

Министерство высшего и среднего специального образования РСФСР  
ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ  
ГОРЬКОВСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ РАДИОФИЗИЧЕСКИЙ  
ИНСТИТУТ

Препринт № 135

**ОСНОВНЫЕ ЭЛЕМЕНТЫ НАУЧНЫХ ПУБЛИКАЦИЙ  
И ИХ АНАЛИЗ**

*В. Т. Федоров*

Горький 1980

Исследуются основные элементы научных публикаций: классификационный индекс, термины, фамилии и инициалы авторов, названия организаций, название источника, цитирование и цитируемость. Проводится анализ каждого из элементов, включая ранговое распределение его конкретных реализаций. Вводится понятие элементной значимости публикации и предлагается метод ее вычисления. Широко используются литературные источники Библ 171 назв.

## СО Д Е Р Ж А Н И Е

1. Основные элементы описания научных публикаций . . . . .	<b>3</b>
2. Классификационный индекс . . . . .	<b>5</b>
3. Термины . . . . .	<b>9</b>
4. Фамилии авторов . . . . .	<b>16</b>
5. Организации . . . . .	<b>20</b>
6. Название источника . . . . .	<b>21</b>
7. Библиографические ссылки . . . . .	<b>23</b>
8. Выводы . . . . .	<b>34</b>
9. Литература . . . . .	<b>34</b>

## 1. ОСНОВНЫЕ ЭЛЕМЕНТЫ ОПИСАНИЯ НАУЧНЫХ ПУБЛИКАЦИЙ

Научная публикация, как и публикация вообще, представляет собой письменный документ, изготовленный при помощи полиграфических или других средств массового размножения документов и предназначенный для широкого распространения [1, с. 58]. После возникновения книгопечатания в Западной Европе в 1440 г., в России в 1553—1564 гг., в Северной Америке в 1640 г. и по мере накопления печатных документов был поставлен вопрос об их описании и учете, что имело решающее значение для судьбы библиографии как науки и практической деятельности [2]. Необходимость унификации этой деятельности в разных странах привела к разработке стандартов на библиографическое описание, или библиографическую запись публикаций [3 4]. Библиографическое описание — совокупность сведений о произведении печати или другом документе (его части или группе документов), дающих возможность идентифицировать документ, а также получить представление о его содержании, читательском назначении, объеме, справочном аппарате и т. д. Библиографическое описание произведений печати состоит из элементов, которые по своим функциям объединены в области. Элементы подразделяются на обязательные и факультативные [3, с. 1. 2]. Библиографические записи разных видов произведений печати или предназначенные внутри одного и того же вида для различных целей могут иметь те или иные отличительные особенности.

В связи с внедрением в национальном и международном масштабах автоматизированных информационных систем разработаны предмашинный и коммуникативный форматы библиографической записи с целью широкого обмена информацией в системах «человек—машина» и «машина—машина» [5, 6].

Элементы публикаций (входящие и не входящие в библиографическое описание), раскрывающие главные, основные аспекты их сущности и идентификации, являются основными элементами. Именно на них ориентировано большинство традиционных и автоматизированных ИПС.

Наиболее древним среди основных элементов следует признать классификационный индекс, поскольку классификация литературы берет свое начало с зарождения первых библиотек. Классификационный индекс не входит в стандартное библиографическое описание, но присваивается публикации при вводе ее поискового образа в целый ряд широко распространенных библиотечных и информационных систем. Сущность этого элемента — в раскрытии (определении, идентификации) содержания публикации в системе человеческих знаний. Содержание публикации, конечно, наиболее полно характеризуется ее текстом. Среди элементов текста в этом отношении наиболее важным являются термины — научные понятия, входящие в заголовок (название) публикации, а также в ее аннотацию и/или реферат. Термины (ключевые слова) заглавия входят в явном виде в библиографическое описание и потому относительно легко могут быть выделены из него. При наличии аналитико-синтетической свертки текста документа в виде аннотации и/или

реферата список терминов, составленный по заголовку, может быть существенно дополнен и расширен. Полученный таким образом список терминов выступает как один из основных элементов, характеризующих содержание публикации, он является основным элементом большинства автоматизированных ИПС, использующих метод координатного индексирования документов.

Помимо элементов, раскрывающих содержание публикации, не менее важны и те, которые служат целям ее идентификации в системе других публикаций<sup>1</sup>. Наиболее важным среди этих элементов является (ютя) автор (ы) публикации. Не случайно, а исторически обусловлено то обстоятельство, что заголовки библиографического описания посвящены авторам. Авторскому аспекту посвящены частично также области заглавия и сведений об авторстве, издания и примечаний в библиографическом описании. В качестве конкретных представителей данного элемента в поисковом образе документа выступает список фамилий и имен лиц — авторов публикации, составителей, редакторов, переводчиков и т. п.<sup>2</sup>, либо список учреждений и организаций, выступающих в роли коллективного автора.

Последний случай имеет важное значение для характеристики публикаций в период интенсивного развития научных коммуникаций. Вот почему название организации (й), хотя и не входящее в состав библиографического описания при наличии авторов, целесообразно выделить в самостоятельный элемент поискового образа публикаций, имея в виду его использование в автоматизированных ИПС. Название организации (й) — важный элемент, т. к. очень часто характеризует уровень публикации, содержит ценнейшую адресную информацию о месте, где работают (ли) авторы публикации, где выполнена работа или получены те или иные результаты.

Идентификация публикации среди других произведений печати невозможна без указания источника. Этой цели в библиографической записи служат области издания, выходных данных, серии, примечаний. Для статьи этим важным идентифицирующим элементом является название сериального издания (журнала), обозначение и порядковый номер его тома, части или выпуска, страниц, на которых напечатана статья.

Дата также является самостоятельным важным элементом, поскольку определяет положение произведения печати во времени. Важность этого элемента связана с вопросами приоритета, право на который распространяется лишь на ученого, который впервые опубликовал (а не достиг!) результат. Наибольший интерес для использования в автоматизированных системах представляют даты поступления рукописи в редакцию<sup>3</sup> и опубликования.

Совершенно особым элементом являются библиографические ссылки, отражающие преемственность публикаций. Их повсеместное и почти обязательное использование в тексте публикаций дает возможность каждой работе сопоставить два списка: список цитирования (перечень цитированных в работе документов) и список цитируемости (перечень документов, цитирующих данную работу).

Совокупность этих двух списков — цитирование и цитируемость — будет рассматриваться в качестве важного элемента научной публикации. К сожалению, этот элемент не входит в стандартную библиографи-

---

<sup>1</sup> Следует отметить, что любой из элементов служит как целям раскрытия содержания, так и идентификации, хотя и не всегда в явной форме.

<sup>2</sup> Подробно об анализе и возможностях этих и других категорий лиц характеризовать публикацию см. [7]

<sup>3</sup> Эта дата, к сожалению, не входит в стандартное библиографическое описание.

ческую запись, его практическое использование реализовано лишь в немногих автоматизированных системах, но там, где это имеет место, значение его трудно переоценить. И действительно, ссылки — особый язык, обладающий большой семантической и идентифицирующей силой: состоя из совокупности многих уже упомянутых элементов, он позволяет вскрыть взаимосвязь публикаций, искусственно разделенных принадлежностью разным разделам классификационных схем, а также использующих различные терминосистемы в своих текстах; позволяет легко устанавливать научные связи авторов, организаций, эффективность самоорганизации информационных каналов (источников), взаимодействие различных элементов во времени.

Важность и сила каждого из указанных элементов не вызывает сомнений, однако хотелось бы подчеркнуть, что только их совместное использование позволяет наиболее полно характеризовать публикацию и определить ее роль и значение в системе других работ. Но прежде следует подробно рассмотреть каждый из названных элементов:

1. Классификационный индекс.
2. Термины
3. Фамилии и инициалы автора (ов).
4. Название организации (й)
5. Название источника.
6. Дата.
7. Цитирование и цитируемость.

Этому и посвящены (за исключением даты) последующие разделы данной работы.

## 2. КЛАССИФИКАЦИОННЫЙ ИНДЕКС

Классификация — это один из способов логического анализа сложных систем [8, с. 26], один из методов системного подхода. Классификация в простейшей форме связана с группировкой сходных и разделением отличных друг от друга объектов или явлений окружающего нас мира.

Центральной проблемой классификации публикаций является определение их содержания (смысла, значения) с помощью классификационного языка.

Документные классификации возникли давно, с появлением первых библиотек [9]. Вместе с развитием библиотечного дела, а позднее и библиографии, совершенствовались и библиотечно-библиографические классификации, основное назначение которых — организация расстановки книг на полках и ссылок на документы в библиографических источниках. Вся история этих классификаций от древности до наших дней свидетельствует об органической, неразрывной их связи с классификациями наук [10—12].

Традиционная (иерархическая) библиотечно-библиографическая классификация иллюстрируется обычно в виде перевернутого дерева [13]. Классифицируемые предметы<sup>4</sup> — это листья или плоды, свисающие с дерева. Они более схожи между собой, находясь на одной из веточек, чем с другими (на других ветках). Смысл иерархической классификации заключается в раскрытии объема предмета путем перечисления всех предметов, которые являются по отношению к нему видовыми, т. е. в делении родового предмета. Основанием деления должен быть не произвольно взятый, а существенный признак классифицируемых предметов.

<sup>4</sup> Предметы могут быть материальными (тело, поле) и идеальными (понятие). В качестве предмета можно рассматривать свойства и отношения, которые в этом случае будут являться абстрактными предметами [1].

Классификация называется естественной, если листья или плоды дерева тем ближе находятся друг к другу, чем более они сходны. Основные правила иерархической классификации [13]:

— каждый признак (основание) деления должен производить по меньшей мере 2 класса;

— только один признак деления может быть использован одновременно, чтобы создать взаимно-исключающие классы;

— виды любого рода должны полностью исчерпывать родительский класс.

Эти довольно жесткие правила не могли длительное время служить целям классификации, вся история развития которой есть история ломки жесткости. Это вполне понятно, если иметь в виду непрерывное развитие науки и научной литературы, когда растет или ослабевает значение ее отдельных областей, непрерывно возникают и развиваются новые направления, совершенно преобразуются целые разделы науки. Средствами жестких систем можно адекватно систематизировать лишь науку и, соответственно, научную литературу, остановившуюся в своем развитии, либо развивающуюся достаточно медленно. Вот почему структурные построения и рамки широко известных и используемых в мире классификаций Дьюи (с 1876 г.), УДК (с 1905 г.), Блисса (с 1935 г.), Ранганатана (с 1933 г.) ныне все более и более не соответствуют революционным процессам в современной науке и научной литературе. Сейчас, как никогда прежде, важнейшее значение приобретают интеграционные процессы.

В настоящее время с учетом возможной нейтрализации этого несоответствия можно предложить следующее описание полностью развитой классификации [13].

1. Весь универсум литературы делится на достаточно широкие предметные области, расположенные в некотором значимом порядке.

2. Каждая предметная область делится на фасеты<sup>5</sup>, расположенные в определенном значимом порядке. Например, в астрономии могут быть следующие фасеты: небесные тела (звезда, планета, комета и т. д.), их части (ось, хвост, поверхность и т. д.), система тел (галактика, созвездие, скопление и т. д.), свойства тел (размер, температура, химический состав и т. д.).

3. Структура каждого фасета представляется в виде иерархической системы, каждая ступень которой есть результат последовательного деления на основе ряда характеристик (признаков). Любой в рамках этой системы набор терминов, образованный последовательным делением, представляет собой цепочку, а любой уровень деления есть группа (множество) терминов (например, в астрономии для уровня «планеты» — Меркурий, Марс, Земля, Венера и т. д.). Термины в пределах группы располагаются в некотором значимом порядке, который должен определяться природой определяемых ими сущностей. В большинстве случаев употребительны следующие порядки следования [13]: 1) логический — от простого к сложному, 2) геометрический — в соответствии с расположением вещей в пространстве, 3) хронологический — в соответствии с положением во времени, 4) генетический — в соответствии со сходством происхождения, 5) исторический — комбинация предшествующих четырех принципов, 6) эволюционный — от простого к сложному, 7) динамический — в порядке возрастания силы, мощности, энергии, 8) алфавитный, 9) математический — в соответствии с цифрами натурального ряда, 10) в порядке уменьшения экстенсии (увеличения, расширения, распространения), 11) в порядке увеличения конкретности.

<sup>5</sup> Фасет — совокупность подклассов, получаемая при делении основного класса по одному и тому же ряду характеристик [1].

4. Формулируются правила комбинации (координации) терминов из каждой группы, различных групп одного и того же фасета, различных фасетов одной и той же предметной области. Правила могут включать специальные операторы отношений или указателей роли.

5. Каждая предметная область, фасет, термин кодируются с целью фиксации их положения в классификационной системе и недвусмысленной комбинации с другими кодами. Система кодов, нотация или индексация является лишь инструментом классификационной системы и никогда не может диктовать порядок и структуру ее организации. С ее помощью структурным единицам классификации присваиваются символы, составленные из знаков языка, знакомого и удобного пользователю, что позволяет осуществлять их поиск. Нотация может определять взаимоотношение между предметами путем сопоставления их индексов, поскольку индексы, обозначающие связанные предметы, имеют общие элементы. Наиболее естественными знаками, снижающими нагрузку на память человека, являются арабские цифры (0—9) и знаки букв (латинские: A—Z, русские A—Я). Именно эти знаки широко используются в наиболее распространенных классификациях [14]. Важно, чтобы нотация позволяла сохранять введенные обозначения и не приводила к существенной перекодировке по мере развития классификационной системы. Это сейчас предусмотрено как в УДК, так и в классификации двоемечием Ранганатана [15].

6. Составляется алфавитный указатель терминов, фасетов, предметных областей с отсылкой к соответствующим индексам нотационной системы.

Другой выход из трудностей классификационной жесткости заключается в широком распространении так называемых специальных или локальных классификаций [13, 16]. Этот подход целенаправлен на решение проблем систематизации литературы в пределах ограниченной, а в ряде случаев довольно узкой предметной области. Очень часто разработка и использование этих классификаций обусловлена интересами научно-исследовательских организаций, коллективов специалистов, объединенных решением одной проблемы или задачи, а в ряде случаев и отдельных лиц.

Первый шаг в разработке специальной классификации состоит в сборе терминов, относящихся к предметной области. При этом должны быть обследованы учебники, обзоры, словари, алфавитно-предметные указатели существующих классификаций, а также отчеты и наиболее важные статьи. Немалую помощь могут оказать специалисты-эксперты, которых можно попросить: просмотреть литературу данной предметной области, индексируя ее наиболее подходящими терминами, дать список вопросов, которые они могут задать поисковой системе, сделать замечания и добавления к готовому списку терминов или разделов классификационной схемы, обсудить терминологию и классификацию в трудных и неясных местах. Следующий шаг начинается с группировки терминов в фасеты с помощью отнесения их к категориям. Например, Ранганатан признает только пять категорий [15]: индивидуальность (персонализация), энергия, материя, место, время. Однако каким бы исчерпывающим и логичным не казался тот или иной конкретный ряд категорий, выбор его не может проводиться а priori, а должен базироваться на конкретном литературном материале данной предметной области. Далее необходимо установить порядок комбинирования категорий. Это лучше всего сделать с помощью специалистов, для которых строится классификационная схема. Часто локальные системы называют локально-десятичными системами (ЛДК), нотация которых строится по принципу УДК. Конечно, деление основной предметной области на ограниченное число (до 10) классов является довольно условной процедурой, в основе кото-

рой, однако, лежат, по-видимому, человеческие возможности. Так, согласно [17, 18], «число позиций на шкале (релевантности) может быть самое большее 6 или 7,— максимум позиций, которые сознательно может различать человеческий разум».

В то время как литературной поддержкой универсальных классификаций являются в основном книги, локальные классификационные схемы опираются в большей мере на журнальные публикации и очень часто представляют собой классификацию фронта исследований предметной области. Здесь нельзя не отметить важную роль обзоров [19], включающуюся, с одной стороны, в рассмотрение самого предмета исследования, а с другой, в рассмотрении литературы, посвященной этому предмету. Структура содержания обзора может служить нередко и научной классификацией проблемы, и классификацией посвященной ей литературы. Вот почему локальные классификации, опирающиеся на обзорную литературу и наиболее важные журнальные статьи, в большей степени стоят на высоте современных требований по сравнению с соответствующими разделами универсальных классификаций.

Следует признать, что наиболее актуально сейчас решение проблем классификации применительно к автоматизированным информационно-поисковым системам (АИПС), состоящим из логико-семантического аппарата, поискового массива, технических средств и людей, взаимодействующих с системой. Логико-семантический аппарат или абстрактная ИПС представляет собой совокупность информационно-поискового языка (ИПЯ), правил индексирования и критерия выдачи. Любой ИПЯ — это не что иное как система классификации. Если традиционные иерархические и алфавитно-предметные классификации относятся к языкам, классы которых строятся путем их логического умножения до индексирования документов, то современные ИПЯ — это языки, классы которых строятся сначала потенциально при индексировании документов, а затем реально при их поиске. Традиционные (включая фасетные) классификации относятся к предкоординированным, а современные — к посткоординированным ИПЯ [20, 21]. Наиболее распространенными посткоординированными языками являются языки дескрипторного типа, в основу построения которых положен принцип координатного индексирования документов, когда координатами документа в некотором  $n$ -мерном пространстве предметно-тематических признаков становятся имена простых классов-дескрипторов, т. е. центральная тема документа и запрос выражается путем логического перемножения дескрипторов. Эффективная реализация ИПС, основанных на использовании дескрипторных ИПЯ, предполагает создание и использование нормативно-поисковых тезаурусов (например, [22]) — нормативных словарей-справочников, в которых эксплицитно выражены важнейшие парадигматические отношения между лексическими единицами языка и которые снабжены различными входами: алфавитно-предметным, систематическим и т. д. и соответствующих средств автоматизации и механизации (перфокарты ручного обращения, счетно-перфорационные машины, ЭВМ) [20]. Несмотря на довольно интенсивное развитие работ в области создания и исследования ИПЯ с целью решения задачи алгоритмического выделения центрального предмета или темы из текста на естественном языке (см. например, [23—26]), задача автоматического индексирования текстов на естественном языке пока не решена и не может быть реализована на машине.

В заключение этого раздела рассмотрим процесс деления предметной области, лежащий в основе классификационного способа анализа сложных систем, с количественной точки зрения [27]. Предположим, что предметная область разбита на  $n$  классов, и каждую из  $t$  публикаций этой предметной области мы будем относить лишь к одному из классов.



При таком подходе можно ввести меру концентрации материала в классах предметной области  $q$  при условии рангового характера распределения публикаций по классам:

1. ранг класса	1	2	3	...	$n$	,
2. число статей в классе при равномерном распределении	$t/n$	$t/n$	$t/n$	...	$t/n$	,
3. число статей в реальном распределении	$a_1$	$a_2$	$a_3$	...	$a_n$	.
4. отношение (реальный случай/равномерный случай)	$\frac{a_1 n}{t}$	$\frac{a_2 n}{t}$	$\frac{a_3 n}{t}$	...	$\frac{a_n n}{t}$	,
5. умножение на ранг и суммирование	$\frac{a_1 n}{t}$	$\frac{2a_2 n}{t}$	$\frac{3a_3 n}{t}$	...	$\frac{na_n n}{t}$	,
6. вычисление среднего (деление на $n$ )	$\frac{a_1}{t}$	$\frac{2a_2}{t}$	$\frac{3a_3}{t}$	...	$\frac{a_n}{t}$	,
7. мера концентрации	$q = \frac{1}{t} (a_1 + 2a_2 + 3a_3 + \dots + na_n)$ .					(1)

Полученная мера концентрации  $q$  может быть нормирована, если принять, что концентрация материала отсутствует в случае равномерного распределения статей по классам предметной области, или равна нулю, а при попадании всех статей в один из классов, она максимальна и равна единице ( $C$  — нормированная мера концентрации),

$$C = \frac{2 [(n + 1)/2 - q]}{n - 1}. \quad (2)$$

С помощью меры  $C$  можно сравнить различные предметные области, а также различные уровни деления внутри одной и той же классификации с точки зрения концентрации материала при наличии его рангового распределения.

### 3. ТЕРМИНЫ

#### Определение, сущность термина

«Термин»<sup>6</sup> в современном понимании есть «слово или словосочетание, призванное точно обозначить понятие и его соотношение с другими понятиями в пределах специальной сферы»<sup>7</sup> [29, с. 140].

<sup>6</sup> Слово «термин» берет свое начало от латинского слова «terminus», первоначально обозначавшего «божество границ», впоследствии «пограничный камень», «рубеж», затем «предел», «окончание». Однако в средневековой латыни среди значений слова были и такие: «определение», «обозначение», что привело к появлению в старофранцузском словаре «terme» со значением «слово». Французский и Большой Оксфордский словари отмечают появление «термина» со значением «слово» во французском и английском языках с XIV в. «В русском языке слово «термин» появляется в первой половине XVIII века». «Вплоть до настоящего времени для обозначения научного понятия «термин» в целом ряде европейских языков широко употребительно латинизированное устойчивое словосочетание «terminus technicus», которое в них заимствуется (ср. франц.: «terme technique», англ.: technical terms, русс.: «технический термин») или калькируется» [28, с. 2, 3].

<sup>7</sup> Имеются и другие, более расширенные определения «термина»: «Термин — это слово или словосочетание (образованное на базе подчинительных связей), имеющее профессиональное значение, выражающее и формирующее профессиональное понятие, которое применяется в процессе (и для) познания и освоения некоторого круга объектов и отношений между ними — под углом зрения определенной профессии» [30, с. 51]. «Термин» — это слово или словесный комплекс, соотносящиеся с понятием определенной организованной области познания (науки, техники), вступающие в системные отношения с другими словами и словесными комплексами и образующие вместе с ними в каждом отдельном случае и в определенное время замкнутую систему, отличающуюся высокой информативностью, однозначностью, точностью и экспрессивной нейтральностью» [28, с. 81; 31, с. 21].



С увеличением содержания понятия — совокупности отличительных признаков, ядром которой являются существенные признаки, — уменьшается его объем — совокупность предметов, отображенных в данном понятии, и наоборот, что является законом обратного отношения между содержанием и объемом понятия. Понятия разделяются на различные классы с точки зрения их содержания и объема, эти классы схематически представлены на рис. 1.

Рассмотрим основные существенные признаки многообразия связей между понятиями и терминами.

1. Соотнесенность «понятие — термин» принадлежит к науке (технике, профессиональной деятельности) и ее языку, но по природе своей глубоко и тесно связана с соотнесенностью «понятие — слово»<sup>11</sup>.

2. Определение (дефиниция, значение) связывает понятие и термин<sup>12</sup>.

3. Термины — члены терминологической системы (терминосистемы), понятия — члены системы понятий. Обе системы теснейшим образом связаны друг с другом, более изоморфны друг другу, чем понятия и слова обыденной жизни, но не до конца, в этом их сила и богатство. Распределение единиц системы понятий среди единиц системы терминов (вместе с неизоморфизмом) приводит к явлениям синонимии, синонимической дублетности, антонимии. Обратный процесс характеризуется явлениями полисемии, омонимии. Терминосистемы образуют иерархию, соответствующую структуре системы знаний о природе и обществе, причем каждая из терминосистем включает также термины из выше- и нижестоящих уровней иерархии.

---

<sup>11</sup> «Нет никакой принципиальной разницы между «обыденными» понятиями и научными понятиями по их форме. Различие может быть только в степени точности и глубине отражения» [40, с. 128]. «В принципе они [«тривиальные понятия, составляющие содержание слов обычной речи»] однородны с абстрактными понятиями науки и отличаются от них только допуском, аппроксимацией, степенью приближения» [41, с. 95]. «Между словарем науки и словарем быта — прямая и тесная связь. Всякая наука начинается с результатов, добытых мышлением и речью народа, и в дальнейшем в своем развитии не отрывается от народного языка» [42, с. 6]

<sup>12</sup> Этой проблеме посвящено специальное Всесоюзное совещание [43]. «Термины не существуют без определений» [44, с. 11] «Определение научного понятия и определение термина (знака этого понятия) — вещи не идентичные». «Все богатство и глубина понятия не укладывается в одной дефиниции» «Фиксируя значение терминов, мы фиксируем и понятия, которые они передают» [46, с. 73]. «Хотя бы выражение «определение термина» — это по существу есть определение понятия, представленного термином» [47, с. 107]. «В подлинной науке понятие — интернационально, термин — часто национален...» «Дефиниции по своему содержанию, так же как и понятия, должны быть интернациональны» [48, с. 157] «Понятие может стать свободным, номинативным значением слова, но и в этом случае семантика слова в целом, рассматриваемого в аспекте системы языка, не исчерпывается и не ограничивается только выражением этого понятия. Что же касается других видов лексических значений слов [имеется в виду «фразеологически-связанных», «функционально-синтаксически ограниченных», «конструктивно организованных или обусловленных»], то эти значения настолько слиты со спецификой данного конкретного языка, что общечеловеческое, понятийное, логическое содержание в них обрывается со всех сторон своеобразными формами и смысловыми оттенками национального творчества данного народа» [42, с. 29] «...дефиниции не имеют значения для науки, потому что они всегда оказываются недостаточными. Единственно реальной дефиницией оказывается развитие существа самого дела, а это уже не есть дефиниция. Для того, чтобы выяснить и показать, что такое жизнь, мы должны исследовать все формы жизни и изобразить их в их взаимной связи. Но для обыденного употребления краткое указание наиболее общих и в то же время наиболее характерных отличительных признаков в так называемой дефиниции часто бывает полезно и даже необходимо, да и не может вредить, если только от дефиниции не требуют, чтобы она давала больше того, что она в состоянии выразить» [49, с. 322]

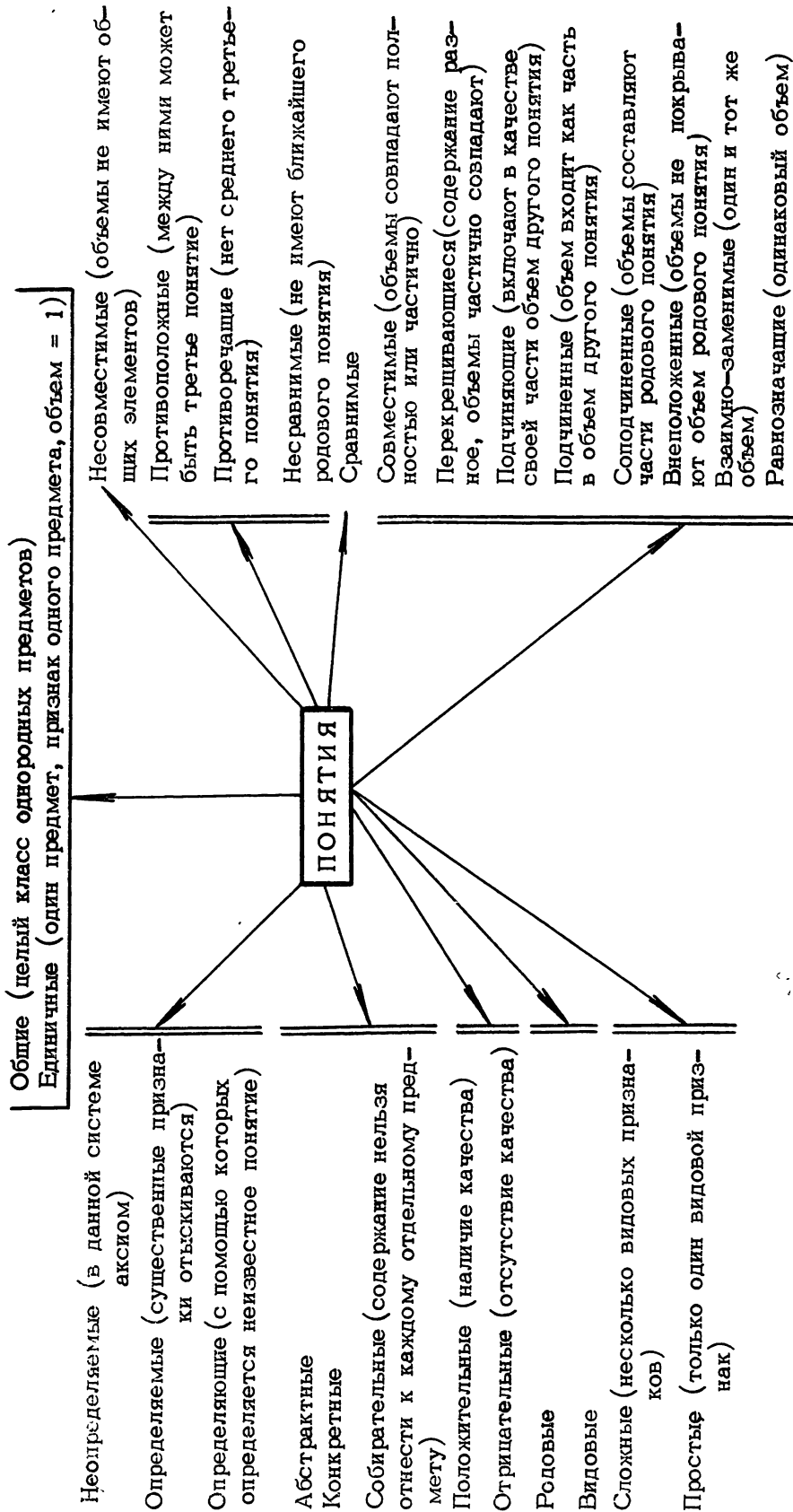


Рис. 1. Основные классы понятий

4. В термине больше мотивированности, чем произвольности<sup>13</sup>.
5. Термин не связан с контекстом в пределах терминосистемы<sup>14</sup>.
6. Термин в сильной степени номинативен, поэтому ему свойственна стилистическая нейтральность, отсутствие экспрессии<sup>15</sup>.
7. Термин одновременно член терминосистемы и лексико-семантической системы национального языка<sup>16</sup>.
8. Термины не вечны, они в меньшей степени непрерывны во времени и в большей — изменчивы, чем слова обыденного языка<sup>17</sup>.
9. «Термин — наиболее информативная единица естественного языка» [31, с. 19].
10. Термин, как правило, краток по своей форме.

### Словари, тезаурусы, стандартизация терминологии

Рассмотренные существенные признаки отношений систем понятий и терминов привлекают все большее внимание исследователей и находят широкое применение и использование в теории и практике информатики и научных коммуникаций [9, 20]. Прежде всего следует отметить фиксацию этих отношений в виде словарей — лексикографию, основы общей теории которой заложил академик Л. В. Щерба [60]. Поскольку многообразие отношений систем понятий и терминов велико, то и число типов словарей неограниченно и устанавливается оно чисто эмпирически» [60, с. 106]. С точки зрения фиксации терминосистем и соответст-

<sup>13</sup> Это свойство вытекает из предыдущего, из системности и относительного изоморфизма понятий и терминов. Вот как об этом говорил Соссюр: «...все, относящееся к языку, как к системе, требует рассмотрения... с точки зрения ограничения произвольности языкового знака». «Не существует языка, где нет ничего мотивированного; но немислимо себе представить и такой язык, где мотивированно было бы все. Между этими двумя крайними точками — наименьшей организованностью и наименьшей произвольностью — можно найти все промежуточные случаи. Во всех языках имеются двоякого рода элементы — целиком произвольные и относительно мотивированные..» [37, с. 165].

<sup>14</sup> «Смысловое содержание термина обусловлено тем понятием, которое данный термин должен выражать; его значение... не может зависеть от того предложения, в котором он употреблен, а должно определяться лишь всей системой понятий и, соответственно, терминологией данной области знания» [50, с. 74] «Термины могут жить и вне контекста, если известно, членами какой терминологии они являются. Однозначность термины получают не через условия контекста, а через принадлежность к данной терминологии» [51, с. 51].

<sup>15</sup> «...особая функция, в которой выступает слово в качестве термина, это функция названия» [52, с. 5, 6].

<sup>16</sup> «Под ближайшим значением слова Потебня разумел значение, которое обычно легко понимается всеми носителями данного языка и фиксируется в толковых словарях. Дальнейшее значение слово приобретает чаще всего в специальных профессиональных или научных областях, с которыми оно может соотноситься» [53, с. 23 см 54, с. 8]

<sup>17</sup> Это легко понять, если вспомнить слова Соссюра: «Язык коренным образом не способен сопротивляться факторам, постоянно меняющим отношения между означаемым и означающим. Это одно из следствий, вытекающих из принципа произвольности знака». «...так как язык существует одновременно и в обществе и во времени, то никто ничего не может в нем изменить; между тем произвольность его знаков теоретически обеспечивает свободу устанавливать любые отношения между звуковым материалом и понятиями. Из этого следует, что оба элемента, объединенные в знаке, живут в небыстрой степени обособленно и что язык изменяется, или, вернее, эволюционирует под воздействием всех сил, которые могут повлиять либо на звуки, либо на смысл. Эта эволюция является неизбежной: нет языка, который был бы от нее свободен» [37, с. 108, 109]. Для языка науки это верно в еще большей степени, особенно в эпоху научно-технической революции. Энгельс писал: «...в науке каждая новая точка зрения влечет за собой революцию в ее технических терминах» [55, с. 31]. Что касается истории развития языка русской науки, то она хорошо изложена в целом ряде работ (см., например, [56—59]).

вующих им систем понятий необходимо указать следующие типы словарей: академический, двуязычный, идеологический, идиоматический, иностранных слов, многоязычный, нормативный, одноязычный, отраслевой, переводной, политехнический, синонимов, словарь-справочник, терминологический, толковый, частотный, энциклопедический и их различные комбинации. Поиск, составление наиболее полного списка и анализ словарей, относящихся к определенной области науки и техники, является одним из первых этапов исследования терминосистемы [61]. Фиксация терминосистемы в виде терминологического словаря, отражающего последние достижения лексикографии, становится все более актуальной и необходимой задачей<sup>18</sup>.

Наиболее перспективным типом словаря для целей автоматизированной обработки информации является информационно-поисковый тезаурус [9, 62—68]. В настоящее время подробно разработана методика построения информационно-поисковых тезаурусов [69—73], а наиболее развитые автоматизированные системы периодически публикуют их новые редакции (см., например, [74—76]). Использование тезаурусов в ИПС основано на идее координатного индексирования документов и в поисковых предписаниях [9, 20], высказанной и разработанной в начале 50-х годов. Основной частью тезауруса является дескрипторный словарь, служащий целям устранения синонимии, полисемии и омонимии терминов определенной тематической области, а также эксплицитного выявления основных семантических отношений между ними (в основном иерархических и ассоциативных).

Тезаурус выступает, таким образом, в качестве инструмента контроля функционирования терминосистемы для целей документального информационного поиска [77], т. е. является одним из звеньев в цепи усилий по стандартизации использования терминов в научных текстах. Проблема стандартизации терминологии науки и техники существует по самой природе функционирования каждого из языков мира, в то же время она имеет и межязыковый характер в силу интернациональности самой науки [78]. В Советском Союзе проблема стандартизации решается на основе общегосударственного подхода, ею занимается ВНИИКИ Госстандарта СССР, Комитет научно-технической терминологии АН СССР, отраслевые НИИ, вузы, научно-технические общества [79]. Здесь следует отметить основополагающие работы Д. С. Лотте [50, 80, 31], на базе которых были разработаны руководства и методические пособия по упорядочению научно-технической терминологии [82—84], уже изданные сборники стандартизованных и рекомендуемых терминов [85], сборники новых терминов [86] и ряд других работ (см., например, [87]). Определенная терминологическая работа проводится и на Западе [88]. Упорядочение и стандартизация научно-технической терминологии требует глубоко научного подхода и несовместима с каким-либо субъективным и односторонним регулированием, даже сама возможность

<sup>18</sup> «В ближайшем будущем можно ожидать появление терминологических словарей, совмещающих в содержании и структуре словарной статьи привычные, традиционные сведения (т. е. прежде всего вид термина и его логическое определение и толкование) со сведениями, которые уже сейчас подсказываются намечившимся многоаспектным изучением терминов (их языковая и логико-понятийная структура, их взаимные логико-семантические связи в системе языка и речи и т. д.), и с частотными показателями последних. В таких словарях читатель должен найти термин (однословный, двухсловный, многословный), сведения о логико-семантических отношениях его частей (если он неоднословный), сведения о его логико-семантической соотнесенности с другими терминами в языке и тексте (синонимия, антонимия, омонимия, родовидовые отношения, отношения различных типов ассоциативной логической сопряженности), о его словообразовательных связях, о структурном месте в пределах терминосистемы и систем, определение значения или значений, указание на отнесенность термина к одной или нескольким областям знания или производства, практической деятельности, рекомендации о целесообразном применении или неприменении термина» [30].

решения проблемы до сих пор является объектом научной дискуссии<sup>19</sup>.

### Закон Ципфа

При исследовании терминосистем и их функционирования весьма полезным и продуктивным является использование количественных методов анализа, в том числе статистических. Исходным материалом исследований служат тексты документов. Терминосистема может рассматриваться как результат статистической обработки суммарного текста достаточно большого числа определенным образом выбранных документов и быть представлена в виде частотного словаря, где каждому из ее членов сопоставлена частота употребления во всем тексте. Между частотой употребления термина (или слова) и его номером (рангом) в частотном словаре имеет место зависимость (в ряде работ называемая законом) Эсту — Ципфа — Мандельброта [92], названная так в честь ученых, впервые обнаруживших ее и внесших определяющий вклад в ее формулировку и изучение [93—95]:

$$F_i = kN (i + \rho)^{-\nu}, \quad (3)$$

где  $F_i$  — частота употребления термина (слова),  $i$  — его ранг (номер в частотном словаре),  $N$  — длина текста,  $k$  — коэффициент относительной частоты наиболее частого термина (слова),  $\rho$  — поправочный коэффициент наиболее частых слов,  $\nu$  — показатель степени (параметр, характеризующий лексическое богатство текста).

Приведенная зависимость носит универсальный характер, она имеет место и для статистики других элементов научных публикаций. Большой вклад в изучение и интерпретацию этой зависимости, или так называемого рангового распределения, внесли работы советских ученых (см., например, [96—101]).

### Значимость термина, терминозначимость документа

Частотный подход к терминам положен в основу составления частотных словарей языка науки и техники [102, 103], всестороннему лингвистическому исследованию которого посвящены многочисленные работы советских и зарубежных авторов [104—110]. Формирование и функционирование терминов существенно связано с их воспроизводимостью в текстах, т. е. с частотой повторения. Таким образом, статистика терминов на множестве научных текстов документов является важной характеристикой как самих текстов, так и документов<sup>20</sup>. Современные ИПС работают с текстами документов на уровне терминов, при этом терминологический образ документа определяется списком значимых терминов из его текста, а если говорить точнее, применительно к существующему положению вещей, то из заголовка документа и/или его аннотации (реферата) [111—113]. Понятия значимости термина и значимо-

<sup>19</sup> См., например, работы [8, 89—91]. Здесь полезно напомнить слова академика Щербы Л. В. из уже упомянутой нами работы: «хороший нормативный словарь не придумывает нормы, а описывает ту, которая существует в языке, и уж ни в коем случае не должен ломать эту последнюю. Может быть норму трудно иногда подметить: но это уже несчастье исследователя и не имеет никакого отношения к принципиальной стороне дела». [Развитие языка] «происходит помимо всяких нормативных словарей: однако эти последние могут помогать естественному ходу вещей, а могут и мешать ему, направляя развитие языка по ложным путям» [60, с. 97, 98].

<sup>20</sup> Здесь полезно привести точку зрения автора работы [89, с. 176]: «термин — любой фрагмент текста..., выражающий (обозначающий, называющий) специальное научно-техническое понятие и не содержащий одновременно субъекта и предиката высказывания».

сти документа связаны между собой. Сущность значимости языкового знака глубоко вскрыта еще Соссюром<sup>21</sup>, который считал это понятие универсальным и за пределами языка называл его «ценностью». Значимость, таким образом, является фундаментальным системным фактором. Солнцев считает [114, с. 15], что «соотносительная значимость и есть наиболее общее и обязательное свойство всякой внутрисистемной связи». Поскольку полная система текстов документов формирует терминосистему, то, следовательно, и значимости терминов. Многочисленные статистические исследования показывают, что воспроизводимость (частотные характеристики) термина в текстах связана с его значимостью в терминосистеме, формируемой этими текстами. Если рассмотреть соотношение «система документов — система терминов», то понятно, что значимости производной системы, т. е. системы терминов, определяют в какой-то мере значимости элементов системы документов. В настоящее время при разработке автоматизированных ИПС большое значение придается количественным методам определения значимостей терминов (см., например, [115, 116]), которые в той или иной форме выражаются через частотные характеристики встречаемости терминов в поисковых образах документов. Под терминозначимостью (терминальной значимостью) документа (ТЗД) мы будем понимать сумму частот употребления в полном наборе документов терминов, входящих в терминообраз документа  $i$ :

$$\text{ТЗД}_i = \sum_j f_{T_j}, \quad (4)$$

$f_{T_j}$  — частота встречаемости термина  $T_j$ , входящего в терминообраз документа  $i$ , в терминообразах всех документов,  $\sum_j$  — сумма  $f_{T_j}$  для терминов, входящих в терминообраз документа  $i$ .

#### 4. ФАМИЛИИ АВТОРОВ

Фамилии авторов как элемент библиографического описания научных публикаций имеет исключительно важное значение [117]. Не случайно развитие библиографического описания в XIX столетии было обусловлено, прежде всего, необходимостью совершенствования авторского алфавитного каталога в библиотеках [118]. По оценкам Прайса [119] 90% ученых всех времен живут и здравствуют в настоящее время. Эта цифра близка к 100%, если иметь в виду ученых, работающих в рамках научного направления, возникшего и развивающегося на наших глазах. Каждый из этих ученых, возможно, имеет свою точку зрения на авторский аспект публикаций научного направления, в котором он работает, и это является источником многочисленных субъективных и нередко противоречивых мнений и оценок. Наша же цель в данном разделе — рассмотреть количественные модели и характеристики, пригодные для автоматизированной обработки документов в современных ИПС

#### Динамика авторского коллектива

Интересной моделью для исследования популяции авторов в дан-

<sup>21</sup> «Входя в состав системы, слово облечено не только значением, но еще главным образом значимостью» [37, с. 148]. «Их [значимостей] наиболее точная характеристика сводится к следующему: быть тем, чем не являются другие» [37, с. 149]. «И понятие и звуковой материал, заключенные в знаке, имеют меньше значения, нежели то, что есть вокруг него в других знаках. Доказывается это тем, что значимость члена системы может изменяться без изменения как его смысла, так и его звуков исключительно вследствие того обстоятельства, что какой-либо другой, смежный член системы претерпел изменение» [37, с. 153].



ном научном направлении является эпидемический процесс. При эпидемическом процессе  $N$  индивидуумов распределены по трем состояниям: восприимчивое ( $S$  индивидуумов), инфекционное ( $I$  индивидуумов), удаленное ( $R$  индивидуумов). Очевидно,  $N = S + I + R$ . В инфекционном состоянии находятся носители инфекции, в восприимчивом — те, которые могут воспринять инфекцию и стать ее носителями, в удаленном — те, которые выбыли из эпидемии, например, переболели или умерли. Гоффман [120] проводит следующую аналогию с авторами научных статей: те авторы, которые публикуют статьи в настоящее время, относятся к состоянию  $I$ , те авторы, у которых в настоящее время нет работ, но появились впоследствии, относятся к состоянию  $S$  и, наконец, те авторы, у которых к настоящему времени уже прошел год с момента опубликования последней работы, относятся к состоянию  $R$ .

Если скорость изменения  $I$  не равна нулю, и изменение этой скорости положительно, то процесс находится в эпидемической стадии, описываемой следующими дифференциальными уравнениями:

$$\begin{aligned} dS &= -\beta SI - \delta S + \mu, \\ \frac{dI}{dt} &= \beta SI - \gamma I + \nu, \\ \frac{dR}{dt} &= \delta S + \gamma I, \end{aligned} \quad (5)$$

где  $\beta$  — скорость распространения инфекции,  $\delta$  — скорость, с которой исчезают восприимчивые члены,  $\mu$  — скорость, с которой новые восприимчивые члены входят в популяцию,  $\gamma$  — скорость, с которой исчезают инфекционные члены,  $\nu$  — скорость, с которой новые инфекционные члены входят в популяцию.

Для эпидемической стадии

$$\frac{dI}{dt} > 0 \text{ или } S > \frac{\gamma - \nu/I}{\beta} = \rho, \quad (6)$$

где  $\rho$  — пороговая плотность восприимчивых членов, выше которой начинается процесс развития эпидемии. В начальный момент  $t = t_0$ ,  $I = I_0$ , если далее  $I$  растет, то  $\nu/I$  падает до нуля, и  $\rho \rightarrow \gamma/\beta$ .

Процесс развития эпидемии достигает пика, когда  $\frac{d^2I}{dt^2} = 0$ , при этом  $\frac{dR}{dt} = \kappa$  (const) в области стабильного участка. Аналитическое решение системы дифференциальных уравнений не всегда возможно. В частном случае, если предположить, что  $I = f(t)$  — экспоненциальная функция, то, разлагая ее в ряд, получим

$$I = C_0 + C_1 t + C_2 t^2 + \dots + C_n t^n + \dots \quad (7)$$

Эпидемиологический подход позволяет моделировать «взрывные» ситуации, когда, например, число авторов по проблеме растет в виде всплеска.

### Коллективизм научного творчества

Имеются неоспоримые доказательства увеличения соавторства в различных областях науки и техники в течение этого столетия [121]: в физике доля работ с одним автором в 20-х годах была 75%, а в 50-х

годах — 39%; в физиологии — 84% и 55% соответственно; в социологии — 98% и 72%. Общая картина, по-видимому, такова: уменьшение доли работ с одним автором приводит к насыщению за счет возрастания доли работ с двумя и тремя авторами, затем соответственно уменьшается доля работ с двумя авторами и возрастает доля работ с тремя и четырьмя авторами и т. д. Среднее количество авторов на статью в данной области свидетельствует о степени «точности» научного направления, его сложности и насыщенности современной техникой, позволяет сравнить научное направление с другими, в целом с научной областью, является одним из параметров, характеризующих так называемые «точки роста» в науке. Как правило, этот параметр выше для нового научного направления, чем для всей научной области в целом [121].

### Статистика продуктивности авторов

Желание публиковать результаты как можно скорее сопровождается у многих ученых желанием публиковаться чаще [121]. По широко распространенному мнению среди ученых (особенно молодых), чем больше статей опубликовано, тем лучше для репутации ученого. Использование количества публикаций в качестве меры научной значимости того или иного лица становится повсеместно заманчивым и привлекательным. Уже накоплен довольно обширный материал [119], свидетельствующий в пользу довольно строгого соотношения между элитностью ученого и его продуктивностью как автора статей. Авторская продуктивность на протяжении жизни ограничена в пределах от 1 до 1000 работ<sup>22</sup>. Количество же авторов, опубликовавших  $n$  работ, пропорционально  $1/n^2$ . Впервые эту довольно строгую закономерность открыл и исследовал Лотка [122]. Рассмотрим более подробно таблицу, составленную на базе 100 авторов, имеющих одну публикацию, и в которой observa-

Т а б л и ц а 1

Продуктивность статья/автор,	Число людей	Число статей	
1	100	100	(75% авторов производят 25% статей)
2	25	50	
3	11,1	33,3	
4	6,2	25	
5	4	20	
6	2,8	16,7	
7	2	14,2	
8	1,5	12,5	
9	1,2	11,1	
10	1	10	
10—11,1	1	10	(Полуитог: 10 авторов производят более 50% статей)
11,1—12,5	1	11,1	
12,5—14,2	1	12,5	
14,2—16,7	1	14,2	
16,7—20	1	16,7	
20—25	1	20	
25—33,3	1	25	
33,3—50	1	33,3	
50—100	1	50	
Более 100	1	100	
Итого:	165	586	

<sup>22</sup> Прайс [119] свидетельствует, что ему не удалось найти человека, опубликовавшего более английского математика Кэли, жившего в XIX столетии, — автора 995 работ. По данным другой работы [121] мировой рекорд продуктивности в течение жизни был установлен Theodore Cockerell (3904 работы за 67 лет) — не очень хорошо известным специалистом в области энтомологии пчел.

ется условие симметрии относительно точки, соответствующей квадратному корню из общего числа авторов или из числа статей самого продуктивного автора [119, с. 322].

Таблица наглядно иллюстрирует принадлежность «закона» Лотка к ранговым распределениям, уже рассмотренным в предыдущем параграфе. Действительно, на основании табл. 1 можно составить список всех 165 авторов, расположив их в порядке убывания продуктивности: первый номер будет иметь автор, написавший 100 работ, номер (или ранг) последнего будет равен 165. Каждому автору ранга  $r$  можно сопоставить частоту встречаемости  $F$  его фамилии среди фамилий других

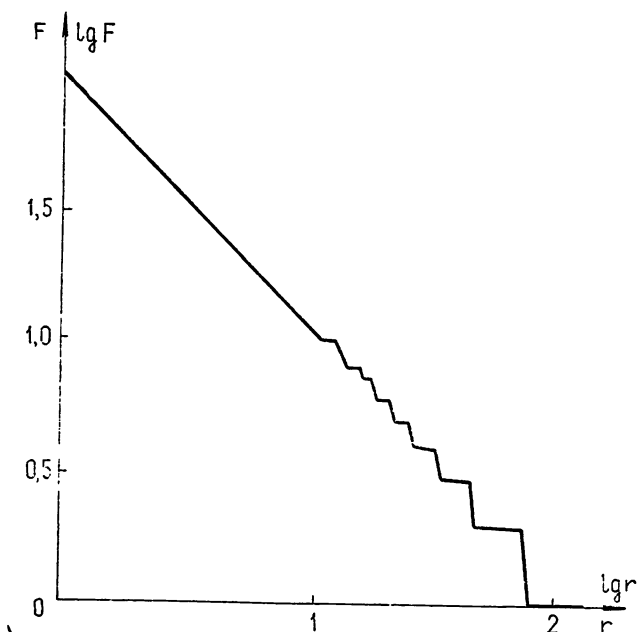


Рис. 2. График функции  $F(r)$  в логарифмическом масштабе (см. табл. 1)

авторов в описании всех 586 документов. График функции  $F(r)$ , выполненный в логарифмическом масштабе, приведен на рис. 2. Нетрудно видеть, что это знакомая уже нам гиперболическая лестница (закон Ципфа):

$$F = \frac{F(1)}{r}, \quad F(1) = 100. \quad (8)$$

#### Авторская значимость документа

Давно подмечено, что, работая с научно-технической литературой, ученым и специалистам далеко не безразлична авторская принадлежность документа, нередко она является решающим критерием стратегии поиска нужных работ. Как правило, отдается предпочтение членам так называемого интернационального «незримого коллектива», идущего в авангарде исследования проблемы. А поскольку научная компетентность или элитность ученого статистически строго связана с числом его работ, то легко понять, что работы наиболее продуктивных авторов чаще оказываются наиболее значимыми для данного научного направления. Учитывая это обстоятельство, можно попытаться определить меру авторской значимости отдельного документа как частоту встречаемости фамилии его автора среди фамилий авторов полного собрания докумен-

тов данного научного направления. Поскольку, как уже отмечалось, авторами отдельной работы нередко является одновременно несколько лиц, то мерой ее авторской значимости будет сумма частот встречаемости фамилий авторов работы. Эта мера достаточно проста и может быть использована в качестве критерия поиска в современных автоматизированных ИПС. В приведенном примере Прайса сумма авторских значимостей всех работ равна  $100^2 + 50^2 + 33,3^2 + 25^2 + 20^2 + 16,7^2 + 14,2^2 + 12,5^2 + 11,1^2 + 10^2 + 10^2 + 9^2 \cdot 1,2 + 8^2 \cdot 1,5 + 7^2 \cdot 2 + 6^2 \cdot 2,8 + 5^2 \cdot 4 + 4^2 \cdot 6,2 + 3^2 \cdot 11,1 + 2^2 \cdot 25 + 1^2 \cdot 100 = 16485$ . Значимость же работ первых 10 авторов, опубликовавших 293 работы (половину общего числа) составляет 15494, относительная (т. е. по отношению ко всем работам) значимость этих работ равна  $15494/16485 \approx 0,94$  или 94%.

В реальных распределениях относительная значимость отдельной работы равна сумме продуктивностей («авторств») ее авторов в данном научном направлении, деленной на сумму абсолютных авторских значимостей всех работ. Распределение авторских значимостей в зависимости от ранга для работ вышеприведенного примера Прайса приве-

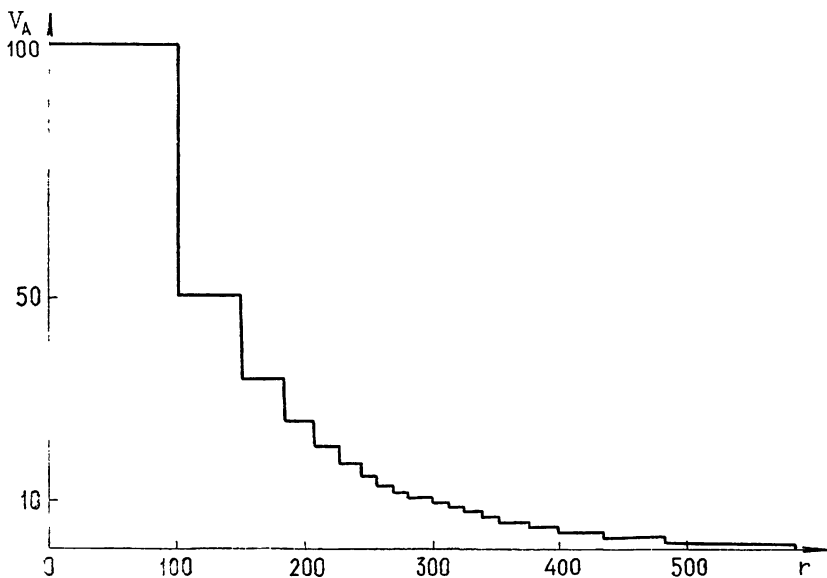


Рис. 3. Ранговое распределение авторских значимостей

дено на рис. 3. Как легко видно, «лестница» довольно хорошо описывается формулой

$$V_A = V_m, \quad 1 \leq r \leq r_m, \quad V_A = \frac{V_m}{(r/r_m)^2}, \quad r > r_m, \quad (9)$$

где  $V_A$  — авторская значимость работы,  $V_m$  — максимальная авторская значимость (для первых 100 работ самого продуктивного автора),  $r_m$  — ранг последней работы самого продуктивного автора. В автоматизированных ИПС, таким образом, в качестве одного из критериев выдачи документов можно использовать заданный уровень авторской значимости работы или совокупности работ.

## 5. ОРГАНИЗАЦИИ

Название организации в настоящее время не используется в стандартном библиографическом описании. Между тем этот элемент легко выделяется из текста публикации (речь идет о периодической литературе), поскольку он проводится либо после фамилий авторов в начале работы, либо в конце текста самой статьи. Наличие в поисковом образе

документа названия организации, от имени которой действовали авторы, направившие рукопись в печать, существенно улучшает возможности научной коммуникации при использовании вторичных источников и автоматизированных систем обработки научно-технической информации. Название организации как элемент поискового образа содержит четкий ответ на такие важные вопросы, как: где работают данные авторы, где выполнено данное исследование или получен тот или иной результат. В нем содержится, как правило, и адресная информация, необходимая при переписке ученых.

В последнее время названию организации как элементу описания стало уделяться большее внимание. Так, один из разделов широко известного Индекса научных ссылок называется «Corporate Index» и содержит сведения об организациях, где выполнены работы, содержащиеся в указателе. В информатике важность этого элемента была переосознана в связи с многочисленными исследованиями «незримых коллективов» — интернациональных отрядов ученых и специалистов, идущих во главе исследования тех или иных проблем науки и техники. Как правило, члены «незримых коллективов» работают в организациях с высокой научной репутацией, где сосредоточена современная исследовательская техника, имеются научные кадры самой высокой квалификации. Вполне понятно, что публикация результатов исследований из таких научных центров имеет заметное преимущество как в количественном (по числу работ), так и в качественном (по уровню) отношении перед работами из других научных учреждений.

Статистика организаций по сравнению со статистикой других элементов описания публикаций исследована слабо. Имеются, правда, общие указания о ранговом характере распределения организаций по числу работ в данной предметной области, однако экспериментальных работ очень мало, поскольку они сопряжены с просмотром *de visu* большого числа первоисточников (см., например, [123]).

Подобно тому, как это было сделано в предыдущих разделах в отношении терминов, введем и рассмотрим понятие значимости публикации с точки зрения значимости организации, где выполнено исследование, среди других организаций, публикующих материалы в данной предметной области. Количественной мерой этой (условно назовем «организационной») значимости публикации является общее число работ от организации, опубликовавшей данную работу, в рассматриваемой предметной области. Чем больше число опубликованных организацией работ, тем больше ее значимость среди других организаций в данной предметной области. «Организационная» значимость каждой работы одной и той же организации одинакова и равна общему числу работ данной организации. Она выступает как мера сравнения работ в аспекте организаций, направивших их в печать. «Организационная» значимость публикации, единственной от данной организации, минимальна и равна единице. В случаях, когда не удается выявить организацию, от имени которой действовали авторы, или когда такая вообще отсутствует, «организационная» значимость работы полагается равной единице, т. е. минимальной величине. Следует отметить, что таких случаев не так уж много. Если же речь идет о фундаментальных исследованиях в современном естествознании, то подавляющая масса работ по научной проблеме имеет четкий адрес, т. е. указана организация, где они выполнены.

## 6. НАЗВАНИЕ ИСТОЧНИКА<sup>23</sup>

Документальные источники информации являются той материальной средой, где фиксируются и передаются последующим поколениям все шаги развития науки и техники. Типология источников, однако, не

<sup>23</sup> Под источником понимается печатный, рукописный или любой другой документ, являющийся основой для составления описания [1, с. 98].

очень разнообразна: так, в работе [124] первичные источники научно-технической информации подразделяются на 13 основных групп. Наиболее продуктивным и в то же время ценным типом первоисточника является статья в научно-техническом журнале. Согласно [121] в естественных науках 82% всех ссылок приходится на статьи, 12% на книги и лишь 6% на прочие виды изданий. «В настоящее время журналы как источники научно-технической информации дают до 70% всей научной информации, используемой учеными и специалистами, и в значительной мере определяют общий темп количественного роста научных документов» [125, с. 216]. Журналы существуют уже более трех веков, их общее количество по естественным, точным и прикладным наукам по некоторым данным достигнет к 1980 г. 40 тыс.<sup>24</sup>

С точки зрения библиометрики одной из важнейших характеристик журнала является его продуктивность, т. е. количество статей в определенной научной области или по определенной проблеме за заданный промежуток времени. Пионером исследования продуктивности научных журналов по праву считается Самуэль Бредфорд. В своей работе [128], впоследствии вошедшей в известную книгу [129], он писал: «Если научные журналы расположить в порядке убывания продуктивности статей по заданной теме, то они могут быть разделены на ядро журналов, в основном посвященных этой теме, и несколько групп или зон, содержащих то же число статей, что и ядро, при этом число журналов в ядре и последующих зонах будет удовлетворять соотношению  $1 : n : n^2$ ».

Исследованию этой закономерности, впоследствии названной законом Бредфорда, посвящено большое количество работ зарубежных<sup>25</sup> и отечественных авторов<sup>26</sup>. Среди них следует отметить работы Викери [138], Кендалла [139], Леймулера [140], Брукса [141, 142], Наранана [143], О'Нейла [144, цит. по 130], Хуберта [145, цит. по 130].

Наибольший интерес представляет собой математическая модель Брукса [142, см. 127]:

$$R(n) = \begin{cases} \alpha n^\beta, & 1 \leq n \leq c \\ K \log(n/s), & c < n \leq N \end{cases} \quad (10)$$

где  $\alpha$  — число статей, опубликованных в наиболее продуктивном журнале,  $\beta < 1$  (экспериментальный факт), его значение постоянно, если библиография ограничена относительно небольшим промежутком времени (2—5 лет),  $K = N$  — ранг последнего журнала (в заданной модели журналы, имеющие одинаковую продуктивность, имеют различные ранги),  $s \leq 1$  (служит мерой широты предметной области),  $c \geq 3$  и увеличивается с расширением предметной области,  $n$  — порядковый номер (ранг) журнала в списке по убыванию продуктивности,  $R(n)$  — суммарное число публикаций в данной предметной области, содержащееся в первых  $n$  журналах. На рис. 4 приведена кривая закона Бредфорда, построенная по математической модели Брукса.

Эта модель интерпретирует закон Бредфорда в интегральной форме. В дифференциальной форме (см. [130]) — это уже обсуждавшееся в предыдущих параграфах ранговое распределение Ципфа:

$$F(r) = ar^{-b}, \quad (11)$$

где  $F(r)$  — количество публикаций по данной предметной области в журнале  $r$  ранга,  $a$  и  $b$  — положительные параметры.

Ранговое распределение журналов в пределах предметной области имеет важное прикладное значение для информатики, широко исполь-

<sup>24</sup> Более подробно см. [9, с. 89—99; 125—127]

<sup>25</sup> См. обзор Хуберта [130], содержащий более 100 ссылок.

<sup>26</sup> См., например, работы [131—137].

зучается при изучении каналов передачи и распространения научно-технической информации [132], комплектовании фондов литературой [127, 146], оценке информационной значимости журналов [134].

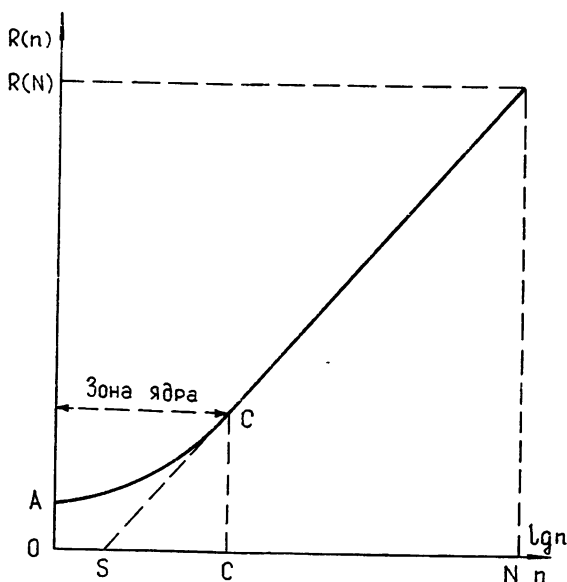


Рис. 4. Кривая закона Бредфорда, построенная по математической модели Брукса

Хорошо известно из практики формальных каналов научных коммуникаций, что подобно «незримым коллективам» ученых в заданной предметной области существует весьма ограниченный список журналов, имеющих высокий международный авторитет и публикующих все наиболее ценное и существенное (это особенно характерно для фундаментальных исследований). Статистически это явление проявляется в ранговом распределении продуктивности журналов. Если оценивать меру значимости научной публикации с точки зрения значимости источника, где она опубликована, то можно ввести понятие «журнальной» значимости публикации и в качестве ее меры взять продуктивность журнала, где она опубликована (см. формулу (11)). При этом «журнальные» значимости работ, опубликованных в ядерных или профильных журналах, будут выше, чем в других журналах, а «журнальная» значимость всех работ из одного и того же журнала будет одинакова и равна продуктивности этого журнала.

## 7. БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ<sup>27</sup>

Разработка научного аппарата сносок и ссылок к концу XIX столетия была обусловлена быстрой эволюцией научной статьи как основного источника фиксации результатов широкого исследовательского фронта естественных наук [147, с. 95]. Путем ссылок каждая опубликованная научная работа связывает себя с сетью всех научных публикаций.

<sup>27</sup> Библиографическая ссылка — краткая запись, содержащая библиографические сведения об источнике, к которому имеет отношение данная часть текста. В зависимости от положения в тексте библиографические ссылки делятся на внутритекстовые, подстрочные (пристраничные) и концевые (т. е. помещаемые в конце статьи, главы или книги). В кратких библиографических ссылках приводят, как правило, только обязательные элементы библиографического описания, в расширенные ссылки допускается включать дополнительно и факультативные элементы [1, с. 87; 3, с. 58].

С помощью цитирования в виде списка ссылок (приводимого, как правило, в конце статьи) публикация связывает себя с предшествующими работами, а с помощью цитируемости, фиксируемой в виде открытого списка кратких библиографических описаний публикаций, сославшихся на данную работу, — с последующими работами. Для каждой публикации, таким образом, цитирование представляет собой множество цитированных в ее тексте работ, а цитируемость — множество публикаций, цитирующих данную работу. Вместе цитирование и цитируемость представляет собой важную характеристику публикации и будет рассматриваться нами как элемент ее описания, который, к сожалению, не входит, да и не может войти в стандартное библиографическое описание, поскольку является открытым, меняющимся списком ссылок (за счет постоянного изменения цитируемости работ).

Механизмы цитирования раскрыты и выявлены не до конца, хотя общепринято утверждение, что в основном цитируются работы, оказавшие непосредственное влияние на появление цитирующей публикации. Однако замечены и иные явления: необоснованное отсутствие в ссылках работ, результаты которых фактически использовались, излишнее самоцитирование, цитирование коллег и научных руководителей. Они, правда, не являются определяющими при наличии достаточно большого статистического материала.

Пионером использования библиографических ссылок для поиска и изучения литературы по праву считается Юджин Гарфилд, предложивший и внедривший в широкую практику указатель библиографических ссылок (Science Citation Index) сначала для естественных наук, а затем и для общественных [148—150]<sup>28</sup>. В настоящее время язык библиографических ссылок стал мощным средством самых различных исследований в теоретической и прикладной информатике (см., например, работы [153—170]). Мы рассмотрим его возможности при исследовании публикаций по определенной научной проблеме.

### Способы записи системы ссылок

Самой реальной ситуацией является та, при которой проблема еще не решена, т. е. поток публикаций — открытая система. Но в каждый заданный момент времени реальная ситуация может быть представлена ее моделью — закрытой системой с фиксированным числом публикаций по проблеме. Такая модель имеет прямое отношение к библиографическим спискам и указателям, составляемым традиционным или автоматизированным путем. Важный момент в такой модели — расположение работ одна за другой на оси времени в порядке дат их публикации<sup>29</sup>. Такое расположение позволяет каждой публикации сопоставить номер натурального ряда чисел, при этом самой ранней работе присваивается номер 1. Это позволяет представить всю систему библиографических ссылок в компактном и удобном для обозрения и анализа виде [171]:

$$1. + 15, 28, 46, 54 \textcircled{58} 76, 142, 315, 418 + \quad (12)$$

Весь материал представляется в виде совокупности записей, расположенных в порядке возрастания номера, заключенного в кружок, — номера публикации.

Слева от знака + до кружка приводятся в порядке возрастания номера цитированных в данной работе документов, а справа (в порядке возрастания от кружка до знака +) — номера цитирующих данную ра-

<sup>28</sup> Более подробно об индексе Гарфилда см., например, [151, 152]

<sup>29</sup> В случае совпадения этих дат можно воспользоваться датами поступления рукописей в редакции журналов, либо их относительным расположением на страницах одного и того же журнала



боту публикаций. Такая запись удобна в различного рода библиографических указателях, поскольку номера работ, заключенные в кружки, легко отыскиваются среди других работ и следуют в порядке натурального ряда чисел.

2. В виде длинной ленты, на которой в центре приводится ось времени, а на ней изображаются квадратами работы, следующие друг за другом в порядке возрастания их номеров. Над и под этими квадратами располагаются квадраты цитирующих данную работу и цитированных в данной работе публикаций в порядке, аналогичном рассмотренному в п. 1 (фрагмент ленты см. на рис. 5).

Для общего обозрения наиболее цитируемых и цитирующих работ диаграмма на ленте может быть сжата по оси времени, при этом столбцы, составленные из квадратов с указанием номеров работ, превратятся в вертикальные отрезки линий, длина которых в выбранном масштабе будет соответствовать общему числу цитирующих и цитированных работ для данной точки оси времени, соответствующей определенной публикации (см. рис. 6).

3. В виде двумерной диаграммы [156, 163] (см. рис. 7а) и б)). На рис. 7 а) по оси абсцисс вправо откладываются номера цитирующих работ, а по оси ординат вниз номера цитированных работ. Рис. 7 б) получается вращением рис. 7 а) против часовой стрелки на  $90^\circ$ . Для построения указанных рисунков, как собственно и всей сети библиографических ссылок, достаточно воспользоваться только цитированием. Действительно, для любой публикации, имеющей номер  $n$ , имеется лишь одна точка на гипотенузе  $AC$ , служащая местом пересечения перпендикуляров к координатным осям в точках  $n$ . Двигаясь от точки  $N$  в сторону оси номеров цитирующих работ, мы получим список цитирования, поскольку номера всех этих работ меньше  $n$ , а двигаясь от точки  $N$  в сторону оси номеров цитированных работ, мы получим список цитируемости, поскольку номера всех этих работ больше  $n$ <sup>30</sup>. Так как одна библиографическая ссылка связывает цитированную и цитирующие работы, то понятно, что число цитированных и число цитирующих случаев одинаково и равно числу библиографических ссылок во всех работах.

4. В виде графа, каждая вершина которого изображает публикацию, а ребро — связь двух публикаций за счет одной библиографической ссылки (рис. 8). С помощью графа можно наглядно представить сеть публикаций и оценить ее густоту в отдельных звеньях, но, как правило, лишь для ограниченного числа работ. Действительно, граф, насчитывающий тысячу и более вершин, очень трудно выполнить на обычном листе бумаги.

### Динамика цитирования и цитируемости

С языком библиографических ссылок связан целый ряд понятий и методов информатики, позволяющих проводить количественные исследования и интерпретировать полученные данные. Рассмотрим динамику цитирования. Для научного направления или предметной области обычно берут для анализа ссылки в работах, опубликованных в течение одного какого-либо года, который является точкой отсчета возраста ссылок. По оси абсцисс откладывается возраст ссылок в годах, а по оси ординат суммарное число ссылок старше  $t$  лет. Очевидно, что при  $t = 0$ ,

<sup>30</sup> Такая модель не всегда может быть адекватна реальной ситуации, когда некоторые номера цитированных работ больше номеров цитирующих их публикаций. Это явление получается из-за цитирования работ, только лишь направленных в печать, и оно особенно свойственно бурно развивающемуся научному направлению. Следует признать, однако, что масштабы явления в целом не позволяют пока отвергнуть принятую в тексте модель.

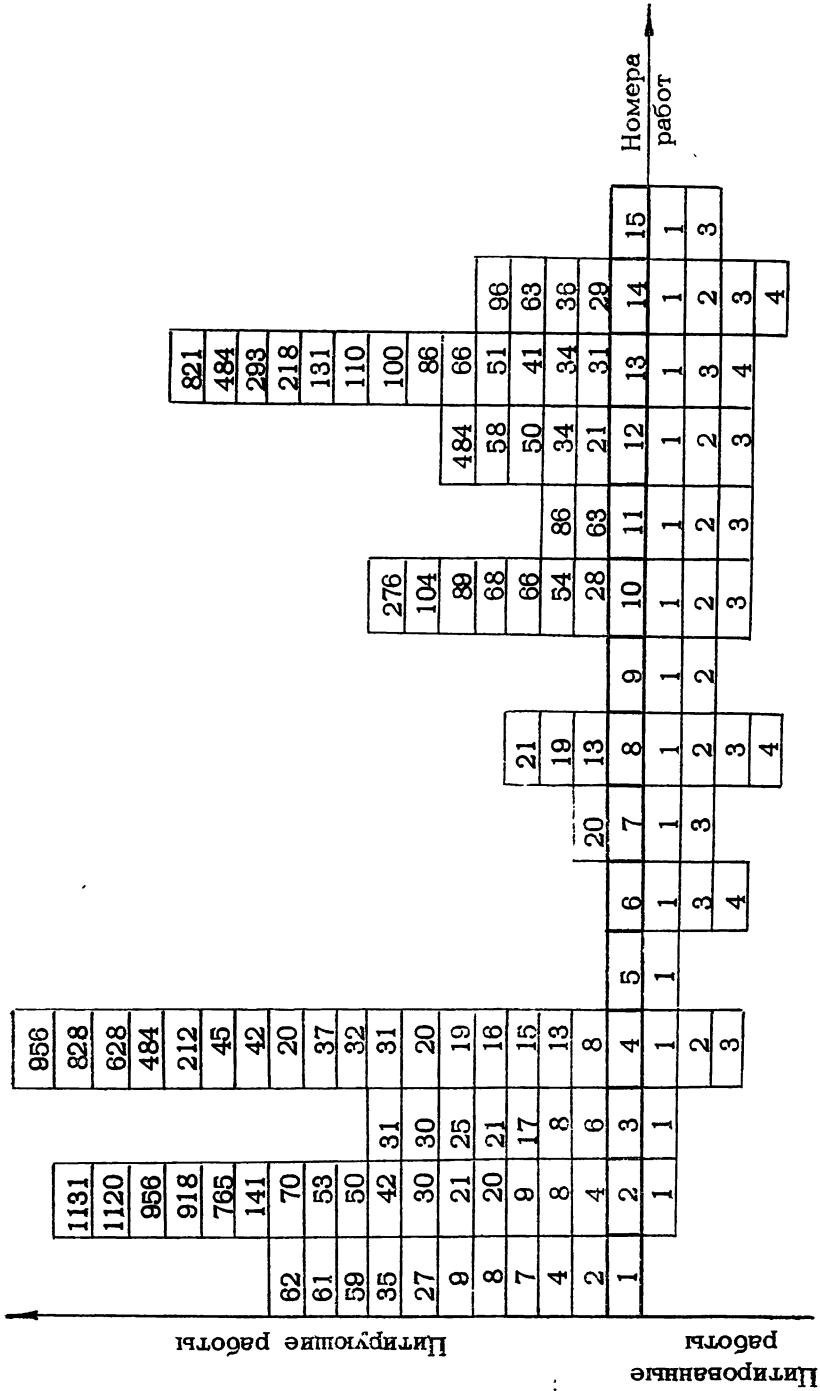


Рис. 5. Фрагмент ленты, иллюстрирующей цитирование и цитируемость работ, расположенных в порядке возрастания их номеров (дат публикации) [171]

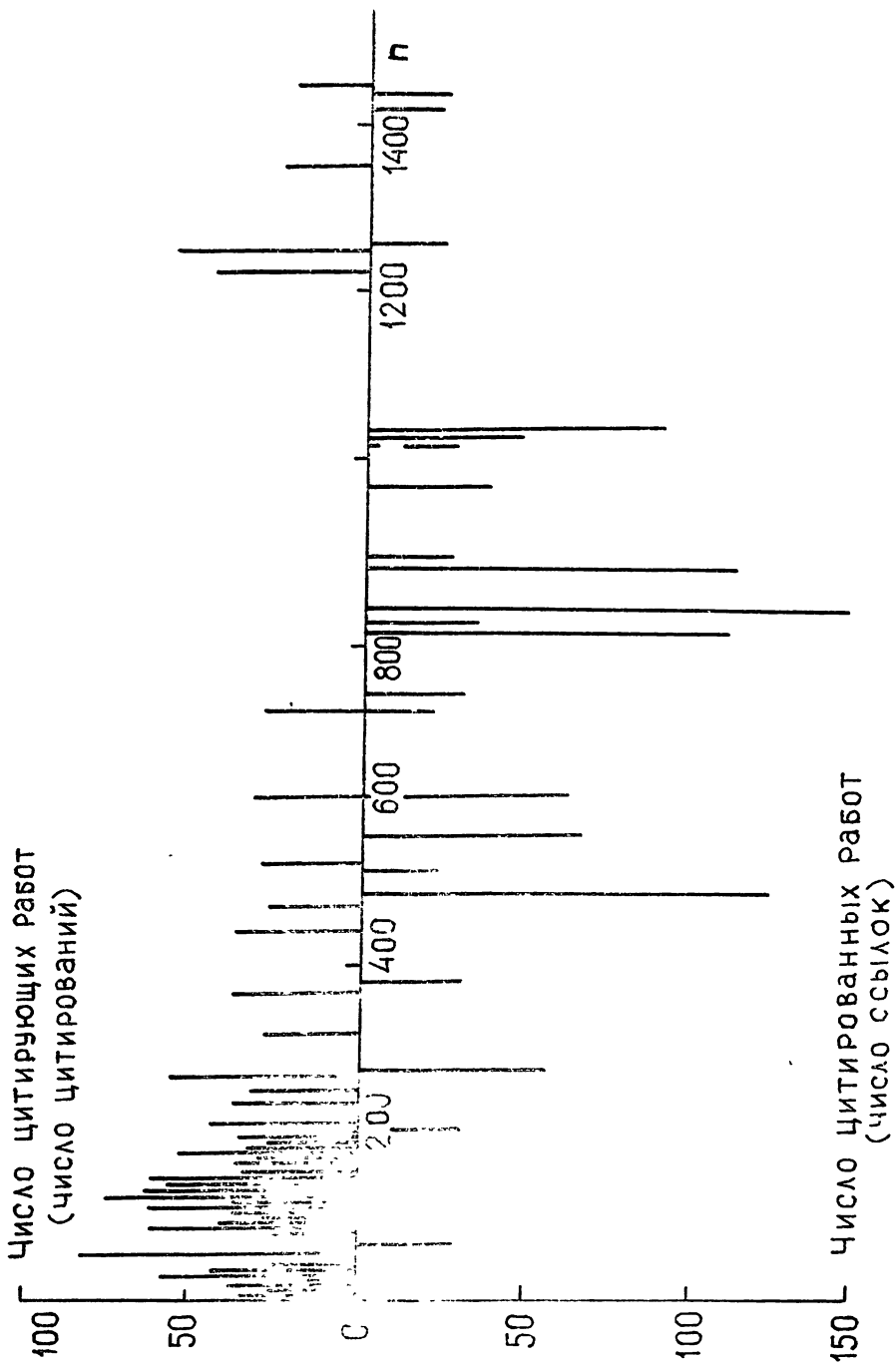
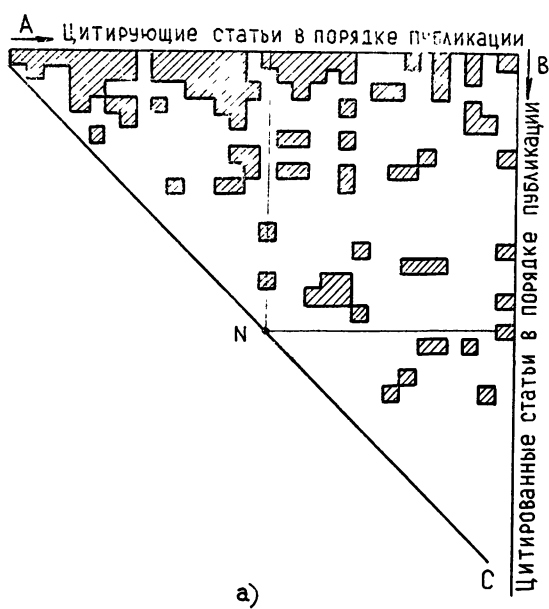
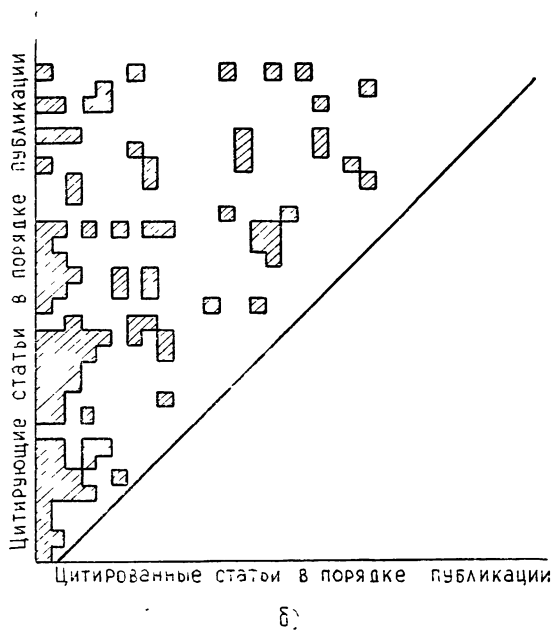


Рис. 6. Графическое представление наиболее цитируемых и цитирующих работ [171]



a)



б)

Рис 7. Двумерная диаграмма цитирования и цитируемости работ научного направления

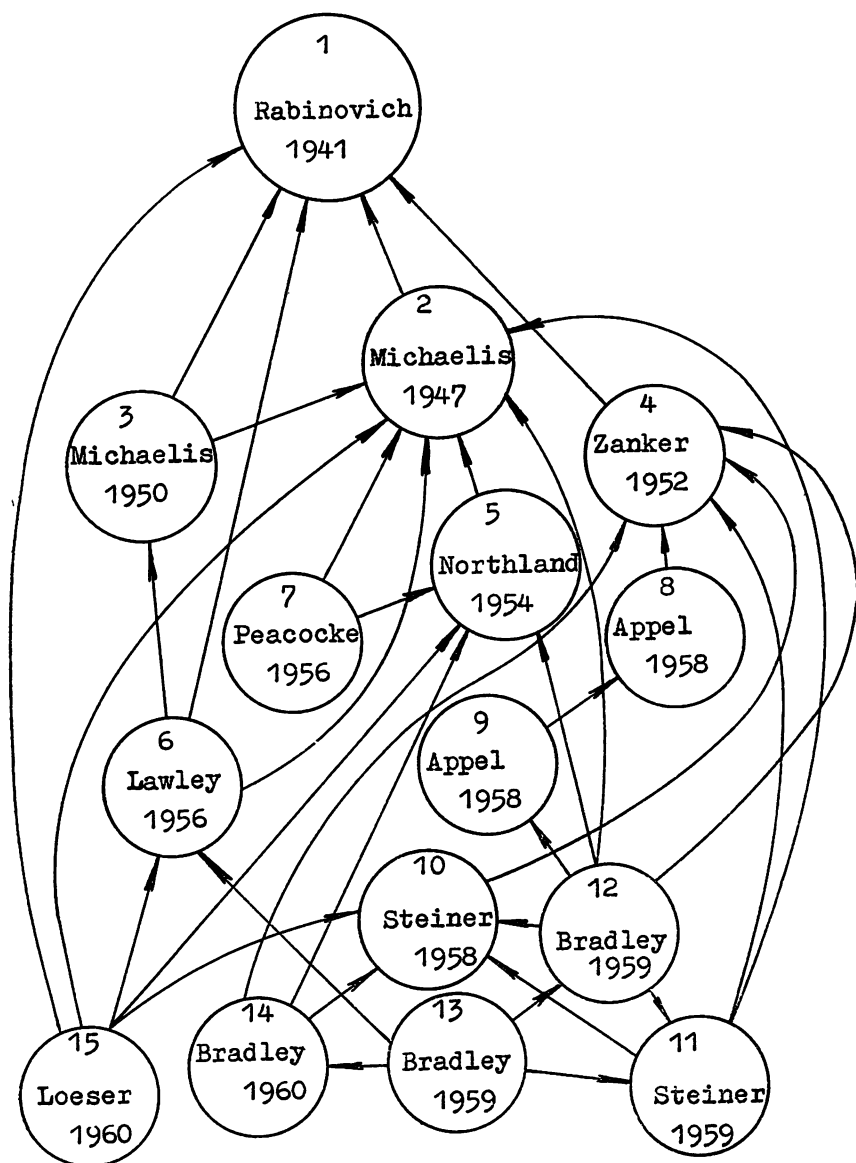


Рис. 8. Граф, составленный американским ученым Г. Алленом для полутора десятков работ по вопросу выявления нуклеиновых кислот, опубликованных в 1941—1960 гг. [9, с. 143]

мы имеем общее количество ссылок для выбранного года, поскольку все они старше  $t = 0$ , а кривая, таким образом, дает нам график изменения суммарного (кумулятивного) числа ссылок (рис. 9а). Многочисленные экспериментальные исследования показали, что кривая эта падает экс-

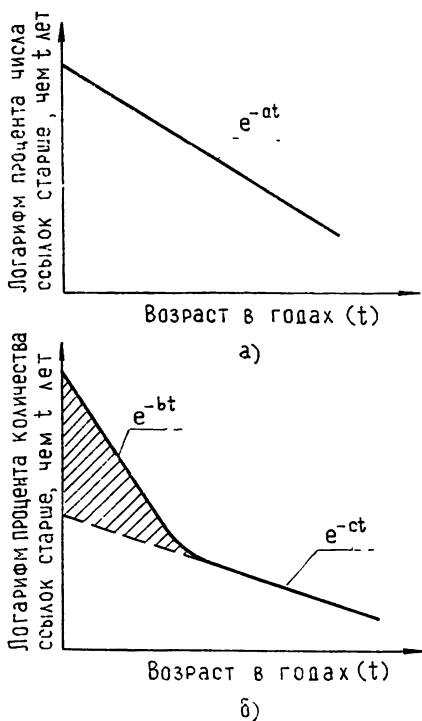


Рис. 9. Кривая цитирования, состоящая  
а) из одной экспоненты,  
б) из двух экспонент.

поненциально, что позволило по аналогии с периодом полураспада для радиоактивного вещества в физике ввести понятие «полужизни» цитированной литературы в данной предметной области, которое соответству-

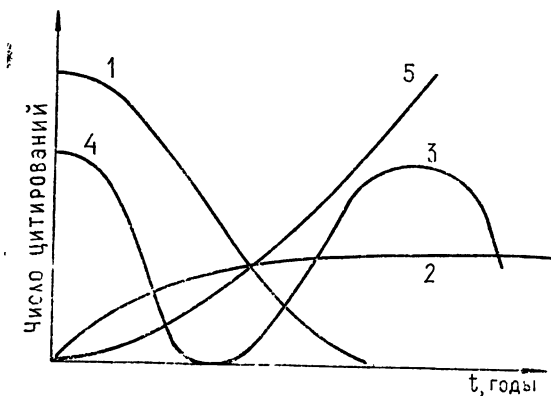


Рис. 10. Изменение во времени числа цитирований научной работы: 1 — обычная работа, 2 — значение работы раскрывается постепенно, 3 — работа прокладывает новые пути в науке и опережает свое время, 4 — работа признана ошибочной, 5 — гениальная работа [160, с. 101, 102].

ет такому временному интервалу, когда кумулятивное число ссылок падает в два раза. Понятие это, конечно, довольно условное, поскольку литература не умирает, она просто не цитируется. Кривая часто состоит из двух экспонент (рис. 9 б), причем первая падает более быстро в силу более частого цитирования работ в пределах фронта исследований по сравнению с более ранними работами (так называемый «эффект непосредственности» — «immediacy effect» [156]).

Динамика цитируемости отдельных работ позволяет классифицировать их по шкале ценности для данного научного направления. Так, например, на рис. 10 проведены кривые цитируемости отдельных работ, которые являются типичными [160].

### Ссылки и связанность публикаций

Механизм цитирования лежит в основе определения и вычисления связанностей публикаций [154, 155, 165, 171]. Рассмотрим эти понятия для публикаций научного направления, номера которых образуют возрастающий ряд натуральных чисел, соответствующих датам опубликования работ. Итак, пусть  $n$  — номер работы ( $n \leq m$ , где  $m$  — общее число публикаций). Далее, пусть  $z$  — множество всех ссылок в  $m$  работах,  $A_n \in z$  — множество публикаций, цитированных в работе  $n$ :

$$A_n = \{a_{n,1}, a_{n,2}, \dots, a_{n,j}, \dots, a_{n,k_n}\},$$

где  $a_{n,j-1} < a_{n,j} < a_{n,j+1}$ ,  $k_n$  — число членов множества  $A_n$ ;  $B_n \in z$  — множество публикаций, цитирующих работу  $n$ :

$$B_n = \{b_{n,1}, b_{n,2}, \dots, b_{n,j}, \dots, b_{n,l_n}\},$$

где  $b_{n,j-1} < b_{n,j} < b_{n,j+1}$ ,  $l_n$  — число членов множества  $B_n$ . Ранее уже было отмечено, что, как правило,  $a_{n,k_n} < n < b_{n,1}$ . Рассмотрим связанность двух работ научного направления, имеющих номера  $n^{(1)}$  и  $n^{(2)}$ , причем,  $n^{(2)} > n^{(1)}$ . Для этого запишем сначала соотношения для этих и цитированных и цитирующих номеров в виде

$$\begin{aligned} a_{n^{(1)},1}^{(1)} < \dots < a_{n^{(1)},k_n}^{(1)} < n^{(1)} < b_{n^{(1)},1}^{(1)} \dots b_{n^{(1)},l_n}^{(1)}, \\ a_{n^{(2)},1}^{(2)} < \dots < a_{n^{(2)},k_n}^{(2)} < n^{(2)} < b_{n^{(2)},1}^{(2)} \dots b_{n^{(2)},l_n}^{(2)}. \end{aligned} \quad (13)$$

Под абсолютной мерой библиографической связи двух документов (bibliographic coupling) мы будем понимать количество элементов в множестве  $A_n^{(1)} \cap A_n^{(2)}$ , а под относительной величиной этой меры (силы) связи отношение

$$\frac{\text{Количество элементов } A_n^{(1)} \cap A_n^{(2)}}{\text{Количество элементов } A_n^{(1)} \cup A_n^{(2)}}. \quad (14)$$

Под абсолютной мерой совместной цитируемости двух документов (co-citation frequency) мы будем понимать количество элементов в множестве  $B_n^{(1)} \cap B_n^{(2)}$ , а под относительной величиной этой меры (силы) связи отношение

$$\frac{\text{Количество элементов } B_n^{(1)} \cap B_n^{(2)}}{\text{Количество элементов } B_n^{(1)} \cup B_n^{(2)}}. \quad (15)$$

Под абсолютной мерой общности цитируемости и цитирования двух документов мы будем понимать количество элементов в множестве  $A_n^{(2)*} \cap B_n^{(1)*}$ , а под относительной величиной этой меры связи отношение

$$\frac{\text{Количество элементов } A_n^{(2)*} \cap B_n^{(1)*}}{\text{Количество элементов } A_n^{(2)*} \cup B_n^{(1)*}}, \quad (16)$$

где

$$B_n^{(1)*} \in B_n^{(1)} \text{ и } b_{n,j}^{(1)*} < n^{(2)},$$

$$A_n^{(2)*} \in A_n^{(2)} \text{ и } a_{n,j}^{(2)*} > n^{(1)}.$$

Библиографическая связанность документов характеризует их общность в использовании одних и тех же работ, связь по совместной цитируемости — общность по их использованию в других работах, связь по цитированию и цитируемости двух документов характеризует преемственную общность двух документов: в ссылках более позднего документа имеются работы, ссылающиеся на более ранний документ.

Общую связанность публикаций научного направления можно характеризовать распределением числа связанных хотя бы одной ссылкой

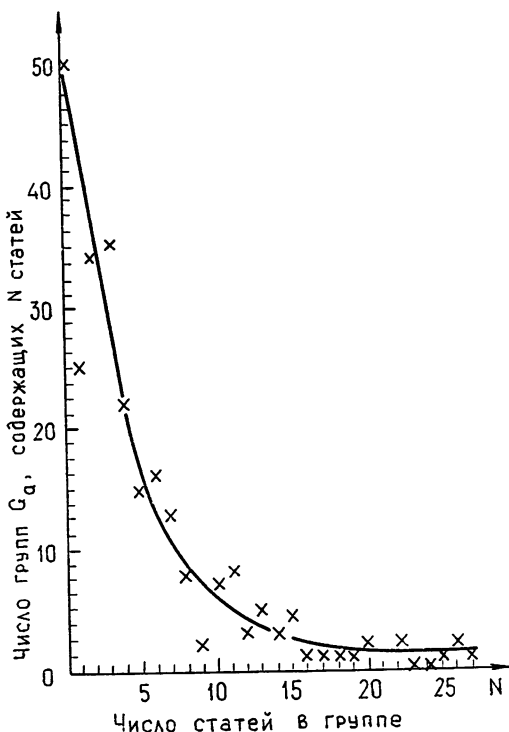


Рис. 11. Кривая связанности 265 статей 97 тома журнала «Physical Review» (по данным [154])

групп, состоящих из  $N$  публикаций (см. рис. 11, полученный по данным работы [154]), а также распределением числа пар документов в зависимости от силы связи (абсолютной меры связанности документов) (см. рис. 12). Связанность данного научного направления (предметной области) с другими научными направлениями (предметными областями) можно количественно характеризовать в виде дроби: в числителе — доля числа цитированных работ, относящихся к данному направлению по отношению ко всем цитированным работам, в процентах, и в знаменателе — доля числа цитирующих работ в данном направлении по отношению ко всем цитирующим работам, в процентах. Вот как выглядят такие дроби для крупных разделов науки [121]: математика — 79/6 (сла-



бо зависит от других областей, но дает огромный вклад во все области), астрономия — 62/93 (себя цитирует не очень сильно, слабо цитируется другими областями), физика — 75/68, химия — 73/30 (заметен вклад

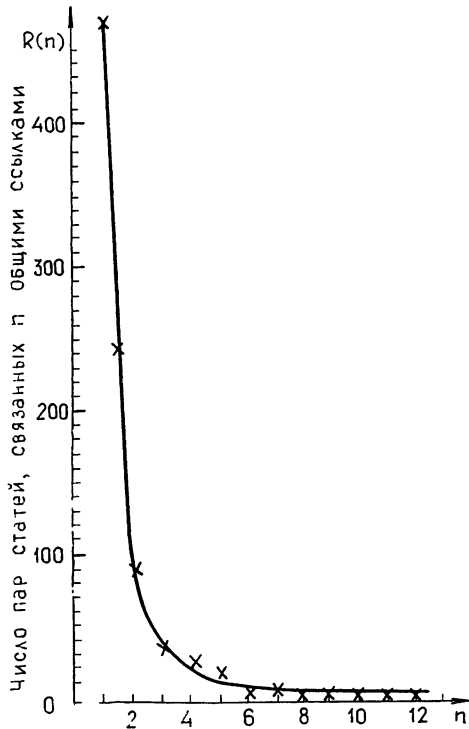


Рис. 12. Кривая силы связи между публикациями (по данным [154])

в другие области), зоология — 38/65, медицина — 61/70, технология — 70/55.

### Статистика ссылок и значимость публикации

Статистика цитирования (цитируемости) может быть представлена в виде кривой, начерченной в декартовых координатах, по оси абсцисс при этом откладывается число ссылок (число цитирований)  $n$ , а по оси ординат — количество работ (или доля этого количества по отношению к общему числу работ), содержащих  $n$  ссылок (число работ, на которые сослались  $n$  раз в других работах). Эта статистика описывается уже неоднократно рассмотренным в предыдущих параграфах возрастным распределением: количество работ быстро убывает с возрастанием  $n$  — в первом случае при больших  $n$  мы имеем дело с обзорами, а во втором случае — с классическими работами. Все более широкое и повсеместное использование языка и указателей библиографических ссылок основано на убеждении, что цитируемость отражает значимость публикации в системе других работ. Действительно, за редкими исключениями обильного цитирования ошибочных работ, чем больше цитируется публикация, тем больше ее влияние на научное направление, тем она важнее, ценнее, значимее. В качестве меры цитируемой значимости публикации мы будем считать количество библиографических ссылок на нее в пределах данного научного направления, в системе публикаций которого и рассматривается значимость этой работы.

1. Основными исторически обусловленными элементами научных публикаций являются: классификационный индекс, термины, фамилии и инициалы авторов, название организаций, название источника, дата, цитирование и цитируемость. Каждый из элементов широко используется в традиционных и/или автоматизированных системах информационного поиска.

2. Большинство элементов в пределах наиболее полной системы публикаций какой-либо предметной области представляет собой систему конкретных реализаций данного элемента. Статистика этих реализаций характеризуется ранговыми распределениями.

3. Значимость элемента является фундаментальным системным фактором и имеет ценностную природу.

4. Значимость публикации в пределах полной системы публикаций предметной области можно характеризовать с помощью набора значимостей ее основных элементов: авторской, журнальной и т. п.

5. Каждая из элементных значимостей публикации определяется как частота встречаемости данной реализации элемента среди других реализаций (основная характеристика рангового распределения).

6. Элементная значимость публикации естественным образом иллюстрирует широко принятое среди ученых, специалистов и информационных работников представление о наиболее активных авторах, организациях, наиболее важных журналах и статьях.

7. Элементная значимость публикации — важное средство, оно может быть широко использовано в автоматизированных ИПС в качестве одного из критериев выдачи, а также в различных научно- и библиометрических исследованиях.

#### ЛИТЕРАТУРА

1. Жданова Г. С., Колобродова Е. С., Полушкин В. А., Черный А. И. Словарь терминов по информатике на русском и английском языках. — М.: Наука, 1971
2. Библиография / Под ред. М. А. Брикмана и А. Д. Эйхенгольца — М.: Книга, 1969.
3. ГОСТ 7. 1—76. Система информационно-библиографической документации Библиографическое описание произведений печати: Взамен ГОСТ 7.1—69. ГОСТ 7.2—69; ГОСТ 7.3—69, ГОСТ 7.8—70 в части п. 9, ГОСТ 7.13—70 — Введ. 01.01.78. — 62 с. Группа Т62 СССР.
4. ISDB (G) General international standard bibliographic description / Prep by the Working group on the General intern. standard bibliogr description set up by the IFLA Comm on cataloguing. — L.: IFLA Intern office for UBC, 1977.
5. Предмашинный формат библиографической записи. Сборник инструкций — М.: ГПНТБ, 1973.
6. Международный коммуникативный формат МСКОФ 2. Содержание и способ представления данных в записи для обмена информацией на магнитной ленте — М.: НТПМ МЦНТИ 19-77, 1977.
7. Хадяров Г. Г. Исследование и разработка методов автоматического поиска в массиве личных имен — Дис соиск. уч. ст. к т. н. М., 1973.
8. Налимов В. В. Вероятностная модель языка. О соотношении естественных и искусственных языков. — М.: Наука, 1974.
9. Михайлов А. И., Черный А. И., Гиляревский Р. С. Основы информатики — М.: Наука, 1968.
10. Шамурин Е. И. Очерки по истории библиотечно-библиографической классификации, т. I — М.: Изд-во Всесоюзной книжной палаты, 1955, т. II — М.: Изд-во Всесоюзной книжной палаты, 1959.
11. Кедров Б. М., Классификация наук, т. I. Энгельс и его предшественники. — М.: ВПШ и АОН при ЦК КПСС, 1961; т. II. От Ленина до наших дней. — М.: Мысль, 1965.
12. Баград М. В. Общая классификация наук и ее связь с библиотечной классификацией — Рига: Звайгзне, 1967.
13. Vickery V. C. Classification and indexing in science. 3rd ed — L.: Butterworth, 1975.
14. Рудельсон К. И. Современные документные классификации. — М.: Наука, 1973.

15. Ранганатан Ш. Р. Классификация двоеточием. Основная классификация — М.: ГПНТБ, 1970
16. Воробьев Г. Г. Документ: информационный анализ.— М.: Наука, 1973
17. Method for priority determination in science and technology.— Paris Unesco, 1978.
18. Miller G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information.— Psychological Rev, 1956, v 63, p. 81—97
19. Woodward A. M. The roles of reviews in information transfer.— J. Am. Soc. Inform Sci., 1977, May, p. 175—180.
20. Черный А. И. Введение в теорию информационного поиска — М.: Наука, 1975.
21. Москович В. А. Информационные языки — М.: Наука, 1971
22. Информационно-поисковый тезаурус по информатике — М.: ВИНТИ, 1974.
23. Карасев С. А. Исследование некоторых алгоритмических методов построения классификаций.— Авт. дис. соиск. уч. ст. к. т. н.— М.: ВИНТИ, 1970
24. Маршак И. В. Формальные классификации динамического документального массива.— Авт. дис. соиск. уч. ст. к. т. н.— М.: ВИНТИ, 1974.
25. Науман П. Методы автоматизации индексирования — Авт. дис. соиск. уч. ст. к. т. н.— М.: Моск. энерг. ин-т, 1974
26. Тафипольская М. Я. Исследование и разработка принципов и методов автоматизации предкоординатного индексирования.— Авт. дис. соиск. уч. ст. к. т. н.— М.: ВИНТИ, 1976.
27. Pratt A. D. A measure of class concentration in bibliometrics — J. Am. Soc. Inform. Sci., 1977, September, p. 285—292.
28. Хаютин А. Д. Термин. терминология. номенклатура.— Самарканд, 1972.
29. Большая Советская Энциклопедия — 3-е изд.— т. 25 — М.: Советская энциклопедия, 1976.
30. Головин Б. Н. О некоторых проблемах изучения терминов — Вестн. Моск. ун-та. Филология, 1972, № 5, с. 51—59.
31. Квитко И. С. Термин в научном документе — Львов: Вища школа, 1976
32. Кондаков Н. И. Логический словарь-справочник — 2-е изд.— М.: Наука, 1975.
33. Ахманова О. С. Словарь лингвистических терминов — М.: Советская энциклопедия, 1966.
34. Якушин Б. В. Алгоритмическое индексирование в информационных системах. Проблематика и методы.— М.: Наука, 1978.
35. Стяжкин П. П. Формирование математической логики — М.: Наука, 1967.
36. Frege G. Uber Sinn und Bedeutung — Zeitschrift fur Philosophie und philosophische Kritik, B. 100, Leipzig, 1892. (пер.: Фреге Г. Смысл и денотат — В кн.: Семантика и информатика. Вып. 8 М., 1977).
37. Соссюр Ферди. ван де. Труды по языкознанию — Пер. с франц. под ред. В. Л. Холодовича.— М.: Прогресс, 1977
38. Формальная логика / Под ред. И. Я. Чупахина и И. П. Бродского — Л.: ЛГУ, 1977.
39. Ленин В. И. Поли. собр. соч., т. 29, с. 190.
40. Войшвилло Е. К. Понятие — М.: МГУ, 1967.
41. Борн М. Физика в жизни моего поколения — М.: Иностранная литература, 1963.
42. Виноградов В. В. Основные типы лексических значений слова — Вopr. язык., 1953, № 5, с. 3—29.
43. Проблематика определений терминов в словарях разных типов — Л.: Наука, 1976.
44. Бархударов С. Г. Актуальные задачи лексикографии в области терминов — В сб.: Проблематика определений терминов в словарях разных типов Л., 1976, с. 5—12.
45. Кутина Л. Л. Термины в филологических словарях (к антитезе: энциклопедическое—филологическое) — В сб.: Проблематика определений терминов в словарях разных типов Л., 1976, с. 19—30
46. Суперанская А. В. Терминология и номенклатура — В сб.: Проблематика определений терминов в словарях разных типов Л., 1976, с. 73—83
47. Климовицкий Я. А. Термин и обусловленность определения понятия в системе — В сб.: Проблематика определений терминов в словарях разных типов. Л., 1976, с. 107—114.
48. Подольская Н. В. Модели дефиниций в словаре ономастических терминов — В сб.: Проблематика определений терминов в словарях разных типов Л., 1976, с. 152—158.
49. Энгельс Ф. Анти-Дюринг.— М.: Госполитиздат, 1957.
50. Лотте Д. С. Основы построения научно-технической терминологии — М.: Изд-во АН СССР, 1961.
51. Реформатский А. А. Что такое термины и терминология.— В сб.: Вопросы терминологии М., 1961, с. 46—54
52. Випокур Г. О некоторых явлениях словообразования в русской технической терминологии — Труды Моск. ин-та истории, философии и литературы, 1939, т. 5, с. 3—54
53. Будагов Р. А. История слов в истории общества.— М.: Просвещение, 1971
54. Потебня А. А. Из записок по русской грамматике — 2-е изд.— Харьков, 1888.
55. Маркс К., Энгельс Ф. Соч., 2-е изд., т. 23, с. 31.

56. П. Н. Денисов Еще о некоторых аспектах изучения языков науки.— В кн: Проблемы языка науки и техники. Логические, лингвистические и историко-научные аспекты терминологии. М., 1970, с 52—89.
57. Кутина Л. Л. Формирование языка русской науки (терминология математики, астрономии, географии в первой трети XVIII в.).— М.—Л.: Наука, 1964.
58. Кутина Л. Л. Формирование терминологии физики в России.— М.—Л.: Наука, 1966
59. Бурдин С. М. Роль Ломоносова в создании русской естественно-научной терминологии.— Уч зап Ташкентского пед ин-та, 1954, вып. 2, с. 49—154.
60. Щерба Л. В. Опыт общей теории лексикографии — В кн.: Избранные работы по языковедению и фонетике, т. 1. Л., 1958, с. 68
61. Наумова Л. Л., Федоров В. Т. О словарях по радиофизике и смежным научным направлениям — В кн.: 1-й Всесоюзный симпозиум «Теория и практика научно-технической информации» (г. Горький, 25—28 ноября 1965 г.).— Тезисы докладов и сообщений. М., 1970, с. 128.
62. Шрейдер Ю. А. Тезаурус. БСЭ.— 3-е изд — т. 25 — М: Советская энциклопедия, 1976, с. 1059—1061.
63. Vickery B. C. Thesaurus-a new word in documentation.— Journal of Documentation, 1960, v. 16, № 4, p. 181—189.
64. Jones K S Some thesauric history.— Aslib. Proc, 1972, v. 24, № 7, p. 400—411.
65. Шрейдер Ю. А. Тезаурусы в информатике и теоретической семантике.— В сб.: Научно-техническая информация, сер. 2, 1971, № 3, с. 21—24.
66. Gilchrist A. The thesaurus in retrieval.— London: Aslib, 1971.
67. Шемакин Ю. И. Тезаурус в автоматизированных системах управления и обработки информации.— М.: Воениздат, 1974
68. Пшеничная А. Э. Тезаурус в документальной ИПС.— Киев: Наукова думка, 1977.
69. Черный А. И. Общая методика построения тезаурусов.— В сб.: Научно-техническая информация. сер. 2, 1968, № 5, с. 9—32.
70. Варга Д. Методика подготовки информационных тезаурусов.— Сборник переводов по вопросам информационной теории и практики.— М.: ВИНТИ, 1970, № 17, с. 107.
71. Aitchison J., Gilchrist A. Thesaurus construction. A practical manual.— London; Aslib, 1972.
72. Методика составления информационно-поисковых тезаурусов.— М.: ВИНТИ, 1973.
73. Soergel D Indexing languages and thesauri. Construction and maintenance — Los Angeles: Melville Publ. Co, 1974.
74. INSPEC thesaurus 1975.— London: IEE, 1975.
75. NASA thesaurus. Vol. 1 Alphabetical listing. Vol. 2 Access vocabulary.— Washington, D C: Scient. Techn. Inf. Office, 1976.
76. INIS: thesaurus «IAEA—INIS».— 1977, № 13 (Rev. 12).
77. Lancaster F. W. Vocabulary control in information retrieval systems — Advances in Labrarianship — New York: Acad. Press, 1977, v. 7.
78. Научно-техническая революция и функционирование языков мира — М: Наука, 1977.
79. Белахов Л. Ю. Проблемы государственной стандартизации терминологии — М.: Изд-во стандартов, 1968.
80. Лотте Д. С. Образование и правописание трехэлементных научно-технических терминов.— М: Наука, 1969.
81. Лотте Д. С. Краткие формы научно-технических терминов. М.: Наука, 1971.
82. Руководство по разработке и упорядочению научно-технической терминологии.— М.: Изд-во АН СССР, 1958
83. Как работать над терминологией. Основы и методы / Сост. С. И Коршунов, Г. Г. Самбунова — М.: Наука, 1968.
84. Краткое методическое пособие по разработке и упорядочению научно-технической терминологии — М.: Наука, 1979.
85. Научно-техническая терминология — Сб. стандартизованных и рекомендуемых терминов. В 10-ти т.— М: Изд-во стандартов, 1969—1971.
86. Тетради новых терминов (АН СССР, ВЦП НТЛ и Д) М, 1975—1978 (20 выпусков).
87. Терминология и норма. О языке терминологических стандартов / Под ред. В. П. Даниленко — М.: Наука, 1972.
88. Суперанская А. В. Терминологическая работа на Западе.— В кн: Совершенствование перевода зарубежной научно-технической литературы. М., 1978, с. 152—154.
89. Кобрин Р. Ю. О формальных критериях терминологичности и методологическом обосновании работ по унификации и стандартизации терминологии.— В сб.: Проблематика определений терминов в словарях разных типов. Л., 1976, с. 174—181.
90. Перерва В. М. О принципах и проблемах отбора терминов и составления словариков терминологических словарей.— В сб.: Проблематика определений терминов в словарях разных типов. Л., 1976, с. 190—204
91. Моисеев А. И. К определению терминов в словарях — В сб.: Проблематика определений терминов в словарях разных типов. Л., 1976, с. 243—250.

92. Пиотровский Р. Т., Бектаев К. Б., Пиотровская А. А. Математическая лингвистика.— М.: Высшая школа, 1977.
93. Estoup J. B. Gammes stenographiques.— Paris: Edu, 1916.
94. Zipf G. K. Human behaviour and the principle of least effort.— Cambridge Mass: Addison-Wesley Publishang, 1949.
95. Mandelbrot B. Structure formelle des textes et communication.— Word, 1954, v. 10, № 1; An information theory of the statistical structure of language— In Proceedings of the Symposium on Applications of communication theory.— L.: Butterworth, 1953.
96. Шрейдер Ю. А. О возможностях теоретического вывода статистических закономерностей текста.— В кн.: Проблемы передачи информации. Вып. 1. М., 1967, с. 57—63.
97. Горькова В. И. Ранговое распределение на множествах научно-технической информации.— В сб.: Научно-техническая информация. сер. 2, 1969, № 7, с. 5—11.
98. Горькова В. И., Наумычева К. И. Частотное распределение множества ключевых слов.— В сб.: Научно-техническая информация. Сер. 2, 1970, № 6, с. 3—8.
99. Шрейдер Ю. А. Что такое устойчивость рангового распределения?— В сб.: Научно-техническая информация, сер. 2, 1972, № 12, с. 21—22
100. Козачков Л. С. Система потоков научной информации.— Киев: Наукова думка, 1973.
101. Козачков Л. С. Информационные системы с иерархической («ранговой») структурой.— В сб.: Научно-техническая информация, сер. 2, 1978, № 8, с. 15—24.
102. Частотный словарь общенаучной лексики / Под общ. ред. Е. М. Степановой.— М.: МГУ, 1970.
103. Денисов П. Н. и др. Комплексный частотный словарь русской научной и технической лексики. 3047 слов.— М.: Русский язык, 1978.
104. Митрофанова О. Д. Язык научно-технической литературы.— М.: МГУ, 1973.
105. Язык научной литературы / Под ред. М. К. Пигальской, Н. М. Разихиной, Е. С. Троянской, В. Е. Шевякова.— М.: Наука, 1975.
106. Ярцева В. П. Развитие языка науки.— В кн.: Наука и человечество.— М.: Знание, 1974.
107. Даниленко В. П. Русская терминология: опыт лингвистического описания.— М.: Наука, 1977.
108. Allen-Widdowson. English in Physical Science — London, 1974.
109. Klasson K. Developments in the terminology of physics and technology — Stockholm: Almqvist & Wiksell intern, 1977.
110. Johanson S. Some aspects of the vocabulary of learned and scientific English — Göteborg, 1978.
111. Петров Ю. Г. Исследование и разработка автоматизированных ИПС на основе заглавий публикации.— Дис на соиск. учен. степ. к. т. н. М., 1973.
112. Кирова Л. Г., Пшеничная Л. Э., Рылова Н. Н. О соотношении тезауруса, заголовка и полного текста в ИПС.— Математическая лингвистика, 1974, № 2, с. 46—53.
113. Кажидеева Л. Ф., Каленов Н. Е. Определение информативности заглавий научно-технических документов на массиве системы АСИОР — В сб.: Научно-техническая информация, сер. 2, 1977, № 4, с. 21, 22.
114. Солнцев В. М. Язык как системно-структурное образование — 2-е изд.— М.: Недра, 1977.
115. Сэлтон Г. Автоматическая обработка, хранение и поиск информации — М.: Советское радио, 1973.
116. Солтон Дж. Динамические : библиотечно-информационные системы — М.: Мир, 1979.
117. Р. С. Гиляревский Библиографическое описание как элемент информации.— Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста. Выпуск 6.— М.: ВИНТИ, 1961
118. Никифоровская И. А. Библиографическое описание. Вопросы теории, истории и методики.— Л.: Наука, 1978.
119. Price D. J. Little science, big science.— N. Y.— L: Columbia University Press, 1963. (Пер в кн.: Наука о науке.— М.: Прогресс, 1966, с. 281—384).
120. Goffman W. Mathematical approach to the spread of scientific ideas — the history of most cell research — Nature, 1966, v. 212, № 5061, p. 449—452.
121. Meadows A. J. Communication in science.— L : Butterworths, 1974.
122. Lotka A. J. The frequency distribution of scientific productivity.— Journal of the Washington Academy of Science, 1926, v 16, № 12, p 317—232.
123. Hawkins D. T. Unconventional uses of on-line information retrieval systems: on-line bibliometric studies — J. Amer. Soc Inform. Sci., 1977, v. 28, p. 13—18.
124. Лебедев Г. А. О документальных источниках информации.— В сб.: Научно-техническая информация, сер. 1, 1975, № 7, с. 3—6.
125. Kronick D. A. A history of scientific and technical periodicals: the origins and development of the scientific and technological press 1665—1790 — N.-Y.: Scarecrow Press, 1962.

126. Hersehman A. The primary journal: past, present and future.— *Journal of Chemical Documentation*, 1970, v. 10, № 1, p. 38.
127. Михайлов А. И., Черный А. И., Гиляревский Р. С. Научные коммуникации и информатика — М.: Наука, 1976.
128. Bradford S. C. Sources of information on specific subjects — *Engineering*, L., 1943, v. 8, Jan 26, p. 85—88.
129. Bradford S. C. *Documentation / Introd. Dr. J. H. Shera, M. E. Egan.*— Reprint ed. Univ. Microfilm Limited, 1971 (first published in 1948 by London, Crosby Lockwood)
130. Hubert J. J. Bibliometric models for journal productivity — *Social indicators research*, 1977, v. 4, № 4, p. 441—473.
131. Теплов Д. Ю. К вопросу о рассеянии информации в советской технической периодике.— В сб.: Научно-техническая информация, сер. 2, 1967, № 1, с. 5—8
132. Барнинова З. Б. и др. Изучение научных журналов как каналов связи. Оценка вклада отдельных стран в мировой научный информационный поток.— В сб.: Научно-техническая информация, сер. 2, 1967, № 12, с. 3—11.
133. Горькова В. П., Меллон С. П. Закономерности распределения публикаций в периодических и продолжающихся изданиях по электронике и энергетике (на примере реферативного журнала «Электроника и энергетика») — В сб.: Научно-техническая информация, сер. 2, 1968, № 11, с. 3—7.
134. Маркусова В. А. О формальном методе оценки информационной значимости научных журналов.— В сб.: Научно-техническая информация, сер. 1, 1973, № 10, с. 8—18.
135. Арапоз М. В., Диллинд А. Н. Изменение продуктивности регулярных источников информации — В сб.: Научно-техническая информация, сер. 2, 1976, № 10, с. 3—14.
136. Смольков Н. А. Уравнение рассеяния публикаций в периодических изданиях.— В сб.: Научно-техническая информация, сер. 2, 1977, № 1, с. 5—8
137. Губанков В. Н., Дронина Н. Л. О распределении научных документов в периодических изданиях — В сб.: Научно-техническая информация, сер. 2, 1978, № 9, с. 1—5.
138. Vickery V. S. Bradford's law of scattering — *Journal of Documentation*, 1948, v. 4, p. 198—203
139. Kendall M. G. The bibliography of operational research — *Operational research quarterly*, 1960, v. 11, p. 31—36
140. Leimkuhler F. F. The Bradford distribution.— *Journal of Documentation*, 1967, v. 23, p. 197—207.
141. Brookes B. C. The derivation and application of the Bradford-Zipf distribution.— *Journal of Documentation*, 1968, v. 24, p. 247—265
142. Brookes B. C. Bradford's law and the bibliography of science — *Nature*, 1969, v. 224, p. 953—956.
143. Naranan S. Bradford's law of bibliography of science: an interpretation — *Nature*, 1970, v. 227, p. 631—632.
144. O'Neil E. T. The generalized Bradford distribution — SUNY at Buffalo, School of Information and Library Science, Research Memorandum № 1.
145. Hubert J. J. Analysis of data by a rank-frequency model — Ph. D. dissertation, SUNY at Buffalo, p. 1—230.
146. Козачков Л. С., Приставка Л. П. Организация подписки на периодическую литературу в научно-исследовательском институте — В сб.: Научно-техническая информация, 1966, № 3, с. 15—17.
147. Price D. J. S. Sommcunication in science: the ends — philosophy and forecast.— In: *Communication in science / Eds. A. Reuck and J. Knight* — Boston. Little, Brown and Company, 1967, p. 199—209.  
(Перевод. Прас Д. Дж. С. Тенденции в развитии научной коммуникации — прошлое, настоящее, будущее — В кн.: Коммуникация в современной науке — М.: Прогресс, 1976, с. 93—109).
148. Garfield E. Citation indexes for science — *Science*, 1955, v. 122, № 3159, p. 108.
149. Garfield E. Science citation index A new dimension in indexing.— *Science*, 1964, v. 144, p. 649—654
150. *Social sciences citation index. An international multidisciplinary index to the literature of the social, behavioral and related sciences* — Philadelphia (Pa): ISI, 1977.
151. Насимов В. В., Мульченко З. М. Индекс Гарфилда — В кн.: Научометрия.— М.: Наука, 1969, с. 79—96.
152. Никитыла Л. Н. «Science Citation Index» и возможности его использования (в информационном поиске). Библиоковедение и библиография за рубежом, 1978, вып. 66, с. 87—98.
153. Tukey J. W. Keeping research in contact with the literature: citation indexes and beyond — *IRE Transactions on Engineering writing and speech EWS-5*, 1962, № 2, p. 78—82.

154. Kessler M. M Bibliographic coupling between scientific papers.— American documentation, 1963, v. 14, № 1, p. 10—21
155. Kessler M. M Comparison of the results of bibliographic coupling and analytic subject indexing.— American documentation, 1965, v. 16, № 3, p. 223—233.
156. Price D. J. Networks of scientific papers.— Science, 1965, v. 149, № 3683, p. 510—515. (Перевод: Прайс Д. Дж Система научных публикаций.— УФН, 1966, т. 90, № 2, с. 349—359)
157. Margolis J Citation indexing and evaluation of scientific papers — Science, 1967, v. 155, № 3767, p. 1213—1219.
158. Атанасиу П. Цитаты как инструмент оценки первичных документов.— М: ВИНТИ, 1968. (Докл. Межд. Конгресса по научн инф. Москва, 16—18 сент. 1968)
159. Groos O. V Citation characteristics of astronomical literature — Jornal of documentation, 1969, v. 25, № 4, p. 344—347.
160. Налимов В. В., Мульченко З. М. Наукометрия — М: Наука, 1969.
161. Мульченко З. М. Исследование информационных потоков в науке на основе библиографических ссылок.— Дис. на соиск уч. ст к. п. н.—Ин-т истории естеств. и техн., 1970.
162. Garfield E. Citation indexing for studying science — Nature, 1970, v. 227, p. 669—671.
163. Meadows A. J., O'Connor J. Bibliographical statistics as a guide to the growth points in science.— Science Studies, 1971, v. 1, № 1, p. 95—99.
164. Тимошин Ю. В., Семиходский Г. С Особенности системы цитирования в литературе по геофизической разведочной технике — В сб: Научно-техническая информация, сер. 1, № 4, с. 24—26.
165. Small H. Co-citation in the scientific literature: a new measure of the relationship between two documents.— JASIS, 1973, v. 24, № 4, p. 265—269.
166. Указатель цитированной литературы по информатике «УЦЛ — информатика». В 3-х частях — Баку: АзНИИТИ, 1978.
167. Мехтиев Д. М., Аракелов Р. К., Меллион С. П., Кульгавина О. Э. Использование указателя цитированной литературы для анализа состояния и тенденций развития информатики.— В сб.: Научно-техническая информация, сер. 2, 1977, с. 9—18.
168. Lieberman M. V. A literature citation study of science-technology coupling in electronics — Proc. IEEE, 1978, v. 66, № 1, p. 5—16
169. Wade N. Send not to know for whom the Nobel tolls: it's not for thee.— Science, 1978, v. 202, № 4365, p. 295—296
170. Ashton S. V., Oppenheim C. A method of predicting Nobel prizewinners in chemistry.— Social studies of science, 1978, v. 8, № 3, p. 341—368
171. Исследование возможностей использования количественных методов (в том числе цифровых с применением ЭВМ) для повышения эффективности обработки документальной информации по радиофизике: Отчет / Научно-исследовательской радиофизический институт (НИРФИ): Руководитель работы В. Т. Федоров; № ГР 71053851, № Б 586475.— Горький, 1976 — 130 с

Дата поступления статьи 15 июня 1980 г.

Федоров Владимир Трифионович

ОСНОВНЫЕ ЭЛЕМЕНТЫ НАУЧНЫХ ПУБЛИКАЦИИ И ИХ АНАЛИЗ

---

Подписано в печать 17.04.80. МЦ 00370. Формат 70 x 108<sup>1</sup>/<sub>16</sub>. Бумага писчая.  
Гарнитура литературная. Печать высокая. Объем 2,5 п. л. Тираж 120 экз.  
Заказ № 1503. Бесплатно.

---

Горьковский ордена Трудового Красного Знамени научно-исследовательский  
радиофизический институт (НИРФИ), г. Горький, 603600, ГСП-51, ул. Лядова 15/14,  
т. 38-90-91, д. 5-09.

Горьковская городская типография областного управления издательств,  
полиграфии и книжной торговли, ул. Свердлова, 37.