

# ПОИСК НАУЧНЫХ ПУБЛИКАЦИЙ ДЛЯ ПРЕДМЕТНЫХ ОБЛАСТЕЙ С НЕУСТОЯВШЕЙСЯ ТЕРМИНОЛОГИЕЙ

Е.И. Моисеев, А.А. Муромский, Н.П. Тучкова

## ВВЕДЕНИЕ

Проблема поиска научных публикаций в интернет неисчерпаема, хотя за последние годы ситуация кардинальным образом изменилась благодаря огромным успехам таких баз данных как, Zentralblatt ([www.zentralblatt-math.org](http://www.zentralblatt-math.org)), ISI WEB Knowledge ([isiknowledge.com](http://isiknowledge.com)), SCOPUS ([scopus.com](http://scopus.com)), РИНЦ ([elibrary.ru](http://elibrary.ru)), Math-Net.ru ([www.mathnet.ru](http://www.mathnet.ru)) и др. Научному сообществу во всем мире стали доступны огромные массивы журнальных публикаций и архивов, о чем и мечтать было невозможно 20-30 лет назад. Все это способствует, несомненно, ускорению научного процесса, усилению конкуренции между отдельными авторами, организациями, журналами. Есть условия для сбора статистики, выявления приоритетных направлений в исследованиях и лидирующих групп в науке. Однако, ни для кого не секрет, что англоязычные ресурсы, включая научные публикации, представляют собой абсолютное большинство. Это объясняется, как историческими причинами, например, проекты Института Научной Информации США (Institute for Scientific Information, ISI, [isiknowledge.com](http://isiknowledge.com)) [1] по сбору коллекций научных публикаций осуществляются с середины XX в., так экономическими, недофинансирование таких проектов, например, для русскоязычных публикаций приводит их, в конечном счете, к малой конкурентоспособности.

## ПРОБЛЕМА ПОИСКА В МЕЖДИСЦИПЛИНАРНЫХ ПРЕДМЕТНЫХ ОБЛАСТЯХ

Особое значение в этих условиях приобретают русскоязычные ресурсы для разделов наук с неустоявшейся лексикой. В эпоху междисциплинарных исследований, которая наступает, когда наука тесно соприкасается с технологией, возникает пересечение предметных областей, что влечет за собой неизбежно смешивание терминологии. Междисциплинарные области науки и технологии стали называться сложными и неоднозначными терминами (наномеханика, наномедицина, биоинформатика и др.), под которыми каждый исследователь понимает то, что его интересует. Всем понятно, например, что нанонаука сейчас наиболее обширная предметная междисциплинарная область, включающая фундаментальные и прикладные исследования целого ряда традиционных (классических) предметных областей [2]. Тем не менее, современному исследователю надо в этом всем как-то ориентироваться, чтобы двигаться вперед в своей конкретной области, а именно, знакомиться с отечественными и зарубежными публикациями. Для реализации поиска в информационных ресурсах научных журналов, различного рода тематических коллекций (архивов, музеев и др.) используются известные признанные классификаторы, такие, как: УДК (<http://www.viniti.ru>), MSC 2010 (<http://www.ams.org/msc/>), DDC (Dewey Decimal Classification, <http://www.oclc.org/dewey/resources/summaries/>) и собственные классификаторы, используемые, например, в INSPEC, Zentralblatt и др. Нано-дисциплины не являются в этом смысле исключением, для них создается множество классификационных ресурсов, таких как добавления в MSC, INSPEC ([inspec.com](http://inspec.com)), Zentralblatt, IUPAC ([iupac.org](http://iupac.org)) и самостоятельных, таких как словари ([nanodic.com](http://nanodic.com), [nanodictionary.net](http://nanodictionary.net), [nanodictionary.com](http://nanodictionary.com)), в том числе и двуязычные версии ([portalnano.ru](http://portalnano.ru), <http://dic.academic.ru/contents.nsf/nanotechnology>) [3].

## ОРГАНИЗАЦИЯ ПОИСКА ПУБЛИКАЦИЙ ДЛЯ ПРЕДМЕТНОЙ ОБЛАСТИ «НАНОНАУКА»

В настоящей работе предлагается один из вариантов организации поиска информации с использованием разработанных авторами терминологических ресурсов по фуллеренам [4] и нанотрубкам, как разделам предметной области наномеханика. Разработанные ресурсы не претендуют на собственную классификацию, но составляют дополнение к имеющимся русскоязычным и двуязычным словарям (например, <http://www.portalnano.ru/read/tezaurus>)

Основу алгоритма составляют «связки», которые задаются благодаря тезаурусным статьям словаря, что позволяет по ключевому слову из запроса пользователя выбрать соответствующее множество терминов и далее использовать их как ключевые слова для поиска публикаций.

Каждая статья словаря (тезауруса) в электронном варианте содержит соответствующие ключевые слова (на русском языке в основном) и снабжена аббревиатурой [4].

Пример состава статьи словаря для термина «Аминокислоты»:

Аббревиатура: А32

Термин: Аминокислоты

Термин ассоциативный (аббревиатура): Б12

Первоисточник (библиографические ссылки в мнемонической кодировке): 208 4.15. 1(2006)(txb)

Примечание (текст, гиперссылки, библиографические ссылки, аббревиатуры связанных статей словаря и пр.): Аминокислоты – органические кислоты, содержащие одну или несколько аминокрупп, а также карбоксильные группы. Карбоксильная группа (карбоксил) – (COOH).

Аминогруппа – соединение, содержащее группу NH<sub>2</sub> (см. 15 8(2009)(dic))

Каждое ключевое слово и аббревиатура статьи образуют при запоминании «связку», содержащую, таким образом, две части:

первая часть связки – ключевое слово,

вторая часть связки – аббревиатура статьи

Эти части при поиске выдаются пользователю вместе, т.е. выдается связка только целиком, а не по частям.

Пример: металлическая проводимость УНТ ↔ K5

Связка содержит только одно ключевое слово, а вторая часть может содержать более одной аббревиатуры статьи (один из вариантов представления связки).

Другой вариант: одно ключевое слово сопровождается только одной аббревиатурой статьи.

Пример: металлическая проводимость УНТ ↔ K5

металлическая проводимость УНТ ↔ K3

Связки имеют одинаковые первые части, но разные вторые части.

Все связки объединяются в некоторые множества:

признак объединения – одинаковая первая буква ключевого слова - первой части связки.

Эти множества удобно различать именно по первой букве первой части.

Пример:

Lnk M

металлическая проводимость УНТ ↔ K5

металлическая проводимость УНТ ↔ K7

В итоге имеем последовательность множеств связок в общем случае:

Lnk A, Lnk B, Lnk B, Lnk Г, ..., Lnk N, N = A, B, B, Г, ...

Все эти множества хранятся и обновляются, а при поиске в зависимости от ключевых слов поискового предписания (запроса пользователя) вызываются для сравнения с ключевыми словами публикации и отбора соответствующих публикаций.

При поступлении поискового запроса алгоритм поиска реализуется следующим образом:

1. Шаг - сравнение поискового запроса с терминами, входящими в определенную алфавитную группу для выявления множества соответствующих связок

2. Шаг – выборка терминов на основе списка связок для получения списка ключевых слов, соответствующих запросу

3. Шаг – сравнение списка ключевых слов с ключевыми словами публикаций для получения списка релевантных документов

4. Шаг – выбор пользователем документов для просмотра из полученного списка релевантных документов

## ЗАКЛЮЧЕНИЕ

В работе предлагается использовать средства тезаурусного представления междисциплинарной предметной области для организации поиска публикаций. Алгоритм основан на связях терминов, которые задаются в статьях словаря, составленного в основном по русскоязычным первоисточникам (разделы наномеханики: фуллерены, нанотрубки). Для представления терминов словаря использованы первоисточники, в которых даются методы получения фуллеренов и нанотрубок и их применение в различных областях науки и технологии. Наличие связей позволяет выявить группы терминов внутри словаря, которые могут использоваться (на основе поискового запроса) как дополнительные ключевые слова для сравнения с ключевыми словами публикаций. Особенность предложенного подхода заключается в том, что он позволяет учесть междисциплинарные связи терминов. Появление новых терминов из смежных предметных областей не нарушает структуры представления междисциплинарной предметной области, а влечет за собой расширение списка связок. При этом учитывается происхождение термина (ссылка на первоисточник) и его распространение (ссылки на использование в приложениях к различным дисциплинам). Таким образом, добавление нового термина влечет за собой добавление ссылки на соответствующие публикации.

По мере развития словаря пополняется список статей и связок. В результате использования множества таких выборок, сокращается количество шума при поиске и уменьшается время поиска пользователем pertinentного [5, 6] документа.

*Работа поддерживается Российским фондом фундаментальных исследований (проекты № 10-07-00040-а, 11-01-12081-офи-м-2011).*

#### ЛИТЕРАТУРА:

1. Cawkell T., Garfield E. Chapter 15. Institute for Scientific Information // A century of science publishing: a collection of essays / Einar H. Fredriksson (Ed.). IOS Press, 2001. С. 149-160. 312 с. ISBN 9781586031480.
2. Williams L., Adams W. Nanotechnology demystified. A self-teaching guide. McGraw-Hill. 2007. 343 с. ISBN 978-0071460231.
3. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Междисциплинарность в системе интернет: представление предметных областей с неустоявшейся терминологией // Труды Международной суперкомпьютерной конференции «Научный сервис в сети Интернет: эксафлопсное будущее», Абрау-Дюрсо, 19-24 сентября, 2011. С. 454-456. ISBN 978-5-211-06229-0
4. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Фуллерены в терминах: Словарь / Отв. ред. С.А. Лурье. М.: МАКС Пресс, 2010. 152 с. ISBN 978-5-317-03548-8
5. Михайлов А.И., Черный А.И, Гиляревский Р.С. Основы информатики. М: Наука, 1965. 655 с.
6. Гиляревский Р.С., Шапкин А.В., Белоозеров В.Н. Рубрикатор как инструмент информационной навигации. СПб.: Профессия, 2008. 352 с.