

ОБ ИНДУКТИВНОЙ КЛАССИФИКАЦИИ НАУК И МЕСТЕ ГЕОЛОГИИ В СИСТЕМЕ НАУК (СТАТИСТИЧЕСКИЙ ПОДХОД). СТАТЬЯ 1

© 2020 г. М. П. Покровский

Ставится задача создания индуктивной классификации наук. Намечаются основные подходы в ее решении. Предлагается методика статистического подхода к созданию индуктивной классификации наук, базирующегося на анализе списков наук в научных текстах, институтах или мероприятиях. Оцениваются объемы таких списков. На основе оценки частоты включения науки в списки и ее доли в этих списках формируется четыре варианта минимизированных ($n-10$ номинаций) списков наиболее социально значимых наук.

0. ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

0.1. Классификация-перечисление наук – один из узлов разработанной автором системы классиологии [Покровский, 2006, 2014]. В этой системе предусматриваются три части – вводная часть, пролегомены (предварительные положения) и собственно классиология (классиология s. s.). В третьей части по степени общности рассмотрения вопросов классификации выделяются три основных раздела – общая, отраслевая и конкретная классиология. Отраслевая классиология рассматривает вопросы классификации в определенной научной дисциплине – математике, астрономии, геологии, биологии, лингвистике, технике, литературоведении и др. Конечно, можно рассматривать проблемы классификации в отдельно взятой отрасли, не принимая во внимание аналогичные ситуации в других отраслях знания. Однако, не вдаваясь в методологические минусы такой позиции, можно считать, что методологически выигрышным и логически естественным было бы предварительное составление перечня отраслей знания, для которых предусматривается рассмотрение классификационных вопросов. Иными словами – предварительное составление классификации-перечисления наук.

0.2. Создавать такой перечень наук предлагается на основе декларированного для всей предложенной системы классиологии принципа апостериорности [Покровский, 2006, с. 96; Покровский, 2014, с. 3740], – индуктивно, с опорой на опыт научного социума, физического носителя Науки. Главная идея, отвечающая такому подходу, – фиксация наук (их номинаций) или списков наук, фигурирующих в социальной научной среде (в публикациях и других материалах), и создание перечня наиболее социально весомых наук на основе анализа собранных данных.

Перечень наук, фигурирующих в научном социуме, огромен (около 1500? номинаций). Такое количество элементов может охватить только иерархическая классификация. Поэтому – если идти «от опыта», индуктивно – для начала разумно было бы сформировать ограниченный, не более нескольких десятков, перечень наук [Покровский, 2014, с. 361], а далее объединять науки этого перечня во все более крупные группы, наращивая классификацию «вверх», выделять разделы этих наук, более частные науки, развивая классификацию «вниз».

Для формирования такого ограниченного перечня наук можно предложить два подхода – статистический и логико-содержательный.

Статистический подход предусматривает какую-либо статистическую оценку распространенности, «упоминаемости», «весомости» науки в научных текстах, институтах или мероприятиях с формированием перечня наиболее социально значимых, «отбираемых» социумом наук.

Логико-содержательный подход предусматривает анализ содержания любой упоминаемой науки с выделением ее компонентов, если наука «гибридная» (геохимия = геология + химия, этноботаника = этнография + ботаника и т. п.) или выявлением более общей, чем данная, науки, если наука «простая» (петрология → геология, кардиология → медицина и т. п.) с формированием перечня «базовых» наук.

Легко понять, что в обоих подходах итоговый перечень будет представлять собой нечеткое *множество*: небольшая доля его элементов окажется включенной в него по случайным причинам, таким же образом некоторые науки, находящиеся в зоне условных границ этого множества, окажутся в него не включенными.

Разумно использовать оба подхода, синтезируя их результаты.

Здесь излагаются результаты статистического подхода.

1. ОБ ИСХОДНЫХ СПИСКАХ НАУК

Основная идея излагаемого статистического подхода – это оценка частоты упоминания и доли науки в различных списках (перечнях) наук. (Оговорим, что в дальнейшем часто вместо «списки наук», «перечни наук» для краткости употребляется формулировка «списки», «перечни».) Ясно, что если наука фигурирует в перечне, содержащем 50 номинаций, ее «социальный удельный вес» один, а если в перечне из 20 номинаций, ее социальная значимость гораздо выше.

Для такого подхода нужны перечни, претендующие (пусть неявно) на перечень «всех» наук, – перечни, явно не ограничиваемые научно-дисциплинарными рамками. Фигурирующие в таких перечнях науки можно свести в единый список и посмотреть, какие науки чаще встречаются в научном социуме.

1.1. Вербальные аспекты использования перечней наук

1.1.1. В работе используются только русскоязычные перечни. Это связано с тем, что понятие науки, обозначаемой определенным термином, в различной языковой среде может иметь различное толкование и различную социальную практику использования. Например, англоязычное *applied sciences* было бы естественно перевести как прикладные науки. Однако оказывается, что в эту рубрику включаются физика, геология, биология, медицина и др. В такой ситуации чисто лингвистический перевод термина теряет смысл. Необходимы специальные и, увы, трудоемкие лингвострановедческие исследования по толкованию и использованию понятия некой науки в соответствующей языковой среде.

1.1.2. В сводный список наук, фигурировавших во взятых в рассмотрение перечнях, включались и номинации, фиксируемые не только собственно названием науки («физика», «тектоника», «биология» и т. п.), но и объектом науки («высокомолекулярные соединения», «государство и право» и т. п.). Хотя в итоговый список наук [Покровский, 2020] вошли как наиболее социально значимые только науки, именуемые собственными названиями.

1.1.3. Лексика и стилистика источника сохранялись. Для номинаций, имеющих в разных источниках отличающуюся формулировку, но одинаковый смысл, считалось позволительным помещать информацию по ним под одним из терминов – более кратким или более принятым. Например, «физические науки» = «физика», «биологические науки» = «биология», «математические науки» = «математика» и т. п.; «культура» = «культурология», «музыка» = «музыковедение», «материалы», «на-

уки о материалах» = «материаловедение», «политика», «политические науки» = «политология» и т. п. Такие случаи редки (менее 30 для списка около 1200 номинаций) и, как можно надеяться, совершенно не искажают смысл номинаций в первоисточнике.

1.2. Об одной содержательной особенности анализируемых списков наук

Основными онтологическими единицами в науке могут считаться объект и метод [Покровский, 2014, с. 48–49]. Наиболее привычно понимание науки как науки об объекте (петрография, минералогия, зоология, ботаника и др.). Однако в ряде перечней, взятых в числе исходных, немалую долю составляют науки о методе (виброметрия, металлургия, нефелометрия, рентгенография, овощеводство и др.). Такие науки о методах, разумеется, тоже включались в сводный список наук, фигурировавших во взятых в рассмотрение перечнях. Хотя в итоговом списке наук [Покровский, 2020] из 60 лишь одна номинация представлена «наукой о методе» (в маргинальной группе).

1.3. Списки наук, использованные для наукометрической статистики

Списки (перечни) наук, послужившие материалом наукометрической статистики в этой работе, в первоисточниках имели явную или контекстуальную форму. Источники, из которых были взяты списки наук, относились к одной из перечисленных далее групп (цифры – число использованных источников, относящихся к этой группе):

- «списки наук» в электронных ресурсах – 4;
- информационные, информационно-поисковые и библиотечно-каталожные системы (в том числе 7 в электронных ресурсах, 1 – библиотечная (УДК)) – 8;
- названия журналов мультинаучных серий – 3;
- рубрики мультинаучных журналов – 12;
- номенклатура ВАК специальностей научных работников – 2;
- секции мультинаучных конференций (в том числе 3 очных, 14 заочных) – 17;
- номинации конкурсов (в том числе 2 на поддержку исследований, 13 на награждение премиями, медалями) – 15;
- анкеты-опросники, регистрационные формы семинаров, конференций – 3;
- социальная структура науки (перечень отделений РАН по областям и направлениям) – 1;
- рубрики тем-планов издательства «Наука» АН СССР – 2.

Всего, таким образом, использовано 67 источников, из которых взято 73 списка наук (в 6 источниках список был двухуровневым, т. е. в нем

содержалось фактически 2 списка наук). Собранные списки охватывают временной интервал 1985–2020 гг., но преобладающее большинство списков относится к 2000–2010 гг.

2. ОБЪЕМЫ СПИСКОВ НАУК

Точное число наук, известных человечеству (какого объема должен был бы быть исчерпывающий перечень наук) установить, по-видимому, невозможно. Для ориентировки по порядку величины приведем следующие данные.

Большой энциклопедический словарь [2004], по нашей оценке, содержит словарные статьи приблизительно о 1260 науках, в том числе о 940 науках об объектах и 320 науках о методах. Вряд ли не будет ошибкой полагать, что не осталось ни одной науки, не вошедшей в это издание. Поэтому общее число наук может быть увеличено до 1500.

В собранных нами 73 перечнях наук фигурирует в общей сложности 1229 номинаций наук. Отдельные перечни имеют объем от 2 до 748 номинаций. Несгруппированная выборка объемов этих перечней выглядит следующим образом (объем перечня – число перечней такого объема):

2 – 1	12 – 1	24 – 4	69 – 1
3 – 3	13 – 1	25 – 1	76 – 1
4 – 2	14 – 1	26 – 3	97 – 1
5 – 1	16 – 4	27 – 5	119 – 1
6 – 1	17 – 1	32 – 2	138 – 1
7 – 1	18 – 2	40 – 1	163 – 1
8 – 5	20 – 3	42 – 2	729 – 1
9 – 5	21 – 1	44 – 2	748 – 1
10 – 1	22 – 5	46 – 1	
11 – 1	23 – 3	47 – 1	

В несгруппированной выборке, сглаженной по стандартной методике, проявлены слабо выраженные максимумы частот (впрочем, заметные и в первичной несглаженной несгруппированной выборке) для объемов 3–4, 8–9, 16–17, 22–23, 26–27, 42–44 номинации в списке.

Группирование выборки производилось с использованием формулы для определения величины класса группирования [Руководство..., 1965]:

$$K \cong \frac{X_{\max} - X_{\min}}{1 + 3,2 \lg N},$$

где K – интервал группирования, $X_{\max}$, $X_{\min}$ – максимальная и минимальная варианты ряда, N – число вариантов в группируемом ряду.

Группирование всего диапазона объемов списков (от 2 до 748) дает своеобразную картину гиперболического распределения (табл. 1).

Огромный диапазон объемов – 214–642 членов перечня, 4 класса из 8 – лагуна, не представ-

Таблица 1. Распределение объемов списков наук (диапазон объемов – 2–748, число списков – 73)

Объем списка наук	Число списков наук с таким объемом	
	n	%
2–107	68	93,15
107–214	3	4,11
214–321	–	–
321–428	–	–
428–535	–	–
535–642	–	–
642–749	1	1,37
749–856	1	1,37
Всего	73	100,00

ленная ни одним перечнем. При этом два последних класса, следующих за этим хиатусом, имеют всего по одному списку. Разумно считать их своего рода аномалией и выявить распределение объемов перечней, сгруппировав выборку с исключением объемов 729 и 748 номинаций. Тогда диапазон группирования – от 2 до 163, число членов ряда – 71 (табл. 2).

Как видно, распределение объемов списков наук и с исключением значений 729 и 748 остается гиперболическим. При этом 48 списков из 71 (~2/3, ~68% списков) имеют объем от 2 до 25 номинаций, 17 списков из 71 (~1/4, ~24% списков) – от 25 до 48 номинаций. Таким образом, 65 списков из 71 (~9/10, ~91%) имеют объем ≤ 48 номинаций. Можно обратить внимание, что два названных «делящих» значения объемов списков (25 и 48) согласуются с перечнем слабо выраженных максимумов, выявляемых на несгруппированной выборке объемов списков.

Таблица 2. Распределение объемов списков наук с исключенными двумя максимальными значениями (диапазон объемов – 2–163, число списков – 71)

Объем списка наук	Число списков наук с таким объемом	
	n	%
2–25	48	67,7
25–48	17	23,9
48–71	1	1,4
71–94	1	1,4
94–117	1	1,4
117–140	2	2,8
140–163	1	1,4
Всего	71	100,0

Таким образом, можно констатировать, что в стихийной одноуровневой группировке наук намечаются следующие – не очень жестко фиксируемые – дискретные значения предпочтительных объемов таких групп, перечней, списков наук:

... 3–4 ... 8–9 ... 16–17 ... 22–23 ... 26–27 ... 42–44 ... 740 ... 1250 ... 1500? ...

Поскольку перечнями столь разных объемов представляется одно и то же множество «всех» наук, можно считать, что чем меньше позиций содержит такой перечень, тем в среднем более крупные группы наук представляет каждая позиция.

3. ВЫЯВЛЕНИЕ ГРУППЫ НАИБОЛЕЕ СОЦИАЛЬНО ЗНАЧИМЫХ НАУК

3.0. Исходные посылки

Науки всех 1229 собранных нами номинаций не могут быть весомы, значимы в равной степени. Необходимо ранжировать их, выделить ядро наук, которым социум уделяет наибольшее внимание. Поэтому в заголовке этого пункта говорится о социально значимых науках. Социум формирует перечни наук полубиологически, де-факто, никак не обосновывая, не аргументируя, не организуя объявляемый перечень, отчасти полустихийно, отчасти рефлекторно улавливая, «за что сейчас согласны платить деньги». Поэтому явные критерии выделения такого ядра социально значимых наук предлагаются здесь.

Для выделения группы таких наук целесообразно использовать две статистические особенности науки: 1) частоту ее включения в перечни наук, 2) оценку доли науки в перечнях, в которые она включена.

3.1. Частота R включения науки в сводные перечни наук

Для каждой из 1229 номинаций фиксируется, в скольких списках из 73 она включена. Различные науки оказались включенными в списки от 1 до 52 раз.

Несгруппированная выборка числа включений науки (R) в единичный список наук приводится далее (число включений науки (R раз) в списки – число наук, R раз включавшихся в списки):

1–343	11–9	23–1	37–1
2–649	12–3	24–1	40–1
3–77	13–5	25–2	41–1
4–44	14–3	26–1	43–2
5–23	15–1	29–1	45–1
6–13	17–4	31–1	46–1
7–5	18–1	32–1	48–2
8–9	20–1	34–1	52–2
9–5	21–1	35–1	
10–7	22–2	36–3	

Таблица 3. Частость включения науки в списки наук (диапазон R – от 1 до 52, число списков – 73)

Число R включений науки в списки наук	Число наук, включенных в списки R раз	
	n	%
1,0–5,7	1136	92,5
5,7–10,4	39	3,2
10,4–15,1	21	1,7
15,1–19,8	5	0,4
19,8–24,5	6	0,5
24,5–29,2	4	0,3
29,2–33,9	2	0,2
33,9–38,6	6	0,5
38,6–43,3	4	0,3
43,3–48,0	3	0,2
48,0–52,7	3	0,2
Всего	1229	100,0

Сгруппированная с помощью приведенной формулы выборка числа R включений науки в списки наук представлена в табл. 3 и на рис. 1. Как видно, частоты включения науки в списки наук также имеют гиперболическое распределение, при котором контрастно выделяются две группы наук: одна, отвечающая первым двум классам сгруппированной выборки ($R = 1,0–10,4$) и почти вертикальной ветви гиперболы; и вторая, отвечающая большинству классов сгруппированной выборки (4–11-й классы, $R = 15,1–52,7$) и практически горизонтальной ветви гиперболы. 3-й класс ($R = 10,4–15,1$) предлагается считать переходным, включающим элементы обеих групп, и делить относимые к нему науки по центральному значению класса: науки с $R < 12,75$ (т. е. 12 и меньше) отнести к первой группе, с $R > 12,75$ (т. е. 13 и более) – ко второй.

Таким образом, по частоте включения науки в перечни можно выделить две группы наук.

1 группа – многочисленны (1187 номинаций из 1229 (96,6%)) малоупоминаемые (12 и менее раз – <17% списков) науки. Из них 343 номинации встречаются из 73 списков лишь в одном, а 649 – лишь в двух.

2 группа – немногочисленные (42 номинации из 1229 (3,4%)) часто включаемые в перечни (13 и более раз (в нашем случае – до 52) – >17% списков) науки. Список этих 42 наук начинают медицина и экономика, фигурирующие каждая в 52 списках из 73, и заканчивают военные науки, литературоведение, органическая химия, физкультура и спорт и этнография, включенные каждая в 13 списков из 73¹.

¹ Полностью список здесь не приводится из-за ограниченного объема публикации.

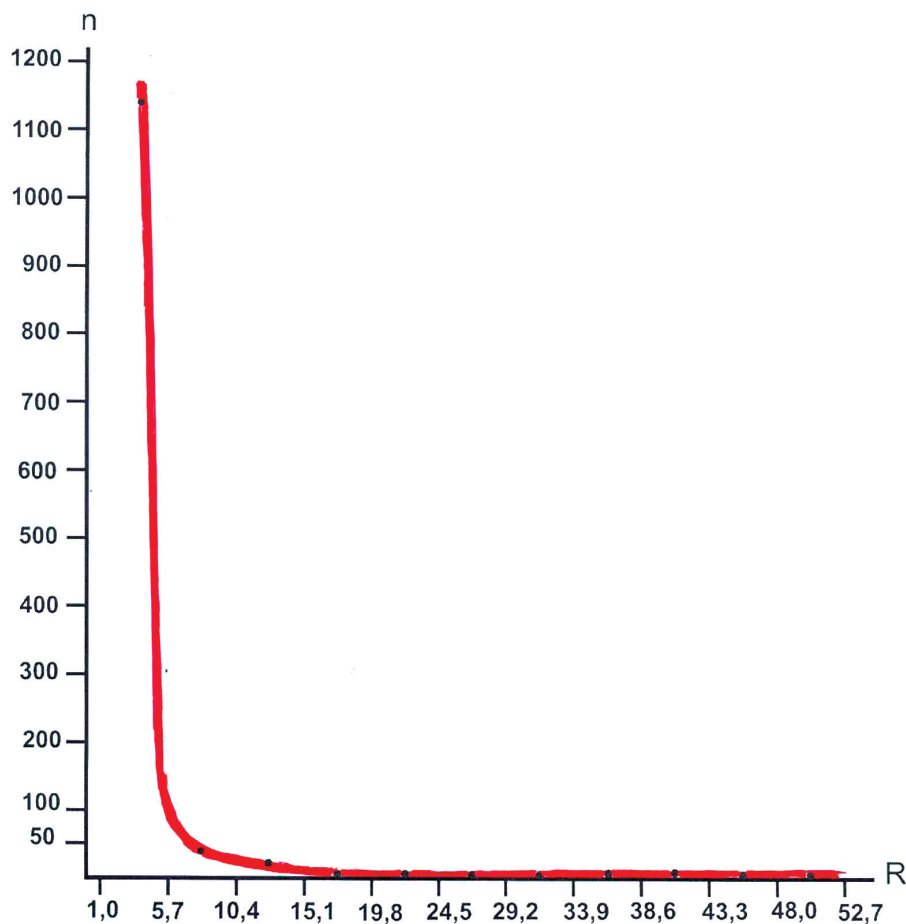


Рис. 1. Частота R включения науки в списки наук.

Итак, по показателю R как наиболее часто включаемые в перечни выделяются 42 науки со значением R от 13 до 52.

3.2. Доля науки в списке наук

Доля науки в перечне — одна из главных идей излагаемого здесь статистического подхода. Можно полагать не нуждающимся в пояснении и обосновании то, что науке, входящей в список из 46, 22 и 3 номинаций следует приписать различный «удельный вес» среди множества наук, охватываемых этими перечнями.

3.2.1. *Определение доли d науки в перечне, различных показателей доли d и методика статистического анализа этих данных.*

3.2.1.1. *Определение доли науки в одном отдельном списке.* В каждом списке, включающем N позиций, каждой позиции приписываются равные доли, выражаемые в процентах ($1/N \cdot 100$, %).

Если позиция содержит одну научную номинацию, то это и будет доля d данной науки в данном списке ($d = 1/N \cdot 100$, %). Для того чтобы доля науки из списков, например, в 1255, 1256, 1257 номина-

ций различалась, ее необходимо вычислять с точностью до 4-го знака после запятой (при необходимости округления оно производится по Гауссу).

Если позиция включает несколько равнозначных номинаций, каждой номинации приписывается равная доля этой позиции. Например, одна из 8 позиций списка (доля позиции — 12,5000) — «Кибернетика. Семиотика. Информатика». Доля каждой из названных наук — 4,1667.

Если позиция фигурирует в такой редакции — «Кибернетика. Семиотика и информатика», она считается состоящей из двух элементов (1 — кибернетика, 2 — семиотика и информатика), каждому из которых приписывается половина доли позиции — 6,2500. Доля второго элемента делится поровну между входящими в него науками. Таким образом, в этом случае доля d кибернетики — 6,2500, семиотики — 3,1250, информатики — 3,1250.

Если источник содержит двухуровневый перечень наук, доля наук на каждом уровне определяется отдельно — как для самостоятельного перечня (перечни наук на разных уровнях обычно различны). Если одна и та же наука фигурирует на разных уровнях (такая не вполне корректная ситу-

ация встречается), доля этой науки подсчитывается для каждого уровня и суммируется; науке – для данного источника – приписывается именно эта, суммарная, доля.

3.2.1.2. Характеристика доли d науки по данным из разных источников.

Всего в 73 взятых в рассмотрение списках наук фигурируют 1229 номинаций науки. Для каждой науки (одна из 1229 строк суммирующего списка) фиксируются:

1) число перечней, содержащих данную науку (частота R включения науки в списки – п. 3.1);

2) сумма Σd единичных значений d науки в разных списках – списках из разных источников (частое включение науки с малым d и редкое включение науки с большим d могут дать близкие величины Σd и свидетельствовать о примерно одинаковой социальной значимости наук);

3) среднее значение d_{cp} для науки ($d_{cp} = \Sigma d/R$);

4) диапазон значений d данной науки в разных источниках – $d_{min}-d_{max}$; если d науки имеет единственное значение (наука встречается только в одном из 73 перечней наук или (редкий случай) все зафиксированные в разных перечнях доли науки одинаковы), вместо диапазона «от – до» ($d_{min}-d_{max}$) фиксируется именно это единственное значение.

3.2.1.3. Статистическая обработка данных по долям d различных наук.

В целях выделения по указанным показателям из 1229 номинаций относительно немногочисленных групп наиболее социально значимых наук проводилась обработка полученных данных – выделялись группы наук, имеющих наибольшие значения Σd , d_{cp} , d_{max} (в последнем случае, если d имело единственное значение (см. ранее), при статистическом анализе d_{max} в расчет бралось оно).

Методика обработки и анализа данных, полученных по названным трем показателям, была аналогична таковой для частоты R включения науки в перечни (п. 3.1).

Для каждого из трех названных показателей фиксировались наибольшее и наименьшее значения в ряду из 1229. С использованием упомянутой формулы формировалась сгруппированная выборка (иногда последовательно, дважды, как при оценке объемов перечней наук (п. 2) – с исключением аномальных значений). Распределение всех трех показателей оказалось также гиперболическим, с четким различием вертикальной и горизонтальной ветвей, и в состав выделяемого ограниченного (и, напомним, нечеткого) множества включались немногие науки с максимальными значениями рассматриваемого показателя («горизонтальная ветвь гиперболы» + половина «переходного» класса, члены которого имеют значения, превышающие срединное для этого класса).

3.2.2. Группы социально значимых наук, выделяемые по особенностям доли d науки в перечнях.

Приведем основные результаты статистического анализа трех упомянутых показателей «удельного веса» наук.

Сумма Σd единичных значений d науки в разных списках колеблется от 0,1244 до 350,6655. Распределение гиперболическое. По показателю Σd как обладающие наибольшими суммами долей науки в разных списках выделяются 37 наук. Перечень их начинают гуманитарные науки ($\Sigma d = 350,6655$) и заканчивает металлургия ($\Sigma d = 48,2560$).

Средняя доля науки d_{cp} у разных наук колеблется от 0,1244 до 33,3333. Распределение гиперболическое. По показателю d_{cp} как имеющие наибольшее значение доли науки в среднем выделяются 16 наук. Перечень их начинают социально-экономические науки ($d_{cp} = 33,3333$) и заканчивает теоретическая физика ($d_{cp} = 5,0207$).

Максимальные единичные значения долей науки d_{max} (максимальное значение d науки в отдельном списке) колеблются от 0,1022 до 50,0000. Распределение показателя – гиперболическое. По показателю d_{max} как имеющие наибольшую долю в отдельном списке выделяются 28 наук. Перечень их начинают гуманитарные и естественные науки ($d_{max} = 50,00000$ каждой группы) и заканчивают (перечисление – в алфавитном порядке) математика, металловедение, металлургия, науки о человеке, общая биология, органическая химия, охрана природы, прикладная математика, теоретическая физика, физика, химия твердого тела, экология, экспериментальная физика, электрофизика, электрохимия (каждая из них имеет $d_{max} = 12,5000$).

ЗАКЛЮЧЕНИЕ

1. В современном научном социуме, по грубой оценке, фигурируют науки около 1500 номинаций. Классификация наук важна как общенаучная проблема и как основа развития и структурирования отраслевой (дисциплинарной) классификации.

2. Строить классификацию наук предлагается индуктивно, «от опыта». При этом возможны два подхода – статистический, оценивающий социальную значимость науки, и логико-содержательный, анализирующий содержание науки.

3. В рамках статистического подхода предлагается оценивать частоту наличия и долю наук в списках наук, претендующих на внедисциплинарность. По предложенной методике (с использованием известного приема группирования выборки) эти особенности были оценены у 1229 наук, входивших в 73 списка наук.

4. На основании 4 социальных особенностей наук (частота R включения в списки, сумма долей Σd науки в разных списках, средняя доля d_{cp} науки в разных списках, единичное максимальное значение d_{max} науки в одном списке наук) были выявлены статистически обособленные 4 группы наук – 42, 37, 16

и 28 наук соответственно. Необходимо каким-либо образом интегрировать эти группы наук, выделенные по родственным, но отдельным особенностям, в единый список.

СПИСОК ЛИТЕРАТУРЫ

- Большой энциклопедический словарь. 2-е изд., перераб. и доп. / гл. ред. А.М. Прохоров. М.: БРЭ; СПб.: Норинт, 2004. 1456 с.
- Покровский М.П.* Введение в классиологию. Екатеринбург: ИГГ УрО РАН, 2014. 484 с.
- Покровский М.П.* Классиология как система // Вопросы философии. 2006. № 7. С. 95–104.
- Покровский М.П.* Об индуктивной классификации наук и месте геологии в системе наук (статистический подход). Ст. 2 // Ежегодник-2019. Тр. ИГГ УрО РАН. Екатеринбург: ИГГ УрО РАН, 2020. В печати.
- Руководство по предварительной математической обработке геохимической информации при поисковых работах / сост. А.А. Беус, С.В. Григорян, М.Т. Ойзерман, П.Г. Чолакян, А.А. Стояновский. М.: Недра, 1965. 120 с.