

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В БИБЛИОТЕКАХ

УДК 026:5+026.6

**Е. В. Ткачёва**

*БЕИ РАН*

## **Обработка постояннодействующих тематических запросов по отдельным таксономическим группам средствами базы данных Web of Science**

На примере тематических запросов по отдельным таксономическим группам, выполняемых для пользователей Библиотеки по естественным наукам РАН, рассмотрены различные ситуации, влияющие на пертинентность результатов поиска по базе данных *Web of Science*. Сформулированы общие рекомендации по составлению таких тематических запросов с использованием поисковых средств БД *Web of Science*, необходимо: а) учитывать синонимичные названия таксонов и неоднозначность определения рангов одних и тех же таксонов разными авторами; б) исключить непрофильные предметные области из результатов поиска, что позволяет отсеять омонимичные термины.

Работа выполнена при поддержке гранта РФФИ № 16–07–00450.

**Ключевые слова:** Библиотека по естественным наукам РАН, избирательное распространение информации, информационный поиск, поисковый запрос, пертинентность результатов поиска, база данных *Web of Science*.

---

UDC 026:5+026.6

**Ekaterina Tkacheva**

*Russian Academy of Sciences Library for Natural Sciences, Moscow, Russia*

## **Processing continuous subject queries in individual taxonomic groups with Web of Science tools**

The author analyzes various situations that influence the pertinence of *Web of Science* database search based on subject queries in several taxonomic groups (the case study of RAS Library for Natural Sciences users). General recommendations on formulating these subject queries using *Web of Science* search instruments are formulated: а) synonym taxon name and ambiguity of taxon range definition by different authors; б) exclusion of non-profile subject areas from search results enables to deselect homonymic terms.

The paper is prepared through the support of RFFR Grant № 16–07–00450.

**Keywords:** Library for Natural Science of the Russian Academy of Sciences, selective distribution of information, information search, search query, pertinence of results, Web of Science database.

---

In this article, we pay attention to the use of the Web of Science database for its original purpose: to find information on a particular topic. We will discuss constantly active searches of new research results which are then transmitted on a regular basis to the interested users (alarm service). One of the situations of thematic search in the biological sciences is searching for information on specific taxonomic groups of organisms. In the simplest case, the user may be interested in articles with floristic or faunal content. As an example, we give the request formulated by the reader: "Bryophytes". The starting point for the creation of thematic inquiry on a particular taxonomic group is the Latin name for the group. It should be borne in mind that some taxonomic groups may have several synonyms like Latin and Russian names (for example, a family of Pulses = butterfly = Fabaceae = Leguminosae = Papilionaceae). To achieve completeness of search results for the Latin names, given not only the synonyms, but also the older taxon choices and options for determining the rank of the taxon. In some articles, the relevant request "Bryophytes" may be absent Latin name systematic group. The authors operate English-language term «mosses». This is characteristic of the publications on the ecology of mosses, their practical use, etc. Step query analysis revealed two problems: synonymous Latin names of biological objects and the ambiguity of the systematic plant ranks. The use of English names of taxonomic groups is also a problem. The user interest in a particular taxonomic group can be quite narrow, and relates to only the phylogenetic relationships within any taxon. Thus, in the preparation of thematic constant query associated with certain particular taxonomic groups of organisms, the first step requires careful analysis of the search terms, referring to the synonym of Latin, English names, as well as the ambiguity of the definition of ranks of taxa by different authors. The second stage – the proposed by the Web of Science data exclusion of results concerning some individual thematic sections. This allows you to get rid of information noise introduced by related fields of science, and from the homonymous terms.

---

Современная информационная среда, в которую интегрированы электронные реферативные и наукометрические базы данных (*Web of Science*, *Scopus*, РИНЦ), открывает огромные возможности для получения научной информации и данных о публикационной активности как конкретного автора, так и научной организации в целом.

Вынужденное увлечение российского научного сообщества проблемами публикационной активности, в частности цитируемости, приводит к несколько однобокому и использованию баз данных цитирования, и обсуждению этой темы в научной печати. На первый план выходят вопросы: как опубликовать свою статью в журнале, индексируемом *Web of Science* [1]; какие инструменты предлагают базы данных цитирования для улучшения отчётных показателей организации [2]. Большинство учёных находится под давлением той «специфической формы», которую приобрёл процесс оценки публикационной активности [3] в настоящее время.

Рассматриваемая проблема не уникальна для российского научного сообщества. Сам создатель оригинального Индекса научного цитирования (*Science Citation Index*) Ю. Гарфилд указывает, что изначальное предназначение этого индекса – информационный поиск – забыто многими [4], а не только российскими учёными.

В этой статье обратим внимание на использование базы данных *Web of Science* по её первоначальному предназначению: для поиска информации по определённой теме. Поисквые запросы, которые будем обсуждать далее, – постояннодействующие, новые результаты по которым на постоянной основе передаются заинтересованным пользователям. Важной спецификой этих запросов является то, что составляются они (на поисковом языке *Web of Science*) библиотекарем – информационным работником, т.е. не специалистом в предметной области.

Одна из ситуаций тематического поиска по биологическим наукам – поиск информации по конкретным таксонам, систематическим группам организмов. В самом простом случае пользователя могут интересовать статьи флористического или фаунистического содержания. В качестве примера приведём запрос, сформулированный читателем: «Мохообразные» (статьи по систематике этой группы организмов от уровня отдельных видов до уровня отдела в целом).

Отправной точкой при создании тематического запроса по конкретной систематической группе (в данном случае – «Мохообразные») является латинское название этой группы. При этом следует иметь в виду, что некоторые систематические группы могут иметь несколько синонимичных как латинских, так и русских названий (например, семейство Бобовые = Мотыльковые = *Fabaceae* = *Leguminosae* = *Papilionaceae*).

Если поисковый запрос касается какой-либо обширной группы, то при выборе её общего названия следует быть особенно внимательным: невозможно предугадать, таксон какого именно ранга разные авторы публикаций

будут использовать в качестве общего названия описываемого ими биологического объекта. Так, в качестве таксона высшего ранга для класса листостебельных мхов (бриевые мхи, или бриопсиды) одни авторы используют название отдела «Bryophyta», а другие – название класса «Bryopsida». В статьях по печёночным мхам (печёночники, или маршанциевые мхи) может вообще не употребляться название отдела «Bryophyta». Нами отмечена и обратная ситуация, когда в работе, посвящённой печёночным мхам, указан только старший таксон уровня отдела («Bryophyta»), а название «Печёночные мхи» (Hepaticopsida = Marchantiopsida = Marchantiophyta) отсутствует.

Наконец, ранг одного и того же таксона, указываемый авторами разных публикаций, также может различаться. Например, для того же таксона «Печёночные мхи» одни авторы указывают ранг как класс Hepaticopsida (характерно для более ранних публикаций), другие – как отдел Hepaticophyta (характерно для современных публикаций и авторов, придерживающихся нового систематического деления). Поэтому достичь полноты результатов поиска по латинским названиям таксонов возможно, если учитывать не только синонимы, но и варианты выбора старшего таксона, а также варианты определения ранга этого таксона.

В некоторых статьях, релевантных запросу «Мохообразные», может вообще отсутствовать латинское название систематической группы. При этом авторы оперируют англоязычным термином «Mosses». Опытным путём было выявлено, что это характерно для публикаций по экологии мхов, их практическому использованию и т.п. Пользователя же по запросу «Мохообразные» интересовали материалы таксономического характера; для таких материалов отсутствие научного (латинского) названия таксона нехарактерно. Поэтому, не используя в качестве синонима англоязычный термин «Mosses», мы сумели отсеять широкий круг публикаций, связанных со смежными дисциплинами (экологией, биогеоценологией и т.п.).

Пошаговый анализ запроса выявил две проблемы: синонимичность латинских названий биологических объектов и неоднозначность систематических рангов растений. Использование англоязычных названий таксономических групп также представляет собой проблему при формулировании поискового запроса. Поскольку большая часть статей, входящих в БД *Web of Science*, – на английском языке, нами было сделано предположение, что использование англоязычного термина «Mosses» должно повысить число pertinentных результатов, но на практике – при поиске статей таксономического содержания – это было опровергнуто.

Второй пример. Интерес пользователя к конкретной систематической группе может быть более узким и касаться, например, только филогенети-

ческих отношений внутри какого-либо таксона: запрос, сформулированный читателем, – «Филогенетика роз» (статьи, содержащие описание эволюционных связей между разными видами семейства Розоцветные – Rosáceae).

Графическое начертание названия рода «Роза» («Rosa») полностью соответствует графическому начертанию названия участка генома «*rosA*», обсуждаемого в публикациях о грибах. Однако приведённый пример – самый яркий из всех встреченных нами. Обычная ситуация – использование одного и того же слова в разных областях знаний в разных значениях: например, «communities» – «(co)общества» в экологии и социальных науках, не имеющие между собой ничего общего.

Единственным выходом из этой ситуации нам представляется работа с *результатами* первичного запроса: отсечение результатов из тех областей знаний, которые явно не соответствуют запросу. В случае с «Филогенетикой роз» такой предметной областью («Research area» в терминологии *Web of Science*) является, например, микология – наука о грибах. Следует помнить: такой подход не служит гарантией того, что документы из других, сторонних, областей знаний не будут попадать во вновь появляющиеся результаты сохранённого запроса при дальнейшем его использовании. Поэтому для своевременного отсечения сторонних результатов в ходе дальнейшей работы с запросом следует просматривать новые результаты (или предметные области, по которым распределяются результаты) на предмет несоответствия заявленным в запросе областям знаний.

Анализ этого запроса показывает, что омонимия поисковых терминов – один из аспектов, которые следует учитывать при конструировании поискового запроса. Фильтрация результатов по областям знаний позволяет избавиться от омонимии, одновременно исключив результаты из смежных областей знаний.

Пример третий. Пользователя могут интересовать статьи по любым вопросам, связанным с конкретной систематической группой. Это могут быть вопросы как собственно систематики (например, описания новых таксонов), филогенетики, так и далекие от систематики (например, экология отдельных систематических групп). Рассмотрим запрос, сформулированный читателем, – «Раковинные амёбы» (публикации по систематике и экологии этих организмов).

В некоторых статьях, пертинентных запросу, может вообще отсутствовать латинское название систематической группы; кроме того, авторы оперируют англоязычным термином. В результате итоговый набор публикаций распадается на две части, в целом равновеликих и при этом не пересекающихся. В одних публикациях авторы используют латинский термин

«*Testacea*», а в других – англоязычный – «*Testate amoebas*». Так как по запросу «Раковинные амёбы» нас интересуют не только материалы таксономического характера, но и широкий круг вопросов, связанных с экологией этих организмов, при этом необходимо учитывать следующее обстоятельство: в работах не таксономической направленности авторы с большой степенью вероятности будут использовать обиходные, англоязычные названия различных организмов. Это утверждение верно для публикаций как по ботанике, так и по зоологии.

Таким образом, при составлении постояннодействующего тематического запроса, связанного с какой-либо конкретной систематической группой организмов, на первом этапе необходима тщательная проработка поисковых терминов: следует иметь в виду синонимию латинских, англоязычных названий, а также неоднозначность определения рангов таксонов разными авторами. Второй этап – предлагаемая БД *Web of Science* функция исключения результатов по отдельным тематическим разделам. Это позволяет избавиться от информационного шума, привносимого смежными областями знаний, и от омонимичных терминов.

Обслуживание пользователей научной библиотеки вторичной библиографической информацией по-прежнему остаётся важным направлением использования библиографических БД, баз данных цитирования, не менее важным, чем простое извлечение из этих БД наукометрических или даже попросту отчётных показателей. Работа с тематическими запросами в библиографических БД требует использования приёмов, специфика которых может зависеть от конкретной предметной области.

Представленные в этой статье процессы обработки тематических запросов по отдельным биологическим таксонам позволяют информационному работнику, не являющемуся специалистом в конкретной области знаний, получить достоверные результаты.

## СПИСОК ИСТОЧНИКОВ

1. Глушановский А. В., Каленов Н. Е. Некоторые подходы к методике выбора зарубежных журналов для российских публикаций // Информ. обеспечение науки: новые технологии : сб. науч. тр. / Каленов Н. Е., Цветкова В. А. (ред.). – Москва : БЕН РАН, 2015. – С. 40–47.  
*Glushanovskiy A. V., Kalenov N. E. Nekotorye podhody k metodike vybora zarubezhnykh zhurnalov dlya rossiyskikh publikatsiy // Inform. obespechenie nauki: novye tehnologii : sb. nauch. tr. / Kalenov N. E., Tsvetkova V. A. (red.). – Moskva : BEN RAN, 2015. – S. 40–47.*

2. **Гуреев В. Н., Мазов Н. А.** Редактирование профиля организаций в Scopus и РИНЦ: сравнение возможностей // НТИ. Сер. 1. – 2016. – № 3. – С. 10–22.

**Gureev V. N., Mazov N. A.** *Redaktirovanie profilya organizatsiy v Scopus i RINTS: sravnenie vozmozhnostey* // *NTI. Ser. 1.* – 2016. – № 3. – С. 10–22.

3. **Цветкова В. А.** Системы цитирования: где благо, где зло // Науч. и техн. б-ки. – 2015. – № 1. – С. 18–22.

**Tsvetkova V. A.** *Sistemy tsitirovaniya: gde blago, gde zlo* // *Nauch. i tehn. b-ki.* – 2015. – № 1. – С. 18–22.

4. **Garfield E.** A Century of Citation Indexing // COLLNET Journal of Scientometrics and Information Management. – 2012. – Vol. 6. – Iss. 1. – С. 1–6.

---

**Ekaterina Tkacheva, Cand. Sc. (Biology), Senior Researcher, Russian Academy of Sciences Library for Natural Sciences;**

**gbsad\_lib@mail.ru**

**11/11, Znamenka st., 119991 Moscow, Russia**