно или опосредованно связано:
с методикой обучения математике;
с апробацией новых подходов к передаче учебных знаний и формированию учебных математических ЗУНов;
с экспериментами по внедрению новых средств обучения и воспитание математике.

Сообразно этому следовало бы решение задачи построения и анализа потока научных статей по проблемам обучения (преподавания), ориентируясь на содержание учебных дисциплин той или иной конкретной предметной области того или иного среднего или высшего учебного заведения, с учётом дальнейшей дифференциации по исходным целям.

Например:

для школы — хотя бы по уровням: начальная школа, основная школа, старшие классы школы;
для вуза — хотя бы по направлениям:
преподавание математики в любом вузе,
методика преподавания математики в педагогическом вузе.

Сделать это оказалось невозможным (см. строчку Общее число публикаций на рис. 9, 1-2).

За время подготовки материалов и написания содержания данного текста было сделано около 100 тематических подборок. Из-за невозможности сохранить их в РИНЦ в первоначальном виде, приходилось их постоянно перепроверять или же отбрасывать.

Утверждения (в разделе О проекте) о том, что большая часть научной информации представлена в РИНЦ «в полнотекстовом виде… в открытом доступе, что позволяет в большинстве случаев ознакомиться с текстом и вынести оценку» на практике не подтвердились.

Более того, как сообщают журналы (например, журнал «Математика в школе») Расширенный Поиск в РИНЦ не нашел (рис. 9.3). Малое количество получаемых в каждой из подборок научных статей в этих условиях ликвидировало всякую возможность не только качественного анализа, но и составления самого потока.

Наличие благополучного завершения данной статьи (когда это исследование сроди только начиналось) было основано на материалах Internet. РИНЦ неоднократно привлекает внимание специалистов из разных областей Научного Знания [2, 3 и 5] (рис. 10.1-2).

Итог:
Предполагалось, что обнаруженные ими ошибки в идее, концепции и реализации РИНЦ уже устранены.
В то же время, в частности, что ещё 26.09.10 руководитель РИНЦ сообщил,

Однако этого не произошло.
Запланированный в начале этой статьи задача «Рассмотрение «неудобной» возможности проведения в РИНЦ анализа потоков научных публикаций по определённому тематическому запросу», осталась нерешённой.
Рис. 10. Рекомендации разработчикам РИИЦ.

Рис. 11. Информация о РИИЦ в 2010 году.

СПИСОК ЛИТЕРАТУРЫ

FACTS OF AUTOMATIC ANALYSIS 
OF THE FLOW OF SCIENTIFIC ARTICLES BASED ON ELIBRARY AND RSCI

N.A. Reznik, Doctor of Pedagogical Sciences, Professor, Senior Researcher 
Laboratory of Scientific and Scientometrical Researches, Federal State Scientific Institution 
"Institute of Scientific and Pedagogical Information" of the Russian Academy of Education (Moscow), Russia

Abstract. The article continues the representation of the research work materials "Information Flow of Publications in the Russian Scientific and Scientific and Methodical Magazines for 2006-2011 by a Technique of Use of Information Resources in Teaching Mathematics".

Keywords: thematic inquiries, h-index, information, mathematics, teaching technique, scientific articles, selections, stream, lists of publications.