

**В. А. Цветкова, Т. Н. Харьбина, Ю. В. Мохначева,
Е. В. Бескаравайная**

Библиотека по естественным наукам РАН

И. Ю. Митрошина

*Институт теоретической и экспериментальной биофизики
Российской академии наук*

Особенности совмещения классификационных систем и формирования массива ключевых слов для определения пространства знаний по микробиологии

Использование различных классификаторов, которые, казалось бы, должны содействовать аналитической работе, структурируя предметные области, на практике её частично затрудняет. В статье представлена методика совмещения различных классификационных систем: Государственного рубрикатора научно-технической информации, Рубрикатора отраслей знаний ВИНТИ РАН, Универсальной десятичной классификации, Международной патентной классификации. Благодаря совмещению этих классификационных систем решается фундаментальная задача – получение максимально детализированного представления о научной проблеме. Решение этой задачи позволяет обозначить границы поисков, отсеивая второстепенную и неактуальную информацию по заданной теме (в данном случае – по микробиологии). В результате проведённого авторами статьи исследования были отобраны авторские ключевые слова, фразы и словосочетания. Благодаря тому, что массивы ключевых слов и фраз были приведены к единообразию, получен необходимый перечень ключевых слов и фраз для базового локального Рубрикатора по микробиологии. Отобранные ключевые слова и фразы будут впоследствии служить основой для формирования модели онтологии предметной области, обеспечивающей возможность автоматизации построения онтологии. Разработанный авторами проект Рубрикатора по микробиологии представляет собой трёхуровневую иерархию. Построение первого уровня рубрикатора основано на классической концепции создания классификационных систем. Рубрикатор по микробиологии разработан на основе классификации областей науки ВИНТИ РАН путём детализации.

Работа выполнена при поддержке гранта РФФИ №18-00-00294-комфи: «Исследование и разработка принципов, методов и средств интеграции естественно-научных информационных ресурсов в единое цифровое пространство научных знаний».

Ключевые слова: информационное пространство, классификационные системы, термины, микробиология, онтология предметной области.

**Valentina Tsvetkova, Tatyana Kharybina, Yuliya Mokhnacheva,
Elena Beskaravainaya**

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

Irina Mitroshina

*Institute of Theoretical and Experimental Biophysics,
Russian Academy of Sciences, Moscow Region, Russia*

Combining classification systems and building the array of keywords for defining the space of microbiological knowledge

Using various classifiers, which are thought to facilitate analytical work through structuring subject areas, in fact impedes this work. The authors present the method of combining several classification systems, namely the State Rubricator of Sci-tech Information (SRNTI); and RAS VINITI's Rubricator of Disciplines; Universal Decimal Classification (UDC); International Patent Classification (IPC). The combination of these classification systems solves the fundamental problem which is to obtain the most detailed understanding of any scientific problem. The solution of this problem enables to determine search boundaries, filter out insignificant and irrelevant information on the subject (in this case – Microbiology). In their study, the authors selected the author's keywords and phrases. On the basis of the arrays of author's keywords and phrases, an essential harmonized list of keywords and phrases for the basic Microbiology local rubricator was designed. Further, the selected keywords and phrases will serve as the foundation for building the ontological model of the discipline, which will enable computerized ontology building. The Microbiology draft rubricator developed by the authors comprises a three-level hierarchy. In particular, the first level is based on the classical concept of classification systems. Rubricator in Microbiology is developed on the basis of VINITI RAS classification of disciplines through the detailing method.

The work was supported by RFBR grant № 18-00-00294-comfi: “Research and development of principles, methods and means of integration of natural science information resources into a single digital space of scientific knowledge”.

Keywords: information space, classification systems, terms, Microbiology, ontology of the subject area.

The purpose of this article is to report scientific project of the Russian Foundation for Basic Research № 18-00-00294, aimed to develop principles, methods and means of integrating natural scientific information resources into a single digital space of scientific knowledge. Within the framework of this project, it is planned to create an ontological superstructure over the most common knowledge classification systems (UDC, State subject heading system of sciotech infor-

mation, LBC, Classification of Patents, etc.) using a bibliometric analysis. This will highlight the subspaces of key terms in the Microbiology. The created project of the Rubricator in Microbiology is a three-level hierarchy. The construction of the first level of the Rubricator consistently reveals the content of the basic concepts in the field of microbiology: from general problems to theoretical aspects and further to issues related to the specifics of the objects being studied – their properties, processes, methods of production, with a transition to the tasks of their application in various fields of knowledge. Fundamentally new in the proposed project is the analysis and use when constructing ontologies of the scientific areas of keywords that are most often found in rating world information resources for a given thematic area. In this regard, to expand and clarify the content and use of the rubricator in the automatic search system, keywords, terms or phrases related to a given topic in Russian, separated by semicolons, were correlated with each subject heading and with all subheadings. The initial criterion for selecting publications was the rating of journals in terms of impact factors for a given field of knowledge. For each section, from 100 to 500 publications were selected, which were sorted in descending order of impact factors. For subsequent comparison of the two arrays, two lists of publications were formed: publications from rating journals on a given topic, and publications with the highest citation for each thematic section. Comparison of the entire repertoire of publications of the two arrays showed 19% of matches.

Цель научного проекта Российского фонда фундаментальных исследований № 18-00-00294 – разработка принципов, методов и средств интеграции естественно-научных информационных ресурсов в единое цифровое пространство научных знаний. Основные аспекты формирования единого цифрового пространства знаний рассмотрены в работах [1, 2]. В [3] представлен один из подходов к извлечению знаний из научных текстов, основанный на определении «научных фронтов» в конкретной тематической области. Изучение состояния терминологической базы и исследовательских фронтов позволяет выявить тенденции в научном знании, что мы и пытаемся показать в рамках этого проекта.

Наука представляет собой сложную, неоднородную, динамичную систему, характеризующуюся многообразием объектов, видов и направлений исследований, внутренних и внешних взаимосвязей. Несмотря на происходящую благодаря прогрессу науки непрерывную трансформацию состава научных направлений и их взаимосвязей, сам тип подобной структуризации носит устойчивый характер, что позволяет считать научную дисциплину ведущим звеном в строении и развитии науки, единицей её организации и изучения.

С появлением новых концепций и подходов к исследованию науки задача классификации наук не стала менее актуальной. Эта задача имеет глубокие корни и продолжает оставаться в центре внимания философии науки, поскольку всегда существует потребность отразить и представить взаимосвязь наук системно – в виде той или иной классификационной системы (схемы).

В мировой науковедческой практике пока не сформулированы точные критерии выделения отраслей науки и научных дисциплин и существует множество вариантов их классификации, которые, преимущественно, строятся по иерархическому принципу, охватывая различные комбинации областей знаний.

Разработка классификаций наук ведёт отсчёт от трудов Платона и Аристотеля; этими проблемами занимались такие выдающиеся учёные и мыслители, как Ф. Бэкон, Р. Декарт, Дж. Локк, Г. В. Ф. Гегель, Ф. Энгельс [4–8]. Наиболее системно теория классификации наук была изложена в работах Ч. Пирса (конец XIX в.) [9], Б. М. Кедрова [10, 11]. Вместе с тем за последние 30 лет работы по формированию классификаций отраслей науки приобрели системный и целенаправленный характер.

Методика исследования

В настоящее время при поддержке РФФИ предполагается создание единого пространства цифровых знаний, включающего онтологию знаний. «Под онтологией понимается формальное описание некоторой области знания, – в данном случае по направлению “Микробиология”, которое может быть использовано для автоматической обработки компьютером» [12].

Б. В. Добров и Н. В. Лукашевич отмечают: «Онтологии являются инструментом системного анализа предметной области, обеспечивая целостное представление совокупности понятий, характеризующих предметную область, и их связей» [13]. Такую среду в рамках настоящего проекта планируется создавать как надстройку над наиболее распространёнными системами классификации знаний (УДК, ГРНТИ, ББК, Классификация изобретений и т.п.) с использованием аппарата библиометрического анализа. Это даст возможность выделить подпространства ключевых терминов, в частности по научному направлению «Микробиология», приводимых в наиболее цитируемых публикациях, которые индексируются в политематических и специализированных отечественных и зарубежных базах данных (*Web of Science Core Collection, Scopus, РИНЦ* и др.).

Для формирования единого пространства научных знаний должны использоваться унифицированные подходы к отбору информации, её представлению, навигации, терминологическому наполнению.

Использование различных классификаторов, которые, казалось бы, должны содействовать аналитической работе, структурируя предметные области, на практике её частично затрудняет. Для решения этой проблемы на примере узкой предметной области «Микробиология» можно было бы использовать в качестве базового один из рубрикаторов: Государственный рубрикатор научно-технической информации (ГРНТИ), Рубрикатор отраслей знаний ВИНТИ РАН (РВИНТИ), Универсальную десятичную классификацию (УДК), Международную патентную классификацию (МПК) и др. Однако они во многом не согласуются между собой, значительно отличаясь друг от друга. Их перманентное эволюционирование приводит к упразднению одних кодов и появлению новых. К тому же далеко не всегда анализируемая предметная область – «Микробиология» – вписывается в таксономию, предлагаемую этими классификаторами.

Б. В. Добров отмечает: «Множественность принятых в современной практике систем описания тематики научных работ создаёт трудности в получении исчерпывающих данных по какой-либо заданной теме исследований, поскольку разные классификационные системы членят предметную область исследований по-разному. С другой же стороны, множественность описаний одной предметной области представляют её онтологию с различных точек зрения, что позволяет надеяться на более объективное отражение физических свойств, явлений и процессов формальной моделью. Необходимость иметь такую модель онтологии в составе поисковых систем для успешного поиска данных является признанным фактом современной информатики» [14]. Таким образом, построение единого пространства научных знаний – задача сложная, многоаспектная и носит комплексный характер.

Приступая к выполнению рассматриваемого проекта, прежде всего необходимо было решить задачу создания локального базового рубрикатора для конкретной научной области на основе использования различных классификаторов. В нашем случае – по микробиологии. Были проанализированы опыт аналогичных проектов и действующие системы российских и международных классификаций сферы науки [15–17].

Изучение опыта построения классификаций позволило выявить направления их совершенствования и развития как в России, так и за рубежом, определить методологические подходы к созданию переходных ключей (таблиц совмещения/связи) между различными классификациями, действующими в России [18–21].

Результаты показали, что российские классификаторы разработаны для обслуживания разных задач в разных системах управления и функционирования науки и образования. Каждая классификация по-разному адаптирована к выполняемой функции и разделяет глобальную область науки и образования по-разному с разной степенью подробности в зависимости от поставленной задачи.

Особенности отдельных классификационных систем

Государственный рубрикатор научно-технической информации. Особенности ГРНТИ как классификационной системы являются: универсальный охват научного и технического знания, сравнительно неглубокая иерархия (3 уровня), десятичный принцип деления классов (позволяющий иметь в сети классов резервные места для развития системы), согласованность структуры классов со структурой управления наукой и народным хозяйством [22]. Коды классов ГРНТИ имеют вид XX.YY.ZZ. На верхнем уровне выделено 69 областей знания.

Рубрикатор отраслей знаний ВИНТИ РАН. В отечественной практике наряду с ГРНТИ применяется другая классификационная система – Рубрикатор отраслей знаний ВИНТИ РАН. Эта система во многом сходна с ГРНТИ, поскольку построена на основе углубления ГРНТИ (3 уровня) до 9-го уровня. Этот рубрикатор предназначен для систематизации информационных продуктов и поиска в БД ВИНТИ РАН; охватывает весь универсум знания, допускает комбинирование классов при индексировании и имеет глубокую иерархию классов [23].

Универсальная десятичная классификация. УДК отвечает требованиям, предъявляемым к классификации: международность, универсальность, мнемоничность, возможность отражения новых достижений науки и техники без каких-либо серьезных изменений в её структуре; охватывает все области знаний. Разделы УДК органически связаны: изменение одного из них влечёт за собой изменение другого, поэтому она и называется универсальной. В полном издании УДК насчитывается более 150 тыс. индексов.

Десятичной УДК называется потому, что для её построения использован десятичный принцип: деление каждого класса на десять (или менее) подклассов. Десятичный принцип структуры УДК позволяет практически неограниченно расширять её путём добавления новых цифр к уже имеющимся, не ломая всей системы в целом. УДК построена по иерархическому принципу, в основе которого лежит деление понятий от общего к частному. При этом основными видами отношений являются подчинение (иерархия) и соподчинение.

Все понятия в УДК делятся на основные и вспомогательные. Основные отражают специфику, присущую отдельным отраслям или областям знания. Систематизированы эти понятия в основной таблице, а соответствующие им индексы принято называть основными, или самостоятельными. Вспомогательные понятия отражают, как правило, повторяющиеся понятия, общие для всех или нескольких разделов или применяемые внутри одного указанного раздела. Вспомогательные, или зависимые понятия служат для ассоциативного уточнения основных индексов [24].

Международная патентная классификация. Важную для естественно-научной сферы область микробиологии охватывает МПК, отражающая изобретательскую деятельность. Она даст выход на управление инновационными процессами. При решении задач совмещения классификационных систем важно включить и МПК.

Принципы разработки структуры рубрикатора по микробиологии

Рубрикатор по микробиологии разработан с учётом общих требований, предъявляемых к подобным инструментам в международной практике, а именно:

- соответствие объективным тенденциям развития науки и образования с учётом специфики отдельных отраслей науки;

- наличие рубрик с достаточно гибкой возможностью их оптимизации, предусматривающей объединение близких и упорядочение дублирующих друг друга научных областей, систематизацию категорий и исключение необоснованной их дифференциации;

- организация рубрик должна быть выражена посредством иерархии, т.е. явным образом отмеченные (обозначенные кодами рубрик) родовые отношения, уровни иерархии подчинения и соподчинения с простой и линейной структурой;

- наличие таблиц перевода в другие общепринятые государственные классификаторы для организации системы взаимосвязанных рубрикаторов;

- названия рубрик должны иметь английский эквивалент;

- для расширения функций рубрикатора текстовые обозначения рубрик должны быть проиндексированы ключевыми словами.

Проект Рубрикатора по микробиологии представляет собой трёхуровневую иерархию. Первый уровень Рубрикатора построен на классической концепции создания классификационных систем. Это позволяет последовательно раскрыть содержание основных понятий в области микробиологии: от общих проблем к теоретическим аспектам и далее – к вопросам, связанным со спецификой изучаемых объектов (свойствами, процессами, методами получения), а затем перейти к задачам их применения в различных областях знания.

Рубрикатор по микробиологии разработан на основе классификации областей науки ВИНТИ РАН путём детализации. Глубина классификации определяется необходимостью наиболее полного и точного описания содержания данной научной области и её спецификой: разные рубрики развиты на разную глубину.

В некоторых случаях индексирование проводилось на основе уже имеющихся кодов других классификаций, применяемых в отечественной практике в настоящее время: коды рубрикатора ГРНТИ, тематические коды УДК, классы МПК.

Рубрикатор представлен в форме таблицы, в которой даны коды и наименования позиций классификации (в русском и английском вариантах). Кроме того, в таблицу включены связи Рубрикатора по микробиологии с другими классификационными системами (УДК, ГРНТИ, РВИНИТИ, МПК).

Рубрики целесообразно выделять так, чтобы они максимально соответствовали классификационным системам, уже действующим в отечественной сфере науки и образования. Поэтому при разработке рубрикатора следует сопоставить его классы с соответствующими и близкими по содержанию классами перечисленных классификационных систем (табл. 1). При сопоставлении можно наблюдать разный характер расхождения между классами, предлагаемыми экспертами, и классами, имеющимися в используемых классификациях.

Таблица 1

Сравнение номенклатуры предложенной классификации, ГРНТИ, РВИНИТИ и УДК

Предложено в проекте	ГРНТИ	РВИНИТИ	УДК
М4 Морфология, физиология, цитология и биохимия микроорганизмов	34.27.17 Морфология и физиология микроорганизмов; 31.27.19 Биохимия микроорганизмов	341.27.17 Морфология и физиология микроорганизмов	579.22 Биохимия и физиология микроорганизмов; 579.23 Морфология и цитология микроорганизмов
М5 Рост и культивирование микроорганизмов	34.27.19 Рост и культивирование микроорганизмов	341.27.19 Рост и культивирование микроорганизмов	579.24 Рост и развитие микробных клеток

Говоря об особенностях экспертного анализа связей Рубрикатора по микробиологии с другими классификациями, следует отметить, что глубина и дробность развития классификаций сильно варьируются – от нескольких десятков рубрик в некоторых классификаторах до возможности широкого

комбинирования иерархии рубрик в УДК. Установление совмещений рубрик Рубрикатора по микробиологии с рубриками РВИНИТИ, УДК, ГРНТИ, МПК проведено параллельно, что позволяет учесть разнообразные аспекты описания информационных ресурсов.

Таким образом, создание таблиц совмещений необходимо. Они обеспечивают взаимодействие различных классификационных систем через систему смысловых соответствий тематических рубрик. В таблице совмещения Рубрикатора по микробиологии, ГРНТИ, РВИНИТИ и УДК рубрики сопоставлялись независимо от их положения в иерархической структуре. С рубрикой верхнего уровня одного рубрикатора могла быть сопоставлена рубрика нижнего уровня другого рубрикатора.

Не во всех случаях содержание понятий сопоставленных рубрик совпадают абсолютно точно. В ряде случаев определённая тематика в области микробиологии может быть отнесена к разным рубрикам ГРНТИ, где она присутствует в контексте разных аспектов рассмотрения. В этом случае в таблице указываются две или несколько рубрик ГРНТИ или УДК. Наблюдаются и противоположные случаи: широкие рубрики УДК или ГРНТИ могут быть совмещены с разными рубриками базового Рубрикатора по микробиологии.

Так же следует поступать, если классы названы разными по значению терминами, но эксперт по логике классификационной системы считает, что в обоих случаях классы должны включать одно и то же множество объектов. Это подтвердилось в ходе практической работы: если авторские ключевые слова в соседних рубриках совпадали более чем на 60%, мы решили, что целесообразно объединить их в общий раздел.

Таблица совмещений представлена в формате Excel (табл. 2). Столбцы Excel представляют: код локальной (базовой) рубрики; название русское; название английское; далее идут связанные с локальной рубрикой индексы рубрикаторов РВИНИТИ, ГРНТИ, УДК, МПК. Перед кодом каждой рубрики в этих столбцах указан обязательный символ их связи с локальной рубрикой. Если рубрики совпадают, то перед индексом ставится «=», если пересекаются – «~», если локальная шире – «>», если уже – «<». Подобный подход рассмотрен в работах [21, 25]. Если локальная рубрика соответствует нескольким рубрикам данного рубрикатора, значения разделяются точкой с запятой, символы связи при этом обязательны, но могут быть как одинаковыми, так и различными.

Примеры совмещения имеющихся рубрикаторов с разрабатываемым

Код локальной рубрики	Название русское	ВИНИТИ	УДК	ГРНТИ	МПК
M4	Морфология, физиология, цитология и биохимия микроорганизмов. Общие проблемы	> 341.27.17	> 579.22; > 579.23	> 34.27.17; > 34.27.19	> A61K; > C12N
M4.1	Морфология, цитология и циклы развития микроорганизмов; иммунохимия микробной клетки	= 341.27.17.07	~ 579.23	< 34,27.17	> A61K; > C12N
M4.2	Физиология микроорганизмов	= 341.27.17.09	< 579.22	< 34.27.17	> C12N; > 61K; > A23K; > C05G
M4.3	Биохимия микроорганизмов	> 341.27.17.09.07; > 341.27.17.09.09; > 341.27.17.09.13; > 341.27.17.09.15	~ 579.222	~ 31.21.19	> A23K; > C02F; > C12N; > C12P; > C05G; > A23L; > A61K; > C12M
M4.4	Коммуникативные межклеточные взаимодействия у микроорганизмов	= 341.27.17.05	< 579.22	< 34,27.17	> C12M; > C02F; > G01N; > C12N

Таким образом, мы имеем первые результаты разработки таблиц совмещения, которые, безусловно, требуют доработки с целью более чёткого соответствия классам РИНИТИ, УДК, ГРНТИ и МПК.

Принципиально новым в предлагаемом проекте является анализ и использование при построении онтологий научных направлений ключевых слов, наиболее часто встречающихся в рейтинговых мировых информационных ресурсах по заданной тематической области. В связи с этим для расширения и уточнения содержания и использования рубрикатора в системе автоматического поиска с каждой рубрикой и со всеми подрубриками были соотнесены ключевые слова, термины или словосочетания по заданной тематике на русском языке, отделённые точкой с запятой.

Методика отбора ключевых слов

Для достижения наиболее объективных результатов разработана методика сбора необходимой информации*, которая включает:

1. Выбор информационных баз данных, наиболее полно соответствующих тематике рубрикатора.
2. Выявление, сбор и создание массива публикаций в области микробиологии из этих БД.
3. Выгрузка созданного массива ключевых слов и фраз из массива публикаций в таблицу Excel.
4. Перевод слов и словосочетаний на русский язык и их редактирование.
5. Подсчёт частотного распределения ключевых слов и фраз во всех рубриках.

В научных публикациях всегда содержатся авторские ключевые слова, которые связаны с тематикой статьи. Поэтому массив таких слов служит ценным материалом для расширения возможностей рубрикатора по микробиологии.

Надо отметить, что рассматриваемая научная проблематика, являясь мультидисциплинарной, сопряжена с целым спектром естественно-научных дисциплин. Поэтому тематический поиск статей с целью сбора ключевых слов для всех рубрик и подразубрик по данной теме сопряжён с определёнными трудностями. Среди них: большой объём информации, очень широкий радиус рассеяния информации, чрезвычайно высокий уровень «шума» при поиске по внешним электронным ресурсам. Поэтому основной задачей в процессе выявления актуальных статей по микробиологии стало максимальное сужение границ поиска для достижения желаемых результатов. Благодаря применению предлагаемой методики мы достигли определённых положительных результатов.

В качестве основной информационной базы была использована *Web of Science Core Collection (WoS CC)*. Объектом исследования стал документный поток по теме «Микробиология». Предмет исследования – выявление «ядерного» круга изданий, а также авторов по заданной теме. В БД *WoS CC* в соответствии с разработанным рубрикатором были отобраны все возможные публикации по микробиологии за 2016–2018 гг.

В результате мы получили значительный по объёму массив информации, из которого отбирались ядерные издания по данной тематике для последующей аналитической обработки. Начальным критерием отбора публикаций выступил рейтинг журналов по показателям импакт-факторов для заданной области знания (табл. 3). Для каждой рубрики, в зависимости

* В обсуждении идеи методики отбора ключевых слов активно участвовал доктор техн. наук, проф. Н. Е. Каленов.

от объёма выявленной информации, отобрано от 100 до 500 изданий, которые были отсортированы в порядке убывания импакт-факторов.

Таблица 3

**Издания с наиболее высоким импакт-фактором,
в которых были опубликованы статьи по микробиологии
за 2016–2018 гг. (по БД WoS CC)**

Издания	Импакт-факторы по JCR-2017 (Clarivate Analytics)
Nature Reviews Drug Discovery	50,167
Nature Biotechnology	35,724
Nature Reviews Microbiology	31,851
Clinical Microbiology Reviews	20,642
Nature Microbiology	14,174
Trends in Biotechnology	13,578
Microbiology and Molecular Biology Reviews	13,439
Trends in Microbiology	11,776
Biotechnology Advances	11,452
Genome Research	10,101

Следующим шагом исследования стал отбор наиболее цитируемых статей по заданной тематической рубрике (табл. 4).

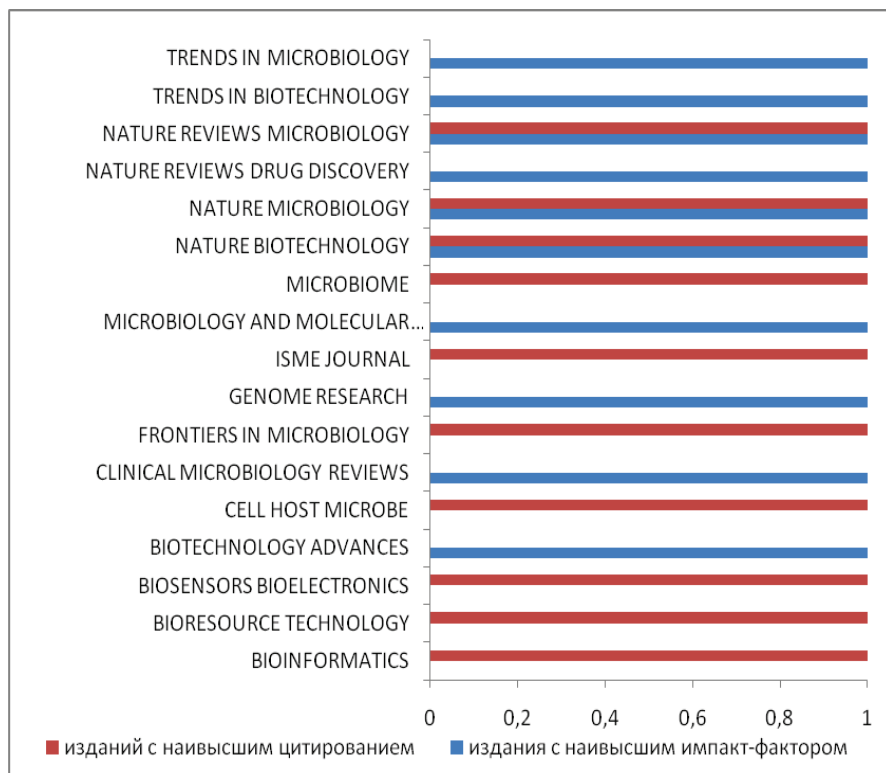
Таблица 4

**Издания, в которых опубликованы статьи по теме «Микробиология»
с наибольшей цитируемостью за 2016–2018 гг. (по БД WoS CC)**

Издания	Число публикаций с наибольшей цитируемостью
Nature Reviews Microbiology	12
Bioresource Technology	8
Frontiers in Microbiology	7
Cell Host Microbe	7
Nature Microbiology	6
Nature Biotechnology	5
Biosensors Bioelectronics	5
ISME Journal	4
Microbiome	4
Bioinformatics	3

Таким образом, для последующего сопоставления двух массивов были сформированы два перечня публикаций: А – публикации из рейтинговых журналов по заданной тематике, Б – публикации с наибольшей цитируемо-

стью по каждой тематической рубрике. Сравнение всего репертуара публикаций двух массивов показало 19% совпадений: например, издания «Nature Microbiology», «Nature Biotechnology», «Nature Reviews Microbiology» присутствуют параллельно в обоих списках (см. рис.).



Журналы с наивысшим импакт-фактором, в которых опубликованы статьи по микробиологии за 2016–2018 гг., и издания с наиболее цитируемыми публикациями за тот же период

Таким образом, публикации, одновременно представленные в обоих массивах, являются профильными по заданным тематическим направлениям. Объединение двух массивов позволило получить более детальное представление о научной проблеме и обозначить границы поисков, отсеивая второстепенную и неактуальную информацию.

Отобранные авторские ключевые слова, фразы и словосочетания были переведены на русский язык, приведены к единообразию и внесены в табл. 5. В итоге мы получили необходимый перечень ключевых слов и фраз для на-

шего Рубрикатора по микробиологии, которые должны стать основой для формирования модели онтологии предметной области, обеспечивающей возможность автоматизации построения онтологии (табл. 5).

Таблица 5

Пример заполнения рубрикатора

Код	Название рубрики		Ключевые слова (частотное распределение в статьях)
	рус. яз.	англ. яз.	
M4	Морфология, физиология, цитология и биохимия микроорганизмов. Общие проблемы	Morphology, biochemistry and physiology of microorganisms. General problems	Активные формы кислорода (1); ассимиляция (2); биологическая гидродинамика (1); бислойные липидные мембраны (2); водные бактерии (1); гибридологический анализ (4); глутаминовая кислота (1); диссимилиация (1); идентификация микроорганизмов (3); капсульные бактерии (4); катаболизм (1); киназы (1); кинетика роста бактериальной популяции (1); кинетохоры (1); клетки грибов (2); клональный анализ (5); культивирование микроорганизмов (6); мейоз (1); мембранные структуры (2); метод отпечатков (2); метод тетрадного анализа (1); митохондрии дрожжей (1); монополин комплекс (2); морские бактерии (1); наружная мембрана (2); нарушение клеточных структур (2); нуклеопротеиды (3); обмен веществ (7); передача наследственных признаков (2); питательные субстраты (2); подвижность бактерий (4); покоящиеся формы прокариот (2); порины (2); порча пищевого сырья (1); поток с малым числом Рейнольдса (1); программирование клеточной смерти (1); продолжительность деления клетки (3); проницаемость перехода (1); процесс расщепления (1); размножение вирусов (6); расщепление сахаров (5); регуляторы роста (2); рекомбинанты (4); рекомбинация (2); синтез белков клетки (4); системы транспорта веществ (2); спиральные движения (1); спорообразующие бактерии (2); структурные компоненты бактериальной клетки (7); типы биологического окисления субстратов (3); умеренные бактериофаги (1); формы бактериальных клеток (3); штамм (1); энергетический метаболизм (4).

Код	Название рубрики		Ключевые слова (частотное распределение в статьях)
	рус. яз.	англ. яз.	
4.1	Морфология, цитология и циклы развития микроорганизмов	Morphology cytology and development cycles of microorganisms	Азотфиксация (2); аксиальная нить (1); археи (1); азотаксис (1); белок синтезирующих систем (4); бинарное деление (1); благоприятные ниши (3); влияние фага (4); галофильные археи (1); генетически модифицированные микроорганизмы (1); гетероморфный рост (1); движение скольжением (3); естественный полиморфизм (4); жгутики (4); жизненный цикл (2); жизнеспособные клетки микроорганизмов (4); изменения хроматинового аппарата (1); истощение источника питания (1); клеточная дифференциация (3); клеточная дифференциации (3); клеточные включения (1); клеточные структуры (1); клеточный цикл (1); концепция клеточной эволюции (1); конъюгация (4); конъюгированный перенос (1); метод криопрепарирования клеток (1); метод цитологического анализа (1); механизм когезии (5); механизм флокулирования (4); микроаэрофильные бактерии (1); микробные пейзажи (4); миксобактерии (1); множественное деление (1); мобилизация белковой системы (1); моделирование природных условий (2); надмолекулярная организация (1); нарастание клеточной поверхности (1); образование биопленок (3); образование экзоспор (3); палочковидные микроорганизмы (1); пассивное движение (1); передача ДНК (1); перенос генетического материала (15); переходная агрегация (3); пили IV типа (3); пищевой субстрат (1); плейоморфизм (9); подвижные клетки бактерий (2); подвижные клетки бактерий (1); покоящиеся клетки (2); почкующиеся формы (5); процесс инициации репликации ДНК (23); разнообразие клеток микроорганизмов (6); роль внеклеточных факторов (4); роль клеточных структур (2); роль фагов в переносе генетического материала (1); серологические исследования (8); сложный цикл развития (1); состав среды (3); споруляция (1); способы питания (1); стадии жизненного цикла (4); структура микробных матов (4); структурно-функциональная адаптация (1); трансдукция (8); углеводный состав (2); удвоения генетического материала (5); ультраструктурная организация (1); ультратонкие срезы (8); устойчивость к высоким температурам (3); криогенные микроорганизмы (2); условия культивирования (4); устойчивость к высушиванию (3).

При построении структуры рубрикатора и составлении перечня ключевых слов учитывалась частотность повторов того или иного термина в массивах информации. Частотность определялась путём анализа результатов последовательных поисков в БД по поисковым предписаниям, в которые включались анализируемые термины. В дальнейшем эти ключевые слова легли в основу глоссария терминов (понятий) для разработки онтологии.

Далее были построены деревья классификации понятий (иерархии классов), которых в онтологии может быть несколько. Именно эти понятия и устанавливают основные связи между классами. Всё это даёт возможность построить алгоритмы для реализации метода автоматизированного построения онтологии конкретной предметной области. Затем эксперты предметной области, для которой разрабатывается онтология (в нашем случае – «Микробиология»), определяют правила логических выводов, предоставляющие возможность использовать полученные данные, представленные в онтологии, и извлекать из неё новые знания.

Подводя итоги, отметим: классификации знания разработаны для решения разных задач в разных системах управления и функционирования науки и образования. Каждая классификация адаптирована к выполняемой функции и разделяет глобальную область науки и образования по-разному в зависимости от угла зрения на поставленную задачу. Общим измеримым параметром классификаций является степень их подробности.

Особенность перечисленных библиографических классификаций – открытый ряд классификационных делений, что не позволяет зафиксировать и полностью представить определённый уровень классификации. Эти системы предоставляют пользователю не только перечень рубрик, но также и механизм для образования новых классов, что влечёт за собой особенности формирования таблиц совмещения с другими классификациями.

Предварительно можно утверждать, что всё многообразие задач управления наукой не ограничивается несложной классификацией областей наук по какому-либо одному признаку. При необходимости приемлемая для всех задач классификационная система должна иметь вид фасетной классификации, где один из фасетных признаков может послужить в качестве основной опорной классификации, от которой должны быть даны ссылки на рубрики других классификаций, применяемые для уточнения содержания основных классов.

СПИСОК ИСТОЧНИКОВ

1. Антопольский А. Б., Ефременко Д. В. К вопросу о едином электронном пространстве знаний // Вестн. Рос. акад. наук. – 2018. – Т. 88. – № 2. – С. 163–170.

Antopolskiy A. B., Efremenko D. V. K voprosu o edinom elektronnom prostranstve znaniy // Vestn. Ros. akad. nauk. – 2018. – T. 88. – № 2. – S. 163–170.

2. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. Точка зрения о едином цифровом пространстве научных знаний // Там же. – 2019. – Т. 89. – № 7. – С. 728–734.

Antopolskiy A. B., Kalenov N. E., Serebryakov V. A., Sotneykov A. N. Tochka zreniya o edinom tsifrovom prostranstve nauchnyh znaniy // Tam zhe. – 2019. – T. 89. – № 7. – S. 728–734.

3. Цветкова В. А., Мохначева Ю. В., Харьбина Т. Н., Бескаравайная Е. В., Митрошин И. А. Пространство знаний: подходы к извлечению знаний из научных текстов // Информ. ресурсы России. – 2019. – № 2 (168). – С. 31–34.

Tsvetkova V. A., Mohnacheva Yu. V., Harybina T. N., Beskaravaynaya E. V., Mitroshin I. A. Prostranstvo znaniy: podhody k izvlecheniyu znaniy iz nauchnyh tekstov // Inform. resursy Rossii. – 2019. – № 2 (168). – S. 31–34.

4. Бэкон Ф. Сочинения: В 2 т. Т. 1. – 1977. – 567 с.; Т. 2. – 1978. – 575 с. – Москва : Мысль.

Bekon F. Sochineniya: V 2 t. T. 1. – 1977. – 567 s.; T. 2. – 1978. – 575 s. – Moskva : Mysl.

5. Декарт Р. Сочинения: В 2 т. Т. 1. – 1989. – 656 с.; Т. 2. – 1989. – 654 с. – Москва : Мысль.

Dekart R. Sochineniya: V 2 t. T. 1. – 1989. – 656 s.; T. 2. – 1989. – 654 s. – Moskva : Mysl.

6. Локк Дж. Избранные философские произведения: В 2 т. Т. 1. – Москва : Соцэкгиз, 1960. – 713 с.

Lokk Dz. Izbrannye filosofskie proizvedeniya: V 2 t. T. 1. – Moskva : Sotsekgiz, 1960. – 713 s.

7. Гегель Г. В. Ф. Энциклопедия философских наук. Т. 1. – 1971. – 452 с.; Т. 2. – 1975. – 693 с.; Т. 3. – 1977. – 471 с. – Москва : Мысль.

Gegel G. V. F. Entsiklopediya filosofskih nauk. T. 1. – 1971. – 452 s.; T. 2. – 1975. – 693 s.; T. 3. – 1977. – 471 s. – Moskva : Mysl.

8. Энгельс Ф. Диалектика природы. – Москва : Политиздат, 1987. – 349 с.

Engels F. Dialektika prirody. – Moskva : Politizdat, 1987. – 349 s.

9. Пирс Ч. С. Избранные философские произведения / пер. с англ.: К. Голубович, К. Чухрукидзе, Т. Дмитриев. – Москва : Логос, 2000. – 448 с.

Pirs Ch. S. Izbrannye filosofskie proizvedeniya / per. s angl.: K. Golubovich, K. Chuhrukidze, T. Dmitriev. – Moskva : Logos, 2000. – 448 s.

10. Кедров Б. М. Классификация наук. Кн. 1. Энгельс и его предшественники. – Москва : Изд. ВПШ и АОН, 1961. – 472 с.

Kedrov B. M. Klassifikatsiya nauk. Kn. 1. Engels i ego predshestvenniki. – Moskva : Izd. VPSH i AON, 1961. – 472 s.

11. Кедров Б. М. Классификация наук. Кн. 2. От Ленина до наших дней. – Москва : Мысль, 1965. – 543 с.

Kedrov B. M. Klassifikatsiya nauk. Kn. 2. Ot Lenina do nashih dney. – Moskva : Mysl, 1965. – 543 s.

12. **Трусова Ю. О., Белоозеров В. Н.** Представление классификационных систем в виде онтологий. Обзор // Науч.-техн. информ. Сер. 1. Орг. и методика информ. работы. – 2015. – № 11. – С. 34–38.

Trusova Yu. O., Belozerov V. N. Predstavlenie klassifikatsionnyh sistem v vide ontologii. Obzor // Nauch.-tehn. inform. Ser. 1. Org. i metodika inform. raboty. – 2015. – № 11. – S. 34–38.

13. **Добров Б. В., Лукашевич Н. В.** Лингвистическая онтология по естественным наукам и технологиям для приложений в сфере информационного поиска // Учёные зап. Казан. ун-та. Физ.-мат. науки. – 2007. – Т. 149. – № 2. – С. 49–72.

Dobrov B. V., Lukashevich N. V. Leengvisticheskaya ontologiya po estestvennym naukam i tehnologiyam dlya prilozheniy v sfere informatsionnogo poiska // Uchenye zap. Kazan. un-ta. Fiz.-mat. nauki. – 2007. – T. 149. – № 2. – S. 49–72.

14. **Онтологи и тезаурусы: модели, инструменты, приложения** : учеб. пособие / Б. В. Добров, В. В. Иванов, Н. В. Лукашевич, В. Д. Соловьев. – Москва, 2012. – 173 с.

Ontologii i tezaurusy: modeli, instrumenty, prilozheniya : ucheb. posobie / B. V. Dobrov, V. V. Ivanov, N. V. Lukashevich, V. D. Solovjev. – Moskva, 2012. – 173 s.

15. **NABS – Nomenclature for the Analysis and Comparison of Scientific Programmes and Budgets.** European Commission. – Luxembourg : Office for Official Publications of the European Communities, 1994.

16. **Handbok** for FoU-statistik. Matning av resurver for Forsknings-och utvecklingsverksamhet. – Stockholm : NORDFORSK, 1974.

17. **Сукиасян Э. Р.** Классификация Библиотеки Конгресса США // Науч. и техн. б-ки. – 1998. – № 5. – С. 59–69.

Sukiasyan E. R. Klassifikatsiya Biblioteki Kongressa SSHA // Nauch. i tehn. b-ki. – 1998. – № 5. – S. 59–69.

18. **Гиляревский Р. С., Шапкин А. В., Белоозеров В. Н.** Рубрикатор как инструмент информационной навигации. – Санкт-Петербург : Профессия, 2008. – 352 с.

Gilyarevskiy R. S., Shapkin A. V., Belozerov V. N. Rubrikator kak instrument informatsionnoy navigatsii. – Sankt-Peterburg : Professiya, 2008. – 352 s.

19. **Гиляревский Р. С.** О сферах применения рубрикатора // Науч.-техн. информ. Сер. 1. Орг. и методика информ. работы. – 1982. – № 1. – С. 22–26.

Gilyarevskiy R. S. O sferah primeneniya rubrikatora // Nauch.-tehn. inform. Ser. 1. Org. i metodika inform. raboty. – 1982. – № 1. – S. 22–26.

20. **Белоозеров В. Н., Малахов А. А., Розина И. А.** Применение ГРНТИ в задачах управления наукой и научно-технической экспертизы // НТИ-99. Материалы конф. – Москва : ВИНТИ, 1999. – С. 44–50.

Belozerov V. N., Malahov A. A., Rozina I. A. Primenenie GRNTI v zadachah upravleniya naukoj i nauchno-tehnicheskoy ekspertizy // NTI-99. Materialy konf. – Moskva : VINITI, 1999. – S. 44–50.

21. **Таблицы** соответствия между Рубрикатором Международной системы научной и технической информации и Международной патентной классификацией. – Москва : МЦНТИ, ТПО «Поиск», 1985.

Tablitsy sootvetstviya mezhdub Rubrikatorom Mezhdunarodnoy sistemy nauchnoy i tehnicheskoy informatsii i Mezhdunarodnoy patentnoy klassifikatsiey. – Moskva : MTSNTI, TPO «Poisk», 1985.

22. **Государственный** рубрикатор научно-технической информации (Рубрикатор ГРНТИ). Изд. пятое. – Москва, 2001.

Gosudarstvennyy rubrikator nauchno-tehnicheskoy informatsii (Rubrikator GRNTI). Izd. pyatoe. – Moskva, 2001.

23. **Государственный** рубрикатор научно-технической информации. – Москва : ВИНТИ, 2001. – 391 с.

Gosudarstvennyy rubrikator nauchno-tehnicheskoy informatsii. – Moskva : VINITI, 2001. – 391 s.

24. **Универсальная** десятичная классификация. Полное четвертое изд. на рус. языке. Т. 1–10. – Москва : ВИНТИ, 2001.

Universalnaya desyatchnaya klassifikatsiya. Polnoe chetvertoe izd. na rus. yazyke. T. 1–10. – Moskva : VINITI, 2001.

25. **Антопольский А. Б., Белоозеров В. Н., Каленов Н. Е., Маркарова Т. С.** О развитии терминологической базы данных в виде комплекса отраслевых информационно-поисковых тезаурусов // Информ. ресурсы России, 2018. – № 5 (165). – С. 22–30.

Antopolskiy A. B., Belozerov V. N., Kalenov N. E., Markarova T. S. O razvittii terminologicheskoy bazy dannyh v vide kompleksa otraslevykh informatsionno-poiskovykh tezaurusov // Inform. resursy Rossii, 2018. – № 5 (165). – S. 22–30.

Valentina Tsvetkova, Dr. Sc. (Technology), Professor, Leading Researcher, Library for Natural Sciences, Russian Academy of Sciences;

vats08@mail.ru

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

Tatyana Kharybina, Senior Researcher, Library for Natural Sciences, Russian Academy of Sciences;

natsl@vega.protres.ru

333/3, Institutskaya st., 142290 Pushchino, Moscow Region, Russia

Yuliya Mokhnacheva, Cand. Sc. (Pedagogy), Head, Scientometric Studies Department, Library for Natural Sciences, Russian Academy of Sciences;

J_v_m@yandex.ru

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

Elena Beskaravaynaya, Senior Researcher, Library for Natural Sciences, Russian Academy of Sciences;

elenabesk@gmail.com

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

Irina Mitroshina, Cand. Sc. (Biology), Researcher, Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences;

xf2@rambler.ru

3, Institutskaya st., 142290 Pushchino, Moscow Region, Russia