

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 9

Москва 2019

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК: [004.89:510.64]:615.9

С.М. Гусакова, Д.А. Добрынин, Н.В. Харчевникова

Сравнение языков представления данных в задаче «структура-активность»*

Памяти Валентины Георгиевны Блиновой

Рассматриваются два языка представления химических структур, и проводится сравнительный анализ результатов прогноза биологической активности, осуществляемого с помощью интеллектуальной ДСМ-системы, использующей эти языки в качестве языков представления данных. Сравнение проводится на трех массивах данных по группе параметров.

Ключевые слова: химическое соединение, биологическая активность, язык представления данных, ДСМ-система, простой метод сходства с запретом на контрпримеры.

ВВЕДЕНИЕ

Одной из целей исследования при выявлении связи структуры вещества с биологической активностью (задача «структура–активность») является выделение

подструктур, которые определяют активность и могут быть использованы при конструировании лекарственных средств или при поиске заменителей вредных соединений (ксенобиотиков), загрязняющих окружающую среду.

При решении этой задачи компьютерными методами химическое соединение должно быть представлено с использованием определенного формального

* Работа выполнена при частичной поддержке проекта РФФИ №17-07-00539

языка. От степени адекватности языка зависят результаты прогноза биологической активности. Использование различных языков в одной и той же компьютерной системе дает разные результаты. Поэтому принципиален выбор языка адекватного и задаче, и системе.

Задача «структура–активность» успешно решалась для различных массивов химических веществ и биологических активностей с помощью интеллектуальной ДСМ-системы, реализующей синтез познавательных процедур: индукции, аналогии и абдукции [1]. В этой системе использовался язык представления химических соединений – ФКСП (фрагментарный код суперпозиций подструктур), разработанный в 1970-х гг. [2] и позже расширенный с целью предоставления возможности кодирования структур новых химических классов, таких, например, как полициклические ароматические вещества [3].

В Научно-исследовательском институте биомедицинской химии имени В.Н. Ореховича был разработан язык *MNA*-дескрипторов (*Multilevel Neighborhoods of Atoms* – многоуровневые атомные окрестности) для использования в системе *PASS*, также решающей задачу «структура-активность» [4].

Для сравнения этих двух языков с помощью ДСМ-системы были проведены эксперименты по прогнозированию биологической активности на трех разных массивах данных, которые будут описаны в работе. Результаты оценивались по совокупности параметров.

ЯЗЫКИ ПРЕДСТАВЛЕНИЯ ХИМИЧЕСКИХ СОЕДИНЕНИЙ

Язык ФКСП описывает химическое соединение в виде дискретного набора всех имеющихся в нем подструктур, представляющих центры локализации π -электронов. В структуре выделяются так называемые дескрипторные центры – атомы или группы атомов, которые могут являться центрами «слабого» взаимодействия или реакционными центрами. Для фиксации химического соединения на рецепторе необходимо наличие в его молекуле, по крайней мере, двух дескрипторных центров, поэтому за слово-дескриптор языка ФКСП приняты цепочки атомов углерода, начинающиеся и кончающиеся дескрипторными центрами. Указывается также сопряжение между атомами этой цепочки. В языке ФКСП имеются также средства для кодировки моно и полициклов.

К преимуществам ФКСП можно отнести: хорошую структурированность данных, что является необходимым условием для работы интеллектуальной ДСМ-системы; простоту представления фрагментов структуры, ответственных за проявление биологической активности; возможность представления таких соединений как полициклические ароматические углеводороды. Недостаток – большое количество циклических дескрипторов.

Дескрипторы в языке *MNA*-дескрипторов имеют несколько уровней. На нулевом уровне перечисляются имена всех атомов, встречающихся в структурной формуле. Некоторые атомы объединены в классы эквивалентности, и если в молекуле встречается атом

из такого класса, то в *MNA*-дескрипторе на нулевом уровне присутствует имя этого класса. На первом уровне каждому атому нулевого уровня в скобках приписываются его ближайшие соседи в каком-нибудь однозначном порядке. На каждом следующем уровне к непосредственным соседям добавляются их соседи и так далее. На практике оказывается достаточно дескрипторов 1-го и 2-го уровней. На 2-м уровне атомы, не входящие ни в какие циклы, помечаются знаком « \leftarrow ».

Методика сравнения языков

Языки ФКСП и *MNA*-дескрипторов сравнивались при решении задачи «структура-активность» с использованием интеллектуальной ДСМ-системы, в основе которой лежит ДСМ-метод автоматизированной поддержки научных исследований [5]. Язык представления данных является одной из главных составляющих модели предметной области, разрабатываемой при создании ДСМ-системы. Таким образом, использование разных языков, по сути, соответствует разным моделям одной и той же предметной области. Для сравнения языков были выбраны три массива. Каждый массив содержит химические соединения, про которые известно, обладают они или не обладают некоторой биологической активностью. Эти химические соединения с разделением их на положительные (соединения, обладающие активностью) и отрицательные (соединения, не обладающие активностью) примеры представляют базу фактов (БФ), состоящую из положительной (БФ⁺) и отрицательной (БФ⁻) частей. При решении реальных задач база фактов содержит еще неопределенные примеры (БФ[?]), включающие соединения, про которые неизвестно, обладают или не обладают они активностью.

Пара (V , W) удовлетворяет прямому простому положительному предикату сходства, если V – результат применения операции сходства [6], определенной на множестве объектов (структур химических соединений), причем рассматриваются все положительные примеры из базы фактов, сходные по V , а W – множество свойств (биологических активностей) общих для всех сходных по V примеров. Аналогично определяются пары, удовлетворяющие прямому простому отрицательному предикату сходства с заменой положительных примеров на отрицательные. Эти пары являются положительными и отрицательными гипотезами о причинах (гипотезы 1-го рода) наличия у соединения биологической активности, если они не совпадают с такой же парой противоположного знака. В случае совпадения гипотеза оценивается как противоречивая.

С использованием гипотез 1-го рода и рассуждения по аналогии ДСМ-система строит гипотезы 2-го рода, позволяющие вычислить истинностные оценки для высказывания «соединение X обладает активностью W ». Так, оценка будет +1, если X обладает W ; -1 – если не обладает; 0 – если имеет место фактическое противоречие (соединение содержит и положительные и отрицательные гипотезы о причинах); τ – если активность соединения доопределить не удалось.

При решении задач прогнозирования биологической активности гипотезы 2-го рода формируются

для неопределенных примеров. Для решения задачи сравнения языков, описываемой в настоящей работе, необходимо точно знать правильно или неправильно доопределено соединение, поэтому для прогноза биологической активности некоторые положительные и отрицательные примеры из БФ⁺ и БФ⁻ переводятся в БФ[?]. В порождении гипотез 1-го рода эти примеры не участвуют.

Для решения задачи в предметной области с помощью ДСМ-системы должна быть выбрана стратегия ДСМ-рассуждений.

Выбрать стратегию – это значит выбрать предикаты положительного и отрицательного сходства, реализующие индуктивные рассуждения, с помощью которых находятся положительные и отрицательные гипотезы о причинах исследуемых явлений. В случае задачи «структура – активность» причинами являются фармакофоры – структурные фрагменты химических соединений, наличие которых в молекуле определяет их активность.

Для проведения сравнения рассматриваемых языков были выбраны две стратегии. Первая содержит прямые простые положительный и отрицательный предикаты сходства, а вторая стратегия отличается от первой добавлением к ней условия запрета на контрпримеры.

Особенность стратегии с запретом на контрпримеры состоит в том, что результат сходства объектов, претендующий на статус гипотезы, бракуется, если такой же фрагмент встречается хотя бы в одном примере противоположного знака. В простом методе без контрпримеров требуется, как минимум, два примера противоположного знака для образования противоречивой гипотезы. Запрет на контрпримеры противоречивую гипотезу не создает.

Соединения из исследуемых массивов были закодированы в языках ФКСП и *MNA*-дескрипторов. Для сравнения языков был запланирован цикл исследований с помощью ДСМ-системы:

- нахождение гипотез о причинах (гипотезы 1-го рода);
- нахождение гипотез о доопределении соединений, активность которых считалась неизвестной (гипотезы 2-го рода) в режиме скользящего контроля^а;
- нахождение гипотез о доопределении каждого соединения с помощью гипотез, полученных на объектах всей базы фактов.

В теории ДСМ-систем принята несингулярная оценка качества рассуждений. Для сравнения языков представления данных в шкалу оценки качества рассуждений были введены параметры:

- 1) количество положительных и отрицательных гипотез 1-го рода;
- 2) характеристика гипотез 2-го рода для представленных на прогноз объектов, в том числе:
 - число правильно предсказанных активных соединений;
 - число правильно предсказанных неактивных соединений;

^а Режим скользящего контроля предполагает поочередное удаление каждого примера из базы фактов и попытку его доопределения с помощью гипотез 1-го рода, полученных из объектов, оставшихся в базе фактов.

- число грубых ошибок – когда объект, обладающий свойством, доопределяется как не обладающий или наоборот;
 - число негрубых ошибок перехода активных или неактивных соединений в противоречивые при доопределении;
 - число отсутствия предсказаний.
- 3) точность (*precision*) и полнота (*recall*) прогноза:
- для класса активных соединений $precision = TP / (TP+FP)$, $recall = TP/P$;
 - для класса неактивных соединений $precision = TN / (TN+FN)$, $recall = TN/N$.

Здесь *TP* и *TN* – количество правильно предсказанных активных (*TP*) и неактивных (*TN*) соединений, *FP* – количество неактивных соединений, предсказанных как активные, *FN* – количество активных соединений, предсказанных как неактивные, *P* – общее число активных соединений в базе фактов, *N* – общее число неактивных соединений в базе фактов.

Точность характеризует в каждом классе долю правильных доопределений среди всех соединений, отнесенных системой к данному классу, а полнота – долю всех правильно доопределенных соединений среди всех соединений данного класса в базе фактов.

Эксперименты по сравнению языков

Эксперименты по сравнению языков проводились на трех массивах химических соединений – массиве нитрозамещенных бензолов, для которых прогнозировалась мутагенная активность в тесте Эймса [7], массиве соединений, тестируемых на способность быть легочными сенсбилизаторами, в частности вызывать астму [8] и массиве незамещенных полициклических ароматических углеводородов по прогнозированию канцерогенности [9].

При кодировке и в одном, и в другом языке объекты представляются в виде множеств, а операция сходства определяется как операция пересечения множеств.

В эксперименте с массивом нитрозамещенных бензолов исследовалось одно свойство – проявление активности в тесте Эймса, причем соединение признается мутагенным, если проявляет эту активность во всех вариантах теста на штаммах *TA100* и *TA98* с активацией и без активации. Обучающая выборка содержала результаты теста Эймса для 149 нитрозамещенных бензолов. Из них 75 соединений не проявили активности и считались отрицательными примерами, остальные 74 активных соединения – положительные примеры в базе фактов. Соединения этого массива были закодированы дескрипторами языка ФКСП и *MNA*-дескрипторами 1+2 уровней.

В табл. 1–3 представлены результаты экспериментов, проведенных с помощью простых методов сходства без запрета и с запретом на контрпримеры.

Массив, содержащий соединения, вызывающие астму (астмагены), содержит 80 соединений, из которых 40 активных и 40 неактивных. Соединения этого массива закодированы дескрипторами ФКСП и *MNA*-дескрипторами 1+2 уровней и только 2-го уровня. С ним были проведены эксперименты с помощью простых методов сходства без запрета и с запретом на контрпримеры.

Результаты экспериментов, проведенных на массиве нитрозамещенных бензолов

Показатели для БФ ⁺ и БФ ⁻	Метод / язык							
	простой / ФКСП		простой с запретом на контрпримеры / ФКСП		простой / MNA-дескрипторов 1+2 уровней		простой с запретом на контрпримеры / MNA-дескрипторов 1+2 уровней	
	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻
Кол-во (+)- и (-)-гипотез	75	74	51	4	291	503	153	256
Кол-во правильно доопределенных соединений	24	34	35	24	0	18	37	46
Кол-во грубых ошибок	3	6	5	3	4	0	7	6
Кол-во негрубых ошибок	40	29	6	6	69	56	10	8
Кол-во недоопределенных соединений	7	6	28	42	1	1	20	15
Точность прогноза, %	80	92	92	83	0/0	82	86	87
Полнота прогноза, %	16	23	23	16	0	12	25	31

Таблица 2

Результаты экспериментов, проведенных на массиве астагенов

Показатели для БФ ⁺ и БФ ⁻	Метод / язык											
	простой / ФКСП		простой с запретом на контрпримеры / ФКСП		простой / MNA-дескрипторов 1+2 уровней		простой с запретом на контрпримеры / MNA-дескрипторов 1+2 уровней		простой / MNA-дескрипторов 2-го уровня		простой с запретом на контрпримеры / MNA-дескрипторов 2-го уровня	
	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻
Кол-во (+)- и (-)-гипотез	94	84	88	67	379	546	246	408	118	201	99	163
Кол-во правильно доопределенных соединений	24	25	28	28	8	2	26	28	17	26	27	29
Кол-во грубых ошибок	2	0	3	1	0	0	2	2	2	1	2	1
Кол-во негрубых ошибок	10	7	2	0	32	38	8	6	20	13	5	6
Кол-во недоопределенных соединений	4	8	7	11	0	0	4	4	10	7	1	0
Точность прогноза, %	100	93	96	90	100	100	93	93	94	93	96	94
Полнота прогноза, %	60	60	70	70	20	5	65	70	40	65	67	73

Таблица 3

Результаты экспериментов, проведенных на массиве полициклических ароматических углеводородов

Показатели для БФ ⁺ и БФ ⁻	Метод / язык							
	простой / ФКСП		простой с запретом на контрпримеры / ФКСП		простой / MNA-дескрипторов 1+2 уровней		простой с запретом на контрпримеры / MNA-дескрипторов 1+2 уровней	
	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻	БФ ⁺	БФ ⁻
Кол-во (+) и (-)-гипотез	44	98	16	34	1	10	1	0
Кол-во правильно доопределенных соединений	5	17	4	9	0	34	0	0
Кол-во грубых ошибок	5	4	0	1	25	1	0	0
Кол-во негрубых ошибок	15	12	4	2	0	0	0	0
Кол-во недоопределенных соединений	2	3	19	24	2	1	27	36
Точность прогноза, %	56	77	80	100	0	58	0/0	0/0
Полнота прогноза, %	0,19	0,47	0,15	0,25	0	94	0	0

Массив полициклических ароматических углеводородов, содержащий 27 активных и 36 неактивных соединений, исследовался на канцерогенную активность. Соединения этого массива были закодированы дескрипторами языков ФКСП и *MNA* 1+2 уровней. С этим массивом были проведены те же эксперименты, что и с массивами замещенных нитробензолов и астагенов.

АНАЛИЗ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

При сравнении шкал оценок можно сделать следующие выводы.

Для массивов замещенных нитробензолов и астагенов язык *MNA*-дескрипторов дает существенно больше гипотез, чем язык ФКСП. Это связано со специфичностью представления химических соединений в языке ФКСП. Количество гипотез в наибольшей степени влияет на два показателя – число негрубых ошибок и число недоопределенных примеров. Значительное число гипотез в простом методе дает больше противоречивых доопределений и меньше недоопределенных примеров. Эти показатели влияют на полноту прогноза. Влияние их на точность прогноза опосредовано – чем больше противоречий и недоопределенностей, тем меньше правильных доопределений.

Эксперименты с массивами замещенных нитробензолов и астагенов показали, что язык ФКСП дает сравнимые по точности и полноте результаты для простых методов без запрета и с запретом на контрпримеры. А для кодировки в языке *MNA*-дескрипторов, особенно 1+2 уровней, простой метод не подходит из-за большого количества гипотез, создающих много шума.

При использовании простого метода с запретом на контрпримеры количество негрубых ошибок больше в языке *MNA*, в то время как число недоопределенных примеров больше в ФКСП.

Результаты для простого метода с запретом на контрпримеры сопоставимы в языках ФКСП и *MNA*-дескрипторов. Близкие значения числа правильных доопределений и грубых ошибок в этих языках дают близкие значения точности и полноты.

Однако необходимо заметить, что не все гипотезы о причинах, полученные разными методами с использованием разных языков, содержательны. Часть из них является шумом. Для решения задачи «структура-активность» очень важно выделить этот шум из общего множества гипотез.

Следует обратить внимание на тот факт, что такие характеристики, как точность и полнота, вводятся, как правило, для оценки алгоритма классификации. При этом и в обучающей выборке, и в полученной классификации присутствуют одни и те же классы.

В задаче, решаемой в настоящей работе, в результате ДСМ-рассуждений появляются еще два класса, отсутствующие в базе фактов, – класс соединений, доопределенных как противоречивые, и класс недоопределенных соединений. При подсчете точности они не учитываются, в то время как их роль в характеристике и ДСМ-рассуждений, и модели предметной области существенна. С целью выявления этой роли было проведено еще по одному эксперименту для массивов нитробензолов и астагенов. Суть этих экспериментов состоит в извлечении гипотез 1-го рода из всей базы фактов без удаления каких-либо соединений, а затем доопределения этими гипотезами каждого соединения. Грубых ошибок при этом не возникает, поэтому точность не определяется. В табл. 4 и 5 представлены результаты эксперимента.

Доопределение активных и неактивных соединений в простом методе в качестве противоречивых говорит о том, что соединение помимо гипотез, соответствующих его активности, содержит гипотезы противоположного знака. Если бы эти гипотезы были реально вызывающими/не вызывающими активность, соединение, их содержащее не могло бы иметь ту активность, которая указана в базе фактов. А это значит, что такие гипотезы неинформативны и являются шумом.

Существенно большее число таких доопределений в языке *MNA*-дескрипторов свидетельствует о высоком уровне шума в множестве гипотез для этого языка. Особенно это относится к дескрипторам 1+2 уровней, что еще раз свидетельствует о том, что язык *MNA*-дескрипторов не подходит для простого ДСМ-метода.

Таблица 4

Результаты эксперимента по доопределению соединений массива нитрозамещенных бензолов гипотезами, полученными из полной базы фактов

Количество	Метод / язык			
	простой / ФКСП	простой с запретом на контрпримеры / ФКСП	простой / <i>MNA</i> -дескрипторов 1+2 уровней	простой с запретом на контрпримеры / <i>MNA</i> -дескрипторов 1+2 уровней
правильно доопределенных соединений	71	101	25	135
негрубых ошибок	66	0	123	0
недоопределенных соединений	12	48	1	14
Полнота прогноза, %	48	68	17	91

Результаты эксперимента по доопределению соединений массива астагенов гипотезами, полученными из полной базы фактов

Количество	Метод / язык					
	простой / ФКСП	простой с запретом на контрпримеры / ФКСП	простой / <i>MNA</i> -дескрипторов 1+2 уровней	простой с запретом на контрпримеры / <i>MNA</i> -дескрипторов 1+2 уровней	простой / <i>MNA</i> -дескрипторов 2-го уровня	простой с запретом на контрпримеры / <i>MNA</i> -дескрипторов 2-го уровня
правильно доопределенных соединений	63	70	11	80	52	80
негрубых ошибок	15	0	69	0	28	0
недоопределенных соединений	2	10	0	0	0	0
Полнота прогноза, %	79	88	14	100	65	100

Недоопределившиеся в этом эксперименте соединения не содержат ни одной гипотезы, а значит в базе фактов нет ни одного сходного с ними соединения, или порожденные ими результаты сходства есть и в примерах противоположного знака. Меньшее количество таких недоопределившихся соединений в языке *MNA*-дескрипторов связано с более мелким дроблением фрагментов в гипотезах 1-го рода.

Простой метод с запретом на контрпримеры уничтожает гипотезы, которые встречаются хотя бы в одном примере противоположного знака из базы фактов, и поэтому в случае нахождения гипотез 1-го рода по всей базе фактов противоречивых доопределений нет. Таким образом этот метод сильно сокращает шум и, соответственно, дает более точные результаты.

Однако это не значит, что все бессодержательные гипотезы удаляются при добавлении к простому методу запрета на контрпримеры. Так, неактивное соединение № 62 в массиве астагенов при использовании языка ФКСП не доопределяется, а при использовании языка *MNA*-дескрипторов доопределяется правильно как неактивное. Оно содержит три отрицательные гипотезы, которые являются путями между дескрипторами ФКСП и, значит, не несут информацию о причинах отсутствия активности. Поэтому результаты работы системы обязательно подлежат экспертному анализу.

Язык ФКСП позволяет выделить гипотезы, состоящие из небольшого количества дескрипторов (в среднем три дескриптора). Гипотезы, полученные из соединений, закодированных в языке *MNA*-дескрипторов, состоят из большого количества дескрипторов (в среднем по 20 дескрипторов), значительная часть которых неинформативна. Так дескриптор CH_3 содержится в большинстве соединений из массива астагенов, но он не является информативным.

Такая структура гипотез о причинах в языке *MNA*-дескрипторов затрудняет идентификацию структурных фрагментов, наличие которых в молекуле определяет ее свойство (*активность*). Эта идентификация чрезвычайно важна при поиске новых соединений, обладающих определенной биологической активностью.

Язык ФКСП включает кратность дескрипторов, что является важным признаком при прогнозе. Так, исследования показывают, что активны соединения с двумя группами $N=C=O$ в молекуле. Присоединение к белкам по этим группам приводит к образованию связей двух групп с разными частями белковой молекулы. Соединения с одной группой такого вида являются неактивными.

На рис 1. приведен пример гипотез 1-го рода в языках ФКСП и *MNA*-дескрипторов для соединения № 50 из массива астагенов.

Эксперименты с массивом полициклических ароматических углеводородов показали, что язык *MNA*-дескрипторов не подходит для данных, содержащих полициклы. Использование *MNA*-дескрипторов не позволяет эффективно кодировать полициклические соединения. Например, для соединений №1 и №2, представленных на рис. 2 и 3, кодирование в *MNA* дает по десять полностью совпадающих дескрипторов. Таким образом, эти соединения в кодировке *MNA* не различимы, хотя и обладают различной структурой. Эта особенность представления полициклических соединений связана со способом кодирования *MNA*-дескрипторов, когда структура химического соединения рассыпается на мелкие фрагменты. В полициклических соединениях важной информацией является порядок присоединения бензольных колец друг к другу. Способ кодирования *MNA*-дескрипторами эту информацию теряет.

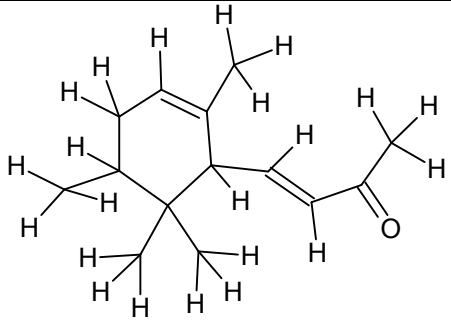
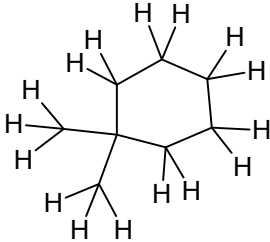
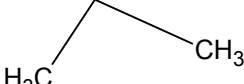
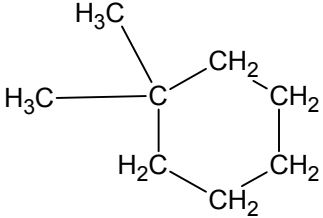
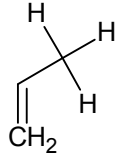
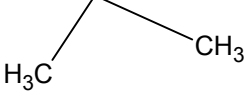
	
Гипотезы ФКСП	Гипотезы MNA
	 H_3C CH_3 $\text{CH}_3\text{-CH}_3$
	$\text{CH}_2=\text{CH-CH}_3$ $\text{CH}_3\text{-CH}_3$
	$\text{CH}_2=\text{CH-CH}_3$ $\text{CH}_3\text{-CH}_3$  H_3C CH_3
	$\text{CH}_2=\text{CH-CH}_3$

Рис. 1. Гипотезы 1-го рода в языках ФКСП и MNA-дескрипторов для соединения № 50 из массива астмагенов

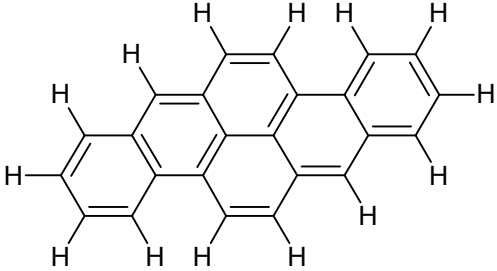
	
Коды MNA	Коды ФКСП
HC CHCC CCCC C(C(CCC)C(CCC)C(CCC)) C(C(CCC)C(CCC)C(CC-H)) C(C(CCC)C(CCC)-H(C)) C(C(CCC)C(CC-H)C(CC-H)) C(C(CCC)C(CC-H)-H(C)) C(C(CC-H)C(CC-H)-H(C)) -H(C(CC-H))	66A6,14 66A6C6,18 66A6C6D6,00 66A6C6B6,22 66B6,14 66A6B6,18 66B6A6Z6,00 66A6C6B6,22 66A6D6,00 66A6D6Z6,00 66A6D6Z6A6,00 66B6D6,00 66A6E6C6,00 66A6Y6C6B6,00 66A6,14 66A6Z6,00 66A6B6,18 66A6,14 66A6Z6,00 66A6C6,18 66D6,00 66D6Z6,00 66A6Z6D6,00 66D6,00 66D6Z6,00 66B6D6Z6,00 66B6,14 66D6,00 66A6D6,00 66D6,00 66A6Z6,00 66A6C6D6,00 66D6,00 66A6,14 66D6,00 66B6D6,00 66A6Z6,00 66B6A6Z6,00

Рис. 2. Соединение № 1

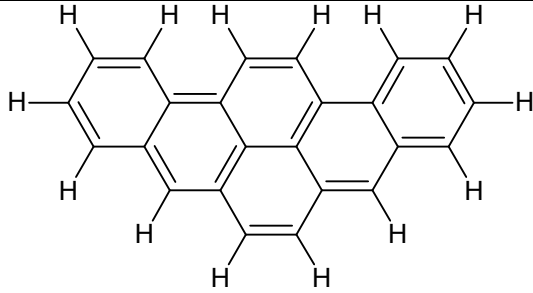
	
Коды MNA	Коды ФКСП
HC CHCC CCCC C(C(CCC)C(CCC)C(CCC)) C(C(CCC)C(CCC)C(CC-H)) C(C(CCC)C(CCC)-H(C)) C(C(CCC)C(CC-H)C(CC-H)) C(C(CCC)C(CC-H)-H(C)) C(C(CC-H)C(CC-H)-H(C)) -H(C(CC-H))	66A6,14 66A6C6,18 66A6C6D6,00 66A6C6A6,22 66B6,14 66A6B6,18 66B6A6Z6,00 66B6A6B6,22 66A6D6,00 66A6D6Z6,00 66A6D6Z6A6,00 66B6D6,00 66A6E6C6,00 66A6C6Y6A6,00 66A6,14 66A6Z6,00 66A6C6,18 66A6,14 66A6Z6,00 66A6B6,18 66D6,00 66D6Z6,00 66A6Z6D6,00 66D6,00 66D6Z6,00 66A6D6Z6,00 66A6,14 66D6,00 66B6D6,00 66D6,00 66A6Z6,00 66B6A6Z6,00 66D6,00 66B6,14 66D6,00 66A6D6,00 66A6Z6,00 66A6C6D6,00

Рис. 3. Соединение № 2

Кодирование в ФКСП тех же соединений дает существенно больше дескрипторов, как совпадающих, так и не совпадающих, что позволяет разделять эти структуры. Для представления полициклических соединений кодом ФКСП было разработано специальное расширение языка. В нем полициклическое соединение кодируется цепочками из 2-х или 3-х циклов, при этом находятся все возможные цепочки. Несмотря на избыточность кодирования таким способом, кодирование в ФКСП дает возможность выявить относительное расположение бензольных колец, что является существенной информацией при представлении структуры полициклических соединений.

ЗАКЛЮЧЕНИЕ

Проведенные исследования по сравнению языков и стратегий ДСМ-рассуждений при решении задачи «структура-активность» показали, что язык ФКСП может быть использован и для стратегии с простым методом без запрета и с запретом на контрпримеры, в то время как язык MNA-дескрипторов дает удовлетворительные результаты только для стратегии с простым методом с запретом на контрпримеры, причем предпочтительнее использовать дескрипторы только 2-го уровня.

При выборе языка представления данных для решения задачи «Структура-активность» надо учитывать особенности структуры исследуемых соединений и приоритеты проводимого исследования. Так, например, для соединений, содержащих полициклы, язык MNA-дескрипторов использован быть не может.

Учитывая достоинства и недостатки каждого языка, имеет смысл разработать вариант ДСМ-метода для решения задачи «структура-активность» с использованием двух языков представления данных.

СПИСОК ЛИТЕРАТУРЫ

1. Финн В.К. Эпистемологические основания ДСМ-метода автоматического порождения гипотез // Научно-техническая информация. Сер. 2. Часть I. – 2013. – № 9. – С. 1-29; Часть II. – № 12. – С. 1-26; Finn V.K. Epistemic foundations of the JSM method automatic hypothesis generation // Automatic Documentation and Mathematical Linguistics. – 2014. – Vol. 48, №2. – P. 96-148.
2. Avidon V.V., Pomerantsev A.B. Structure-activity relationship oriented languages for chemical structure representation // J. Chem. Inf. Comput. Sci. – 1982. – Vol. 22, № 4. – P. 207-214.
3. Блинова В.Г., Добрынин Д.А. Языки представления химических структур в интеллектуальных системах для конструирования лекарств // Научно-техническая информация. Сер. 2. – 2000. – № 6. – С. 14-21; Blinova V.G., Dobrynin D.A. Structure representation languages in intelligent drug-design systems // Automatic Documentation and Mathematical Linguistics. – 2000. – Vol. 34, №3. – P. 53-63.
4. Филимонов Д.А., Лагунин А.А., Глориева Т.А., Рудик А.В., Дружиловский Д.С., Погодин П.В., Поройков В.В. Предсказание спектров биологической активности органических соединений с помощью веб-ресурса PASS ONLINE // Химия гетероциклических соединений. – 2014. – № 3. – С. 483-499.
5. Арский Ю.М., Финн В.К. Принципы конструирования интеллектуальных систем // Информационные технологии и вычислительные системы. – № 4. – 2008. – С. 4-37.

6. Гусакова С.М., Финн В.К. Сходство и правдоподобный вывод // Известия АН СССР. Сер. Тех. киб. – 1987. – Вып. 5. – С. 42-63.
7. Харчевникова Н.В., Блинова В.Г., Добрынин Д.А., Журков В.С. Использование интеллектуальной ДСМ-системы для анализа связи структуры нитрозамещенных бензолов с их мутагенной активностью в тесте Эймса // Научно-техническая информация. Сер. 2. – 2015. – № 3. – С. 6-11.
8. Харчевникова Н.В., Блинова В.Г., Добрынин Д.А. Сопоставление различных моделей исследования связи «структура – астмагенная активность» // Научно-техническая информация. Сер. 2. – 2016. – № 2. – С. 23-28.
9. Блинова В.Г., Добрынин Д.А., Максин М.В., Жолдакова З.И., Харчевникова Н.В. Прогноз канцерогенности полициклических ароматических углеводородов с использованием квантовохимического модуля генерации метаболитов интеллектуальной ДСМ-системы // Научно-техническая информация. Сер. 2. – 2003. – № 11. – С. 12-17; Blinova V.G., Dobrynin D.A., Maksin M.V., Zholdakova Z.I., Kharchevnikova N.V. Carcinogenicity forecasting for polycyclic aromatic hydrocarbons from the quantum-chemical module for metabolite generation in a JSM

systems // Automatic Documentation and Mathematical Linguistics. – 2003. – Vol. 37, №6. – P. 6-12.

Материал поступил в редакцию 05.06.19

Сведения об авторах

ГУСАКОВА Светлана Марковна – кандидат физико-математических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН, Москва
e-mail: svem45@yandex.ru

ДОБРЫНИН Дмитрий Анатольевич – кандидат технических наук, научный сотрудник Российского Государственного Гуманитарного Университета (РГГУ), старший научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН, Москва
e-mail: rabota51@mail.ru

ХАРЧЕВНИКОВА Нина Вениаминовна – доктор биологических наук, ведущий научный сотрудник ФГБУ «Центр стратегического планирования и управления медико-биологическими рисками здоровью» Минздрава России
e-mail: kharchevnikova_n@mail.ru

Ещё один вероятностный алгоритм для вычисления сходств*

Описан новый вероятностный алгоритм вычисления гипотез как результатов сходства между обучающими примерами для задачи машинного обучения, основанного на бинарной операции сходства. В отличие от ранее предложенных вероятностных алгоритмов порядок учета обучающих примеров является фиксированным для всех гипотез. Этот алгоритм полезен для реализации на GPGPU. Основным результатом этой статьи состоит в независимости вероятности появления каждого сходства от порядка появления обучающих примеров в выборке.

Ключевые слова: сходство, вероятностный алгоритм, контрпример

ВВЕДЕНИЕ

Феномен переобучения, проявлявшийся в порождении «фантомных» сходств при интеллектуальном анализе данных с помощью исчерпывающего порождения всех сходств между обучающими примерами (ДСМ-метод [1] и анализ формальных понятий [2]) был исследован в работах [3, 4]. Для исчерпывающего порождения всех сходств дополнительной проблемой служит потенциальная возможность наличия экспоненциально большого числа таких сходств по сравнению с числом обучающих примеров (и с числом бинарных атрибутов, применяемых для их описания). В качестве альтернативы в статьях [5, 6] был предложен и исследован подход, когда рассматривается только малая доля вероятностно порождаемых сходств. Этот подход позволил принять точку зрения статистической теории обучения.

В статье [7] получен результат о достаточном числе порождаемых сходств в духе классической теоремы Вапника-Червоненкиса [8]. Этот результат показывает, что если для описания обучающих и тестовых (предъявленных для прогнозирования) примеров используется n бинарных атрибутов, то для того, чтобы с вероятностью, не меньшей $1 - \delta$, все ϵ -важные положительные тестовые объекты были предсказаны правильно, доста-

точно вероятностно породить
$$N \geq \frac{2(n+1) - 2 \log \delta}{\epsilon}$$

сходств. Здесь важность тестового примера определяется как вероятность порождения хотя бы одного сходства, которое вкладывается в этот пример (и служит причиной положительного предсказания).

При анализе полученного результата обращает на себя внимание независимость приведенной нижней

границы на достаточное число порождаемых сходств от объёма обучающей выборки (числа обучающих примеров). Это особенно интересно в связи с тем, что на практике объём обучающей выборки может на несколько порядков превышать размерность признакового пространства (т. е. числа признаков, используемых для описания объектов).

Алгоритмы, основанные на спаривающих цепях Маркова, допускают хорошее распараллеливание (из-за независимости траекторий) на современных многоядерных компьютерах, однако возник вопрос о существовании вероятностного алгоритма, который может хорошо распараллеливаться для вычислений на графических картах общего назначения (*GPGPU* – *General Purpose Graphical Processor Unit*).

Сегодня подавляющее большинство современных центральных процессоров (*CPU* – *Central Processor Unit*) располагает 4–16 ядрами с возможностью выполнять одновременно 8–32 нитей вычислений (вычислительных потоков), тогда как графические карты фирм *nVidia* и *AMD* могут вести вычисления одновременно в 4096 нитей и даже более. Основная проблема вычислений на *GPGPU* – малый объём локальной памяти, из-за чего необходимо подгружать данные из (глобальной) памяти компьютера. Случай алгоритмов из статей [5, 6], основанных на спаривающих цепях Маркова, является плохим для вычислений на *GPGPU* в том смысле, что номер следующего обрабатываемого обучающего примера определяется датчиком случайных чисел и не может быть предсказан и загружен в локальную память заранее.

Идеальный случай для вычисления на *GPGPU* – когда обучающие примеры подаются на вход алгоритма в некотором фиксированном порядке, тогда некоторый пакет этих примеров может быть загружен в локальную память *GPGPU*, и они могут обрабатываться подмножествами вычислительных устройств

* Работа выполнена при частичной поддержке гранта РФФИ № 18-29-03063мк

внутри *GPGPU*, называемыми *рабочими группами*, по порядку. Из оценки статьи [7] можно извлечь требуемое количество сходств между обучающими примерами, которое и будет вычислять наш новый алгоритм. Точнее, каждая рабочая группа будет вычислять одно такое сходство между обучающими примерами независимо от других таких рабочих групп и параллельно им, если число требуемых сходств не превысит числа поддерживаемых *GPGPU* рабочих групп. В случае превышения числом сходств числа доступных рабочих групп новый алгоритм вычисления сходств должен быть запущен еще несколько раз.

Когда приходит время для обработки следующего обучающего примера, тогда каждая рабочая группа (подмножество вычислительных устройств внутри *GPGPU*) с некоторой вероятностью (зависящей только от самого примера) пересекает ранее вычисленное сходство с вновь поступившим примером независимо от других таких групп, работающих параллельно с ним.

Здесь ключевой проблемой является установление факта, что вероятность порождения сходства описанным способом не зависит от порядка предъявления обучающих примеров. Доказательство этого утверждения и составляет содержание настоящей работы.

Возможна ситуация, когда у сходства окажется контрпример (пример противоположного знака, в описание которого входит рассматриваемое сходство). Принцип «запрета контрпримеров», введенный В.К. Финном [3], отбрасывает такие сходства. Чтобы не разрушить основной результат настоящей статьи и по причине сложности проверки этого принципа в процессе порождения сходств (из-за ограниченности локальной памяти *GPGPU*) предлагается проверять его после этапа порождения сходств.

ВСПОМОГАТЕЛЬНАЯ ЛЕММА

Предположим, что наши алгоритмы работают на примерах, извлеченных из некоторой нижней полурешетки [4]. Без снижения общности (см. [4, 9]) можем считать, что примеры и их сходства закодированы битовыми строками таким образом, что бинарная операция совпадает с побитовым умножением. В статье [10] был изложен алгоритм такого кодирования (и доказана его корректность) – когда объекты описываются признаками, на значениях каждого из которых определена своя собственная нижняя полурешетка (с возможным добавлением наименьшего элемента, соответствующего случаю, когда у двух объектов на этом признаке сходства не имеется).

Рассмотрим случай перестановки обучающих примеров: исходное сходство обозначим h , обучающие примеры через o_1 и o_2 , а возникающие сходства через

$$h_1 = h \cap o_1, h_2 = h \cap o_2, \\ h_3 = h_1 \cap o_2 = h_2 \cap o_1 = h \cap o_1 \cap o_2.$$

Лемма 1. Возможны следующие случаи:

- 1) $h_3 = h_2 = h_1 = h$;
- 2) $h_3 = h_2 \subset h_1 = h$;
- 3) $h_3 = h_1 \subset h_2 = h$;
- 4) $h_3 = h_2 = h_1 \subset h$;
- 5) $h_3 = h_1 \subset h_2 \subset h$;

$$6) h_3 = h_2 \subset h_1 \subset h;$$

$$7) h_3 \subset h_2 \subset h \wedge h_3 \subset h_1 \subset h \wedge h_2 \not\subset h_1 \wedge h_1 \not\subset h_2.$$

Доказательство. Имеем очевидные включения $h_3 \subseteq h_2 \subseteq h$, $h_3 \subseteq h_1 \subseteq h$. Если $h_3 = h$, то имеем первый вариант для нашей леммы. Рассмотрим случаи нарушения только одного равенства в каждой из двух цепочек. Здесь возможны три случая: $h_3 \subset h_1 = h \wedge h_3 = h_2 \subset h$, $h_3 = h_1 \subset h \wedge h_3 \subset h_2 = h$ и $h_3 = h_1 = h_2 \subset h$, так как случай $h_3 \subset h_1 = h \wedge h_3 \subset h_2 = h$ – невозможен. Он невозможен, так как тогда $h_3 = h_1 \cap h_2 = h \cap h = h$, что противоречит допущению этого случая, так как одно из равенств нарушено. Первый из наших случаев соответствует второму варианту леммы, второй случай – третьему варианту, третий случай – четвертому варианту. Остаются случаи, где в одной и цепочек оба равенства нарушены, а в другой – только одно: $h_3 = h_1 \subset h_2 \subset h$, $h_3 = h_2 \subset h_1 \subset h$. Это, соответственно, пятый и шестой варианты леммы. Оставшиеся случаи – $h_3 \subset h_2 \subset h_1 = h$, $h_3 \subset h_1 \subset h_2 = h$ – невозможны: из $h_1 = h$ следует $h_3 = h_1 \cap o_2 = h \cap o_2 = h_2$, а из $h_2 = h$ следует $h_3 = h_2 \cap o_1 = h \cap o_1 = h_1$. Остаётся последний случай, когда нарушены все равенства: $h_3 \subset h_2 \subset h$, $h_3 \subset h_1 \subset h$. Это седьмой вариант леммы, если мы докажем, что в этом случае $h_2 \not\subset h_1 \wedge h_1 \not\subset h_2$. Но если $h_2 \not\subset h_1$, то $h_3 = h_1 \cap h_2 = h_2$, а если $h_1 \not\subset h_2$, то $h_3 = h_1 \cap h_2 = h_1$, что исключается условиями этого случая. Лемма доказана.

Напомним, что имеется список контрпримеров – объектов противоположного знака, которые используются для выкидывания сходства h обучающих примеров, если $h \subseteq c$ для некоторого контрпримера c .

Описание алгоритма

На шаге t k ранее вычисленному сходству h_{t-1} при появлении обучающего примера o_t применяется следующий базовый шаг: $h_t = h_{t-1} \cap o_t$ с вероятностью p_t и $h_t = h_{t-1}$ с вероятностью $1 - p_t$. При этом вероятность перехода на пересечение должна зависеть только от примера o_t . Она может быть постоянной (например, $p_t = 1/2$) или определяться долей наличествующих в o_t бинарных атрибутов по отношению к общему их числу n .

Если запрет контрпримеров проверяется сразу, то алгоритм таков: $h_t = h_{t-1} \cap o_t$ с вероятностью p_t и $h_t = h_{t-1}$ с вероятностью $1 - p_t$, если $h_{t-1} \cap o_t$ не имеет контрпримеров, и $h_t = h_{t-1}$, если $h_{t-1} \cap o_t$ обладает контрпримерами. К сожалению, этот подход требует загрузки контрпримеров в локальную память *GPGPU*, что не практично с вычислительной точки зрения. Более того, это разрушает утверждение основного результата настоящей статьи. Поэтому мы проверяем запрет контрпримеров после завершения этапа порождения сходств.

ОСНОВНОЙ РЕЗУЛЬТАТ

Теорема 1. Вероятность появления сходства не зависит от порядка появления обучающих примеров.

Доказательство. Рассмотрим сначала случай без запрета контрпримеров. Так как любая группа перестановок порождается транспозициями соседних элементов, достаточно рассмотреть случай перестановки двух ближайших обучающих примеров. Проанализируем варианты, описываемые Леммой 1. В первом варианте мы окажемся в $h_3 = h_2 = h_1 = h$ с вероятностью 1 независимо от порядка поступления обучающих примеров. Во втором варианте леммы с вероятностью $1 - p_2$ окажемся в $h_1 = h$ и с вероятностью p_2 окажемся в $h_3 = h_2$. В третьем варианте леммы с вероятностью $1 - p_1$ окажемся в $h_2 = h$ и с вероятностью p_1 окажемся в $h_3 = h_1$. В четвертом варианте леммы с вероятностью $(1 - p_1) \cdot (1 - p_2)$ окажемся в h и с вероятностью $p_1 + p_2 - p_1 \cdot p_2$ окажемся в $h_3 = h_2 = h_1$. В пятом случае леммы с вероятностью $(1 - p_1) \cdot (1 - p_2)$ окажемся в h , с вероятностью $p_1 \cdot (1 - p_2)$ окажемся в h_1 и с вероятностью p_2 окажемся в $h_3 = h_2$. В шестом случае леммы с вероятностью $(1 - p_1) \cdot (1 - p_2)$ окажемся в h , с вероятностью $(1 - p_1) \cdot p_2$ окажемся в h_2 и с вероятностью p_1 окажемся в $h_3 = h_1$. В седьмом варианте леммы с вероятностью $(1 - p_1) \cdot (1 - p_2)$ окажемся в h , с вероятностью $p_1 \cdot (1 - p_2)$ окажемся в h_1 , с вероятностью $(1 - p_1) \cdot p_2$ окажемся в h_2 и с вероятностью $p_1 \cdot p_2$ окажемся в h_3 .

Случай с запретом контрпримеров может быть рассмотрен аналогично. Но справедливость теоремы в этом случае вытекает из предыдущего рассуждения с помощью формулы условной вероятности, так как запрет контрпримеров – это условие, что пересечение сходства с новым примером не имеет контрпримеров. Этим завершается доказательство теоремы.

ЗАКЛЮЧЕНИЕ

В настоящей работе описан вероятностный алгоритм порождения сходства обучающих примеров, хорошо подходящий для реализации вычислений на графических картах общего назначения *GPGPU*. Ключевая теорема настоящей работы гарантирует независимость вероятности порождения сходства от порядка предъявления элементов обучающей выборки. Нижнюю границу на число необходимых сходств предъявляет главная теорема из статьи [7].

* * *

Автор благодарит своих коллег из ФИЦ «Информатика и управление» РАН за конструктивную критику и полезные замечания.

СПИСОК ЛИТЕРАТУРЫ

1. ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания / ред. В.К. Финн, О.М. Аншаков. – М.: Эдиториал-УРСС, 2009. – 432 с.
2. Ganter B., Wille R. Formal Concept Analysis. – Berlin: Springer-Verlag, 1999. – 284 p.
3. Виноградов Д.В. Скорость сходимости к пределу вероятности порождения случайного сходства при наличии контр-примеров // Научно-техническая информация. Сер. 2. – 2018. – № 2. – С. 21-24; Vinogradov D.V. The Rate of Convergence to the Limit of the Probability of Encountering an Accidental Similarity in the Presence of Counter Examples // Automatic Documentation and Mathematical Linguistics. – 2018. – Vol. 52, № 1. – P. 35-37.
4. Vinogradov D.V. Accidental formal concepts in the presence of counterexamples // Proceedings of International Workshop on Formal Concepts Analysis for Knowledge Discovery (FCA4KD 2017): CEUR Workshop Proceedings. – 2017. – Vol. 1921. – P. 104-112.
5. Виноградов Д.В. Вероятностное порождения гипотез в ДСМ-методе с помощью простейших цепей Маркова // Научная и техническая информация. Сер. 2. – 2012. – № 9. – С. 20-27; Vinogradov D.V. Random Generation of Hypotheses in the JSM Method using Simple Markov Chains // Automatic Documentation and Mathematical Linguistics. – 2012. – Vol. 46, № 5. – P. 221-228.
6. Vinogradov D.V. Machine Learning Based on Similarity Operation // Communication in Computer and Information Science. – 2018. – Vol. 934. – P. 46-59.
7. Виноградов Д.В. Надежность предсказания по аналогии // Научно-техническая информация. Сер. 2. – 2017. – № 7. – С. 11-15; Vinogradov D.V. The reliability of analogy-based prediction // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 4. – P. 191-195.
8. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов (статистические проблемы обучения). – М.: Наука, 1974. – 416 с.
9. Davey B.A., Priestley H.A. Introduction to Lattices and Order. 2nd eds. – Cambridge: Cambridge University Press, 2002. – 298 p.
10. Виноградов Д.В. О представлении объектов битовыми строками для ВКФ-метода // Научная и техническая информация. Сер. 2. – 2018. – № 5. – С. 1-4; Vinogradov D.V. On Object Representation by Bit Strings for the VKF-Method // Automatic Documentation and Mathematical Linguistics. – 2018. – Vol. 52, № 3. – P. 113-116.

Материал поступил в редакцию 22.04.19

Сведения об авторе

ВИНОГРАДОВ Дмитрий Вячеславович – доктор физико-математических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН и доцент Российского Государственного Гуманитарного Университета, Москва
e-mail: vinogradov.d.w@gmail.com

Ю.М. Брумштейн, Н.В. Васильев

Академии наук стран Западной Европы: анализ направлений деятельности, информационного наполнения сайтов и их вебметрических показателей

Рассмотрена проблематика, связанная с информационным присутствием в Интернете академий наук (АН) ряда стран Западной Европы; охарактеризованы юридические статусы этих академий, особенности их административного, финансового и информационного взаимодействия с органами государственного управления, корпоративными структурами, национальными и международными научными организациями, отдельными исследователями.

Проанализирован состав информации, размещенной на сайтах этих АН; особенности интерфейсов этих сайтов; номенклатура используемых языков; доступные для пользователей сервисы; поисковые системы сайтов; ключевые вебметрические показатели; средства рассылки бюллетеней.

Сделаны выводы о роли изученных сайтов АН в обеспечении доступности и распространении научно-технической информации, в том числе журнальных статей; о формировании национальных и международного научного информационного пространства; о степени полезности сайтов этих АН для российских исследователей.

Ключевые слова: *западноевропейские страны, национальные академии наук, интернет-сайты, агрегация информации, информационное наполнение, сервисы пользователей, поисковые системы, языки интерфейсов, вебметрические показатели*

ВВЕДЕНИЕ

Ранее авторами было выполнено исследование особенностей деятельности и сайтов академий наук трех категорий стран: постсоветских государств, возникших на территории бывшего СССР; государств, ранее относившихся к "социалистическому лагерю"; стран Американского континента [1,2]. В настоящей статье рассматриваются вопросы, относящиеся к деятельности академий наук (АН) ведущих стран Западной Европы (включая скандинавские страны, а также Финляндию), исследуется информационное наполнение, функциональные возможности и вебметрические показатели (ВМП) их сайтов. Для российских исследователей эти вопросы важны по следующим причинам: территориальная близость рассматриваемых стран к России и, как следствие, удобные возможности реализации совместных научных проектов, поддержки процессов академической мобильности исследователей [3] и студентов; традиционное наличие связей между АН России и указанных стран, а также прямых связей с исследовательскими академическими организациями и вузами; систематическая публикация Российским фондом фундаментальных исследований предложений о грантовой поддержке совместных проектов российских исследователей и ученых из ряда стран Западной Европы. Однако в

существующих русскоязычных публикациях слабо отражена тематика, связанная с условиями деятельности академий наук стран Западной Европы, информационным наполнением их сайтов, практикой использования этих сайтов для подготовки и проведения международных научных мероприятий, формирования и поддержки деятельности «рабочих групп» исследователей, обеспечения доступности и распространения научно-технической информации. Основная цель настоящей статьи – устранение указанных недостатков.

При подготовке статьи нами были использованы следующие источники информации: сайты самих АН, включая размещенную на них информацию о деятельности академий; опубликованные академиями отчеты, иные документы; научные статьи, доступные в репозиториях научной информации; информационные материалы в Интернете; общедоступные статистические данные; Википедия [4]. Для получения данных о вебметрических показателях сайтов академий наук (за период апрель – июнь 2019 г.) применялись методики, описанные в [1, 2, 5]. Учитывался также опыт других исследователей [6, 7]. Сведения о численности населения стран взяты с сайта <https://all-populations.com/ru/list-of-countries-by-population.html> для периода 2018-2019 гг. Цитируемые фрагменты текста в настоящей статье даны курсивом.

ОБЪЕДИНЕНИЯ (АССОЦИАЦИИ), В ДЕЯТЕЛЬНОСТИ КОТОРЫХ УЧАСТВУЮТ АКАДЕМИИ НАУК СТРАН ЗАПАДНОЙ ЕВРОПЫ

Изучение опыта европейских стран в вопросах интеграции науки [8–10] является актуальным для России. Кроме того, полезно знать, как финансируются научные исследования и производится оценка их результативности [11–13], как используются гранты для поддержки конкретных научных исследований [14]; как доводится до целевой аудитории информация о предоставлении грантов на проведение исследований интернациональными научными коллективами.

Еще в 2000 г. Европейская комиссия выступила с инициативой создания «Европейского научного пространства» с целью уменьшения фрагментарности исследований, проводимых в Европейском Сообществе. Отметим рамочные программы Европейского Союза (http://www.bio-economy.ru/ramochnaya_programma_es/7_ramochnaya_programma_es/), в которых прямо отражена цель создания «Общего Европейского Научного Пространства» (ERA): (1) шестую (на 2002–2006 гг.) с бюджетом 17,5 млрд евро; (2) седьмую (FP7) – на 2007–2013 гг. с бюджетом 54,582 млрд евро; (3) восьмую (Horizon 2020), направленную на «превращение Евросоюза в основанную на знаниях конкурентоспособную и динамичную экономику» с бюджетом 80 млрд евро; (4) девятую (Horizon Europe) на 2021–2027 годы с бюджетом 100 млрд евро.

Основными направлениями интеграции науки в европейских странах можно считать: создание национальных организаций для выполнения исследований по конкретным направлениям (например, Европейское космическое агентство, Европейский центр ядерных исследований); проведение совместных научных изысканий разового или регулярного характера коллективами (группами) исследователей из разных стран; грантовую поддержку национальными фондами работ, проводимых местными исследователями совместно с зарубежными учеными; издание международных научных журналов, включение в состав членов их редколлегий и привлечение в качестве авторов исследователей из разных стран; обеспечение доступности опубликованной научной информации в условиях глобализации экономики и развития международного научно-информационного пространства [15–18]; международный обмен стажировками научных сотрудников, аспирантов, студентов [3]; присуждение национальным и зарубежным исследователям, международных премий по результатам научной деятельности.

Ключевую роль в интеграции науки играет развитие информационно-коммуникационных технологий, агрегация научной информации на интернет-сайтах [1, 5, 19, 20]; формирование глобального научно-информационного пространства [15, 21, 22]; обеспечение известности в нем исследований, проведенных национальными научными организациями [7, 23] и отдельными исследователями.

Пока в рамках ЕС нет централизованной структуры типа «Европейской АН». Однако действуют мно-

гочисленные объединения европейских и международных научных организаций [5], в работе которых принимают участие АН стран Западной Европы.

На сайты указанных объединений имеются ссылки с различных интернет-ресурсов, включая сайты отдельных АН, рассматриваемых далее. Список этих организаций (объединений, ассоциаций, советов и пр.) приводится в алфавитном порядке:

ALLEA (All European Academies) – Европейская Федерация национальных АН (более чем из 40 стран) и гуманитарных исследований – сайт <http://www.allea.org/>

CAETS (International Council of Academies of Engineering and Technological Sciences) – сайт <http://www.caets.org/>

CIOMS (Council for International Organizations of Medical Sciences) – сайт <https://cioms.ch/>

EACME (European Association of Centres of Medical Ethics) – сайт <http://www.eacmeweb.com/>

EASAC (European Academies Science Advisory Council) – Научный консультативный Совет Европейских Академий – сайт <https://www.easac.eu/>

ESF (European Science Foundation) – сайт <http://www.esf.org/>

Euro-CASE (European Council of Applied Sciences and Engineering) – сайт <http://www.euro-case.org/>

IAC (InterAcademy Council) – сайт <http://www.inter-academies.org/>

IAMP (InterAcademy Medical Panel) – сайт <http://www.iamp-online.org/>. Доступ к этому сайту ограничен.

IAP (The interacademy partnership) – глобальная сеть академий наук, действующих по направлениям «наука, исследования, здоровье» – сайты: <http://www.interacademies.org/> (относится также к *IAC*); www.interacademies.net (по последней ссылке открывается сайт <http://www.interacademies.org/>).

ICSU (International Council For Science) – Международный совет по науке – сайт <http://www.icsu.org>.

IFS (International Foundation for Science) – сайт <http://www.ifs.se/>.

I.H.R. Network, International Health Regulations – Международная правозащитная сеть академий и научных обществ – сайт www.nationalacademies.org/humanrights

ISC (International Scientific Council) – сайт <https://council.science/>

UAI (Union Académique Internationale) – Международный союз Академий – сайт www.uai-uaa.org

Количество объединений (ассоциаций), связанных с научно-издательской деятельностью, редактурованием, переводом научных произведений достаточно велико. Поэтому мы отметим только те из них, в названиях которых прямо отражена «принадлежность к Европе».

European Association of Science Editors (EASE) – сайт <http://www.ease.org.uk/>

European Medical Writers Association (EMWA) – сайт <http://www.emwa.org>

MET: Mediterranean Editors & Translators – сайт <http://www.metmeetings.org>

В табл. 1 и 2. представлены вебметрические показатели сайтов перечисленных нами организаций.

Таблица 1

Вебметрические показатели для сайтов объединений (ассоциаций) академий наук, в деятельности которых участвуют АН стран Западной Европы (первая часть)

№	Сокращенное название организации, которой принадлежит сайт и адрес сайта	АТ, сек.	Count, URLs			Size, Mb	AVD, чч:мм:сс
			text/html	image	application		
1	ALLEA (https://www.allea.org)	19,7	621	262	368	270,5	0:08:48
2	CAETS (http://www.caets.org)	3,0	2 767	33	125	176,5	0:17:15
3	CIOMS (https://cioms.ch)	12,1	374	279	210	112,5	0:01:41
4	EACME (http://www.eacmeweb.com)	1,0	24	10	32	13,9	0:00:33
5	EASAC (https://www.easac.eu)	3,7	1	0	0	0,04	0:00:22
6	ESF (http://www.esf.org)	8,0	227	535	79	248,03	0:00:42
7	Euro-CASE (http://www.euro-case.org)	4,7	10 573	489	192	1 028,02	0:27:02
8	IAC (https://www.interacademies.org)	8,8	3 169	1 495	179	1 175,8	0:16:48
9	IAMP (http://www.iamp-online.org)	Нет доступа к серверу (Сообщение. Forbidden. You don't have permission to access / on this server)					
10	ICSU (http://www.icsu.org)	Не удается получить доступ к сайту					
11	IFS (http://www.ifs.se)	2,3	127	90	75	159,9	0:00:11
12	I.H.R. Network (http://nationalacademies.org)	5,5	9 262	161	57	689,7	0:12:30.
13	ISC (https://council.science)	3,7	602	19	171	314,4	0:04:54
14	UAI (http://www.uai-iaa.org)	2,7	481	46	9	14,8	0:01:34
15	EASE (http://www.ease.org.uk)	7,2	450	70	251	95,7	0:10:14
16	EMWA (https://www.emwa.org)	5,3	492	300	234	229,3	0:04:11
17	MET (https://www.metmeetings.org)	4,3	Нет доступа к серверу (Сообщение: еггг 12157 ошибка поддержки безопасных каналов)				

Таблица 2

Вебметрические показатели для сайтов объединений академий наук, в деятельности которых участвуют АН стран Западной Европы (вторая часть)

№	Сокращенное название организации, которой принадлежит сайт	Абсолютные количества ссылок				КУнПос. за месяц	КПросм. за месяц	Scholar Google (SG)
		Входящие	Внут.	Исходящие				
				Ω	Ψ			
1	ALLEA	776	569	4	125	4 338	17 340	1 230
2	CAETS	165	985	5	37	н	н	19 600
3	CIOMS	859	1 113	9	64	7 416	29 670	14 100
4	EACME	48	91	0	23	1 479	5 910	8 680
5	EASAC	323	4 993	30	63	4 227	16 920	1 340
6	ESF	5 582	15 659	7	91	7 976	31 890	40 500
7	Euro-CASE	121	432	25	134	1 292	5 160	225
8	IAC	38	1 849	6	145	7 812	31 260	1 130
9	IAMP	88	1 707	0	0	н	н	202
10	ICSU	2 624	17 600	24	138	9 120	36 480	20 500
11	IFS	697	1 894	8	39	11 999	48 000	55 400
12	I.H.R. Network	15 772	46 531	52	50	503 404	1 036 740	18 800
13	ISC	16	250	6	136	9 632	38 520	25 600
14	UAI	96	205	2	61	1 359	5 430	762
15	EASE	800	2 591	17	143	4 729	18 930	303 000
16	EMWA	373	1 439	14	247	7 183	28 740	1 270
17	MET	113	959	54	28	1 605	6 420	26 000

ПРИМЕЧАНИЯ: символ «н» означает, что невозможно получить информацию в связи с отсутствием на исследуемом сайте средств сбора соответствующих статистических данных.

Величины АТ соответствуют показателю скорости доступа к стартовым страницам (СС) сайтов и оцениваются как средние значения времен доступа из разных «точек» планеты. Из табл. 1 видно, что по этому показателю рассмотренные сайты значительно различаются. Быстрее всего открывается сайт *EASME*. По соотношению *text/html* лидерами являются сайты *Euro-CASE* и *I.H.R. Network*. Наиболее высокий показатель для *image* (соответствует количеству графических объектов) у сайта *IAC*. По показателю *application* (количество «приложений») лидерство у сайта *ALLEA*.

В табл. 1 даны оценки объемов только тех частей сайтов, которые доступны для входа без аутентификации пользователей, а также без «личных кабинетов», требующих ввода логинов и паролей. Самый большой размер сайтов у *IAC* и *Euro-CASE*. Наименьший размер, оцененный внешними программными средствами, у *EASAC*.

Для ряда сайтов (прежде всего *Euro-CASE*, *CAETS*, *IAC*) оценки средней продолжительности работы пользователей за один визит (колонка *AVD* в табл. 1) весьма высокие. К возможным причинам можно отнести большие объемы полезной информации на этих сайтах и значительное количество их посещений «корпоративными пользователями», использующими материалы сайтов при выполнении служебных работ (это могут быть и пользователи из организаций – владельцы сайтов).

Из данных табл. 2. можно сделать такие выводы.

1. По количеству входящих ссылок значительно преобладает сайт *I.H.R. Network*. У него также есть наибольшее по сравнению с другими сайтами количество внутренних ссылок. Однако, по объему он находится только на третьем месте – в табл. 1 этот показатель приведен в предпоследней колонке. **2.** Количество исходящих ссылок на всех сайтах, отраженных в табл. 2, невелико. **3.** По количеству уникальных посетителей и просмотров за месяц значительно лидирует сайт *I.H.R. Network*. **4.** В тоже время количество документов, обнаруживаемых на нем поисковой системой *Scholar Google*, сравнительно невелико. Последний показатель (количество обнаруживаемых документов) у сайта *EASE* многократно больше, чем у других сайтов.

ОБЩАЯ ХАРАКТЕРИСТИКА УСЛОВИЙ ДЕЯТЕЛЬНОСТИ АКАДЕМИЙ НАУК ОТДЕЛЬНЫХ СТРАН ЗАПАДНОЙ ЕВРОПЫ, ОСОБЕННОСТИ ИХ САЙТОВ

В существующих публикациях рассматриваются как общие вопросы «Европейской интеграции в сфере науки» (например, [9, 10, 16, 24, 25]), так и тематика, относящаяся к России [26] и отдельным европейским странам. Упомянем, в частности, следующие работы: в отношении Германии, Швейцарии, Австрии – [8]; опыт Германии и Бельгии сравнивается в [27]; в [19] из Европейских стран рассматривается опыт Нидерландов; вопросы развития научного сотрудничества России и Германии отражены в [8]; особенности национальной научно-технической политики проанализированы для Франции в [29]; для Испании – в [30]; для Италии – в [31]; для Велико-

британии – в [32]; для скандинавских стран – в [33]. Таким образом, в русскоязычной научной периодике в той или иной степени рассмотрены вопросы управления наукой во всех крупных западноевропейских странах. Однако тематика, связанная с деятельностью академий наук в этих странах, отражена достаточно фрагментарно.

Перечислим характерные особенности для лидирующих в научном отношении стран Западной Европы.

1. Традиционно важная (а иногда и ведущая) роль академий наук в создании и управлении распространением национальной научно-технической информации – в отличие от ряда стран американского континента [2], где ключевую роль в этих процессах обычно играют университеты.

2. Во многих случаях издание академиями наук не только национальных, но и «интернациональных» научных журналов, в которых велика доля иностранных членов редколлегий и статей из-за рубежа, а также совместных работ авторов из разных стран. Такие журналы как правило публикуются на английском языке, а статьи нередко отражают результаты исследований/разработок, проводимых в международных исследовательских центрах или интернациональными рабочими группами – при этом исследователи в процессе работы часто взаимодействуют друг с другом через Интернет.

3. Ряд научных журналов в странах Западной Европы издается на национальных языках – преимущественно страноведческого, социологического, лингвистического характера и др.

4. Развитые связи национальных академий наук стран Западной Европы с вузами и иными типами организаций – по некоторым направлениям научной деятельности.

5. Во многих странах Западной Европы академии наук являются непосредственными организаторами важнейших научных мероприятий национального и международного характера.

Как правило, сайты академий наук стран Западной Европы размещены в «национальных доменных зонах» и соответствуют «доменам 2-го иерархического уровня». Подчиненные академиям наук организации (исследовательские институты, центры и пр.) могут иметь группы страниц на интернет-ресурсах АН и/или свои отдельные сайты с доменными именами 2-го уровня. Во многих случаях на сайтах АН стран Западной Европы и/или подчиненных им организаций размещаются гиперссылки на сайты научных журналов. Такие сайты часто являются отдельными интернет-ресурсами по отношению к сайтам АН или организаций-издателей.

Основные функции, выполняемые сайтами рассматриваемых академий наук.

1. Представление информации о самих АН, их членах (действительных членах, членах корреспондентах, зарубежных членах и пр.) – обычно с разбивкой по тематическим направлениям деятельности. Во многих случаях по гиперссылкам на сайтах АН можно перейти на личные «интернет-страницы» членов академий. Однако их личных адресов электронной почты на таких страницах, как правило, нет. Часто на сайтах АН приводится также ретроспективная ин-

формация о предыдущих президентах АН; о членах и членах-корреспондентах АН за последние годы или даже начиная с момента основания АН.

2. Агрегация научной и научно-технической информации, включая ту, которая создана в рамках деятельности самих АН и подчиненных им организаций.

3. Создание условий для нахождения (обнаружения) этой информации поисковыми системами Интернета на национальных и английском языках; штатными поисковыми системами сайтов АН.

4. Обеспечение для интернет-пользователей дистанционной доступности размещенной информации, включая научно-техническую, отчетов о деятельности АН, их комитетов и пр.

5. Поддержка «связности» научно-информационного пространства за счет использования входящих и исходящих гиперссылок между интернет-сайтами, внутренних гиперссылок на сайтах.

6. Формирование «информационно-коммуникационных площадок» для установления/поддержания научных контактов между отдельными исследователями и научными организациями, в т.ч. из разных стран. С этой целью используется, в частности, формирование «рабочих групп» по конкретным направлениям исследований и публикаций, подготовке научных мероприятий, выпуску научных изданий и пр.

7. Размещение информации о присуждении отдельным национальным ученым «научных наград», премий, медалей и пр., что может быть важным средством стимулирования личной научной активности исследователей, оценки результативности их научной деятельности. Отметим, что в ряде случаев используются «номинации» наград, специально предназначенных для молодых исследователей.

8. В некоторых случаях на сайтах АН размещается информация и по грантам.

При создании и эксплуатации сайтов академий наук, анализируемых в данной статье, используются современные технологии. Однако решения сайтов по интерфейсам интернет-страниц, расположению на них функциональных элементов, графических объектов и пр. не унифицированы, что несколько затрудняет работу для пользователей.

ХАРАКТЕРИСТИКА АКАДЕМИЙ НАУК И ИХ САЙТОВ В СТРАНАХ ЗАПАДНОЙ ЕВРОПЫ

Рассмотрим (в алфавитном порядке) сайты академий наук только наиболее крупных стран Западной Европы.

Австрия (8 774 тыс. чел.). Сайт Австрийской Академии наук (*ÖAW – Österreichische Akademie der Wissenschaft*) имеет версии на двух языках: немецком (<https://www.oeaw.ac.at/>) и английском (<https://www.oeaw.ac.at/en/>). На стартовой странице (СС) сайта основное меню организовано в виде верхнего горизонтального меню-ленты (МЛ). Наиболее важные вкладки в англоязычной версии: (1) *Research Institute*, в том числе содержащий подпункты *Library*, *Archive of the OEAW*. (2) *Events Communications*. Содержит, в частности, сведения о различных лекциях, а также о возможности «подписки» на *OEAW newsletter* – только на немецком языке. (3) *Academy of Sciences Press*. Включа-

ет следующие подпункты: (а) *Institutional repository of the OEAW* – с возможностью перехода на веб-портал репозитория (<http://epub.oeaw.ac.at/epub>). На его СС размещена развитая поисковая система по отдельным научным публикациям; (б) *Open Access* – с возможностью перехода на портал открытых публикаций (<http://www.austriaca.at/oa/>). На нем через подпункт «*OA Publications*» обеспечивается открытый доступ к различным материалам: книгам, журналам, отдельным публикациям, рабочим материалам, базам данных и пр. Кроме того, важен подпункт «*News and Events*», который включает, в том числе, информацию по некоторым опубликованным научным материалам.

Бельгия (8 774 тыс. чел.). Королевская Академия наук и искусств Бельгии (*RASAB – The Royal Academies for Science and the Arts of Belgium*) имеет сайты на трех языках: французском (<http://www.rasab.be/index.php/fr/>); английском (<http://www.rasab.be/index.php/en/>); голландском (<http://www.rasab.be/index.php/nl/>). Верхнее горизонтальное меню-лента содержит следующие пункты (на английском языке): *RASAB* (включает подпункты *Annual reports*, *News* и др.); *National Committees* (включает подпункты *The Committees*, *International Unions*); *International Relations* – содержит гиперссылки на сайты ряда организаций-контрагентов *RSAB*, включая следующие (см. выше) *ALLEA*; *EASAC*; *ICSU*; *UAI*; *I.H.R. Network* и др.

Великобритания (67 809 тыс. чел.). Роль академии наук в этой стране выполняет Лондонское Королевское Общество (по развитию знаний о природе), созданное в 1660 г. Оно является самоуправляющейся частной организацией, но использует субвенции правительства страны. Насчитывает около 1450 национальных и иностранных членов, в том числе 80 лауреатов Нобелевской премии. Лондонское Королевское Общество издает научные журналы, присуждает престижные научные премии. Сайт <https://royalsociety.org/>. На стартовой странице есть два горизонтальных меню-ленты. Для первого отметим пункты «*Contact Us*», а также средство поиска по сайту.

Во втором меню-ленте важными вкладками (их содержимое в виде наборов пунктов открывается при наведении курсора на вкладку без щелчка на ней) для нашей статьи являются: (1) *Fellows* – информация о членах общества с указанием годов их избрания, с возможностями перехода на «личные странички», ознакомления с научными специализациями, биографиями и пр. (2) *Events* – информация о различных научных мероприятиях, включая научные выставки. (3) *Grants, Schemes & Awards* – включает, в частности, сведения о двух типах грантов. (4) *Journals*. Для этой вкладки отметим пункты: *Authors* – это подробная информация для авторов публикаций; *Publishing Times* – указывается длительность этапов рассмотрения работ и их опубликования для шести категорий научных журналов; *Free Content* – сведения о возможности открытого доступа к текстам статей; *Ethics and policies* – подробно структурированная информация по правилам публикационной этики; *Publishing metrics* – отражены основные наукометрические показатели изданий АН, в том числе для журналов по «Биологии» и «Мультидисциплинарным наукам».

(5) *Collections* – представлены коллекции материалов, находящихся в открытом и платном доступе; на странице размещена отдельная поисковая система по материалам «коллекций», в т.ч. находящихся в архивах.

Упомянем также Британскую академию – Национальную академию для гуманитарных и социальных наук, объединяющую более 800 учёных. Она является самоуправляемой и независимой организацией. Сайт (<https://www.britac.ac.uk/>) имеет два верхних меню-ленты, пункты которых дублируются внизу стартовой страницы. В середине СС по высоте располагается подробный тематический рубрикатор, а также информация о предстоящих событиях и новостях. Предоставляется возможность «подписки» на «newsletter». Во втором МЛ отметим, в частности, пункты. (1) *Fellows* – только информация о сотрудничестве с примерно 1400 национальными и международными академиями в области гуманитарных и социальных наук. (2) *Funding* – информация о видах предоставляемых грантов и условиях их получения. (3) *Search* – поисковая система со свободным вводом запросов и возможностью фильтрации по укрупненным тематическим направлениям исследований.

Германия (82 850 тыс. чел.). Союз немецких академий наук – это, по существу, «зонтичная» организация, объединяющая восемь академий наук, включая и АН по гуманитарным исследованиям. Согласно сведениям из Википедии «*Академии наук выступают в роли посредника между наукой и обществом, а также способствуют международному научному сотрудничеству. Финансирование немецких Академий наук осуществляется из бюджетных средств федеральных земель*». Таким образом, в силу федеративного устройства в ФРГ отсутствует единая (национальная) академия наук, а действующие имеют «местное финансирование». Сайты Союза немецких АН – <http://www.akademienunion.de/> (на немецком) и <http://www.akademienunion.de/en/> (на английском). На англоязычной версии стартовой страницы сайта размещена карта Германии с указанием мест расположения отдельных АН – членов Союза. При наведении курсора на каждую из иконок на этой карте появляются всплывающие «расшифровки» названий АН. При щелчке левой кнопкой «мыши» по каждой из иконок открывается более подробная информация о соответствующих АН, в том числе содержащая гиперссылки для перехода на интернет-страницы сайтов отдельных академий наук, входящих в Союз. Интерфейсы этих страниц унифицированы.

На сайте «Союза немецких АН» основное меню размещено в нижней части стартовой страницы и содержит четыре колонки. Две последние из них – контактная информация, включающая гиперссылки. Отметим, в частности, следующие пункты меню. (1) Общая информация о направлениях исследовательской деятельности. (2) *Working groups* – фактически размещены сведения по направлениям «Информационные технологии в гуманитарной сфере». При этом ряд указанных на странице сайта <https://www.akademienunion.de/en/working-groups/working-group-on-ehumanities/> направлений прямо относится к публикационной деятельности, технологиям распространения научно-технической информации, «интерфейсам взаимодействия»

с пользователями и пр. (4) *Publications* – доступен аннотированный перечень изданий организации за разные годы на немецком языке. (5) *Events* – обзор «событий» на немецком языке. (6) *Data privacy* – информация представлена только на немецком языке.

В Германии помимо, указанных академий наук, действуют многочисленные научные общества и научно-промышленные объединения, включающие исследовательские организации.

Греция (10 847 тыс. чел.). Афинская Академия наук является наиболее крупной научно-исследовательской организацией страны. Номинально её деятельностью руководит Министерство образования и религии Греции. Стартовая страница сайта Афинской АН имеет версии на английском (<http://www.academy-ofathens.gr/en>) и греческом (<http://www.academy-ofathens.gr/el>) языках. В верхнем меню-ленте на стартовой странице сайта для вкладки «*Academy*» отметим, в частности, персональные сведения о членах академии. Они представлены по трем «секциям», причем для каждой из них используется разбивка на три группы: члены, иностранные члены, члены-корреспонденты. Во всех случаях указываются годы рождения и избрания в члены Академии, научные специализации. Однако контактная информация по членам АН не приводится. Для вкладки «*Publications*» в пункте «*Printed*», подпункте «*Proceedings*» представлен список «Трудов Академии» на греческом языке, но без возможности ознакомления даже с их содержанием. Во вкладке «*Library*» есть только общие сведения, причем 14.10.2018 г. информация была размещена не на всех открывавшихся страницах. При использовании вкладки «*Digital Academy*» с единственным пунктом «*Digital Collection*» открывается страница, в нижней части которой размещены гиперссылки на 12 «Порталов репозиторий», включая «Центральный». Собственно доступ к информации для «Центрального» (и большинства других порталов) предусмотрен через страницы поисковых систем, открывающиеся через пункты «*Items*» во вкладке третьего верхнего меню-ленты «*Services*». Альтернативный вариант доступа к поисковой системе – через объекты «*Items List*» в центрах страниц для порталов. Особенность поисковых систем – выдается информация по «истории поиска», т.е. для ранее выполненных запросов. Большие возможности отбора информации обеспечивает пункт «*Advanced Search*», расположенный на уровне 2-го МЛ. Для него отметим, в частности, пункт «*Category*», позволяющий отфильтровать источники по их «типам». Помимо «английского» и «греческого» для страниц, относящихся к порталам, предусмотрена возможность использования и французского языка.

Дания (5 668 тыс. чел.). Датская Королевская Академия Наук и Литературы (*Royal Danish Academy of Sciences and Letters*) имеет сайт на английском (<http://www.royalacademy.dk/en>) и датском (<http://www.royalacademy.dk/da>) языках. Меню на стартовой странице сайта организовано в виде двух верхних горизонтальных лент. В первом меню-ленте наиболее полезен пункт «*Contact*» (Контактная информация). Во втором МЛ наиболее важны вкладки. (1) *Events* – в отношении двух типов лекций: «*Royal*

Academy Lectures» и «*Nobel Laurate Lectures*». Однако названия предстоящих лекций есть только в версии сайта на датском языке. (2) *Recent Lectures (in English)*. (3) *Publications* с пунктом *WebShop* (<https://www.youreconomicwebsite.com/det-kongeligedanske-videnskabernes-selskab/lang-EN>). С этой страницы можно перейти на интернет-страницы тематических серий книг для их приобретения через Интернет. (4) *Members* – информация по примерно 500 исследователям сегментирована по первым буквам их фамилий. По каждой «букве» приводятся совместные списки датских и зарубежных (в том числе и из России) ученых, причем по гиперссылкам можно перейти на их персональные странички. Информация на этих страничках очень краткая, но есть личные адреса электронной почты. На СС сайта отметим также иконку (и гиперссылку) на *International Human Rights Network of Academies and Scholarly Societies*. Ссылок на сайты каких-либо научных журналов на сайте Датской Академии наук нет.

Датская королевская академия изящных искусств – это вуз. В нем проводится как обучение искусству, так и некоторые исследования в отношении его теоретических и технических аспектов. Сайт имеет версии на датском (<https://kunstakademiet.dk/>) и английском (<https://kunstakademiet.dk/en>) языках. Обе версии не содержат меню в общепринятом понимании термина; есть только ограниченное количество «новостной» информации.

Испания (46 529 тыс. чел.). Испанская Королевская Академия Наук – ведущая научная организация страны, входит в состав «Института Испании». Сайт АН имеет версии на английском (http://www.rac.es/0/0_1.php?i=2) и испанском (http://www.rac.es/0/0_1.php?i=1) языках. В правом верхнем углу стартовой страницы по кнопке *Links* открывается список «смежных» организаций с гиперссылками на их сайты. Отметим в частности ссылки на следующие типы организаций. (1) Испанские научные организации: *Instituto de Eespaña; Real academia de ingeniería; Real academia de doctores de España*. (2) Международные научные объединения: *The Interacademy pane, EASAC – European Academies Science Advisory Council*. (3) Зарубежные АН, причем представлены ссылки на сайты академий во всех испаноязычных странах.

На стартовой странице сайта в верхнем меню-ленте во вкладке «*Library*» представлены только общие сведения об академической библиотеке. На вкладке «*Publications*» размещены пять основных пунктов, а также правое боковое меню-лента, которое дублирует эти пункты. (1) *Bibliographical and Documentary File* – открывается страница с поисковой системой по публикациям на испанском языке (без англоязычной версии). (2) *Journal of the Royal Spanish Academy of Sciences*. Представлены два «вида» журналов: *Journal General Series; Journal Serie A Matemáticas (RACSAM)*. В обоих случаях через боковое МЛ по пунктам *Indexes* можно получить доступ к оглавлениям отдельных номеров журналов. В случае *RACSAM* из оглавлений номеров (выпусков) также «действуют» гиперссылки, обеспечивающие открытие текстов отдельных статей. (3) *Publications Catalogue* – это, преимущественно, перечень книг,

изданных АН. (4) *Reports* – научные доклады (сообщения) по различным темам, чаще всего достаточно узким. (5) *Electronic Publications* – это каталог электронных публикаций, включая некрологи. Тексты публикаций доступны по соответствующим гиперссылкам.

На стартовой странице в верхнем меню-ленте отметим также вкладку «*Mediateca*», которая имеет собственную поисковую систему по видеозаписям. Ниже прямоугольной панели, используемой для размещения изображений, располагается второе горизонтальное меню-лента, содержащее восемь пунктов, размещенных «в два этажа». Среди них наиболее интересен пункт «*Bibliographical and Documentary File*» – при его выборе открывается страница для задания условий поиска публикаций.

В целом размер стартовой страницы рассматриваемого сайта по вертикали и горизонтали очень невелик – возможно, что это специально сделано для удобства просмотра на смартфонах.

Италия (59 589 тыс. чел.). Национальная академия деи Линчеи была основана в 1603 г. и является старейшей в Европе. Источники финансирования её работы: поступления от Министерства народного просвещения Италии; средства, поступающие от Национального совета исследований Италии; денежные перечисления от ряда частных фондов. Основные направления деятельности академии: координация национальных научных исследований; организация национальных и международных научных мероприятий; поддержка взаимодействия с зарубежными АН, исследовательскими центрами, крупными зарубежными учеными; присуждение научных премий.

Сайт АН имеет версии на итальянском (<https://www.accademiaxl.it>) и английском (<http://www.accademiaxl.it/en/>) языках. Меню на стартовой странице сайта организовано в виде верхней ленты с несколькими пунктами. 1) *Home* – соответствует стартовой странице сайта. На ней содержится новостная информация, а также фотографии нескольких ведущих ученых и видеозаписи их лекций на итальянском языке. 2) *Academy* – открывается вкладка с рядом пунктов, относящихся к истории АН, текущей деятельности академии, составу ее членов и т.п. 3) *Publications* – вкладка имеет четыре пункта. Отметим, в частности следующие. (а) *Rendiconti*: регулярные выпуски по двум сериям «Математика», «Физика и Естественные науки». Однако доступ к опубликованным в них материалам платный. (б) *Scrittii e documenti (Writings and Documents)*. Приведен перечень различных документов (включая некоторые каталоги) на итальянском языке, доступ к «бумажным копиям» которых также является платным. (в) *Off series Volumes*. Перечень опубликованных научных материалов на итальянском языке – доступ к «бумажным копиям» платный. (г) *Thematic Hypertext* – на 18.01.2019 имелось пять гиперссылок на страницы с различными материалами, в том числе по развитию национальной науки в разных сферах. 3) *Library* – вкладка имеет два пункта: *History u Catalogue*. При выборе последнего открывается форма «поискового предписания» с набором полей, определяющих условия селективного выбора изданий. 4) *Archival Funds* –

вкладка имеет четыре тематически специализированных пункта, один из которых относится к квантовой физике. 5) *Prizes* – информация о научных наградах, в том числе носящих имена выдающихся ученых. Специально отметим ежегодные награды за «диссертации на получение степеней, относящиеся к истории биологии, химии, физики». 6) *Seats* – место расположения АН, контактная информация и пр.

Нидерланды (Голландия – 17 217 тыс. чел.). «Королевская Академия наук и искусств» имеет бюджетное финансирование и согласно сведениям из Википедии «... оказывает правительству консультации по вопросам науки, ведёт экспертную деятельность, координирует рецензирование научных статей, проведение научных конференций и форумов ...», присуждает научные награды. Под руководством АН действуют научные институты. Нидерландская АН имеет версии сайта на национальном (<http://www.knaw.nl/nl>) и английском (<http://www.knaw.nl/en>) языках. Для англоязычной версии сайта размер стартовой страницы невелик по ширине и высоте. Основное меню представлено в виде верхнего меню-ленты, содержащего ряд пунктов, по которым открываются вкладки. В центральной части СС размещена оперативная информация по направлениям *News* и *Events*. В нижней части стартовой страницы дублируются пункты верхнего меню-ленты с отображением всех подпунктов.

В верхнем МЛ имеются такие пункты. (1) *News*. Особый интерес представляет подпункт *Publications*. При его выборе открывается страница с различными вариантами «рубрикации» опубликованных материалов. К сожалению, переводы на английский язык носят только частичный характер, а качество переводов с голландского языка на русский с использованием *Google*-переводчика оставляет желать лучшего. (2) *About us* – содержит 10 подпунктов с возможностями перехода по гиперссылкам на соответствующие интернет-страницы. Особо отметим подпункт «*The Young Academy*» – эта структура имеет 50 членов и предназначена для «молодых ученых и обучающихся лиц». (3) *Members*. Включает пункты, относящиеся к информации о членах, членах-корреспондентах, иностранных членах, а также о «прошлых членах» академии наук. По иностранным членам рассматриваемой АН приводится информация о направлениях их деятельности, но не о гражданстве; по гиперссылкам открывается более подробная информация, включая адреса электронной почты. (4) *Institutes* – на странице, открывающейся при выборе этого пункта, размещена информация о деятельности академических институтов, подчиненных АН. (5) *Advisory Work* (консультативная деятельность) – отражена работа «консультативных советов» и «консультативных комитетов», которые готовят соответствующие отчеты. Особо отметим ссылку на страницу <https://www.knaw.nl/en/international/cooperation-with-international-organisations>, на которой перечислены «организации партнеры» АН Нидерландов в виде международных организаций европейского и глобального уровней. (6) *Funds and Awards*. По пункту *Funds* открывается Интернет-страница с достаточно большим перечнем фондов-грантодателей, часть из которых предоставляет гран-

ты и для исследований вне Нидерландов. (7) *International* – имеет три пункта в выпадающем списке: *Cooperation with International Organizations* (отражены *ALLEA, EASAC, IAP, ICSU*); *Scientific Cooperation with China*; *Scientific Cooperation with Indonesia*.

Норвегия (5 405 тыс. чел.). Академия наук и литературы Норвегии – «автономная общенациональная организация». Согласно сведениям из Википедии – ее цель «содействие развитию науки в Норвегии, и презентация норвежской науки за рубежом». При этом академия «имеет 487 мест для норвежских членов и 408 мест для иностранных», а также имеет «почетных членов»; состоит из «двух отделов: естественного и историко-философского». Норвежская Академия наук и литературы является членом Научного консультативного совета европейских академий *EASAC*.

Сайт норвежской АН и литературы (*Norwegian Academy of Science and Letters*) имеет версии на норвежском (<http://www.dnva.no/>) и английском (<http://english.dnva.no/>) языках. Основное меню организовано в виде левого бокового меню-ленты. Из пункта «*Publications*» этого МЛ открывается страница, содержащая только перечень видов изданий академии наук, включая серийные публикации (содержит только одну гиперссылку). На стартовой странице сайта содержатся также гиперссылки на архивы публикаций, отсортированные по годам (2007-2019 гг.). Однако лишь часть этих публикаций можно отнести к научной или научно-технической информации.

В левом меню-ленте стартовой страницы сайта отметим также вкладку «Лекции». По ней открывается страница с гиперссылками на научные лекции, прочитанные на заседаниях АН в 2010-2018 гг.

Португалия (10 374 тыс. чел.). Лиссабонская Академия наук – высшее научное учреждение Португалии. Включает два отделения: естественных и точных наук (7 секций) и гуманитарных наук (7 секций); два основных «института». Сайт Академии имеет версию на португальском языке (<http://www.acad-ciencias.pt/>) – переключатель на английский язык не просматривается. На стартовой странице сайта имеется левое вертикальное меню-лента. В нем есть вкладка «*Livraria*», через которую можно попасть на страницу с каталогом «книжных» публикаций (<http://www.acad-ciencias.pt/academia/livraria-papel>), находящихся в платном доступе. Эта вкладка имеет собственную поисковую систему.

С помощью вкладки в меню-ленте «*Academia digital*» можно открыть дополнительное меню с несколькими пунктами. Особый интерес представляет пункт (в переводе на русский) «Цифровые публикации». При его выборе открывается страница <http://www.acad-ciencias.pt/academia/publicacoes-digitais>, содержащая список публикаций, доступных для бесплатной загрузки пользователям.

Также в меню-ленте через пункт «Информационный бюллетень» (в переводе на русский) можно обеспечить получение на адрес указанной интернет-пользователем электронной почты соответствующих рассылок бюллетеня АН.

Финляндия. Организация под названием «Академия Финляндии» – это правительственный орган

по финансированию научных исследований в стране, который является элементом структуры «Министерства образования и культуры» и занимается распределением средств для выполнения научно-исследовательских проектов. Академия Финляндии имеет сайт на финском (<http://www.aka.fi/>), шведском (<http://www.aka.fi/sv/>) и английском (<http://www.aka.fi/en/>) языках.

В верхнем меню-ленте последнего сайта отметим вкладки *Funding, Review & Funding Decisions*, непосредственно связанные с грантовой деятельностью академии. В правой части СС подробно представлена номенклатура различных «академических программ». С помощью них, в частности, обеспечивается сбор сведений (заявок) о «*намерениях об участии*».

Финский титул «академик» является почетным званием, присуждаемым президентом Финляндии. При этом количество ныне живущих финских граждан, являющихся академиками, является ограниченным, а зарубежных граждан – не ограниченным. В Интернете (https://ru.wikipedia.org/wiki/Академия_Финляндии) указывается, что «*Академия ежегодно распределяет более 260 миллионов евро для проведения научно-исследовательских проектов, над которыми работают более 5 тысяч исследователей*».

В Финляндии действуют также «Финская Академия Наук и Литературы», «Финское общество Наук и Литературы». Они являются финскими национальными «*почётными академиями*». Первая – для учёных, «говорящих» на финском языке, а вторая – на шведском. Отметим также полезную сводку ссылок на «академические организации» Финляндии (раздел «Образование, наука») на сайте «<https://polpred.com/?cat=5&cnt=168>».

Франция (68 860 тыс. чел.). Французская Академия наук – одна из пяти академий, составляющих «Институт Франции» (*Institute de France*). Сайт имеет версии на французском (<http://www.academie-sciences.fr/fr/>) и английском (<http://www.academie-sciences.fr/en/>) языках. Верхнее горизонтальное меню-лента содержит, в частности, следующие вкладки. (1) *Scientific Community* – в ней есть пункт *Colloquia, Conferences and Debates*. (2) *Transmitting Knowledge* – в том числе включает пункты *Public Sessions, Specific Programs*. (3) *International Collaboration*. Для этой вкладки отметим пункты. (а) *International networks* – содержит ссылки на две сети. (а1) *International Scientific Council* – «*наиболее важная неправительственная научная организация в мире. Включает 121 национального члена и 32 международных научных союза*». (а2) *French Committee for International Scientific Unions*. (б) *Interacademic networks* – отражены только *IAP, EASAC, ALLEA*. (в) *Bilateral cooperation* (двухсторонне сотрудничество), включая «*Межакадемические действия*» и секции, относящиеся к отдельным континентам. (4) *Expertise & Advice* – имеются, в основном, пункты по «тематическим комитетам» с перечислением их персонального состава. При этом по гиперссылкам можно перейти на личные странички отдельных исследователей.

Швеция (10 006 тыс. чел.). Шведская Королевская Академия наук имеет сайты на шведском (<http://www.kva.se/sv/startside>) и английском (<http://www.kva.se/en/startside>) языках.

Для начала работы с сайтом требуется обязательно согласиться с политической использованием *cookies*.

На стартовой странице сайта в верхней правой части размещен интерфейс поисковой системы по сайту. Также вверху расположено горизонтальное меню-лента, в котором наиболее важны следующие пункты. (1) *Events* – с собственной поисковой системой, содержащей совокупность «фильтров» для отбора необходимой информации. На этой страничке содержится также средство оформления подписки на автоматическое получение «*новостей о событиях*». (2) *Grants/Funding* – содержит, в том числе, отдельную вкладку для «*обучающихся*». (3) *Published* – открывается страница с собственной поисковой системой по всем документам, которые академия опубликовала на *web*-сайте (отчеты, книги, популярная научная информация, брошюры).

На стартовой странице сайта комитета по Нобелевским премиям <https://www.nobelprize.org/> также декларируется использование обработки *cookies* для улучшения работы с пользователями (последние должны дать на это согласие). В верхней части СС расположено горизонтальное меню-лента. По теме статьи в нем наиболее интересны две вкладки. (1) *Events* – информация о мероприятиях, доступных для свободного посещения, которые имеют высокую социально-экономическую значимость. Большинство таких мероприятий декларируются в качестве «диалогов». (2) *Education Network* – включает пункты: *Nobel Prize Lessons, Nobel Prize Teacher Summit*.

В нижней части СС содержится информация о «востребованности» интернет-ресурса в четырех социальных сетях, а также средство оформления подписки на «*newsletters*», формируемые организацией.

Швейцария (8 236 тыс. чел.). Общий сайт «Швейцарских академий наук и искусств» (*Swiss Academies of Arts and Sciences*) имеет версии на трех языках: немецком (<http://www.swiss-academies.ch/de/index/Aktuell/News.html>); французском (<http://www.swiss-academies.ch/fr/index/Aktuell/News.html>) и английском (<http://www.swiss-academies.ch/en/index/aktuell/news.html>). Отметим, что во всех трех случаях сразу же открываются странички, содержащие новостную информацию, при этом на англоязычной версии сайта часть сообщений показывается на немецком языке.

В левой части стартовой страницы расположено вертикальное меню-лента. Наибольший интерес представляет вкладка *Publications*. При ее выборе открывается ряд пунктов. По пункту «*Swiss Academies Reports*» обеспечивается доступ к интернет-страничке, с которой возможно бесплатное скачивание всех номеров издания за последние годы (в виде *pdf*-файлов). При выборе пункта «*Partners*» открывается ссылка со списком «партнеров».

В правой части СС размещены новостные сообщения швейцарских АН; средство бесплатной подписки на журнал «*Horizon*» (освещающий тематику «научные исследования в Швейцарии»), а также на рассылку новостных сообщений этого издания. Специально отметим возможность перехода на сайт <http://p3.snf.ch/>, отражающий возможности как получения «грантов по дисциплинам», так и международного сотрудничества. Сокращение *P³* означает «*pro-*

jects, people, publications», a snf – swiss national science foundation.

На стартовой странице сайта <http://www.swiss-academies.ch/en/index/Aktuell/News.html> в верхнем меню-ленте при нажатии на кнопку «Links» открывается страничка, содержащая коллекцию гиперссылок на образовательные и научные организации страны, включая «Национальные институты» (в т.ч. университеты), *Swiss National Science Foundation (SNF)* и «швейцарские академии»: *Swiss Academy of Sciences (SCNAT)*; *Swiss Academy of Humanities and Social Sciences (SAHS)*;

Swiss Academy of Medical Sciences (SAMS); *Swiss Academy of Engineering Sciences (SATW)*. Приводятся также гиперссылки на сайты международных организаций, в т.ч. европейских, в которых эти «академии» являются членами. Набор ссылок включает почти все указанные ранее международные организации.

Сведения по вебметрическим показателям сайтов для отдельных академий наук, рассмотренных в данном разделе, представлены в табл. 3 и 4. Для пунктов 2в и 9а в этих таблицах обозначение «нид.» соответствует нидерландскому (голландскому) языку.

Таблица 3

Вебметрические показатели для сайтов отдельных АН в странах Западной Европы (первая часть)

№	Название организации или группы организаций, которой принадлежит сайт; адрес сайта	АТ, сек.	Count, URLs			Size, Mb	AVD, чч:мм:сс
			text/html	image	application		
1а	Австрийская АН (https://www.oeaw.ac.at/) (нем.)	9,1	11 233	8 097	18 123	5 677,5	0:03:39
1б	Австрийская АН (https://www.oeaw.ac.at/en/) (анг.)	8,5	6344	0	5321	270	0:02:15
2а	Королевская АН и искусств Бельгии (http://www.rasab.be/index.php/fr/) (фра.)	3,6	682	0	2	9,3	0:00:43
2б	Королевская АН и искусств Бельгии (http://www.rasab.be/index.php/en/) (анг.)	2,0	679	0	2	9,4	0:00:45
2в	Королевская АН и искусств Бельгии (http://www.rasab.be/index.php/nl/) (нид.)	2,2	682	0	2	9,7	0:00:52
3	Лондонское Королевское Общество (https://royalsociety.org/)	3,8	533	46	47	235	0:00:35
4а	Союз немецких АН (https://www.akademienunion.de/) (нем.)	6,8	554	1 011	390	522,6	0:01:29
4б	Союз немецких АН (https://www.akademienunion.de/en/) (анг.)	6,5	43	0	0	0,9	0:00:08
5а	Афинская АН (http://www.academyofathens.gr/en/) (анг.)	11,8	2 828	325	509	558,4	0:14:42
5б	Афинская АН (http://www.academyofathens.gr/el/) (гре.)	8,5	2 828	325	509	558,4	0:13:54
6а	Датская Королевская АН и Литературы (http://www.royalacademy.dk/en/) (анг.)	6,0	2 540	2 374	84	3 720,5	0:06:46
6б	Датская Королевская АН и Литературы (http://www.royalacademy.dk/da/) (дат.)	19,8	1 734	2 139	148	3 188,4	0:04:54
7а	Испанская Королевская АН (http://www.rac.es/0/0_1.php?i=2) (анг.)	1,0	3	2	0	0,02	0:00:01
7б	Испанская Королевская АН (http://www.rac.es/0/0_1.php?i=1) (исп.)	0,8	3	2	0	0,02	0:00:01
8	Итальянская НА деи Линчеи (http://www.accademiaxl.it/en/)	19,3	558	0	282	14,2	0:10:21
9а	Королевская АН и искусств Голландии (https://www.knaw.nl/nl/) (нид.)	8,7	4 919	2 280	826	2 977,5	0:10:34
9б	Королевская АН и искусств Голландии (https://www.knaw.nl/en/) (анг.)	8,5	3 321	1 564	515	1 526,3	0:07:05
10а	АН и литературы Норвегии (http://www.dnva.no/) (нор.)	1,1	н	н	н	н	н
10б	АН и литературы Норвегии (http://english.dnva.no/) (анг.)	1,1	н	н	н	н	н
11	Лиссабонская АН (http://www.acad-ciencias.pt/)	3,3	489	697	260	309,4	0:00:59
12а	Академия Финляндии (http://www.aka.fi/) (фин.)	3,3	4 393	1 416	1 832	2 557,2	0:04:39
12б	Академия Финляндии (http://www.aka.fi/sv/) (шве.)	3,5	7 242	2 956	3 121	4 001,5	0:08:37

№	Название организации или группы организаций, которой принадлежит сайт; адрес сайта	АТ, сек.	Count, URLs			Size, Mb	AVD, чч:мм:сс
			text/html	image	application		
12в	Академия Финляндии (http://www.aka.fi/en) (анг.)	3,6	7 241	2 956	3 121	4 001,5	0:08:07
13а	Французская АН (https://www.academie-sciences.fr/fr/) (фра.)	13,2	17 707	0	2 581	843,3	0:26:44
13б	Французская АН (https://www.academie-sciences.fr/en/) (анг.)	10,0	19 156	0	2 059	887,5	0:00:13
14а	Шведская Королевская АН (https://www.kva.se/sv/startside) (шве.)	10,5	1	0	0	0,006	0:00:21
14б	Шведская Королевская АН (https://www.kva.se/en/startside) (анг.)	10,6	1	0	0	0,006	0:00:31
15	Комитет по Нобелевским премиям (https://www.nobelprize.org)	6,7	478	112	62	71,2	0:01:36
16а	Общий сайт швейцарских АН и искусств (http://www.swiss-academies.ch/de/index/Aktuell/News.html) (нем.)	11,9	1	0	0	0,1	0:00:21
16б	Общий сайт швейцарских АН и искусств (http://www.swiss-academies.ch/fr/index/Aktuell/News.html) (фра.)	11,8	8	0	6	2,5	0:00:06
16в	Общий сайт швейцарских АН и искусств (http://www.swiss-academies.ch/en/index/Aktuell/News.html) (анг.)	11,7	8	0	7	2,6	0:00:05

ПРИМЕЧАНИЯ. 1. Символ «н» означает, что невозможно получить информацию в связи с отсутствием на исследуемом сайте средств для сбора необходимых статистических данных. 2. Информация по «Союзу немецких АН» и «Академии Финляндии» размещена на сайтах «вышестоящих организаций». При этом доменное имя не изменяется, а полный URL не зависит от содержимого. При таких условиях в силу особенностей использованных авторами настоящей статьи программных средств определить (выделить) показатели, относящиеся только к указанным «Союзу» и «Академии» не удастся. Поэтому для этих двух организаций показатели в таблице 3 приводятся для сайтов в целом.

Таблица 4

Вебометрические показатели для сайтов АН в странах Западной Европы (вторая часть)

№	Название организации или группы организаций, которой принадлежит сайт	Абсолютные количества ссылок				КУнПос. за месяц	КПросм.	Scholar Google (SG)
		Входящие	Внутр.	Исходящ.				
				Ω	Ψ			
1а	Австрийская АН (нем.)	6 230	46 436	13	175	229 809	1 585 217	15 600
1б	Австрийская АН (анг.)			5	171			
2а	Королевская АН и искусств Бельгии (франц.)	29	277	4	26	н	н	35
2б	Королевская АН и искусств Бельгии (анг.)			3	21			
2в	Королевская АН и искусств Бельгии (нид.)			4	26			
3	Лондонское Королевское Общество	14 033	60 939	14	177	370 916	896 052	2 130 000
4а	Союз немецких АН (нем.)	619	1 065	6	124	3 600	14 400	20 300
4б	Союз немецких АН (анг.)			1	93			
5а	Афинская АН (анг.)	896	1 732	3	94	н	н	658 000
5б	Афинская АН (греч.)			4	148			
6а	Датская Королевская АН и Литературы (анг.)	151	657	2	143	н	н	41 900
6б	Датская Королевская АН и Литературы (дат.)			2	151			
7а	Испанская Королевская АН (анг.)	665	1 419	н	н	н	н	188 000
7б	Испанская Королевская АН (исп.)			4	25			

№	Название организации или группы организаций, которой принадлежит сайт	Абсолютные количества ссылок				КУнПос. за месяц	КПросм.	Scholar Google (SG)
		Входящие	Внутр.	Исходящ.				
				Ω	Ψ			
8	Итальянская НА деи Линчеи	80	287	1	105	н	н	41 100
9а	Королевская АН и искусств Голландии (нид.)	4 111	13 128	4	115	389 747	1 239 098	164 000
9б	Королевская АН и искусств Голландии (анг.)			4	114	н	н	
10а	АН и литературы Норвегии (норвеж.)	190	563	7	63	4 791	19 170	60 900
10б	АН и литературы Норвегии (анг.)	58	135	6	42			
11	Лиссабонская АН	126	1 863	0	122	2 622	10 500	94 600
12а	Академия Финляндии (фин.)	2 152	6 155	18	221	18 570	74 280	1 510 000
12б	Академия Финляндии (швец.)			16	173	н	н	
12в	Академия Финляндии (анг.)			16	207	н	н	
13а	Французская АН (франц.)	3 674	11 076	9	759	60 917	268 159	2 810 000
13б	Французская АН (анг.)			9	264	н	н	
14а	Шведская Королевская АН (швец.)	2 472	4 218	н	н	6 407	25 620	246 000
14б	Шведская Королевская АН (анг.)			н	н	н	н	
15	Комитет по Нобелевским премиям	43 353	179 063	13	116	1 424 682	4 357 123	51 800
16а	Общий сайт швейцарских АН и искусств (нем.)	39	989	0	7	н	н	85 100
16б	Общий сайт швейцарских АН и искусств (франц.)							
16в	Общий сайт швейцарских АН и искусств (анг.)							

ПРИМЕЧАНИЯ. 1. Для столбцов с «исходящими ссылками» символ «н» отражает отсутствие данных, находящихся в открытом доступе. 2. Для столбцов «количество уникальных посетителей» и «количество просмотров сайтов» символ «н» означает отсутствие на сайтах систем сбора соответствующей статистики или ее недоступность для использования внешними программными средствами. 3. Для «Союза немецких АН» и «Академии Финляндии» делается та же оговорка, что и в п. 2 для табл. 3.

Из данных табл. 3 можно сделать следующие выводы. 1. По показателю «среднее время доступа» (*Average Time, AT*) сайты значительно различаются. В тоже время оценить этот показатель при доступе только из Европейских стран не представляется возможным. 2. По соотношению *text/html* лидируют сайты французской АН на английском и французском языках, на третьем месте находится сайт Австрийской АН. 3. По показателям «*image*» и «*application*» абсолютным лидером является Австрийская АН. 4. Лидером по размеру сайта является австрийская АН. Второе и третье места делят сайты «Академии Финляндии» на шведском и английском языках. При этом версия того же сайта на финском языке значительно меньше по объему. 5. Лидером по средней длительности сеанса работы посетителей (колонка *AVD*) является сайт Французской АН на французском языке. В тоже время для англоязычной версии того же сайта этот показатель невелик. 6. Очень небольшие значения *AVD* для некоторых сайтов в табл. 3 могут, вероятно, объясняться тем, что автоматически срабатывает переадресация входящих на сайты пользователей на другие ресурсы (страницы).

Из данных табл. 4. можно сделать такие выводы. 1. По количеству входящих ссылок абсолютным лидером является «Комитет по Нобелевским премиям»,

на втором месте – сайт «Лондонского Королевского Общества». Для оценки долей ссылок на приведенные в таблице сайты изнутри и извне соответствующих стран требуется проведение дополнительных работ. 2. По количеству внутренних ссылок на сайте лидерами являются те же два сайта, что и в п. 1. 3. Количество исходящих ссылок с сайтов, приведенных в табл. 4, на порядок меньше чем входящих. Только у Французской АН на французском языке таких ссылок достаточно много, при этом для англоязычной версии сайта их в три раза меньше. Указанная ситуация (по соотношению числа входящих и исходящих ссылок) является достаточно характерной для сайтов-агрегаторов научно-технической информации. 4. По количеству уникальных посетителей за месяц лидирует сайт «Комитета по Нобелевским премиям» – почти 1,5 млн человек, на втором месте – сайт «Королевской АН и искусств Голландии» (нидерландский язык) и лишь на третьем – сайт «Лондонского Королевского Общества». 5. По общему количеству посещений (просмотров) лидирует сайт «Комитета по Нобелевским премиям», на втором месте – сайт «Австрийской АН» (версия на немецком языке), на третьем – сайт «Лондонского Королевского Общества». 6. По количеству документов, которые «видит» поисковая система *Scholar Google*, лидирует

«Лондонского Королевского Общества», на втором месте – сайт французской АН, на третьем – сайт «Академии Финляндии». Отметим, что оценки посещаемости можно получить только для сайтов, которые добровольно подключены к средствам сбора статистики количества посетителей.

ВЫВОДЫ

1. В странах Западной Европы сайты академий наук (и иных организаций, выполняющих их функции) играют важную роль в агрегации национальной научно-технической информации, обеспечении её доступности на национальном и международном уровнях.

2. Большинство рассмотренных в настоящей статье сайтов имеют версии на национальном и английском языках. В качестве объектов, на которые указывают гиперссылки, размещаемые на российских интернет-ресурсах, целесообразно использовать англоязычные страницы сайтов зарубежных АН. При выборе объектов для размещения таких гиперссылок на российских сайтах необходимо учитывать степень известности в научно-информационном пространстве соответствующих АН, в том числе по отдельным тематическим направлениям деятельности; содержание размещенной на их сайтах информации; наличие связей российских исследователей с академическими организациями в соответствующих странах и пр.

3. Возможности автоматического мониторинга информации, размещенной на сайтах зарубежных академий наук, ограничиваются различиями в номенклатурах интернет-страниц, подходов к размещению на них информации о научных мероприятиях и пр.

4. На рассмотренных нами интернет-ресурсах ссылки на сайты других академий наук встречаются, но относительно редко.

5. Большинство сайтов АН имеют вкладки типа «Publications». Однако во многих случаях это не научные публикации, а отчеты и аналогичные им документы о деятельности самих АН или их подразделений.

6. Сайты многих АН предоставляют возможность подписки на «Новостные бюллетени», но без каких либо возможностей тематической фильтрации рассылаемых сообщений.

7. Некоторая часть материалов научных публикаций, размещенных на сайтах АН, дублируется на иных интернет-ресурсах.

СПИСОК ЛИТЕРАТУРЫ

1. Брумштейн Ю.М., Васильев Н.В. Анализ структуры сайтов Академий наук некоторых зарубежных стран // Научно-техническая информация. Сер. 2. – 2019. – № 1. – С. 16-30; Brumshteyn Yu.M., Vasil'yev N.V. Academies of Sciences Websites in Selected Foreign Countries: Structural Analysis // Automatic Documentation and Mathematical Linguistics. – 2019. – Vol. 53, № 1. – P. 16–31.
2. Брумштейн Ю.М., Васильев Н.В. Академии наук стран американского континента: анализ информационного наполнения сайтов и их вебметрических показателей // Научно-техническая информация. Сер. 2. – 2019. – № 4. – С. 13-24;

- Brumshteyn Yu.M., Vasil'yev N.V. Academies of Sciences in the American Continents Countries: Analysis of Their Information Presence in the Internet Space // Automatic Documentation and Mathematical Linguistics. – 2019. – Vol. 53, № 2. – P. 91–102.
3. Титова Т.П. Развитие международной мобильности научных кадров: опыт Европейского союза // Социология науки и технологий. – 2015. – Т. 6, № 1. – С. 90-97.
4. Комаров С.Ю. Википедия как средство продвижения информационных ресурсов // Библиосфера. – 2012. – № 5. – С. 38-40.
5. Брумштейн Ю.М., Васьковский Е.Ю. Сайты международных ассоциаций организаций, работающих в научно-технической сфере: анализ функциональности, вебметрических показателей, роли в научном информационном пространстве // Научно-техническая информация. Сер. 2. – 2018. – № 7. – С. 1-19; Brumshteyn Yu.M., Vas'kovskii E.Yu. The Websites of International Association Organizations in the Science and Engineering Area: Analysis of Their Functionality, Webometric Ranks, and Role in the Scientific Information Space // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 52, № 4. – P. 157-174.
6. Редькина Н.С. Направления развития инструментов веб-аналитики // Научно-техническая информация. Сер. 2. – 2017. – № 5. – С. 5-10.
7. Скородумов П.В., Холодов А.Ю. Анализ популярности веб-сайта научной организации с помощью различных систем сбора статистических данных // Вопросы территориального развития. – 2016. – № 1(31). – С. 7.
8. Авдонин В.С. Центры изучения интеграции науки в Германии, Швейцарии и Австрии (обзор) // Метод. – 2014. – № 4. – С. 413-423.
9. Балякин А., Мун Д.В. Формирование системы открытой науки в Европейском Союзе // Информация и инновации. – 2017. – № S1. – С. 39-44.
10. Балякин А.А., Мун Д.В. Особенности развития европейской науки: программа Горизонт 2020 // Информация и инновации. – 2017. – № 1-2. – С. 75-78.
11. Белоусова Е.А. Финансирование и результативность исследований и разработок в странах Европейского Союза // Экономические науки. – 2016. – № 134. – С. 163-170.
12. Воронцовская Л.Г. Сравнительный анализ финансирования науки и оплаты труда научных работников в России, Беларуси и других странах // Вестник Прикамского социального института. – 2018. – № 1(79). – С. 62-71.
13. Феоктистова О.А., Фокина Т.В. Особенности планирования и выделения государственных средств на науку за рубежом // Финансы и кредит. – 2015. – № 39(663). – С. 23-40.
14. Лебедев К.В., Леонова Т.Н., Маланичева Н.В., Кольцов А.В. Развитие системы грантового финансирования науки в Европейском Союзе // Инноватика и экспертиза: научные труды. – 2013. – № 2(11). – С. 191-199.

15. Кабанов Ю.А. Информационное пространство как новое (гео)политическое пространство: роль и место государств // Сравнительная политика. – 2014. – Т. 5, № 4(16-17). – С. 54-59.
16. Пипия Л.К., Дорогокупец В.С. Изменение подходов к управлению наукой в индустриально развитых странах // Наука за рубежом. – 2017. – № 58. – С. 1-35.
17. Семенюк Э.П. Глобализация информационного пространства и человечество // Научно-техническая информация. Сер. 1. – 2015. – № 1. – С. 1-13; Semenyuk E.P. Globalization of the Information Area and its Significance for Mankind // Scientific and Technical Information Processing. – 2015. – Vol. 42, № 1. – P. 1-12.
18. Сютюренко О.В., Гиляревский Р.С. Задачи информационного обеспечения инновационного развития экономики и роль инжиниринга // Научно-техническая информация. Сер. 1. – 2017. – № 5. – С. 1-14.
19. Биктимиров М.Р., Глебский В.Л., Долгов Б.В., Поликарпов С.А. Использование информационных технологий и инфраструктур для агрегации научной информации. Опыт Канады, Нидерландов, США // Моделирование и анализ информационных систем. – 2015. – Т. 22, № 1. – С. 114-126.
20. Гуров А.Н., Гончарова Ю.Г., Бубякин Г.Б. Открытый доступ к научным знаниям: состояние, проблемы, перспективы развития // Научно-техническая информация. Сер. 1. – 2016. – № 4. – С. 10-16; Gurov A.N., Goncharova Yu.G., Bubyakin G.B. Open Access to Scientific Knowledge: Its State, Problems, and Prospects of Development // Scientific and Technical Information Processing. – 2016. – Vol. 43, № 2. – P. 88-94.
21. Клименко О.А. Модели представления академического веб-пространства // Информационные и математические технологии в науке и управлении. – 2016. – № 2. – С. 103-110.
22. Лопатина Н.В. Информационная инфраструктура общества: проблемы изучения и управления // Научно-техническая информация. Сер. 1. – 2016. – № 5 – С.1-4.
23. Aguillo I. F. Measuring the institutions' footprint in the web // Library Hi Tech. – 2009. – Vol. 27(4). – P. 540-556.
24. Власенко А.С. Модели государственного управления научно-технической и инновационной политикой: зарубежный опыт // Государственное и муниципальное управление в XXI веке: теория, методология, практика. – 2013. – № 10. – С. 13-17.
25. Зорников И.Н. Россия – Европейский Союз: на пути к общему научному пространству // Вестник Воронежского государственного университета. Серия: Проблемы высшего образования. – 2009. – № 2. – С. 140-146.
26. Михалева О.М., Матюшкина И.А., Игольникова И.В. Тенденции развития науки в России и за рубежом // Международный научный журнал. – 2015. – № 3. – С. 22-26.
27. Овчинникова О.А., Босердт Н.Ю., Белюсова Е.Н. Сравнительный анализ Германии и Бельгии по уровню развития науки // В сб.: Лучшая студенческая статья 2018. Сборник статей XV Международного научно-исследовательского конкурса / отв. ред. Г.Ю. Гуляев. – г. Пенза: Международный центр Научного Сотрудничества «Наука и Просвещение», 2018. – С. 110-115.
28. Зорников И.Н., Плященко Т.Е. Научно-технологическое и образовательное сотрудничество России и Германии: опыт и перспективы // Вестник Воронежского государственного университета. Серия: Проблемы высшего образования. – 2011. – № 2. – С. 155-159.
29. Мехдиев Э.Т., Зиновьева Е.С. Научная политика Франции: существующие проблемы и пути их решения // Евразийский юридический журнал. – 2016. – № 10(101). – С. 263-266.
30. Новикова В.М. Научная политика в Испании: государственные программы и восприятие науки в обществе // Вестник МГИМО Университета. – 2016. – № 5(50). – С. 128-135.
31. Райнхардт Р.О., Маслова Е.А. Государственная политика Италии в области науки и образования // Вестник Брянского государственного университета. – 2017. – № 1(31). – С. 72-76.
32. Серёгин Р.С. Инновационная и научно-технологическая политика Великобритании // Новая наука: Современное состояние и пути развития. – 2017. – № 1-1. – С. 258-261.
33. Талагаева Д.А. Скандинавские страны: научная политика как баланс национальных приоритетов и задач общеевропейского развития // Вестник Брянского государственного университета. – 2018. – № 2(36). – С. 203-216.

Материал поступил в редакцию 20.06.19.

Сведения об авторах

БРУМШТЕЙН Юрий Моисеевич – кандидат технических наук, доцент, доцент Астраханского государственного университета
e-mail: brum2003@mail.ru
ORCID <http://orcid.org/0000-0002-0016-7295>

ВАСИЛЬЕВ Никита Вячеславович – программист отдела *Internet*-технологий Астраханского государственного университета
e-mail: nikivas@mail.ru
ORCID <http://orcid.org/0000-0002-4937-3305>

СПРАВОЧНО-ИНФОРМАЦИОННЫЙ РАЗДЕЛ

УДК 81'322'374:025.48

С.В. Лесников

Виды разметок текстовых корпусов русского языка

Рассматриваются виды разметок текстовых корпусов русского языка для формализованной разметки цифрового корпуса академических толковых словарей русского языка в форме информационно-поискового гипертекстового сводного тезауруса. Цифровой корпус разрабатывается в виде гипертекстовой системы оцифрованных лексикографических материалов русского языка в нелинейной форме с учетом реляционных, иерархических и сетевых парадигматических связей посредством реализации синтагматических связей в интерактивном режиме на персональном компьютере. Это позволит на основе цифровых технологий посредством соответствующей классификации и систематизации, дигитализации и ретроцифровки справочников и энциклопедий русского языка структурировать, систематизировать и объединить лексикографические материалы в единый словарный свод (корпус, фонд, кортеж, множество, набор, собрание, совокупность, подборка, библиотека, ресурс), обеспечив тем самым их оперативный ввод в научный оборот с целью оптимизации научных исследований в современной лексикографии.

Ключевые слова: база данных, гизаурис, гипертекстовый тезаурус, метаязык, искусственный интеллект, корпус, лексикография, лингвистика, поиск, русский язык, словарь, справочник, текст, энциклопедия

Развитие русского языка предполагает фиксацию, описание и репрезентацию академических толковых словарей и других лексикографических материалов.

Задача формализации лексикографических материалов на современном этапе развития современной прикладной, структурной, математической, компьютерной лингвистики остается чрезвычайно актуальной.

В работах Ю.Д. Апресяна, А.Н. Баранова, В.П. Захарова, Ю.Н. Караулова, И.С. Кудашева, С.Е. Никитиной, В.В. Рыкова представлены и апробированы на практике базовые виды разметок текстовых корпусов как основа для формализованной разметки цифрового корпуса академических толковых словарей русского языка [1–11].

Сегодня наряду с национальными корпусами (Мангеймский корпус немецкого языка (*COSMAS corpora* или *DeReKo* – <https://cosmas2.ids-mannheim.de/cosmas2-web/>), Национальный корпус шотландского языка – www.scottishcorpus.ac.uk/), Британский национальный корпус (*British National Corpus* – www.natcorp.ox.ac.uk/), Корпус современного американского английского (*Corpus of Contemporary American English, COCA* – <http://corpus.byu.edu/time/>), Коранический корпус арабского языка (*Tresor de la langue francaise* – <http://corpus.quran.com/>), Корпус Хамшахри – <http://dbrg.ut.ac.ir> /Hamshahri/, Чешский национальный

корпус – <http://ucnk.ff.cuni.cz/>, Корпус эсперанто, Эсперантский текстовый корпус (*Tekstaro de Esperanto* – <http://tekstaro.com/>), Корпус японской разговорной речи – www.kokken.go.jp/katsudo/seika/corpus и др.) можно констатировать создание следующих корпусов текстов русского языка:

а) Национальный корпус русского языка (НКРЯ) – онлайновый корпус русских текстов (www.ruscorgora.ru/). Открыт 29 апреля 2004 г.;

б) Корпус русского литературного языка (КРЛЯ) – универсальный инструмент, обеспечивающий более эффективную работу всех, кто связан по роду своей деятельности с современным русским литературным языком или просто интересуется состоянием и функционированием этого языка (www.nagusco.ru/);

в) Машинный фонд русского языка (<http://cfl.ruslang.ru/>). Работы по его созданию были начаты в 1985 г. по инициативе академика А.П. Ершова, после проведенной в 1983 г. специальной всесоюзной конференции. Тогда же был создан отдел Машинного фонда русского языка в Институте русского языка РАН [12];

г) Проект по созданию размеченного корпуса текстов с лингвистической разметкой (<http://open-corpora.org>).

Корпус – это информационно-справочная поисковая система, основанная на совокупности размечен-

ных текстов в цифровой форме. Оцифрованные лексикографические материалы, по факту создания в гипертекстовой форме, имеют нелинейную структуру. Основной характеристикой цифрового, электронного корпуса является лингвистическая разметка, которая и отличает корпус от простых электронных коллекций текстов или электронных библиотек (например, библиотека Максима Мошкова, Русская виртуальная библиотека, Фундаментальная электронная библиотека, Научное наследие России, Альдебаран, Викитека, КиберЛенинка, Либрусек, ЛитРес, Наука права, Национальная электронная библиотека, Национальная электронная детская библиотека, Президентская библиотека имени Б.Н. Ельцина; Проекты «Авалон», «Гутенберг», «Розетта», «Рунеберг»; Российская государственная библиотека, Российская национальная библиотека, Руниверс, Русская фантастика, Тибетская и гималайская библиотека, Флибуста).

Рассмотрим основные виды разметок текстовых корпусов.

Разметка (*tagging, annotation*) корпуса лексикографических материалов заключается в приписывании источникам и их компонентам специальных лингвистических и экстралингвистических тэгов (токенов). Именно разметка позволяет идентифицировать первоисточники и обеспечивает семантический, осмысленный, расширенный поиск по различным тегам, параметрам и критериям.

Различают следующие виды разметок:

автоматическая – это лексикографическая база данных, сконструированная в процессе использования специальных программ в интерактивном режиме;

анафорическая – фиксирует референтные связи, например, местоименные (как правило, антецедент, в роли которого обычно выступает именное словосочетание, берется в пронумерованные скобки, а рядом с местоимением-заместителем ставится особый знак, отсылающий к антецеденту с соответствующим номером);

внешняя (интеллектуальная) – включает библиографические данные; типологические, идеографические, тематические и социологические характеристики;

грамматическая (парсер (*parser* – разборщик, анализатор) – программа, реализующая процесс сбора необходимой информации для грамматической разметки из заданных источников; морфологическая, частеречная);

дискурсная – служит для обозначения коммуникативных актов, оговорок, пауз, повторов, реплик, хезитаций;

жанровая (например, описание, песня, поговорка, пословица, рассказ-повествование, рассуждение, сказка);

издательская (оригинал-макет);

имплицитная (служебная);

интеллектуальная (внешняя);

лингвистическая (лексикографическая – форма выражения не учитывается) – включает несколько видов: анафорическая, грамматическая, морфологическая, семантическая, синтаксическая, просодическая и др.;

метаразметка (экстралингвистическая разметка, паспортизация источников);

метатекстовая – включает параметры, характеризующие источник в целом (например, автор, редактор, составитель (фамилия, имя, отчество); название (заголовок, подзаголовок); аббревиатура (сокращение, используемое для цитирования и иллюстраций); хронология (даты создания, опубликования, переиздания); род и жанр);

морфологическая (*part-of-speech tagging*; грамматическая, частеречная; парсер) – включает не только признак части речи, но и признаки грамматических категорий, собственных данной части речи, т.е. в том или ином структурированном виде включаются лемма, признак части речи и признаки грамматических категорий (граммемы – падеж, переходность, род, одушевленность, число). Морфологический анализ корпуса текстов может заключаться: в токенизации (разбиении на орфографические слова, элементарные знаки – токены); в сегментации (на предложения, клаузулы и др. виды сегментов); в лемматизации (нормализация словоформ, т.е. сведение различных словоформ к некоторому единому представлению – к исходной форме, или лемме); в стемминге (нормализация словоформ, когда разные словоформы приводятся к "псевдооснове" или усекаются до их корневой части, для отражения инвариантного значения слова); в частеречном тэгинге (*pos-tagging*, т.е. указание части речи для каждой словоформы в тексте) или в полном морфологическом анализе – приписывании грамматических характеристик всем словоформам; в построении гипотез для нераспознанных слов;

просодическая – использует тэги, обозначающие ударение и интонацию, паузы, повторы. В корпусах устной разговорной речи просодическая разметка сопровождается дискурсной;

редакторская (специфика типографского макета не учитывается);

синтаксическая – является результатом парсинга, выполняемого программно в автоматизированном режиме на основе данных морфологического анализа, описывает синтаксические связи между лексическими единицами и различные синтаксические конструкции;

семантическая – приписывает единицам текста один или несколько семантических и словообразовательных признаков; предусматривает спецификацию значения слов, разрешение омонимии и синонимии, категоризацию слов (разряды, родовые и видовые дескрипторы), выделение тематических классов, признаков каузативности, оценочных и деривационных характеристик, семантических ролей (например, агенс, инструмент, пациенс, результат [1]);

ситуационная (разметка отношений, т.е. ситуаций – многоместных отношений, в которых размеченные сущности выступают в определенных семантических ролях);

служебная (имплицитная) – включает: 1) «текст-стиль» (академический, архаизованный, диалектный, индивидуально-авторский, научно-популярный, научный (в зависимости от области знаний может встречаться большое количество подробных расчетов и формул; логичность и однозначность; наличие доказательств правдивости и правильности изложенных

фактов и сведений; передача четкой информации по конкретной сфере знаний; присутствие относительно большого количества терминов), нейтральный, официально-деловой (логичность; однозначность мысли; сухое изложение; отсутствие сленга и явного выражения эмоций), публицистический (может содержать образные выражения, эмоциональные оттенки; понятен для большинства читателей, а не только специалистам; передача понятных для пользователя сведений и формирование определенного отношения к рассматриваемому объекту или ситуации), разговорный (эмоциональность; использование сленга; частое использование в дружеских диалогах), сниженный, сниженный с элементами грубого просторечия и жаргона, художественный (пробуждение эмоций и яркого воображения) и пр.), 2) информацию об аудитории (возраст; размер, уровень образования);

специальная – приписывание источникам специальных, формальных меток (внешних, экстралингвистических) [3];

структурная – эксплицирует логическую структуру источника и предназначена для выделения структурных элементов (том, книга, часть, глава, действие, явление, реплика, сноска, ремарка, строфа, стих, а также: абзац, предложение, словоформа и текстоформа; таблица, формула и др.);

терминологическая – предназначена для фиксации терминов и особенности использования общепотребительной лексики.

техничко-технологическая – кодировка, дата обработки, исполнители, источник электронной версии;

формальная – отражает имя файла, параметры кодирования, версию языка разметки; авторов, составителей, редакторов и исполнителей проекта;

форматная – описывает параметры оформления (параметры шрифта: размер, жирность, курсив, разрядка, верхние/нижние индексы; параметры абзаца: выравнивание, отступы, межстрочный интервал; полиграфические и декоративные элементы: колоннотулы, виньетки; номера страниц и стихов);

частеречная (морфологическая);

экстралингвистическая (метаразметка метаданных) – включает внешнюю разметку, формальную структурную разметку, а также технико-технологическую разметку (кодировку, даты обработки, исполнителей, источник электронной версии).

Все виды разметок осуществляются в соответствии с принципами: 1) описание (обоснование) схемы разметки; 2) общепринятая система лингвистических понятий; 3) известная для пользователя схема анализа; 4) мотивированность введения параметров; 5) теоретически нейтральная (традиционная) схема разметки; 6) следование международным стандартам [4, с. 32-49], но при этом ни одна разметка не может априорно считаться единственным стандартом; 7) возможность убрать разметку источника или вычленить собственно разметку.

Для представления корпусов текстов в Интернете в настоящее время необходимо учитывать спецификации *HTML* (*HyperText Markup Language* – язык разметки гипертекста), *XHTML* (*eXtensible HyperText Markup Language* – расширяемый язык разметки гипертекста), *XML* (*eXtensible Markup Language* – рас-

ширяемый язык разметки), *CSS* (*Cascading Style Sheets*, каскадные таблицы стилей), *PDF* (*Portable Document Format* – межплатформенный открытый формат электронных документов), *RTF* (*Rich Text Format* (англ. *rich* – богатый) – формат обогащённого текста), *PostScript* (язык описания страниц, в основном используемый в настольных издательских системах), вики-разметка (используется для оформления текста в вики-проектах) и аналогичные.

Академический словарный корпус русского языка, фактически являясь информационно-поисковым гизаурусом (гипертекстовым сводным тезаурусом [13, с. 29; 14; 15]), даст богатый материал для различного рода научных исследований не только в области русского языка, но и других современных языков, что позволит на основе цифровых технологий при соответствующей классификации и систематизации словарей объединить лексикографические материалы, обеспечить их оперативный ввод в научный оборот для оптимизации научных исследований в современной лексикографии¹.

СПИСОК ЛИТЕРАТУРЫ

1. Апресян Ю.Д., Богуславский И.М., Иомдин Б.Л. и др. Синтаксически и семантически аннотированный корпус русского языка: современное состояние и перспективы // Национальный корпус русского языка: 2003–2005. – М.: Индрик, 2005. – С. 193-214. – URL: <http://www.ruscorga.ru/old/sbornik2005/12apresyan.pdf>.
2. Баранов А.Н. Введение в прикладную лингвистику. – М.: Эдиториал УРСС, 2001. – 347 с.
3. Захаров В.П. Электронный обменный формат для словарей TEI. – СПб: СПбГУ, 2013. – 80 с.
4. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – СПб: СПбГУ, 2013. – 148 с.
5. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Наука, 1981. – 367 с.
6. Кудашев И.С. Проектирование переводческих словарей специальной лексики. – Helsinki, 2007. – 445 с.
7. Никитина С.Е. Семантический анализ языка науки: на материале лингвистики. – М.: Книжный дом, 2010. – 140 с.
8. Рыков В.В. Корпус текстов как отражение состояния русского языка // Труды Международного конгресса «Русский язык: исторические судьбы и современность». – М.: МГУ им. М.В. Ломоносова, 2001. – С. 421.
9. Лесников С.В. Конструирование терминологического корпуса русского языка // Известия Волгоградского государственного педагогического университета. – 2019. – №4(137). – С. 94-110.
10. Загоровская О.В., Лесников С.В. Виды лексикографической информации в автоматическом словаре русских говоров Коми АССР и сопредельных областей // Машинный фонд русско-

¹ Подробную библиографию и научные работы С.В. Лесникова см. на сайте <https://lsw.ru/!lesnikow/article/>.

- го языка: Предпроектные исследования. – М.: ИРЯз АН СССР, 1988. – С. 64-70.
11. Лесников С.В., Загоровская О.В. Формальная грамматика словарной статьи автоматического словаря русских говоров Коми АССР и сопредельных областей (АСРГКА) // Вторая Всесоюзная конференция по созданию Машинного фонда русского языка: Материалы конференции. – М.: ИРЯз АН СССР, 1988. – С. 107–119.
12. Лесников С.В. Владислав Митрофанович Андрущенко – научный руководитель и консультант, главный конструктор МФРЯ // Тезисы всероссийской конференции «От языковых машинных фондов к лингвистическим корпусам: памяти В.М. Андрущенко». – М.: МГУ им. М.В. Ломоносова, ИРЯз РАН, 2018. – С. 58–60.
13. Лесников С.В. Словарь русских словарей. – М.: Азбуковник, 2002. – 334 с.
14. Лесников С.В., Вилежанинов В.М. ГИЗАУРУС (гипертекстовый тезаурус): автоматизированная гипертекстовая разметка текстов по лексикографической базе данных // Свидетельство о государственной регистрации программы для ЭВМ № 2013613020. Зарегистрировано в Реестре программ для ЭВМ 21 марта 2013 г. Выдано Федеральной службой по интеллектуальной собственности. – М.: Роспатент, 2013.
15. Булыгина Д.С., Лесников С.В. Алгоритм автоматизированного конструирования гипертекстового тезауруса (гизауруса) русского языка на основе оцифрованных словарей и справочников новых слов и значений для интерактивного лексикографического корпуса «Лексико-семантическая неология в русском языке начала XXI века» // Достижения вузовской науки 2019: сборник статей IX Международного научно-исследовательского конкурса (20 мая 2019 г. в г. Пенза): в 2 ч. Ч. 1. – Пенза: МЦНС «Наука и Просвещение». 2019. – С. 119-125.

Материал поступил в редакцию 22.05.19

Сведения об авторе

ЛЕСНИКОВ Сергей Владимирович – кандидат филологических наук, доцент, доцент кафедры математического моделирования и кибернетики, ведущий специалист (программист) Отдела лексикографии современного русского языка. Группа Большого академического словаря Института лингвистических исследований Российской академии наук, Санкт-Петербург.
e-mail: serg@lsw.ru

ВНИМАНИЮ ПОДПИСЧИКОВ!

С 2018 года возобновляется издание информационного бюллетеня «Иностранная печать об экономическом, научно-техническом и военном потенциале государств-участников СНГ и технических средствах его выявления» серии «Экономический и научно-технический потенциал» (56741) взамен информационного бюллетеня «Экономика и управление»

Периодичность выхода – 12 номеров в год. Объем 48 уч.-изд. л. в год.

В бюллетене освещаются материалы иностранной печати по широкому спектру вопросов, касающихся сфер экономического и научно-технического развития России и стран СНГ: общие вопросы, финансы, промышленность, рынки, сельское хозяйство, космос, транспорт и связь, природные ресурсы, трудовые ресурсы, внешние торгово-экономические и научные связи

Оформить подписку на информационный бюллетень, начиная с любого номера, можно в ВИНТИ РАН по адресу: 125190, Россия, Москва, ул. Усиевича, 20,

Телефоны: (499) 151-78-61; (499) 155-42-85

Факс: (499) 943-00-60;

E-mail: contact@viniti.ru; sales@viniti.ru

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ РОССИЙСКОЙ АКАДЕМИИ НАУК

предлагает научным работникам, аспирантам и другим специалистам в области естественных, точных и технических наук, желающим быстро и эффективно опубликовать результаты своей научной и научно-производственной деятельности, использовать способ публикации своих работ через *систему депонирования*.

Депонирование (передача на хранение) – особый метод публикации научных работ (отдельных статей, обзоров, монографий, сборников научных трудов, материалов научных конференций, симпозиумов, съездов, семинаров), разрешенных в установленном порядке к открытому опубликованию.

Подготовка и передача на депонирование научных работ происходит в соответствии с «Инструкцией о порядке депонирования научных работ по естественным, техническим, социальным и гуманитарным наукам» (М., 2014).

Депонированные научные работы находятся на хранении в депозитарии ВИНТИ РАН, копии работ предоставляются заинтересованным организациям и специалистам на бумажном и электронном носителях и являются официальной публикацией.

Информация о депонированных научных работах включается в информационные издания ВИНТИ РАН: Реферативный журнал, Базу данных и Аннотированный библиографический указатель «Депонированные научные работы».

Направить научную работу на депонирование можно, обратившись в Группу депонирования ЦНИО ВИНТИ РАН по адресу:

125190, Москва, ул. Усиевича, 20.

ВИНТИ РАН, Группа депонирования ЦНИО

Тел.: 499-155-43-28, 499-155-43-76, 499-155-42-43, Факс: 499-943-00-60,

E-mail: cnio@viniti.ru, dep@viniti.ru

С инструкцией о порядке депонирования можно ознакомиться на сайте ВИНТИ РАН:
<http://www.viniti.ru>