

# НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

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

Издается с 1961 г.

№ 4

Москва 2019

## ИНФОРМАЦИОННЫЕ СИСТЕМЫ

УДК 004.72:004.89

А.И. Водяхо, В.Ю. Осипов, Н.А. Жукова, М.А. Червонцев

### Когнитивные технологии в управлении мониторингом

*Определено понятие когнитивного мониторинга. Рассмотрены возможные подходы к построению когнитивных систем мониторинга и приведена обобщенная их структура. Введено понятие когнитивной машины мониторинга. Предложено использование архитектурно-когнитивного подхода к проектированию систем мониторинга, основной идеей которого является возможность генерации загружаемых архитектур. Описывается структура платформы, ориентированной на реализацию этого подхода. Приведен пример реализации когнитивной системы мониторинга*

**Ключевые слова:** когнитивные системы, системы мониторинга, архитектурно-когнитивный подход к проектированию систем мониторинга

#### ВВЕДЕНИЕ

Когнитивные информационные технологии активно внедряются в повседневную жизнь. Все более широкое применение находят такие технологии как смартфоны, «умный» телевизор, «умный» дом, системы мониторинга состояния больного и многие другие. Быстрое расширение сферы применения интер-

нета вещей (*Internet of Things – IoT*) приводит к появлению внушительных по размеру гетерогенных сетевых структур, включающих множество различных элементов, использующих большое число разнообразных стеков протоколов. При этом создается огромный объем структурированных и неструктурированных данных, а сама структура таких систем постоянно меняет-

ся. Для эффективного управления огромными информационными системами (ИС) с постоянно изменяющейся структурой связей и составом оборудования, требуется использование принципиально новых парадигм управления. Одна из основных подзадач управления – определение состояния системы, т.е. задача мониторинга.

Термин «когнитивность» происходит от латинского *cognito*, что означает «думаю». В «Кембриджском словаре» когнитивность определяется как сознательная умственная деятельность. При таком определении внимание уделяется тому, как человек обучается, запоминает и рассуждает, а не на отдельные факты. Термин «когнитивность» пришел из философии, а затем стал применяться и психологами, в частности, «когнитивная психология», которая изучает то, как люди думают.

Понятие когнитивности самым тесным образом связано с понятием «искусственный интеллект» (*Artificial Intelligence – AI*). По существу, системы искусственного интеллекта (ИИ) позволяют понимать ситуацию, обучаться и на основании этого принимать решения о необходимых действиях. Классические системы ИИ решают те задачи, которые могут быть четко сформулированы и запрограммированы. В связи с этим когнитивные системы, способные реализовывать обработку информации подобно тому, как это делают люди, можно рассматривать как расширение традиционных систем искусственного интеллекта.

Одна из целей создания когнитивных систем – это уменьшение семантического зазора между человеком и машиной, что является ключевой проблемой практической реализации когнитивного подхода. Уменьшая семантический зазор между человеком и машиной, мы увеличиваем его между прикладным программным обеспечением, осуществляющим когнитивную обработку, и используемыми программно-аппаратными платформами, что влечет появление дополнительных уровней обработки и, как следствие, – серьезных проблем в плане быстрой реакции.

## КОГНИТИВНЫЙ КОМПЬЮТИНГ

Когнитивные системы принципиально отличаются от традиционных информационных систем, а также от систем искусственного интеллекта [1, 2], они не должны жестко программироваться, так как подобно человеку обучаются в процессе работы и общения с людьми, что со временем приведет их к более эффективному функционированию [3]. В соответствии с этим когнитивные системы (КС) базируются не на заложенных априорно в них алгоритмах и правилах, а, преимущественно, на обучении [2].

Использование когнитивных технологий позволит расширить способности компьютеров – им будут доступны задачи, традиционно решаемые людьми и требующие реализации таких процедур как планирование, рассуждения, обучение, работа с неполными и недостоверными данными и т.п. [4].

*Когнитивный и традиционный компьютеринг.* Возникает закономерный вопрос: Где проходит граница между традиционным и когнитивным компьютерингом? Ответ очевиден – четкую границу провести невозможно. В подобных ситуациях обычно принято

использовать некоторые дискретные или непрерывные меры приближения (модели зрелости), которые могут быть определены в терминах уровней зрелости соответствующих технологий. Эти уровни определяют: какие когнитивные возможности и в каком порядке могут быть добавлены при построении информационной системы и образуют некоторую дорожную карту, демонстрирующую переход от традиционных ИС к когнитивным.

Выделим два альтернативных подхода к определению границ когнитивного компьютеринга: определить границы в терминах способности решать задачи подобно тому, как это делают люди или в терминах способов решения задач.

Первый подход предполагает сравнить поведения машины и человека при решении целевых задач. Например, если предложить эксперту различить подходы к решению задачи человеком и машиной, и он не будет способен их уверенно различить, то можно говорить о когнитивном компьютеринге. Также возможно определить насколько интенсивно используются «человеческие» подходы к решению задач. Например, для сервисно-ориентированных ИС степень когнитивности определяется исходя из того, какая часть сервисов рассматривается как когнитивные. Проблема в том, что далеко не всегда известно, как человек решает задачу. Заметим, что для практики такой подход практически ничего не дает.

Большой практический интерес представляет второй подход, в соответствии с которым когнитивный компьютеринг определяется через используемые архитектурные решения. При применении этого подхода когнитивный компьютеринг можно определить как семейство архитектур, которые в процессе функционирования используют модели внешнего мира и самой системы, и эти модели должны строиться в терминах знаний.

Архитектуру информационной системы, в соответствии со стандартом [5], определяют как базовую организацию системы, воплощенную в её компонентах, их отношениях между собой и с окружением, а также принципы, на которых основывается проектирование и развитие системы. Таким образом, понятие архитектуры ИС включает и её проектирование.

Понятие архитектуры можно применить к семейству информационных систем, которое в процессе функционирования и/или проектирования использует модели внешнего мира и самой системы, представленные в терминах знаний. Если когнитивность реализуется только на этапе проектирования, то требуется наличие инструментов для трансформации когнитивных решений для класса ИС в некогнитивные решения для конкретного экземпляра ИС. Такой подход будет полезен в случаях, когда использование когнитивных подходов в режиме *run time* затруднительно по причине жестких требований по быстротедействию и/или по доступным ресурсам. Когнитивные системы можно рассматривать как подкласс ИС, отличительной особенностью которых является: способность к самообучению, адаптации, обработке неструктурированных данных, масштабированию, управлению собственной структурой; динамизм, интерактивность и др. [6].

## ПОДХОДЫ К РЕАЛИЗАЦИИ КОГНИТИВНЫХ СИСТЕМ

Четыре альтернативных подхода к построению систем на базе искусственного интеллекта, основанных на следующих тезисах, выделяются в [7]: ИС, которые думают как человек; ИС, которые думают рационально (разумно); ИС, которые действуют как человек; ИС, которые действуют рационально (разумно). Отметим, что понятия «думать» или «действовать как человек» и «действовать рационально» – это разноплановые понятия.

Существующие КС используют все четыре подхода. В частности, для исследования вопросов о том, как думает человек, может оказаться полезным подход, при котором анализируется характер ошибок, совершаемых компьютером и человеком. Если они делают похожие ошибки, то следует считать, что система и человек используют сходные модели решения задачи [8]. Очевидно, что такой подход более полезен физиологам, а не ИТ-специалистам, поскольку для них важно, чтобы машина решала задачи с требуемыми характеристиками.

Более важным является то, что когда человек думает, он оперирует знаниями. Поэтому с практической точки зрения полезным является определение когнитивной системы, думающей рационально, как системы, которая оперирует с моделями, представленными в терминах знаний.

Подробный и свежий анализ (относящийся к 2018 г.) состояния исследований и разработок в области когнитивного компьютеринга представлен в [8].

*Когнитивный компьютеринг и мультиагентные системы (МАС)* – имеют много общего. МАС – это, прежде всего технология, которая интегрирует многие технологии и платформы. Понятие МАС – многоплановое. Эти системы можно рассматривать, руководствуясь, по крайней мере, тремя точками зрения:

1) как один из возможных подходов к построению систем распределенного искусственного интеллекта; 2) как отдельный стиль программирования, либо комбинацию нескольких архитектурных стилей (для прикладных программистов МАС) [9]; 3) как аппаратно-программную платформу (для системных программистов МАС).

## ОБЫЧНЫЙ И КОГНИТИВНЫЙ МОНИТОРИНГ

Мониторинг следует определить как процесс сбора информации о состоянии некоторой целевой системы, её обработки и/или трансформации с целью представления сведений заинтересованным сторонам, в частности лицам, принимающим решения (ЛПР). Процесс мониторинга можно рассматривать как бизнес-процесс. Системы мониторинга (СМ), реализующие процесс мониторинга, могут представлять собой либо отдельную подсистему, либо быть одной из подсистем более высокого уровня. В последнем случае речь идет о самомониторинге. Результаты мониторинга используются для решения задач, связанных с управлением целевой системой: диагностика, саморемонт, оптимизация производительности, реконфигурирование системы и др. [10].

*Обобщенная структура системы мониторинга.* Целевой объект и система мониторинга рассматриваются либо как единая система, либо целевой объект рассматривается как контекст, в котором функционирует система мониторинга. Выбор варианта зависит от специфики решаемой задачи. Если система мониторинга не может выдавать управляющие воздействия, т.е. оказывать влияние на целевой объект (ЦО), то последний следует рассматривать как контекст. СМ и ЦО образуют единую систему, когда система способна оказывать управляющие воздействия на целевой объект, т. е. может быть системой мониторинга и управления (СМУ).

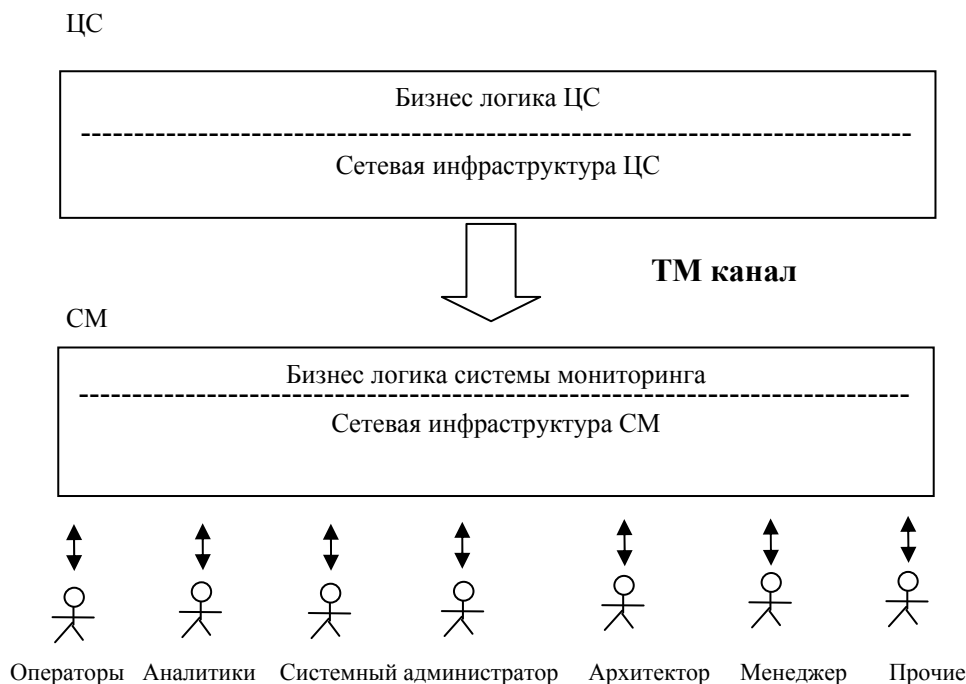


Рис. 1. Обобщенная структура системы мониторинга

Целевые объекты, как правило, имеют связи с другими объектами. Совокупность ЦО и связанных с ними объектов образуют целевую систему, которая, как и СМ включает программный (бизнес-логику) и инфраструктурный (датчики, исполнительные механизмы, серверы и сетевое оборудование) компоненты. На рис. 1 показана обобщенная структура системы мониторинга.

Целевая система может быть представлена в виде множеств:  $\{D, IM, EO, BP, C, L\}$ , где:  $D$  – датчики;  $IM$  – исполнительные механизмы;  $EO$  – элементы оборудования;  $BP$  – реализуемые бизнес-процессы;  $C$  – сигналы (сообщения), которыми обмениваются перечисленные элементы;  $L$  – записи в лог-файлы о событиях типа «начало» и «конец» бизнес-процесса.

$D$  – источники сигналов, которые генерируются в контексте конкретного бизнес-процесса.

$IM$  – приемники сигналов. Принимаемые сигналы имеют формат  $\{I, Z, T_{np}\}$ , где  $I$  – имя параметра,  $Z$  – значение параметра,  $T_{np}$  – временная метка приема сигнала.

$EO$  – элемент, которому можно поставить в соответствие состояние (исправен, неисправен, частично исправен). В качестве  $EO$  могут выступать  $D, IM$ , серверы, элементы сетевой инфраструктуры.

$BP$  – процесс обработки информации, имеющий множество входов и выходов, по которым поступают и выдаются сигналы ( $C$ ) – в общем случае, это сообщения, которыми обмениваются  $BP$ .

$C$  имеет формат:  $\{BP, VM\}$ , где  $BP$  – идентификатор бизнес-процесса,  $VM$  – временные метки, каждая из которых имеет формат:  $\{I, Z, T_{выд}\}$ , где  $I$  – имя параметра,  $Z$  – значение параметра,  $T_{выд}$  – временная метка выдачи сигнала.

$L$  – записи в лог-файлы о системных событиях обычно с привязкой ко времени. Информация из лог-файлов также доступна для анализа.

## ОСНОВНЫЕ ТИПЫ СИСТЕМ МОНИТОРИНГА

Системы мониторинга и управления (СМУ) на верхнем уровне обработки информации можно классифицировать по трем параметрам: структуре целевой системы (централизованная, распределенная),

структуре системы мониторинга и управления (централизованная, распределенная) и по способу обработки (отложенная или в масштабе реального времени).

На рис. 2 в общем виде представлена СМУ, в которой информация от объектов ( $O$ ) поступает по каналам телеметрии (КТ). По полученным данным строятся модели ( $M$ ) объектов. Представленная система имеет структуру дерева, но допустимо использование и других топологий. Распределенные целевые объекты (РЦО) и распределенная система мониторинга и управления (РСМУ) могут быть связаны между собой несколькими каналами телеметрии. Данный вариант соответствует случаю, когда целевая система является системой РЦО. Это могут быть, в частности, подвижные объекты. СМУ также представляет собой распределенную систему мониторинга и управления.

Следует выделить решаемые СМУ основные задачи:

- анализ телеметрии в отложенном режиме;
- обработка телеметрии в масштабе реального времени и управление целевой системой;
- управление сетевой инфраструктурой СМУ;
- отладка процесса работы с телеметрией;
- моделирование работы СМУ на этапе проектирования.

Когнитивные системы мониторинга (КСМ) определим как системы, построенные на базе принципов когнитивного компьютеринга.

Принципиально можно говорить о трех аспектах реализации концепции когнитивного мониторинга. В этом случае речь идет об ответах на три вопроса.

1. Каким образом организовывается когнитивная обработка?
2. Каким образом функциональность, обеспечивающая когнитивность, распределяется между отдельными элементами системы мониторинга?
3. Каким образом организована среда передачи данных мониторинга?

Ответом на первый вопрос является описание механизмов когнитивной обработки, которые используются при построении КСМ и с помощью которых они реализуются.

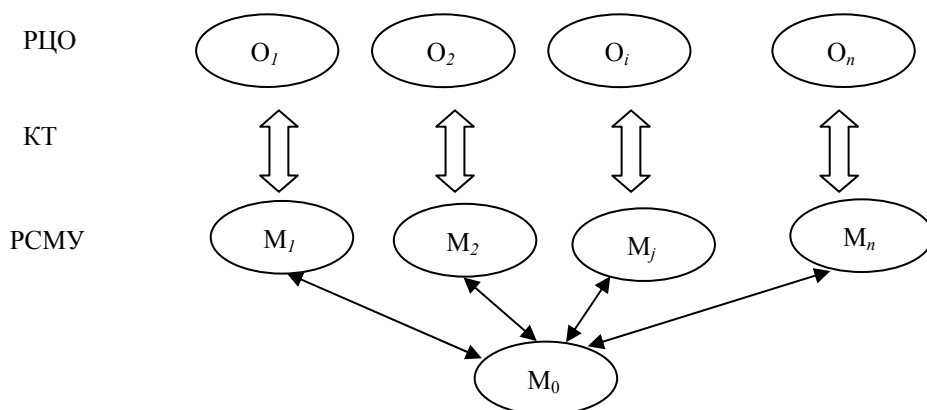


Рис. 2. Общий вид системы мониторинга и управления

Ответ на второй вопрос позволяет определить, из каких подсистем состоит ИС. К таким подсистемам можно отнести централизованную ИС с клиент-серверной архитектурой; распределенную систему, построенную по принципу *fog computing*; мультиагентные архитектуры и др. Следует заметить, что у самих источников информации возможно наличие встроенных средств когнитивной обработки данных.

Ответ на третий вопрос дает возможность понять, как организована среда передачи данных мониторинга. Принимая во внимание тот факт, что целевые системы все чаще представляют собой совокупность распределенных подвижных объектов, естественным является построение среды передачи данных по принципу когнитивного Интернета вещей (*Cognitive Internet of Things – CIoT*).

### ОБОБЩЕННАЯ СТРУКТУРА КОГНИТИВНОЙ СИСТЕМЫ МОНИТОРИНГА

Целевая система может иметь встроенные датчики, которые реализовывают отдельные элементы когнитивной обработки, преимущественно на уровне сигналов. Обобщенная структура КСМ показана на рис. 3. Основу КСМ составляет когнитивная машина, которая отвечает за формирование требуемых моделей, поддержание их в актуальном состоянии, трансформацию моделей и принятие решений на моделях. Связь между когнитивной машиной и полем когнитивных датчиков осуществляется посредством когнитивной сети [8].

Подходы к реализации КСМ можно классифицировать на основании трех признаков: 1) используемые архитектурные решения; 2) применяемые структурные решения; 3) подходы к использованию моделей.

Для первого признака (архитектурные решения) выделим четыре основных альтернативных подхода: использование классических (традиционных) решений, используемые вертикальных решений, использование горизонтальных решений и использование гибридных решений.

Традиционные решения предполагают, например, использование решений на базе протокола сетевого управления *SNMP*, при этом когнитивная обработка может выполняться на сервере.

Вертикальные решения – это многослойные решения, в которых когнитивная обработка реализуется на нескольких уровнях. Примером такого подхода могут служить туманные (*fog*) структуры.

Горизонтальные решения предполагают *P2P* решения. В этом случае когнитивная обработка реализуется коллективом вычислителей, работающих на одном уровне. Пример такого подхода – мультиагентные решения.

Гибридные решения – это комбинация горизонтального и вертикального подхода. Примером служат мультиагентные системы, работающие на нескольких уровнях.

Как указывалось ранее, когнитивная обработка предполагает использование двух основных типов моделей: модели целевой системы и собственной модели. Если речь идет о скомпилированных архитектурах, модели могут не применяться в конечных решениях когнитивных систем мониторинга. В этих системах возможно использование только отдельных моделей, например, модели целевой системы. Модели бывают либо статическими, либо динамическими. Статические модели «зашиваются» в КСМ на этапе проектирования. Динамические модели формируются и/или изменяются в процессе функционирования.

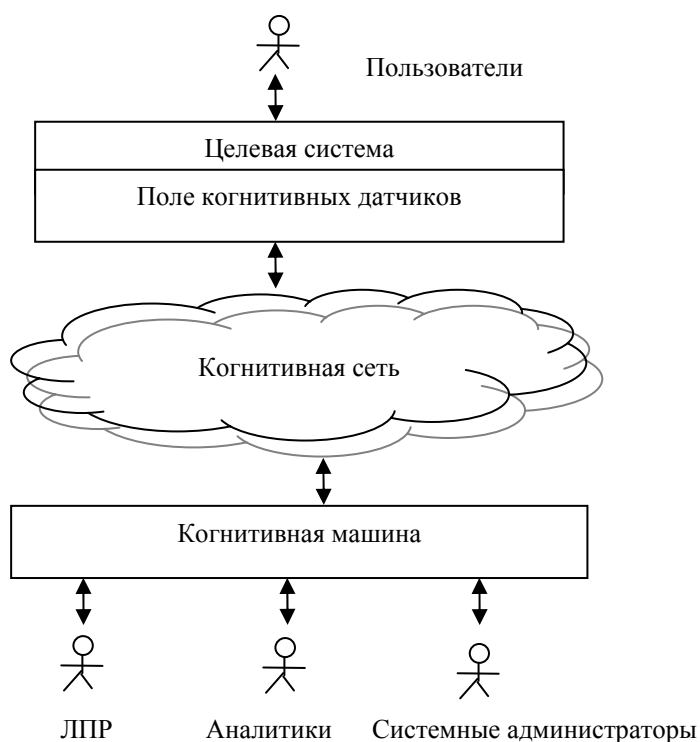


Рис. 3. Обобщенная структура когнитивной системы мониторинга

При построении КСМ применяются различные архитектурные решения. В простейшем случае это могут быть фиксированные архитектуры, как спроектированные вручную, так и полученные из архитектурных метамоделей, в соответствии с подходом, который будет описан далее. Если в информационной системе реализуются механизмы виртуализации, то появляется возможность применения загружаемых архитектур, которые хранятся в библиотеке архитектур. Наиболее эффективное и сложное решение состоит в использовании автоматически генерируемых динамических архитектур, способных адаптироваться как к контенту, так и к контексту.

### КОГНИТИВНАЯ МАШИНА МОНИТОРИНГА

Основой КСМ является когнитивная машина мониторинга, которая может рассматриваться либо как архитектурный фреймворк, либо как платформа, либо как среда проектирования когнитивной системы мониторинга и представляет собой набор сервисов разного уровня, используемых для построения конкретных экземпляров КСМ. На рис. 4 показана концептуальная структура когнитивной машины мониторинга.

В состав когнитивной машины мониторинга (КММ) входят семь основных подсистем:

- 1) модуль построения и поддержания моделей в актуальном состоянии (*Model Mining Module – MMM*);
- 2) модуль трансформации моделей, т.е. приведения их к нужному представлению (*Model Transformation Module – MTM*);
- 3) модуль обработки моделей – отвечает за процессинг моделей, в частности, реализует рассуждения на моделях (*Model Processor – MP*);

4) генератор архитектур – формирует загружаемые модули и используется при работе с динамическими архитектурами;

5) репозиторий – предназначен для хранения моделей, данных и знаний; набор адаптеров;

6) адаптеры – используются для поиска информации и знаний во внешних источниках;

7) генератор отчетов и реакций – выполняет несколько функций. В режиме ручного управления он выдает информацию о состоянии целевого объекта для ЛПП, а в режиме автоматического управления – выдает управляющие воздействия на систему мониторинга и на целевой объект. В случае использования динамической загружаемой архитектуры этот модуль отвечает также за загрузку архитектурных модулей в распределенную среду КСМ.

Когнитивная машина мониторинга функционирует следующим образом.

На вход КММ поступают данные мониторинга. Если модель неизвестна или требует уточнения и/или верификации, то для этого используется модуль *МММ*. Типовой режим – это проверка текущих моделей на актуальность и их перестройка при поступлении сообщений о наступлении некоторых событий. Местом хранения моделей является репозиторий. Принятие основных решений реализуется на уровне моделей, которые строятся в терминах знаний. Для этого из входного потока данных требуется извлечь знания. Эту процедуру реализует модуль *МММ*, в состав которого включены сервисы более низкого уровня, позволяющие извлекать знания из данных и событий. Такой процесс строится в соответствии с хорошо известной *JDL*-моделью [11]. Модуль *МР* выполняет функцию процессора, к которому можно обращаться с помощью языка запросов к данным *SPARQL* [12].

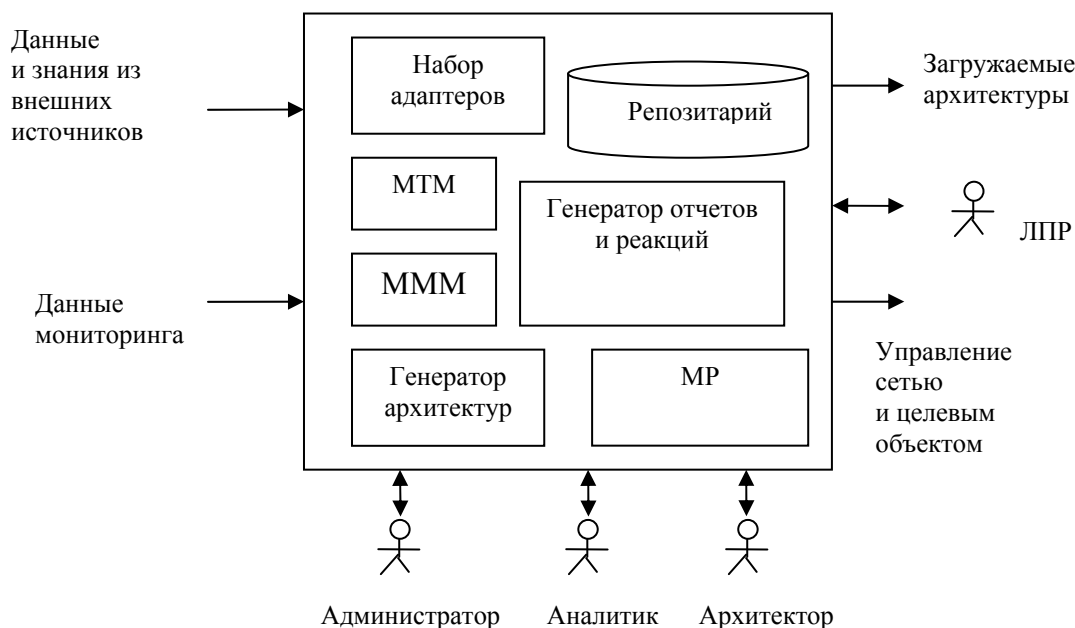


Рис. 4. Концептуальная структура когнитивной машины мониторинга

Генератор отчетов и реакций решает три основные задачи: 1) при работе в ручном режиме он выдает ЛПР информацию, необходимую для принятия решений; 2) при работе в автоматическом режиме генератор самостоятельно формирует управляющие воздействия, как для управления когнитивной сетью, так и для управления целевым объектом (ЦО), например, с целью реконфигурации; 3) если в рамках когнитивной машины мониторинга реализуется динамическая архитектура, то генератор архитектур синтезирует требуемые модули. Такое решение может быть актуально в случае, когда требуется сгенерировать тестовую процедуру для заданной конфигурации. Подобные решения актуальны и в случаях, когда ЦО и/или соединительная сеть построены по принципу интернета вещей. Если структура обслуживаемого целевого объекта известна и мало изменяется, загружаемые модули могут храниться в библиотеках, которые находятся в репозитории. Использование загружаемых архитектур (*on demand*) [13] актуально, прежде всего, для туманных структур, когда ресурсы модулей промежуточного уровня ограничены.

Как указывалось выше, когнитивный компьютеринг и, соответственно, когнитивный мониторинг, предполагает самое широкое использование моделей, в том числе моделей, представленных в терминах знаний.

На сегодня построение моделей при проектировании информационных систем весьма проблемно и в большинстве случаев выполняется вручную силами аналитиков, что требует больших затрат и не исключает ошибок. Если структура моделируемой информационной системы постоянно изменяется, построение адекватных моделей становится еще более сложным, поэтому актуально – автоматическое формирование моделей.

Для когнитивных систем мониторинга модели должны обеспечивать решение следующих основных задач: оценивать текущее состояние целевой системы и прогнозировать её поведение; оценивать и прогнозировать собственное состояние; синтезировать архитектурные модели, которые позволяют переходить от архитектурного описания к реализации. Применительно к КСМ первоочередной интерес представляют модели в терминах знаний. Это, прежде всего, онтологические модели. Таким образом, для построения КСМ требуются разноплановые модели.

В настоящее время известны различные методы получения моделей из данных и информации о событиях. В этом контексте представляют интерес три технологии: извлечение знаний из данных (*data mining*) [14], построение моделей процессов из лог-файлов в форме сетей Петри [15], построение автоматных, в частности, многоуровневых автоматных моделей [16]. На сегодня это разрозненные и слабо связанные между собой технологии. Представляется целесообразным рассматривать эти технологии как единое целое – автоматическое построение моделей (*Model Mining – MM*), которое можно определить как  $MM = \{Mod, Met, Tmet, I, R\}$ , где *Mod* – модели, *Met* – методы майнинга моделей из данных и сообщений о

событиях, *Tmet* – методы трансформации моделей, *I* – инструментальные средства майнинга моделей, *R* – средства реализации моделей. *MM* можно рассматривать как интеграционную технологию. Введение этого понятия позволяет взглянуть на проблему автоматического синтеза моделей интегрально и понять какие элементы, необходимые для практической реализации систем автоматического синтеза моделей, отсутствуют; накапливать знание в форме моделей (модельное знание).

## ТРАНСФОРМАЦИЯ МОДЕЛЕЙ

Модельный подход предполагает, что система описывается в терминах используемых моделей и алгоритмов. Этот подход известен как проектирование на моделях (*Model Driven Engineering – MDE*). Идея состоит в построении стека моделей. Ключевыми являются такие понятия: как модель (*model*), метамодель (*metamodel*), связывание моделей (*model weaving*), преобразование моделей (*model transformation*). Представление набора моделей в виде стека требует определения процедуры перехода от модели верхнего уровня к моделям нижнего уровня. Преобразование моделей – это процесс, который определяет, каким образом можно получить целевую модель из одной или нескольких исходных моделей. Модели могут трансформироваться либо «по горизонтали», либо «по вертикали». В первом случае трансформируемая модель преобразовывается в модель, которая принадлежит одному с ней уровню; во втором – является моделью более высокого или более низкого уровня. Для связывания разнотипных моделей предназначен специальный тип – модели связывания, основное назначение которых состоит в установлении соответствия между отдельными элементами моделей (низкоуровневое связывание) [17].

Заметим, что применительно к КСМ практический интерес представляют как горизонтальные, так и вертикальные преобразования моделей, поскольку при реализации процедур *Model Mining* не всегда возможно автоматическое построение требуемого типа моделей. Например, могут быть доступны сервисы, позволяющие получить результаты майнинга в форме многоуровневого автомата, а для реализации системы требуется её описание на языке *UML*.

## АРХИТЕКТУРНО-КОГНИТИВНЫЙ ПОДХОД К ПРОЕКТИРОВАНИЮ СИСТЕМ МОНИТОРИНГА

При создании реальных КСМ разработчики сталкиваются, по крайней мере, с двумя проблемами.

1. Одно из требований, предъявляемых к КСМ – это способность работать в режиме реального времени. Применительно к этим системам основная сложность – реализация интеллектуальных элементов поведения требует перехода на работу со знанием. Манипуляции со знаниями предполагают реализацию логического вывода, что, в свою очередь, требует большого числа машинных операций, которые не всегда можно выполнять параллельно, что затрудня-

ет использование КСМ не только в системах жесткого реального времени, но и в системах с умеренными требованиями к соблюдению сроков выполнения задач. Кроме того, время, необходимое для реализации логического вывода, не всегда можно точно предсказать. Эта проблема решается за счет увеличения производительности используемых серверов, распараллеливания процесса вычислений, выполнения вычислений «по месту», т.е. непосредственно в интеллектуальных контроллерах. При использовании современных платформ всё это даёт эффект только при очень умеренных требованиях к соблюдению сроков выполнения задач.

2. Построение КСМ предполагает реализацию механизмов работы со знаниями, представленными, чаще всего, в форме онтологии. Следует отметить, что создание подобного рода систем требует от аналитика и программиста достаточно высокой квалификации, т.е. приходится привлекать высокооплачиваемых специалистов, что приводит к увеличению стоимости разработки и, соответственно, серьезно ограничивает сферу применения КСМ.

Для решения этих проблем может быть предложен архитектурно-когнитивный подход к проектированию систем мониторинга. Идея подхода состоит в создании доменно-ориентированной платформы для проектирования КСМ (КСМ-платформы), которая включает архитектурный фреймворк (типовое доменно-ориентированное решение) и инструментарий для проектирования на его базе конкретных экземпляров систем мониторинга. Обобщенная структура КСМ-платформы представлена на рис. 5.

Платформа представляет собой набор средств, необходимых для проектирования когнитивной системы мониторинга различного назначения. В ее состав входят: репозиторий, предназначенный для хранения моделей, правил, скриптов и данных; подсистема импорта-экспорта; набор сервисов; инструментальные средства разработки и проектирования – билдер,

позволяющий строить бизнес-процессы обработки с использованием языков высокого уровня (ЯВУ) и объектно-ориентированного программирования (ООП); графический пользовательский интерфейс (*Graphical User Interface – GUI*) а также модели. Кроме этого могут использоваться стандартные средства проектирования и объектно-ориентированного моделирования. Основу платформы составляют сервисы: извлечения данных (*data mining*), автоматического построения моделей (*model mining*), трансформации моделей, генерации модулей загружаемых архитектур, сервисы-модели, сервисы-движки, сервисы доступа к данным и др.

В платформу заложена возможность на основе когнитивных моделей генерировать модули целевой системы, использующие более простые решения, например, на основе автоматов или нейронных сетей. Если имеются жесткие требования по времени, то автоматы зашиваются в код, а в более сложном случае, при изменении контекста, их можно повторно генерировать и загружать в форме модулей. Отдельные модули реализуют критичные ко времени функции, которые могут генерироваться из архитектурного описания, как в статике, так и в динамике. Для загрузки модулей возможно предусмотреть отдельные временные слоты. В последнем случае следует говорить об адаптивной (*agile*) архитектуре [18]. Предложенный подход к использованию моделей определим как *Cognitive Supported Architecture (CSA)*. Перспективы практического применения этого подхода определяются доступностью сервисов трансформации, которые дают возможность переходить от моделей в форме знаний к другим описаниям, таким как автоматы, таблицы, деревья решений, и другим моделям, позволяющим получить высокое быстродействие.

Предлагаемый подход направлен на интеграцию доступных решений и позволяет использовать КСМ-платформу как средство разработки и моделирования. В настоящее время она находится в стадии проектирования.

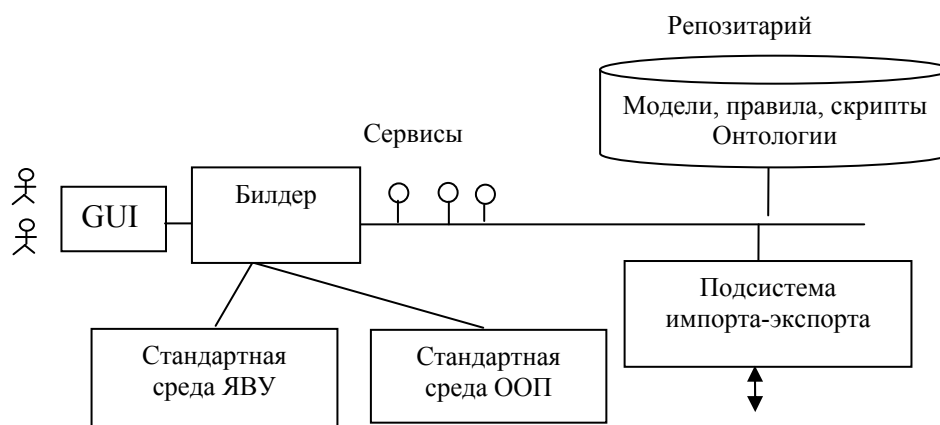


Рис. 5. Обобщенная структура КСМ-платформы



## Пример когнитивной системы мониторинга ресурсов в сетях кабельного телевидения

В качестве одного из примеров практического применения описанного выше модельного подхода при построении КСМ приведём систему оперативно-го управления сетями кабельного телевидения.

**Обобщенная структура.** Системы кабельного цифрового телевидения (СКЦТ) – одно из традиционных направлений развития телекоммуникационных технологий. Современные СКЦТ – это сложные распределенные системы, насчитывающие сотни тысяч абонентов, они включают как оборудование, установленное у пользователей сети передачи данных, так и серверное оборудование, которое находится в Центрах сбора и обработки данных (ЦСОД) операторов и представляет собой кластеры серверов, осуществляющие сбор и передачу данных на устройства конечных пользователей. При этом сеть разделяется на сегменты, в каждом из которых имеются собственные серверы. На стороне абонента устанавливается ресивер цифрового телевидения (ТВ-тюнер, ТВ-приставка, *Set-top Box – STB*), обеспечивающий базовую и расширенную функциональность телевизора. Для управления СКЦТ используются системы оперативного мониторинга и управления сетями (*Operation Management System – OMS*) абонентов. Типовая структура СКЦТ показана на рис. 6. Центры управления осуществляют мониторинг ТВ-приставок сети для определения их текущего состояния и обеспечивают техническую поддержку. Кроме того, на них возлагается решение задач, связанных с распределением информационных и вычислительных ресурсов системы.

**Реализуемая функциональность.** Типовая система оперативного мониторинга решает две группы задач: задачи, связанные с технической поддержкой самой системы, и задачи, связанные со сбором статистики в интересах аналитических и маркетинговых отделов.

Первая группа задач включает: мониторинг; непрерывное отслеживание состояния сервиса; оперативное информирование оператора в случае возникновения сбоев и отказов; регистрацию возникающих ошибок, их идентификацию и локализацию; оценку возможных путей исправления ошибок; оперативное устранение неполадок; выявление причин возникновения проблемных ситуаций; анализ ситуации по сети в целом или по отдельным подсетям; выявление зависимостей между возникающими ошибками на различных компонентах, зависимостей возникновения ошибок и состоянием сети или действий пользователя; прогноз возникновения ошибок; раннюю диагностику и предотвращение возникновения ошибок.

Вторая группа – задачи, связанные со сбором и обработкой информации, относящейся к мониторингу деятельности пользователей (пользовательские предпочтения, стиль работы каждого отдельного пользователя, паттерны поведения пользователей и т.п.). Эти задачи, в большинстве своем, относятся к задачам бизнес-аналитики и нами рассматриваться не будут.

Основная задача первой группы этой группы – предоставление операторам информации о состоянии системы кабельного цифрового телевидения, обнаружение и локализация возникающих в ее среде неисправностей.

**Требования, предъявляемые к системе оперативного мониторинга:** возможность работы с большими данными, гибкая логика работы (возможность адаптации в режиме реального времени к состоянию среды, действиям пользователей), минимальные требования к техническим характеристикам ресивера *STB* и параметрам сети, высокие показатели готовности и т.п. При этом система оперативного мониторинга должна оперативно обновляться с учетом новой реализованной логики и новых потребностей в данных для решения аналитических задач.

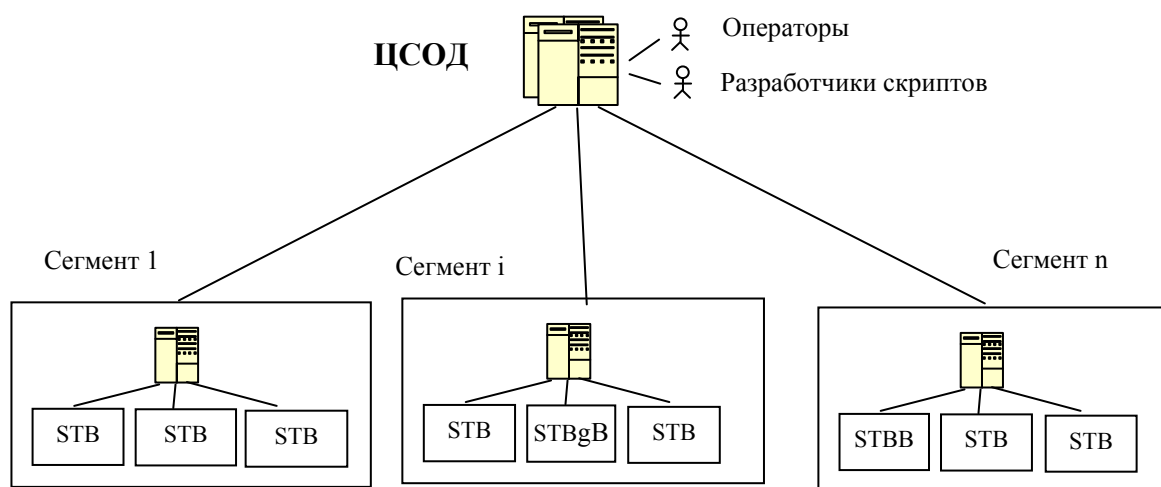


Рис. 6. Структура системы кабельного цифрового телевидения

*Типовые проблемы создания и использования системы оперативного мониторинга.* В процессе функционирования этой системы наблюдается высокая динамичность среды. Загрузка сети и устройств непрерывно меняется в зависимости от поведения пользователей, исправности технических средств, а прогноз состояния практически невозможен. Как следствие, вероятно ситуация, когда дополнительная нагрузка со стороны системы мониторинга приведет к перегрузке. Всегда есть опасность эффекта «лавин», когда при неполадке возникает значительное увеличение трафика локальным сервером и ресивером *STB* вследствие многократных попыток получения / передачи данных.

Разработчики системы оперативного мониторинга часто сталкиваются с проблемами:

- длительный сложный цикл внедрения программного обеспечения. Тестирование и поставка новой версии ПО занимает около шести месяцев. Следовательно, часто менять программный код нежелательно в связи со сложностью процесса и высокой стоимостью выпуска новых версий;

- масштабы и структура сети, актуальные задачи конкретного оператора определяют логику, которая должна реализовываться ресивером *STB*. Как следствие, требуется разработка многих модификаций систем сбора и анализа данных, при этом используется ограниченное число систем каждого отдельно типа. При проведении работ по проектированию и реализации каждой модификации системы потребуются неоправданно большой объем ресурсов.

*Возможные походы к решению задачи мониторинга состояния системы кабельного цифрового телевидения.* Задачу мониторинга состояния этой системы можно рассматривать как задачу установления причинно-следственных связей на множестве событий, где причина – это неисправность, а следствие – сообщение о событии, которое может быть либо послано *STB*, либо сформировано на серверной стороне как артефакт мониторинга. Сложность решения этой задачи применительно к СКЦТ состоит в том, что её структура фактически является динамической, поскольку постоянно изменяется не только по причине появления новых абонентов, но и из-за используемого на стороне сервера оборудования. Появляются новые типы *STB*, которые, с одной стороны, обладают большими функциональными возможностями, а с другой стороны, используют новые архитектурные решения (набор состояний, форматы сообщений). При этом далеко не всегда удается ограничиться разработкой драйвера. Довольно часто при появлении нового типа *STB* приходится разрабатывать диагностические скрипты и появляется потребность в их разработке, что приводит к повышению совокупной стоимости владения системой.

В связи с этим желательно автоматизировать процедуру составления скриптов, или, по крайней мере, предельно упростить её. Одним из возможных способов минимизации стоимости создания скриптов является использование доменно-ориентированных языков (*Domain Specific Languages – DSL*) [19].

*Онтологический подход.* Основная идея онтологического подхода состоит в том, что диагностический

скрипт формируется на базе онтологических описаний элементов, в частности, *STB*. При использовании онтологического подхода для добавления в систему нового типа элемента аналитик составляет его онтологическое описание и добавляет его в рабочую онтологию, которая выполняет функцию базы знаний. Онтологическое описание используется для генерации автоматной модели. В определенном смысле онтологический подход можно рассматривать как надстройку над другими подходами. Это, в первую очередь, касается автоматного подхода. Онтология при этом выступает как средство описания динамической структуры. Онтологический подход может также использоваться совместно с модельным подходом, при этом для генерации скриптов вместо объектных моделей используются онтологические модели.

Системы, использующие онтологии, не отличаются высокой скоростью работы, но для рассматриваемого нами случая это не является критичным.

Использование онтологического подхода, с одной стороны, упрощает работу аналитика, поскольку позволяет строить онтологические описания элементов на базе существующих, но с другой стороны, требует от аналитика навыков работы с онтологиями.

*Алгоритм работы системы диагностики.* Обобщенный алгоритм работы системы мониторинга при решении задач диагностики и устранения ошибочных ситуаций включает четыре шага.

Шаг 1. Выполнение алгоритма инициируется при возникновении ошибочной ситуации на ресивере *STB*.

Шаг 2. Определяются параметры, значения которых требуются для устранения ошибки в работе ресивера.

Шаг 3. На *STB* с использованием стандартных команд отправляются запросы на получение параметров. Результатом выполнения этих запросов являются информационные сообщения со значениями параметров или файлы. Проводится анализ полученных значений и при необходимости, формируются дополнительные запросы.

Шаг 4. На основании собранных данных определяется и реализуется путь устранения неисправности.

Формализованный алгоритм диагностики и устранения ошибочных ситуаций на ресивере *STB* представлен в [20]. Для применения математического аппарата при решении задач синтеза, а также поисковых и оптимизационных задач предложены показатели эффективности, система моделей описания данных. Детально рассмотрен этап определения параметров состояния *STB*, необходимых для идентификации, локализации и устранения ошибочной ситуации. Предложен алгоритм автоматизированного синтеза программ сбора дополнительных параметров с ресиверов для диагностики неполадки, построенный на основе аппарата условно-конечных операционных автоматов.

Для решения задач диагностики используются модели, описывающие ресиверы и среду, в которой они функционируют, а также действия конечных пользователей. Модели системы диагностики включают: функциональную модель – описывает функции, предоставляемые пользователям; модель программ-

ных компонентов – описывает программное обеспечение ресивера; модели параметров компонентов – определяют параметры состояния компонентов программного обеспечения.

В состав этих моделей входят: модель платформо-независимых параметров, модели платформо-зависимых параметров (определяется типом модели производителя *STB*), контекстные модели (модели среды и способов использования *STB*), определяющие контекст формирования алгоритмов мониторинга. К контекстным моделям относятся: модель сети передачи данных; модель поведения пользователя; модели состояния *STB* и их компонентов. Модели состояния *STB* описывают состояние ресивера для заданного контекста в соответствии с объемом поддерживаемого пользовательского функционала и модели ошибочных ситуаций [20].

Опыт разработки и применения системы кабельного цифрового телевидения продемонстрировал, что, несмотря на ограниченные когнитивные возможности, использование когнитивного подхода, основанного на применении моделей, приводит к минимизации совокупной стоимости системы. Тестирование системы мониторинга СКЦТ показало, что она позволяет создавать достаточно эффективные диагностические скрипты. Дальнейшее направление работ по совершенствованию этой системы связано с расширением её когнитивных возможностей и с созданием инструментальных средств, облегчающих работу аналитика и разработку механизмов учета положительного и отрицательного опыта. Следует отметить, что данная когнитивная система мониторинга СКЦТ находится на достаточно низком уровне зрелости, с точки зрения реализуемой когнитивности, в частности в ней заложены, но пока не реализованы механизмы обучения.

## ЗАКЛЮЧЕНИЕ

Когнитивные технологии находят все более широкое применение при решении реальных задач, но их использование еще не стало повседневной практикой. Не в последнюю очередь это касается когнитивных систем мониторинга, поскольку во многих случаях именно наличие этих систем делает сами системы когнитивными, т.е. реализующими элементы разумного поведения.

Одна из основных целей создания когнитивных систем – повышение уровня «интеллекта» систем мониторинга. Однако, это может быть достигнуто только за счет увеличения сложности используемых программных решений, что приводит к ужесточению требований к самой КМС-платформе. При этом увеличивается число уровней информационной системы, и, соответственно увеличивается сложность и стоимость её разработки. В этом контексте на первый план выступают задачи, связанные с разработкой промежуточного программного обеспечения, фреймворков и инструментария проектирования, использование которого превратит создание когнитивных информационных систем из некоторой экзотики в один из архитектурных стилей, что должно привести к уменьшению затрат на проектирование когнитивных систем и, соответственно, к расширению сферы их применения.

## СПИСОК ЛИТЕРАТУРЫ

1. Kelly J.E. Computing, cognition, and the future of knowing. IBM Research. – URL: <https://cra.org/crn/2016/09/computing-cognition-future-knowing-humans-machines-forging-new-age-understanding/>
2. Kelly J.E. Smart machines: IBM's Watson and the Era of Cognitive Computing. Columbia Business School Publishing. – URL: <https://www.delaat.net/smartnetworks/files/watson%20papers/146316241-Smart-Machines-IBM%E2%80%99s-Watson-and-the-Era-of-Cognitive-Computing.pdf>
3. Balani N. Cognitive IoT (2015). – URL: <http://navvenbalani.com/>
4. Schatsky D., Muraskin C., Gurusurthy R. Cognitive technologies: the real opportunities for business // Deloitte review. – 2015. – № 16. – P. 115–129.
5. International Standard ISO/IEC/IEEE 42010 Systems and software engineering: Architecture description. – URL: <https://www.iso.org/standard/50508.html>
6. How is cognitive computing different from big data and NLP? – URL: <https://coseer.com/blog/how-is-cognitive-computing-different-from-big-data-and-nlp/>
7. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 3-rd ed. – NJ: Upper Saddle River, 2010. – 1132 p.
8. Sangaiah A.K., Thangavelu A., Sundaram V.M.S. Cognitive computing for Big Data systems over IoT. Frameworks, tools and applications. – Cham (Switzerland): Springer, 2018 – 375 p.
9. Bass L., Clements P., Kazman R. Software Architecture in Practice. 3-rd ed. – Upper Saddle River. – NJ: Addison-Wesley, 2013. – 661 p.
10. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга состояния и управления структурной динамикой сложных технических объектов. – М.: Наука, 2005. – 291 с.
11. Blasch E., Bosse E., Lambert D. High-Level information fusion management and system design. – Norwood, MA: Artech House Publishers, 2012. – 364 p.
12. Gasevic D., Djuric D., Devedzic V. Model driven architecture and ontology development. – Berlin-Heidelberg: Springer-Verlag, 2006. – 311 p
13. Sommerville I. Software engineering. – Boston, MA: Addison-Wesley, 2011. – 773 p.
14. Zaki M., Meira W. Data mining and analysis: fundamental concepts and algorithms. – Cambridge: Cambridge Univ. Press, 2014. – 606 p.
15. Aalst van der W. Process mining. Data Science in action. 2nd ed., Berlin-Heidelberg: Springer-Verlag. – 2016. – 467 p.
16. Осипов В.Ю. Автоматический синтез программ действий интеллектуальных роботов // Программирование. – 2016. – № 3. – С. 47–54.
17. Zivin B. E., Jouault J., Valduriez P. On the need for megamodels. – URL: <https://scinapse.io/papers/195085068>.
18. Babar M. A., Brown A. W., Mistrik I. Agile software architecture. – Waltham, MA: Elsevier, 2014. – 392 p.

19. Kelly S. Tolvanen J. Domain-specific modeling: Enabling full code generation. – London: John Wiley & Sons, 2008. – 340 p.
20. Водяхо А.И., Мустафин Н.Г., Жукова Н.А. Онтологический подход к построению систем мониторинга ресурсов в сетях кабельного телевидения // Известия СПбГЭТУ. – 2017. – № 2. – С. 29–38.

*Материал поступил в редакцию 11.01.19*

#### **Сведения об авторах**

**ВОДЯХО Александр Иванович** – доктор технических наук, профессор кафедры Вычислительной техники Санкт-Петербургского государственного электротехнического университета  
e-mail: aivodyaho@mail.ru

**ОСИПОВ Василий Юрьевич** – доктор технических наук, заведующий лабораторией информационно-вычислительных систем и технологий программирования Санкт-Петербургского института информатики и автоматизации Российской академии наук  
e-mail: osipov vasilii@mail.ru

**ЖУКОВА Наталия Александровна** – кандидат технических наук, старший научный сотрудник лаборатории информационно-вычислительных систем и технологий программирования Санкт-Петербургского института информатики и автоматизации Российской академии наук  
e-mail: nazhukova@mail.ru

**ЧЕРВОНЦЕВ Михаил Александрович** – главный конструктор проекта, Научно-инженерный центр Санкт-Петербургского электротехнического университета  
e-mail: chervontsev.mikhail@nicetu.spb.ru

## Академии наук стран американского континента: анализ информационного присутствия в интернет-пространстве

*Исследована проблематика, связанная с информационным присутствием в Интернете академий наук (АН) стран американского континента. Указаны юридические статусы этих АН, а также направления их взаимодействия с органами государственного управления, корпоративными структурами, национальными и международными научными организациями, отдельными исследователями и пр.*

*Подробно проанализирован состав размещенной на сайтах АН информации; особенности применяемых интерфейсов; возможности выбора языков; совокупность доступных сервисов; основные вебометрические показатели.*

*Сделаны выводы о роли сайтов рассмотренных АН в обеспечении доступности и распространения научно-технической информации, их значении в формировании национальных и международного научного информационного пространства, о потенциальной полезности сайтов этих АН для российских исследователей.*

**Ключевые слова:** страны американского континента, национальные академии наук, Интернет, информационное присутствие, сайты, информационное наполнение, агрегация информации, сервисы пользователей, языки интерфейсов, вебометрические показатели.

### ВВЕДЕНИЕ

Функционально-тематическое исследование номенклатуры и возможностей сайтов некоторых групп организаций, связанных с научной и научно-технической деятельностью, было выполнено авторами в работах [1–4], а в настоящей статье исследована проблематика, относящаяся к информационному присутствию в Интернете академий наук (АН) и функционально аналогичных им организаций в странах американского континента (кроме республики Куба, материал по которой уже был приведен в [4]). Рассмотрение этой проблематики имеет целью решение следующих задач. (1) Анализ зарубежного опыта управления научной деятельностью [5–7], в том числе и в отношении размещения информации о предоставлении грантов [8, 9] организациям, отдельным исследователям, студентам и др. В русскоязычной научной периодике из всех стран американского континента основное внимание уделяется опыту США [10–12], а другим странам посвящены лишь немногочисленные работы [13, 14]. (2) Информационно-аналитическая поддержка деятельности российских исследователей и научных организаций, включая обеспечение известности для них мест размещения полезной информации на зарубежных сайтах. В свою

очередь это необходимо для информационного обеспечения процессов инновационного развития экономики России [15]. (3) Поддержка процессов интеграции российских исследовательских организаций и отдельных ученых в международное научно-информационное веб-пространство [16]. Это пространство сейчас обычно трактуется как часть «академического» [17], включающего еще и различные образовательные интернет-ресурсы. (4) Оценка состояния и тенденций развития двусторонних связей Российской АН и входящих в нее организаций с академиями наук стран американского континента [18]. (5) Формирование международных инновационных и образовательных инфраструктур, в том числе для ассоциаций и консорциумов вузов из группы стран, входящих в БРИКС [19, 20]. (6) Повышение эффективности научной деятельности университетов России, в рамках международных ассоциаций и иных объединений вузов.

В имеющихся публикациях тематика, связанная с использованием для решения этих задач сайтов академий наук стран американского континента, практически не исследована. Поэтому цель настоящей статьи – попытка устранить указанный недочет.

## МАТЕРИАЛЫ И МЕТОДИКА ИССЛЕДОВАНИЙ, ОСОБЕННОСТИ ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИИ

Объектами анализа в нашем исследовании были академии наук (и некоторые смежные организации) стран американского континента; место этих АН в организации национальной и международной научной деятельности; информационное наполнение, структура и вебметрические показатели (ВМП) сайтов этих АН. Источниками приводимых далее данных были информация, размещенная на сайтах АН; результаты выдач поисковых систем Интернета и информационно-поисковой системы на сайте [www.elibrary.ru](http://www.elibrary.ru). При необходимости использовались сведения и из других источников, включая Википедию [20]. Фрагменты текста, заимствованные из Википедии, иных интернет-источников или научных статей, далее обычно даны курсивом и/или в кавычках.

Сведения о численности населения стран взяты с сайта <https://all-populations.com/ru/list-of-countries-by-population.html> на 2018-2019 гг. Однако прямые сопоставления численностей населения государств со сведениями о популярности сайтов АН могут быть не показательными по следующим причинам: помимо граждан этих стран указанные сайты востребованы и зарубежными пользователями; различные страны имеют разную структуру научных учреждений, поэтому различается и состав размещенной на сайтах АН информации; в национальной экономике рассматриваемых стран место научных исследований и разработок значительно различается.

Для анализа вебметрических показателей сайтов использовались внешние программные средства (ПС). Сами показатели и методики их исследования подробно описаны в [1–3], а применительно к деятельности сайтов библиотек – также в [22].

Оценки количества страниц сайтов, полученные с помощью информационно-поисковой системы *Google Chrome*, приведены внутри < > [2–4]. Далее в составе ВМП объемы сайтов, полученные с использованием внешних ПС, указываются округленно. Такие ПС не позволяют в полной мере оценить объемы и количества гиперссылок на сайтах в следующих случаях. (1) Вход на сайт защищен с помощью «логина-пароля», который пользователь получает при регистрации на сайте в индивидуальном порядке и, возможно, последующего одобрения модератором проведенной регистрации. При таком условии в автоматическом режиме внешние ПС оценить ВМП сайтов не могут вообще. (2) Вход в «личные кабинеты» отдельных исследователей или их групп на сайте защищен «логином-паролем». При оценках ВМП сайтов внешние ПС такие «кабинеты» обработать не могут. Как следствие – получаемые с их помощью ВМП сайтов (и, прежде всего их объемы) будут заниженными. (3) В некоторых случаях из-за применяемых на сайтах программных решений внешние ПС не позволяют оценивать количество входящих ссылок и некоторые иные вебметрические показатели. В таких случаях мы указываем «0» или «-» в соответствующих клетках таблиц.

Отсутствие модерации информации в «личных кабинетах» отдельных пользователей и/или их групп

представляет потенциальную угрозу в отношении возможностей распространения деструктивной или тенденциозной информации, в том числе направленной на подрыв усилий по международному научному сотрудничеству. Однако размещение такой информации на рассматриваемых в настоящей статье сайтах, судя по всему, не характерно.

В случае отдельных сайтов академий наук их ВМП в таблицах указываются без {}. Если у АН есть «группа страниц» на сайте «вышестоящей» организации, то в приводимых далее таблицах даются объемы «сайтов в целом», заключенные в {}. Внутри комбинации фигурных и угловых скобок {< >} указывается количество страниц для «сайта в целом», полученное с помощью *Google Chrome*.

### Общая характеристика академий наук и их сайтов для стран американского континента

Отметим, что термин «Академия» в рассматриваемых странах широко используется не только для собственно АН, но и для специализированных научных обществ, иногда – для образовательных организаций.

В общем случае юридический статус традиционных академий наук может быть различен. (1) АН являются автономными по отношению к государству некоммерческими (бесприбыльными) организациями. При этом по крайней мере часть финансов может поступать в АН от корпоративных структур, различных фондов и пр. Также АН могут принимать «пожертвования» (*donations*) от частных лиц, в том числе с использованием программных средств, размещенных на сайтах – как пример укажем Канадскую академию наук. (2) АН полностью или частично финансируются государством, но формально не входят в состав государственных структур. (3) АН (или аналогичные им организации, подразделения организаций) являются составной частью структур государственного управления в сфере научных исследований, разработок и образования.

В большинстве стран, по крайней мере Центральной и Южной Америки, основная масса научных исследований выполняется не в «академических институтах», а в университетах и автономных исследовательских центрах, в том числе имеющих государственное финансирование. Часть научных исследований осуществляется и в «корпоративных» исследовательских центрах (однако это не характерно для стран Центральной и, с оговорками, Южной Америки).

Сайты рассматриваемых в настоящей статье академий наук играют важнейшую роль в обеспечении их информационного присутствия в научно-информационном пространстве; в восприятии этих организаций национальными и зарубежными исследователями, другими лицами, а также в обеспечении информационного взаимодействия АН с ними. Как правило, эти сайты размещены в «национальных зонах» соответствующих стран. Подчиненные АН организации (институты, центры и пр.) могут иметь как группы страниц на сайтах АН, так и свои собственные (отдельные) сайты. На сайтах АН и/или подчиненных им организаций размещаются гиперссылки на сайты их изданий. Однако собствен-

ные научные журналы издают лишь немногие из рассматриваемых нами академий наук.

Основные функции сайтов рассматриваемых АН. (1) Представление информации об АН, их членах (действительных членах, членах-корреспондентах и пр.) – обычно с разбивкой по тематическим направлениям деятельности. Во многих случаях по гиперссылкам можно перейти на индивидуальные интернет-страницы членов АН, однако их личные адреса электронной почты на таких страницах, как правило, отсутствуют. В ряде случаев на сайтах АН приводится ретроспективная информация: о предыдущих президентах АН; о членах и членах-корреспондентах АН за последние годы или начиная с момента основания АН. (2) Агрегация научной и научно-технической информации, прежде всего созданной в рамках деятельности самих АН. (3) Обеспечение «видимости» информации, размещенной на сайтах АН, для поисковых систем Интернета на национальных и иностранных языках. Особо отметим «видимость» на сайтах информации, которая в дальнейшем отражается в показателях, определяемых с помощью информационно-поисковой системы ресурса *Scholar Google* – русскоязычная версия сайта «Академии Google» расположена по адресу <https://scholar.google.ru/>. (4) Обеспечение для интернет-пользователей доступности размещенной на сайтах информации, включая научную и научно-техническую. Для улучшения доступности, наряду с англоязычными сайтами, часто используются версии и на национальных языках. Однако часть информации на некоторых сайтах находится в «платном доступе». (5) Поддержка «связности» национального и международного научно-информационного пространства за счет наличия входящих и исходящих гиперссылок между сайтами, внутренних гиперссылок на сайтах, а также «тематических коллекций» гиперссылок на сайты организаций-партнеров рассматриваемых АН. (6) Обеспечение отдельных исследователей и научных организаций (национальных и зарубежных) необходимыми «коммуникативными площадками» для установления/поддержания научных контактов. (7) Информационно-техническая поддержка формирования и деятельности рабочих групп по конкретным направлениям исследований, подготовке и проведению научных мероприятий, выпуску научных изданий и пр. (8) Размещение информации о присуждении исследователям научных наград, медалей и пр. Это важно для стимулирования личной научной активности ученых, оценки «научным сообществом» результативности их творческой деятельности и пр. (9) Информационно-техническая поддержка грантовой деятельности, в том числе предоставления грантов для отдельных исследователей и студентов (например, в Бразилии). Однако во многих странах, включая США, фонды-грантодатели [8, 9] и иные организации, предоставляющие гранты, являются автономными по отношению к АН и имеют собственные сайты.

Структура и интерфейсы рассматриваемых сайтов АН в целом соответствуют современным тенденциям «сайтостроительства». В тоже время их интерфейсы, состав страниц, расположение на них информации, функциональных элементов, графических объектов и

т.п. в основном не унифицированы. Это затрудняет работу пользователей с сайтами разных академий, а также возможность создания программных средств автоматического мониторинга сайтов в отношении появления информации, представляющей интерес для конкретных пользователей [23]. Однако возможность «парсинга» содержимого сайтов (прежде всего в отношении появления новостной информации) имеется практически во всех случаях.

## Академии наук стран Северной Америки

### 1. США (население 330 604 тыс. чел.).

Национальная Академия наук США (*National Academy of Sciences*) – частная, некоммерческая организация. Члены АН США избираются и работают на «общественных началах» (на 01.01.2019 их было примерно 2200 граждан США и около 400 иностранных). Около 200 членов этой академии – Нобелевские лауреаты (это максимум среди всех АН в мире). Деятельность АН США финансируется фондами Форда, Рокфеллера и др., собственных исследований она не ведет и поэтому у нее нет подчиненных научных институтов и аналогичных им организаций. В некоторых публикациях эта АН трактуется как «консультативный орган по вопросам науки».

Сайт национальной АН США (<http://www.nas-online.org/>) в верхней части стартовой страницы имеет два горизонтальных меню-ленты (МЛ). В верхнем МЛ отметим пункты *Newsroom, Meetings and Events*. На втором МЛ есть вкладка *Publications*, при нажатии на которую открывается ряд пунктов: *Proceedings of the National Academy of Sciences* (можно получить доступ к содержаниям выпусков, но сами статьи находятся в платном доступе); *National Academy Press; Issues in Science and Technology* – ежеквартальное издание с собственным сайтом (<http://www.issues.org/>). В центральной части стартовой страницы (СС) под заголовком *Activities & Programs* расположен набор иконок с подписями-гиперссылками, по которым открываются страницы соответствующих «программ» со своими собственными меню. Внизу СС сайта находится копия второго горизонтального МЛ со всеми пунктами, отображающимися постоянно (т.е. не в виде выпадающего списка) и копия первого меню-ленты. Целью двукратного размещения этих МЛ на СС является, очевидно, повышение удобства доступа к ключевым интернет-страницам сайта.

Национальная Академия инженерных наук США (*National Academy of Engineering – NAE*) – позиционируется как «частное, независимое, некоммерческое учреждение, которое обеспечивает стране техническое лидерство». Сайт *NAE* (<https://www.nae.edu/>) имеет два горизонтальных МЛ. Верхнее из них содержит пункты: *Giving* (в отношении «инвестиций» в *NAE*); *About* (общая информация о *NAE*); *Contact* (контактная информация); *MediaRoom* (средство поиска по сайту). Во втором меню содержатся вкладки: (а) *Publications* – обеспечивается доступ, в основном, к внутренним изданиям этой академии и к некоторым статьям обзорного характера. Доступа к статьям в научных журналах с этой страницы сайта нет; (б) *Events*; (в) *Activities*; (г) *Members* – есть, в частности, информация о недавно умерших ученых (с разбивкой

по тематическим направлениям их деятельности и датами смерти). В нижней части СС сайта дублируется второе МЛ с воспроизведением всех пунктов для вкладок. Кроме того, слева в дополнительном меню размещены пункты, относящиеся непосредственно к сайту: *Terms of Use, Privacy Statement, DMCA policy, Low Graphics* (для отображения СС крупным шрифтом и с увеличенной контрастностью символов).

Национальная Академия медицины США (National Academy of Medicine – NAM) – является «независимой организацией выдающихся профессионалов из разных областей деятельности». Позиционирует себя как «консультант для нации и международного сообщества». Сайт (<https://nam.edu/>) имеет в верхней части СС два МЛ. На первом из них отметим вкладку *Events*, а на втором – вкладки *Programs, Initiatives, Perspectives, News*. Внизу СС есть возможность подписаться на получение информации АН по электронной почте (sign up for NAM email updates).

Существует также общий сайт этих трех академий США – <http://www.nationalacademies.org/>. Его основное меню – верхнее горизонтальное, содержащее пункты *Events and Activities, Resources* (из пунктов выпадающего списка обеспечивается доступ к некоторым публикациям), *Newsroom*.

Важную роль в США играет также Национальный научный фонд [8, 9] – независимое агентство при правительстве США. Этот фонд осуществляет финансовую поддержку развития науки и технологий путем предоставления грантов – в основном индивидуальным исследователям или их небольшим группам. В меньшей степени финансируются «развитие человеческих ресурсов» и «дорогостоящие исследования». Потенциальными получателями грантов фонда могут быть различные категории лиц, включая «undergraduate» и «postgraduate» students (по российской терминологии – студенты и аспиранты). В фонде имеется семь «директоратов» по тематическим направлениям.

Сайт фонда – <https://www.nsf.gov/>. Верхнее МЛ содержит пункты, по каждому из которых открываются выпадающие меню: *Research Areas; Funding; Awards; Document Library* – отметим, в частности, подпункты *All documents* (с собственной поисковой системой) и *Obtaining documents; News; About NSF*.

Для российских исследователей особый интерес представляет каталог гиперссылок (<https://www.nsf.gov/about/partners/external.jsp>) на многочисленные организации-партнеры фонда, а также на «тематические коллекции гиперссылок» в отношении организаций, работающих в определенных сферах науки.

Специально отметим пункт *Plain Language* в МЛ, расположенном в самом низу СС сайта. На интернет-странице, открывающейся при выборе этого пункта, поясняется, почему для представления информации используется «простой язык». На практике простота формулировок важна для их понимания не только коренными жителями США, но и лицами, недавно переехавшими в эту страну; зарубежными стажерами; студентами и др.

2. Канада (население 35 853 тыс. чел.). Королевское общество Канады (*Royal Society of Canada* –

*RSC*) выполняет функции АН и включает в себя три организации: Академия искусств и гуманитарных наук, Академия общественных наук, Академия наук (направление – точные, естественные и прикладные науки, технологии). Сайт Королевского общества Канады – <http://rsc-src.ca/>. При переходе по этому адресу появляется страница, содержащая только эмблему организации и средство выбора языка – английского (сайт – <http://rsc-src.ca/en>) или французского (сайт – <http://rsc-src.ca/fr>). После выбора открывается стартовая страница сайта, в верхней части которой имеются два горизонтальных МЛ. В первом МЛ присутствуют пункты *Giving and Funding, Member login* (вход для членов академии); *Contact us*.

Во втором МЛ размещены пункты: *News & Events, Research & Reports, Awards, Fellows & Members, About us*. Выбор каждого из них приводит не к появлению выпадающих списков, а к открытию отдельных страниц с цветовыми решениями, отличающимися от того, что используется на СС. В середине СС эти пункты верхнего меню дублируются. Пункты верхних меню-лент и содержание середины СС продублированы также в «нижнем меню» СС, представленном в виде колонок с несколькими пунктами.

В правой части СС расположена кнопка *Donate* (пожертвования), которая остается неподвижной при вертикальном скроллинге («прокручивании») страницы. Аналогичная кнопка (*Donate to RSC*) есть и внизу СС.

В нижней части СС сайта имеются: средство подписки (*Sign up to receive regular updates from RSC*); дополнительное меню с пунктами *Feedback, Privacy Policy, LinkedIn Group Rules, Contact Us, Media, Site Map*; кнопки для перехода на страницы ведущих социальных сетей, включая *LinkedIn*.

3. Мексика (Мексиканские Соединённые Штаты) – население 124 884 тыс. чел. Согласно данным из Википедии в этой стране на долю государства приходится «свыше 90% расходов на науку», в том числе на государственные НИИ, исследовательские учреждения вузов. К государственным относятся «академические научные учреждения: Национальная академия наук, Академия юридических наук, Мексиканская академия истории, Национальная академия истории и географии, Мексиканская академия языка, Национальная академия медицины». Однако доля этих академий в общем объёме проводимых в стране научно-исследовательских работ, согласно сведениям из Википедии, незначительна.

Сайт мексиканской АН (<http://www.amc.unam.mx/>) только испаноязычный. В верхнем МЛ наиболее полезны для настоящей статьи пункты (на испанском): *Programas; Assuntos internacionais* (открывается список академий и фондов); *Publicaciones*. При выборе последнего пункта открывается список книжных публикаций академии, в том числе достаточно давних по времени выхода «в свет». Для некоторых элементов списка присутствуют гиперссылки на pdf-файлы.

Основные вебметрические показатели для сайтов этих трех стран представлены в табл. 1 и 2. Данные в этих и последующих таблицах относятся ко второй половине января 2019 г.



## Вебметрические показатели для сайтов АН Северной Америки (первая часть)

Первая часть таблицы							
№	Название ресурса	АТ, сек.	Count, URLs			Size, Mb	AVD, чч:мм:сс
			text/html	image	application		
<b>США</b>							
1	Национальная АН США <a href="http://www.nasonline.org/">http://www.nasonline.org/</a>	0,63	1633	1673	569	665	0:06:22
2	Национальная Академия инженерных наук США <a href="https://www.nae.edu/">https://www.nae.edu/</a>	10,9	34158	3364	1150	8331	0:08:57
3	Национальная Академия медицины США <a href="https://nam.edu/">https://nam.edu/</a>	1,07	15980	4676	1257	2860	0:17:45
4	Общий сайт трех АН США <a href="http://www.nationalacademies.org/">http://www.nationalacademies.org/</a>	16,2	95906	1584	1068	7021	0:25:55
5	Национальный научный фонд США <a href="https://www.nsf.gov/">https://www.nsf.gov/</a>	6,83	7491	515	611	721	0:15:21
<b>Канада</b>							
6	Королевское общество Канады <a href="http://rsc-src.ca/en">http://rsc-src.ca/en</a> (англ. яз.)	16,71	2	46	0	219	0:01:02
7	Королевское общество Канады <a href="http://rsc-src.ca/fr">http://rsc-src.ca/fr</a> (фр. яз.)	18,15	1252	162	13	43,3	0:07:02
<b>Мексика</b>							
8	Мексиканская АН <a href="http://www.amc.unam.mx/">http://www.amc.unam.mx/</a>	1,28	2	0	0	0,001	0:00:02

Таблица 2

## Вебметрические показатели для сайтов АН Северной Америки (вторая часть)

№	Название ресурса	Абсолютные количества ссылок				КУнПос. за месяц	КПросм., месяц	Scholar Google (SG)
		Входящие	Внутр	Исх				
				Ω	Ψ			
<b>США</b>								
1	Национальная АН США <a href="http://www.nasonline.org/">http://www.nasonline.org/</a>	7132	98	35	98	19 441	77 760	1 650
2	Национальная Академия инженерных наук США <a href="https://www.nae.edu/">https://www.nae.edu/</a>	4103	252	0	0	12 259	49 050	2 020
3	Национальная Академия медицины США <a href="https://nam.edu/">https://nam.edu/</a>	2307	32	12	31	15 312	61 260	1 090
4	Общий сайт трех АН США <a href="http://www.nationalacademies.org/">http://www.nationalacademies.org/</a>	14859	50	51	50	61 214	244 860	13 300
5	Национальный научный фонд США <a href="https://www.nsf.gov/">https://www.nsf.gov/</a>	42707	187	21	171	146 091	584 370	37 400
<b>Канада</b>								
6	Королевское общество Канады <a href="http://rsc-src.ca/en">http://rsc-src.ca/en</a> (англ. яз.)	1039	42	4	39	3 571	14 280	183
7	Королевское общество Канады <a href="http://rsc-src.ca/fr">http://rsc-src.ca/fr</a> (фр. яз.)		39		36			
<b>Мексика</b>								
8	Мексиканская АН <a href="http://www.amc.unam.mx/">http://www.amc.unam.mx/</a>	361	0	0	0	662 681	2 650 710	141

Из данных табл. 1 можно сделать следующие выводы: имеется значительный (более чем на порядок) разброс по длительности открытия стартовых страниц рассматриваемых сайтов; высокие и весьма высокие средние длительности сеансов работы пользователей на сайтах АН США, могут свидетельствовать о комплексном характере размещенной на этих сайтах информации и удобстве пользования ими.

Из данных табл. 2 можно сделать следующие выводы. 1) По количеству входящих ссылок абсолютно преобладает сайт Национального научного фонда США. Это, вероятно, является следствием высокой востребованности информации о грантах в англоязычном научно-информационном пространстве. На втором месте по количеству входящих ссылок находится общий сайт трех АН США. На сайтах других АН США входящих ссылок значительно меньше. 2) Сайт Национального научного фонда США и общий сайт трех АН лидируют и по таким показателям, как: общее количество просмотров и количество уникальных посетителей за месяц; количество документов, которые обнаруживает на них ИПС ресурса *Scholar Google*. 3) Для сайтов АН всех трех стран Северной Америки количество исходящих ссылок на порядки меньше чем входящих. 4) Для АН Мексики, несмотря на малый объем сайта (см. табл. 1), месячные количества «уникальных посетителей» и «просмотров» весьма высокие. Это может быть связано либо с методическими недочетами в использованных программных средствах аналитики, либо с интересом к сайту испаноязычных пользователей – вероятно, из Латинской Америки.

## **Академии наук стран Центральной Америки и Карибского бассейна**

Прежде всего, отметим, что в некоторых из этих стран академий наук просто нет. В других странах они существуют (упоминаются в русскоязычных текстах, посвященных конкретным странам), но ссылок на их сайты в Рунете найти не удается.

Республика Гаити (9 894 тыс. чел.) – информации о научных учреждениях в Рунете найти не удалось, но есть сведения о нескольких университетах, расположенных в столице страны.

Республика Гватемала (25 176 тыс. чел.) – укажем сайт «академической сети» ([www.vanderbilt.edu/gsn/](http://www.vanderbilt.edu/gsn/)), который, судя по заголовку на стартовой странице сайта, принадлежит образовательной организации *Vanderbilt College of Arts and Science*

Республика Гондурас (8 725 тыс. чел.). Функции управления наукой выполняет, видимо, Министерство образования (сайт [www.oei.es/noti30.htm](http://www.oei.es/noti30.htm)). Упоминаний об АН Гондураса в Рунете найти не удалось.

Доминиканская республика (10 649 тыс. чел.). В Рунете упоминаются «Доминиканская академия» и «Доминиканская академия истории», адреса сайтов которых не указываются.

Республика Коста-Рика (4 773 тыс. чел.). На русскоязычных интернет-ресурсах упоминаются «Национальная Академия наук» и «Академия истории и языка», но без адресов сайтов.

Республика Никарагуа (6 198 тыс. чел.). В Рунете упоминается Академия наук, основанная в 1934 г., но без адреса сайта.

Республика Панама (3 764 тыс. чел.). На русскоязычных интернет-ресурсах есть сведения о наличии в стране «Панамской академии», «Панамской академии истории», «Национальной академии наук Панамы», «Национального индеанистского института». Однако адресов сайтов этих организаций не приводится.

Республика Парагвай (7 113 тыс. чел.). В русскоязычном сегменте Интернета для этой страны упоминаются «Национальная Академия изящных искусств», «Академия Наук», «Академия языка и культуры гуарани» – но также без адресов сайтов этих организаций.

Восточная Республика Уругвай (3 416 тыс. чел.). В Рунете сведений о наличии в этой стране АН, в традиционном понимании, не встречается, но упоминаются «Национальная Академия литературы» и «Национальная Академия медицины» – сайт [https://web.archive.org/web/20170701100323/](https://web.archive.org/web/20170701100323/http://www.anm.org.uy/), <http://www.anm.org.uy/>.

Республика Эквадор (15 839 тыс. чел.). Научные центры связаны преимущественно с университетами страны, но есть и государственные НИИ. Академии наук в традиционном понимании – нет, но в названиях некоторых научно-исследовательских институтов «не государственного характера» присутствует термин «академия». В частности «Академия Медицинских наук» и «Эквадорский институт политической экономии», имеют *независимый* (не государственный) статус. Адреса сайтов этих организаций в Рунете не встречаются.

Ямайка (2 930 тыс. чел.). В русскоязычном Интернете упоминаний о сайте АН найти не удалось. Судя по всему функции «управления наукой» выполняют следующие организации: Министерство образования (сайт [www.moe.gov.jm](http://www.moe.gov.jm)), Совет по научным исследованиям (доступ по адресам [www.src-jamaica.org](http://www.src-jamaica.org) и [src.gov.jm](http://src.gov.jm)), Национальная научно-техническая комиссия (сайт [www.ncst.gov.jm](http://www.ncst.gov.jm)).

Для некоторых других стран Центральной Америки и Карибского бассейна (Республика Эль-Сальвадор и др.) в русскоязычном Интернете не удалось найти интернет-сайтов, принадлежащих АН или «органам управления наукой», но имеются ссылки на сайты отдельных университетов.

Как и следовало ожидать, размеры сайтов организаций, представленных в табл. 3, – относительно невысоки. Исключением является сайт Министерства образования Ямайки, на котором содержится не только научно-техническая, но и много иной информации. Для этого сайта отметим также высокое значение AVD.

По количеству просмотров и уникальных посетителей за месяц в табл. 4 лидирует сайт «академической сети» Гватемалы. Это, вероятно, происходит из-за большого количества посещений сайта лицами, получающими образование. В тоже время популярность у пользователей сайта Министерства образования Ямайки сравнительно невысока – несмотря на его большой объем (см. табл. 3).

**Вебметрические показатели для сайтов АН и «смежных с ними» организаций  
в некоторых странах Центральной Америки и Карибского бассейна (первая часть)**

Первая часть таблицы							
№	Название ресурса	АТ, сек.	Count, URLs			Size, Mb	AVD, чч:мм:сс
			text/html	image	application		
1	<b>Гватемала</b> – сайт академической сети <a href="http://www.vanderbilt.edu/gsn">www.vanderbilt.edu/gsn</a>	1,21	638	3	1	8,92	0:02:21
2	<b>Гондурас</b> – сайт Министерства образования <a href="http://www.oei.es/noti30.htm">www.oei.es/noti30.htm</a>	-	-	-	-	-	-
3	<b>Уругвай</b> – Национальная академия медицины <a href="https://web.archive.org/web/20170701100323/http://www.anm.org.uy/">https://web.archive.org/web/20170701100323/http://www.anm.org.uy/</a>	1,217	16	10	5	0,3	0:02:03
4	<b>Ямайка</b> – сайт Министерства образования <a href="http://www.moe.gov.jm">www.moe.gov.jm</a>	16,14	2016	1030	127	860,6	0:10:33
5	<b>Ямайка</b> – сайт Совета по научным исследованиям <a href="http://www.src-jamaica.org">www.src-jamaica.org</a>	0,86	17	45	25	2,8	0:03:32
6	<b>Ямайка</b> – сайт Национальной научно-технической комиссии <a href="http://www.ncst.gov.jm">www.ncst.gov.jm</a>	26,87	101	150	112	102,7	0:06:21

Таблица 4

**Вебметрические показатели для сайтов АН и «смежных с ними» организаций  
в некоторых странах Центральной Америки и Карибского бассейна (вторая часть)**

№	Название ресурса	Абсолютные количества ссылок				КУнПос. за месяц	КПром., месяц	Scholar Google (SG)
		Входящие	Внутр	Исх				
				Ω	Ψ			
1	<b>Гватемала</b> – сайт академической сети <a href="http://www.vanderbilt.edu/gsn">www.vanderbilt.edu/gsn</a>	17817	74	9	71	321 171	1 284 690	87 500
2	<b>Гондурас</b> – сайт Министерства образования <a href="http://www.oei.es/noti30.htm">www.oei.es/noti30.htm</a>	5108	121	76	21	58 871	235 470	20 300
3	<b>Уругвай</b> – Национальная академия медицины <a href="https://web.archive.org/web/20170701100323/http://www.anm.org.uy/">https://web.archive.org/web/20170701100323/http://www.anm.org.uy/</a>	47894	22	134	7	-	-	52 100
4	<b>Ямайка</b> – сайт Министерства образования <a href="http://www.moe.gov.jm">www.moe.gov.jm</a>	172	292	0	0	1 508	6 030	84
5	<b>Ямайка</b> – сайт Совета по научным исследованиям <a href="http://www.src-jamaica.org">www.src-jamaica.org</a>	55	0	0	0	1 111	4 440	30
6	<b>Ямайка</b> – сайт Национальной научно-технической комиссии <a href="http://www.ncst.gov.jm">www.ncst.gov.jm</a>	51	98	8	96	-	-	27

### Академии наук в странах Южной Америки

Аргентина (Аргентинская республика) имеет население 55 132 тыс. чел. Национальная Академия наук Аргентины (НАНА) «поддерживается» государством. В работе [14] (и других статьяx тех же авторов) дана оценка места АН в «научном комплексе» Аргентины по различным группам «отраслей наук». Сайт академии имеет версии на испанском языке (<http://www.anc-argentina.org.ar/es/>) и на английском (<http://www.anc-argentina.org.ar/en/>). Отметим, что

«англоязычные» версии интернет-страниц, открывающиеся по гиперссылкам из стартовых страниц, часть текста имеют на испанском языке.

Верхнее меню-лента СС сайта, имеет относительно небольшие размеры. Для него отметим следующие элементы. (1) *Institutional*, включая пункт *Academics*, в котором фамилии и фото членов академии сегментированы по пяти категориям. (2) *Publications*, включает вкладки *Proceedings*, *Bulletins* и пр. Гиперссылки на сайты каких-либо конкретных научных журналов не просматриваются. (3) *Catalogue* – толь-

ко краткая информация о каталоге НАНА. (4) *Rooms*, включая пункты *Main Conference Room* и *Reading Room*. (5) *Activities*, включая пункты *Congresses*, *Conferences* и др. (6) *News*, включая информацию о конференциях (Однако 05.01.2019 сайт НАНА был недоступен).

Отметим также Национальный научно-технический исследовательский совет Аргентины (сайт <https://www.conicet.gov.ar/>). Его фактические функции во многом совпадают с функциями академией наук в традиционном понимании. В стране существует ещё Национальная Академия медицины Аргентины (<https://web.archive.org/web/20100216140311/http://www.acamedbai.org.ar/>).

Боливия – (официальное название — Многонациональное Государство Боливия) имеет население 17 411 тыс. чел. По данным сайта [https://latin\\_america.academic.ru/548](https://latin_america.academic.ru/548) Национальная Академия наук Боливии была создана в 1960 г. и «выполняет функции консультативного органа при правительстве и отвечает за разработку основ научной политики страны». Впрочем, АН Боливии имеет и собственную «научно-исследовательскую базу». В Рунете встречаются упоминания о «Боливийской академии», «Национальной академии истории», «Национальной Академии юридических наук Боливии», однако найти адреса сайтов этих организаций нам не удалось.

Федеративная Республика Бразилия. По численности населения (209 533 тыс. чел.) является крупнейшей страной Южной Америки. Ключевую роль в отношении управления наукой и ее финансирования играет «Национальный исследовательский совет» и «Национальное агентство по финансированию образования и исследований» (обе эти организации подчиняются Министерству образования и науки Бразилии) [15]. Считается, что основными исследовательскими центрами в стране являются университеты и, особенно, университет Сан-Паулу (сайт <https://www5.usp.br/>). Этот университет кроме г. Сан-Паулу имеет подразделения и в нескольких других городах страны.

Однако и Академия наук Бразилии играет в стране важную роль. Сайт АН имеет версии на португальском языке (<http://www.abc.org.br/centenario/>) и английском (<http://www.abc.org.br/en/>). По ссылке <http://www.abc.org.br/?-English-> открывается испаноязычная версия СС, с которой можно перейти на англо- или португалоязычные версии (кнопка перехода на испаноязычную версию на англо- и португалоязычных версиях сайта нет).

В верхней части СС сайта размещены два пункта меню, относящиеся к реализуемым «программам» исследований. Ниже располагаются еще два пункта, один из которых соответствует «контрактам», заключаемым АН.

Под этими пунктами располагается верхнее горизонтальное МЛ, при выборе пунктов в котором, появляются дополнительные горизонтальные меню, содержащие вкладки с пунктами. Отметим в этом меню следующие вкладки (в переводе на русский язык). (1) «Учреждение» (организация), включая пункты «Структура» (сведения о структуре академии и лицах, занимающих руководящие должности); «Память» (сведения по истории АН); «Награды и

отзывы». (2) «Члены» – сведения о членах АН, ее партнерах и пр. (3) «Повестка дня» – приводятся сведения о предстоящих мероприятиях. (4) «Новости». (5) «Публикации». Наиболее важны пункты дополнительного меню: а) «Стратегические исследования» – открывается доступ к странице с файлами, содержащими агрегированные материалы по ключевым направлениям исследований; б) «Журналы» – для этого пункта показываются списки не научных журналов, а отчетов, ежегодников и других материалов, публикуемых АН Бразилии.

В центральной части СС сайта представлена различная информация, относящаяся к конференциям, научным лекциям, научным проектам и пр., с использованием графических объектов и текстов. При условии регистрации пользователей имеется возможность подписки на рассылку «Бюллетеня АН Бразилии».

В нижней части СС дублируются элементы верхнего горизонтального МЛ, причем на экране отображается «подчиненное содержание» для всех элементов указанного меню. В верхней и нижней частях стартовой страницы сайта есть также гиперссылки для перехода на страницы социальных сетей.

В русскоязычном Интернете упоминается еще «Национальная Академия медицины Бразилии» (<http://www.anm.org.br/>).

Отметим важную роль сайта <http://www.scielo.br/> (*Scientific Electronic Library OnLine*). На нём накапливаются электронные версии работ, опубликованных в ведущих научных журналах Бразилии и ряда других стран – прежде всего, Латинской Америки.

В Боливарианской Республике Венесуэла (32 040 тыс. чел.) функционирует (согласно данным на [https://ru.wikipedia.org/wiki/Венесуэльская\\_академия\\_языка](https://ru.wikipedia.org/wiki/Венесуэльская_академия_языка)) пять академий наук, размещенных в одном здании (Дворце Академий): «Венесуэльская Академия языка» (<http://avelengua.org.ve/>), «Национальная академия истории» (<http://www.anhvenezuela.org.ve/>), «Национальная академия медицины» (<http://www.anm.org.ve/anm/saciindex.php>), «Академия политических и социальных наук» (<http://www.acienpol.org.ve/>), «Академия физико-математических и естественных наук» (<https://www.acfiman.org/>). Сайты всех этих организаций – испаноязычные, причем их оформление различается.

Венесуэльская Академия языка – общественный культурный институт Венесуэлы по испанскому языку. В верхнем горизонтальном меню СС её сайта отметим такие вкладки. (1) «Физические лица». (2) «Члены-корреспонденты» – национальные и зарубежные. (3) «Почетные члены». (4) «Публикации» – для этой вкладки открывается достаточно короткий список, содержащий ссылки как на отдельные бюллетени, так и на каталоги. (5) «Новости» – содержит достаточно разнородную информацию, связанную с научной деятельностью.

В левом вертикальном МЛ на стартовой странице сайта дублируются названия вкладок из верхнего горизонтального меню. При этом для каждой вкладки отображаются все входящие в неё элементы – в виде гиперссылок на соответствующие страницы сайта. Например, для вкладки «Публикации» такими

элементами являются (в переводе на русский язык) «Информационные бюллетени», «Авторы», «Различные публикации».

На СС сайта Венесуэльской Академии физико-математических и естественных наук (ACFIMAN) в верхнем МЛ для темы статьи наиболее важны пункты (в переводе на русский). (1) «Академия» (официальные сведения об организации и отчеты по управлению). (2) «Ученые», включает и подпункт «умершие ученые». (3) «Программы» – с достаточно большим количеством подпунктов, включая «сеть молодых исследователей». (4) «Публикации», включая подпункты «Информационный бюллетень ACFIMAN», «Книги и другие материалы». (5) «Награды», включая подпункт «Женщины в науке». Гиперссылки на конкретные научные журналы на этом сайте не просматриваются.

Республика Колумбия (49 334 тыс. чел.). По сведениям из Интернета, научная деятельность сосредоточена, в основном, в университетах. В частности «Национальный Университет Колумбии» (г. Богота) имеет пять научно-исследовательских институтов. Также в Интернете встречаются упоминания об академиях гуманитарного направления: «Колумбийская академия истории», «Картахенская академия истории», «Колумбийская академия юридических наук». Выявить для Колумбии сайт АН «политематического» или «естественнонаучного» характера в Рунете не удалось. Поэтому был рассмотрен сайт «Колумбийского института оценки образования» (<http://www.icfes.gov.co/>). В его вкладке «Исследова-

тели и аспиранты» отметим пункты: «Календарь важных дат» и «Исследовательские проекты».

В Рунете есть также ссылка на «Национальную Академию медицины Колумбии» (сайт <http://anmde-colombia.net/>), но 05.01.2019г эта ссылка не работала.

Республика Перу (39 489 тыс. чел.). В русскоязычном Интернете упоминается «министерство образования», которое в определенной степени выполняет функции управления наукой (сайт [www.minedu.gob.pe](http://www.minedu.gob.pe)). Действует также «Национальная Академия медицины» (сайт <http://www.acadnacmedicina.org.pe/>).

В республике Чили (18 572. тыс. чел.) «Институт Чили», выполняет функцию объединения шести тематически ориентированных академий наук, каждая из которых имеет свой собственный сайт. Со СС каждого из этих шести сайтов по гиперссылкам можно перейти на сайты других пяти академий, но 28.06.2018 полностью открывались не все стартовые страницы этих сайтов. Внешний вид СС всех шести сайтов в значительной степени унифицирован. На СС сайтов в левом вертикальном МЛ есть пункт «Публикации», который призван обеспечивать доступ к спискам «бюллетеней» по тематике деятельности соответствующих академий. Подчеркнем, что все шесть указанных сайтов на 05.01.2019 г. были размещены на [www.arxiv.org](http://www.arxiv.org).

Существует также сайт АН с адресом <http://www.academia-ciencias.cl/>, однако он содержит только относительно небольшое количество новостной информации.

Таблица 5

**Вебометрические показатели для сайтов АН и «смежных с ними» организаций в некоторых странах Южной Америки (первая часть)**

Первая часть таблицы							
№	Название ресурса	АТ, сек.	Count, URLs			Size, Mb	AVD, чч:мм:сс
			text/html	image	application		
1	<b>Аргентина</b> – Национальная академия наук <a href="http://www.anc-argentina.org.ar/es/">http://www.anc-argentina.org.ar/es/</a> (исп. яз.)	4,05	461	0	125	10	0:06:10
2	<b>Аргентина</b> – Национальная академия наук <a href="http://www.anc-argentina.org.ar/en/">http://www.anc-argentina.org.ar/en/</a> (англ. яз.)	6,68	2528	0	760	59	0:11:59
3	<b>Бразилия</b> – Академия наук <a href="http://www.abc.org.br/?-English-">http://www.abc.org.br/?-English-</a> (англ. яз.)	1,04	18957	468	4472	3140	0:27:36
4	<b>Бразилия</b> – Академия наук <a href="http://www.abc.org.br/centenario/">http://www.abc.org.br/centenario/</a> (порт. яз.)	1,30	55	0	0	2	0:00:12
5	<b>Чили</b> – Институт Чили <a href="http://www.academia-ciencias.cl/">http://www.academia-ciencias.cl/</a>	0,96	18	20	23	3	0:00:08
6	<b>Венесуэла</b> – Венесуэльская Академия языка <a href="http://avelengua.org.ve/">http://avelengua.org.ve/</a>	3,65	467	246	46	24	0:02:02
7	<b>Колумбия</b> – Колумбийский институт оценки образования <a href="http://www.icfes.gov.co/">http://www.icfes.gov.co/</a>	1,93	3455	339	3600	15273	0:05:36
8	<b>Перу</b> – сайт Министерства образования <a href="https://www.gob.pe/minedu">https://www.gob.pe/minedu</a>	1,238	1633	55	79	62	0:03:01
9	<b>Перу</b> – сайт Национальной Академии медицины <a href="http://www.acadnacmedicina.org.pe/">http://www.acadnacmedicina.org.pe/</a>	0,449	99	65	544	926	0:00:16

**Вебметрические показатели для сайтов АН и «смежных с ними» организаций  
в некоторых странах Южной Америки (вторая часть)**

№	Название ресурса	Абсолютные количества ссылок				КУнПос. за месяц	КПросм., месяц	Scholar Google (SG)
		Входящие	Внутр	Исх				
				Ω	Ψ			
1	<b>Аргентина</b> – Национальная академия наук <a href="http://www.anc-argentina.org.ar/es/">http://www.anc-argentina.org.ar/es/</a> (англ. яз.)	96	58	2	57	1 172	4 680	10
2	<b>Аргентина</b> – Национальная академия наук <a href="http://www.anc-argentina.org.ar/en/">http://www.anc-argentina.org.ar/en/</a> (исп. яз.)							
3	<b>Бразилия</b> – Академия наук <a href="http://www.abc.org.br/?-English-">http://www.abc.org.br/?-English-</a> (англ. яз.)	1275	106	9	96	5 184	20 730	2 950
4	<b>Бразилия</b> – Академия наук <a href="http://www.abc.org.br/centenario/">http://www.abc.org.br/centenario/</a> (порт. яз.)		91	11	83			
5	<b>Чили</b> – Институт Чили <a href="http://www.academia-ciencias.cl/">http://www.academia-ciencias.cl/</a>	99	36	18	41	1575	6300	36
6	<b>Венесуэла</b> – Венесуэльская Академия языка <a href="http://avelengua.org.ve/">http://avelengua.org.ve/</a>	31	43	0	43	1791	7170	6
7	<b>Колумбия</b> – Колумбийский институт оценки образования <a href="http://www.icfes.gov.co/">http://www.icfes.gov.co/</a>	1363	0	0	0	70448	281790	1 680
8	<b>Перу</b> – сайт Министерства образования <a href="https://www.gob.pe/minedu">https://www.gob.pe/minedu</a>	763	60	73	57	-	-	3860
9	<b>Перу</b> – сайт Национальной Академии медицины <a href="http://www.acadnacmedicina.org.pe/">http://www.acadnacmedicina.org.pe/</a>	0	46	2	46	2240	8970	78

В Южной Америке есть ещё две небольшие страны, для которых в Рунете информацию о наличии в них АН или иных академий найти не удалось – Кооперативная Республика Гайана (802 тыс. чел.); Республика Суринам (548 тыс. чел.). В последней работает университет, в рамках которого действуют пять НИИ. Также в Южной Америке имеются три «зависимые территории» с невысокой численностью населения.

На основании данных из табл. 5 можно сделать следующие выводы. 1) По объёму (размеру) преобладает сайт Академии наук Бразилии на английском языке. Португальскоязычная версия этого сайта имеет весьма скромные размеры, хотя португальский – государственный язык страны. Для португальскоязычной версии средняя продолжительность работы пользователей с сайтом невелика (12 сек.), поэтому можно предположить, что они быстро переключаются на англоязычную версию. 2) Преобладание по размеру англоязычной версии (по сравнению с испаноязычной) отметим и для сайта АН Аргентины. 3) Очень высокий объём (размер) сайта «Колумбийского института оценки образования» может быть связан с тем, что на нем размещено много материалов, не от-

носящихся к НТИ. 4) Отметим относительно большой объём сайта «Национальной Академии медицины Перу» и низкую среднюю продолжительность нахождения на нем пользователей – всего 16 сек. 5) По величине средней продолжительности работы пользователей с сайтом абсолютное лидерство у АН Бразилии (для сайта на англ. яз.).

Выводы из данных табл. 6: 1) По количеству входящих ссылок в этой группе стран лидируют сайты АН Бразилии (на англ. яз.) и «Колумбийского института оценки образования». Однако для АН Бразилии количество входящих ссылок значительно меньше, чем для сайтов АН США (см. табл. 2) – при сопоставимой численности населения этих двух стран. 2) По количеству просмотров (визитов) абсолютное лидерство у сайта «Колумбийского института оценки образования». Однако, очевидно, что большое количество посещений этого сайта связано с многочисленностью учащихся в этой довольно крупной (примерно 50 млн чел.) стране. 3) Для сайта Бразильской АН его ежемесячную востребованность (посещаемость) с учетом численности населения этой страны (более 200 млн) следует считать весьма невысокой. 4) Невысоки показатели «посещаемости» и для сайтов других АН в

странах Южной Америки – всего по несколько тысяч просмотров в месяц. 5) Количество документов, обнаруживаемых ИПС ресурса *Scholar Google* на сайтах АН, включенных в табл. 6, также можно считать достаточно небольшими.

## ВЫВОДЫ

1. В большинстве крупных стран американского континента существуют академии наук. Однако основная масса исследований и разработок выполняется вне рамок этих академий – прежде всего в университетах и их подчиненных структурах, в автономных исследовательских центрах, в том числе имеющих бюджетное финансирование со стороны государств.

2. Для сайтов академий наук стран американского континента, адреса которых удается найти в Рунете, отметим очень малое количество ссылок на издаваемые научные журналы. Зато в ряде случаев подробно отражена тематика, связанная с представлением грантов, в том числе и студентам.

3. Новостная тематика, а также информация о проведении научных мероприятий, присутствуют на всех рассмотренных сайтах.

4. Также типичным для сайтов является использование «личных кабинетов» (или «кабинетов для членов»), вход в которые защищен «логинами-паролями».

5. Как правило, на сайтах академий наук кроме версий на национальных языках есть и англоязычные версии. При этом объемы (размеры) англоязычных версий сайтов обычно значительно преобладают. Вебметрические показатели рассмотренных в настоящей статье сайтов АН значительно различаются и они не пропорциональны численностям населения стран. Такая ситуация является, очевидно, следствием разной востребованности сайтов АН со стороны национальных исследователей и международного научного сообщества.

## СПИСОК ЛИТЕРАТУРЫ

1. Брумштейн Ю.М., Васьковский Е.Ю. Анализ вебметрических показателей основных сайтов, агрегирующих политематическую научную информацию // Научно-техническая информация. Сер. 2. – 2017. – № 11. – С. 16-32; Brumshiteyn Yu.M., Vas'kovskii E.Yu. Analysis of the Webometric Indicators of the Main Websites that Aggregate Multithematic Scientific Information // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 6. – P. 250-265.
2. Брумштейн Ю.М., Васьковский Е.Ю. Исследование функциональности и вебметрических показателей специализированных сайтов, связанных с научной деятельностью // Научно-техническая информация. Сер. 2. – 2018. – № 1. – С. 16-30; Brumshiteyn Yu.M., Vas'kovskii E.Yu. Studying the Functionality and Webometric Indicators of Specialized Science-Related Websites // Automatic Documentation and Mathematical Linguistics. – 2018. – Vol. 52, № 1. – P. 7-23.
3. 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; Brumshiteyn 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. – 2018. – Vol. 52, № 4. – P. 157-174.
4. Брумштейн Ю.М., Васильев Н.В. Анализ структуры сайтов Академий Наук некоторых зарубежных стран // Научно-техническая информация. Сер. 2. – 2019. – № 1. – С. 16-30.
5. 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.
6. Пипия Л.К., Дорогокупец В.С. Изменение подходов к управлению наукой в индустриально развитых странах // Наука за рубежом. – 2017. – № 58. – С. 1-35.
7. 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.
8. Райнхардт Р.О. Роль и место национальных научных фондов в обеспечении кадрового потенциала науки (на примере США, Германии и России) // Вестник Академии. – 2017. – № 1. – С. 92-98.
9. Рогов С.М., Супян В.Б., Рей А.И., Супян В.Б., Бибикова Т.С. Национальный научный фонд и национальные Академии Наук США: роль и место в развитии научного потенциала страны / руководитель и отв. ред. д.э.н. проф. В.Б. Супян. – М.: ИСКРАН, 2015. (Институт Соединённых Штатов Америки и Канады РАН).
10. Истомина И.А. Государственная поддержка научных исследований в США: современный этап эволюции в контексте политической борьбы // Сравнительная политика. 2016. – Т. 7, № 3(24). – С. 121-134.
11. Коннов В.И., Балышев А.В. Научная политика США: от концепций к практикам // Вестник Российского университета дружбы народов. Серия: Международные отношения. – 2012. – № 2. – С. 5-12.
12. Райнхардт Р.О. Взаимосвязь между финансированием науки в США и численностью американского научного сообщества: опыт национального научного фонда // Научный диалог. – 2016. – № 10(58). – С. 261-273.
13. Петровский А.Б., Проничкин С.В., Стернин М.Ю., Шепелёв Г.И. Организация и управление наукой: опыт Бразилии // Труды Института системного анализа Российской академии наук. – 2017. – Т. 67, № 1. – С. 41-53.
14. Дорогокупец В.С., Пипия Л.К. Оценка научного комплекса Аргентины: физико-математические науки // Наука за рубежом. – 2016. – № 51. – С. 1-52.

15. Сютнюрено О.В., Гиляревский Р.С. Задачи информационного обеспечения инновационного развития экономики и роль инжиниринга // Научно-техническая информация. Сер. 1. – 2017. – № 5. – С. 1-14.
16. Жуковская И.Ф. Факторы и риски интеграции российских интеллектуальных ресурсов в глобальную инновационную систему // Вестник Ивановского государственного университета. Серия: Экономика. – 2016. – № 3(29). – С. 19-30.
17. Клименко О.А. Модели представления академического веб-пространства // Информационные и математические технологии в науке и управлении. – 2016. – № 2. – С. 103-110.
18. Лавёров Н.П., Евсеев В.В., Шиян Ю.К. Контроль над вооружениями: 30 лет сотрудничества академий наук России и США // Вестник Российской академии наук. – 2012. – Т. 82, № 7. – С. 653.
19. Леонова Т.Н., Маланичева Н.В. Создание совместной инновационной инфраструктуры стран БРИКС // Вестник Университета (Государственный университет управления). – 2018. – № 2. – С. 27-30.
20. Меморандум о создании ассоциации научно-образовательных центров БРИКС. – URL: <http://old.fgr.msu.ru/2016/06/14/5896/76/> (дата обращения 05.01.2019).
21. Комаров С.Ю. Википедия как средство продвижения информационных ресурсов // Библиосфера. – 2012. – № 5. – С. 38-40.
22. Редькина Н.С. Направления развития инструментов веб-аналитики // Научно-техническая информация. Сер. 2. – 2017. – № 5. – С. 5-10; Redkina N.S. The Development Tendencies of Web Analytics Tools // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 3. – P. 112-116.
23. Кравец А.Д., Петрова И.Ю., Кравец А.Г. Агрегация информации о перспективных технологиях на основе автоматической генерации интеллектуальных агентов мультиагентных систем // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 4(32). – С. 141-148.

*Материал поступил в редакцию 07.02.19.*

#### **Сведения об авторах**

**БРУМШТЕЙН Юрий Моисеевич** – кандидат технических наук, доцент, доцент Астраханского государственного университета  
e-mail: [brum2003@mail.ru](mailto: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](mailto:nikivas@mail.ru)  
ORCID <http://orcid.org/0000-0002-4937-3305>



# АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

---

УДК 81'322.2:[025.43:(088.8)]

С.О. Шереметьева

## Экстракция многокомпонентных терминов и ключевых слов из многоязычной патентной документации

*Представлены метод и инструмент экстракции многокомпонентных терминов и ключевых слов для обработки патентной документации на разных языках. Инструмент, состоящий из двух модулей – экстрактора именных групп и модуля взвешивания, не требует предварительно построенного корпуса документов предметной области, работает на текстах любого размера, не пропускает низкочастотные единицы, обеспечивает высокую скорость обработки текстов и может быть адаптирован к различным предметным областям, типам лексики и языкам, включая русский.*

**Ключевые слова:** автоматическая экстракция, ключевые слова, патенты, многокомпонентные термины, многоязычность

### ВВЕДЕНИЕ

Разработка методов автоматической экстракции лексических единиц различного типа, к числу которых относится ключевая и неключевая терминология, сегодня является предметом многих научно-исследовательских и опытно-конструкторских проектов. Решение этой задачи напрямую связано с повышением уровня всех видов автоматической обработки информационных потоков: информационного поиска, индексации и категоризации документов, автоматического аннотирования и реферирования, построения терминологических баз данных и т. д. Особо значимым, хотя и более сложным, является автоматическое извлечение многокомпонентных ключевых терминов, значительно повышающих точность решения задач обработки текста [1, 2]. Еще более проблематичными оказываются попытки автоматического извлечения лексических групп определенного типа, например, именных групп, которые наилучшим образом отражают контент документа [3].

Качество автоматической экстракции многокомпонентных ключевых слов имеет большое значение для оперативной обработки патентов, как основного объекта интеллектуальной собственности, стимулирующего прогресс во всех технологических сферах деятельности [4–9]. Несмотря на большое количество работ, автоматическое извлечение терминологии и ключевых слов всех типов представляет серьезную

проблему, которая до сих пор не решена. Это отрицательно влияет на экспертизу русскоязычных патентов, поскольку, за исключением очень немногих работ [10–11], подавляющее большинство исследований по проблеме лексической экстракции, даже отечественными авторами, выполняется для английского языка.

В настоящей статье делается попытка внести вклад в решение проблем обработки патентной информации на различных языках, включая русский, на основе автоматической экстракции многокомпонентной именной терминологии. Предлагается универсальный метод решения этой проблемы и его реализация в виде программного инструмента извлечения ключевой терминологии, состоящего из двух модулей – экстрактора многокомпонентных именных групп и модуля взвешивания. Первый модуль не зависит от процедур взвешивания и может быть реализован в виде отдельного инструмента. Модули экстрактора не требуют предварительно построенного корпуса патентных документов, хорошо работают на текстах любого размера, не пропускают низкочастотные уникальные единицы, имеют вычислительно привлекательные свойства и могут быть достаточно просто адаптированы к различным предметным областям, лексическим типам и языкам, включая русский.

## ОБЗОР ОСНОВНЫХ МЕТОДОВ ЛЕКСИЧЕСКОЙ ЭКСТРАКЦИИ

Анализ литературы показывает, что приемы экстракции ключевой и неключевой лексики, в том числе многокомпонентной, при всем их разнообразии, имеют много общего, а именно, используют статистику и включают два основных этапа обработки текстового материала: формирование множества «кандидатов» и фильтрацию этого множества для получения результирующего ранжированного лексического списка. При этом в перечень процедур, используемых в разной комбинации на указанных этапах, входят:

- отфильтровывание стоп-слов, к которым относят либо не несущие никакой смысловой нагрузки слова (артикли, предлоги, союзы, частицы, местоимения, вводные слова, междометия и т. д.) [12], либо слова с низкой для конкретной предметной области информационной значимостью [13];
- нормализация лексем путем морфологической лемматизации или эвристического стемминга для подсчета общей частоты слов, имеющих общий корень;
- автоматическое разбиение текста на предложения для расчета веса слов;
- фильтрация и взвешивание кандидатов в ключевые слова для оценки их релевантности. При этом частотность как единственный параметр релевантности ключевого слова, несмотря на его широкое использование в практических приложениях, не считается эффективной. Не редко именно уникальные термины более точно сигнализируют о теме документа, например, о новизне изобретения в патентных документах [8].

Самой популярной мерой для учета частотности и уникальности лексем текста является произведение  $tf * idf$ , где  $tf$  – частота слова (термина) в документе,  $idf$  – обратная частота слова в коллекции документов. Последнее означает количество документов в корпусе, где термин употреблен, по крайней мере, один раз [14].

Методики экстракции отличаются одна от другой конкретными комбинациями указанных процедур, техниками обработки текста, количеством необходимых знаний и обязательным [14] или необязательным [15, 16] использованием корпуса соответствующей тематики. Более корректные результаты дают гибридные методики экстракции, в которых статистические методы обработки документов дополняются одной или несколькими лингвистическими процедурами (морфологическим, синтаксическим и семантическим анализами) и лингвистическими базами знаний различной глубины (словарями, онтологиями, Википедией, и т. д.) [16–19]. Разрабатываются методы экстракции ключевых слов на основе графов [20], машинного обучения [21] и нейросетей [22]. В качестве меры релевантности популярно использование комбинации значений различных параметров, например, частоты лексемы в документе/корпусе, ее расположения в определенной части текста, статистики совместной встречаемости слов в документе/корпусе, их дисперсии [15] или отношения логарифмического правдо-

подобия [14]. В работах по экстракции многокомпонентных ключевых слов, как правило, двухкомпонентных [1, 2, 10, 17, 18, 21–23], понятие лексической группы рассматривается статистически или грамматически. При статистическом подходе лексическая группа определяется как последовательность слов без ограничений на синтагматику, а экстракция основана на вычислении  $n$ -грамм [24] или совместной встречаемости слов в тексте [14].

При извлечении грамматически правильных ключевых лексических групп, например, именных групп, статистика дополняется лингвистическими методами, например, тэгами частей речи и мелким парсингом. При этом достаточно часто идентификация лексической группы как именной и отнесение ее к ключевым лексемам представляет две последовательные процедуры. Например, в [23] экстракция ключевых именных групп включает поиск в тексте документа базовых именных групп с использованием морфо-синтаксического анализа на основе словарей, а затем вычисление их релевантности и выделение ключевых именных групп.

Конкретные гибридные методики экстракции именных групп отличаются порядком применения статистических и лингвистических процедур и глубиной последних. Например, достаточно часто используется предварительно тэггированный корпус. На его основе в одних случаях, используя статистическую информацию о совместной встречаемости, строится множество кандидатов. К этому множеству для исключения «плохих» кандидатов применяется парсинг. В других случаях лингвистические знания в виде ограниченного числа синтаксических шаблонов или парсинга применяются уже на первом этапе обработки текста для построения множества кандидатов, которое затем фильтруется с использованием статистической информации о встречаемости шаблонов, синтаксических конфигураций или логарифмического отношения правдоподобия [16].

В [17] предлагается сочетать статистические приемы поиска многокомпонентных лексических групп с семантическим тэггером, что вызывает проблемы покрываемости. При разработке методов экстракции из патентных документов делаются попытки получить более релевантные результаты на основе лингвистических ограничений подязыков предметных областей и структурных особенностей патентов [4–9].

## ПОСТАНОВКА ЗАДАЧИ ИЗВЛЕЧЕНИЯ КЛЮЧЕВЫХ СЛОВ

При выборе подхода к извлечению ключевых слов при обработке информации, в нашем случае патентной, необходимо решить вопросы:

- о текстовой единице, которую следует рассматривать в качестве ключевого слова для патентного анализа;
- об идентификации этой единицы в необработанном патентом тексте;
- о метриках, используемых для ранжирования ключевых слов по их релевантности для патентного документа;

- об уменьшении вычислительных затрат на процедуру извлечения ключевых слов для обеспечения более высокой скорости обработки документов;
- о том, как сделать процедуру экстракции достаточно универсальной с возможностью ее адаптации к патентным документам на разных языках и, таким образом, значительно сократить время и усилия при разработке новых инструментов.

Принимая решения по сформулированным вопросам, мы руководствовались тем, что в сфере профессиональной патентно-информационной деятельности окончательное суждение о релевантности анализируемого патента всегда выносит патентный аналитик. Это означает, что полученное в результате автоматической экстракции ключевое слово должно быть грамматически корректной единицей, а не усеченной строкой букв, как это часто бывает.

Поддерживая мнение о том, что контент документа наилучшим образом отражается в многокомпонентных именных группах [24], в качестве ключевых мы выбрали единицы именно этого типа лексических групп. Отметим, что такие группы в патентном тексте считаются терминами.

Меры для ранжирования ключевых именных групп патента по релевантности должны как основываться на комбинации нескольких параметров, так и учитывать количественные характеристики лексических единиц (например, частоту, длину фразы и др.) и их принадлежность к определенным структурным частям патентного документа. Таким, как ОПИСАНИЕ, подробно представляющее изобретение, ФОРМУЛА ИЗОБРЕТЕНИЯ, которая является наиболее важной частью патента, так как имеет юридическое значение и, как следствие, придает больший вес включенным в нее терминам. Кроме того, в процессе патентной экспертизы важны термины ПРЕАМБУЛЫ – части текста формулы изобретения, где излагается цель или назначение изобретения, что также должно учитываться при разработке мер взвешивания входящих в этот раздел терминов.

Что касается универсальности алгоритмов извлечения ключевых слов и отсутствия необходимости в трудоемких и времязатратных процедурах построения лингвистических баз знаний, то это относится к преимуществам статистических методик. Однако статистические методы в чистом виде не обеспечивают высокого качества результатов, требуемого при обработке патентной информации. При этом область их применения в значительной степени ограничена языками с бедной морфологией, такими как английский, где частотность словоформ одной лексемы велика. Статистические приемы, как правило, не пригодны для языков с богатой морфологией, в частности, для русского языка, где каждая лексема характеризуется большим количеством словоформ с низкой частотностью в каждом конкретном тексте. Высокое качество результатов можно было бы ожидать от лингвистических приемов обработки текста, но времязатратность и трудоемкость построения необходимых ресурсов делает их применение нереалистичным. Поэтому мы сделали выбор в пользу гибридной методологии экстракции, сочетающей статистические приемы с мелкими лингвистическими

знаниями, исходя при этом из недоступности корпуса патентных документов требуемой предметной области, что в общем случае соответствует реальной ситуации.

Наша система экстракции ключевых фраз состоит из двух основных последовательных процедур:

1. Экстракции именных групп.
2. Вычисления их релевантности и выделения ключевых именных групп.

Ниже проиллюстрируем нашу методику на примере экстрактора ключевых именных групп из англоязычных патентов (он был разработан первым), а затем продемонстрируем адаптацию англоязычного экстрактора к другим языкам, в первую очередь, к русскому.

## ИЗВЛЕЧЕНИЕ ИМЕННЫХ ГРУПП ИЗ ПАТЕНТНОГО ТЕКСТА

При разработке методологии извлечения именных групп из патентного текста мы руководствовались следующим:

- Именные группы распознаются на основе общепринятых критериев, касающихся их морфологического и синтаксического контекста.
- По умолчанию именные группы, которые в обрабатываемом тексте встречаются только в составе более длинных именных групп, не извлекаются.
- При экстракции не теряются низкочастотные и уникальные именные группы.
- Экстракцию следует осуществлять без опоры на корпус предметной области.
- Алгоритм экстракции должен быть по возможности универсален и прост в вычислениях.

Для выполнения указанных критериев разработана новая гибридная модель лексической экстракции, сочетающая статистические методы обработки текста и мелкие лингвистические знания.

Что касается конкретных приемов обработки, то, прежде всего, мы отказались от идеи начать процедуру экстракции с удаления слов, входящих в обычно используемый стоп-лист, поскольку это может привести к «плохим» комбинациям слов. Например, если рассматривать фрагмент англоязычного патента «...*a table in which the wireless location system continuously maintains a copy of the status of transmitters...*», удаление стоп-слов (выделены жирным шрифтом) на этапе предварительной обработки приведет к извлечению таких «именных фраз» как «*\*table wireless location system*» и «*\*copy status transmitters*». Подобные комбинации в нашей модели не отфильтровываются автоматически (см. далее), и могут иметь высокую частоту из-за специфики подязыка патентов.

В отличие от процедур, применяемых в большинстве экстракторов, мы не применяли морфологическую нормализацию на этапе предварительной обработки текста, а перенесли ее на последний этап экстракции, поскольку эвристическое определение основ (стемминг) может привести к извлечению неправильно комбинированных и/или усеченных строк, которые невозможно понять, что противоречит нашим намерениям извлекать грамматически корректные фразы, а лингвистическая лемматизация на ос-

нове правил на первом этапе обработки текста очень затратна в ресурсном и вычислительном отношениях. Отнесение лексической нормализации на последний этап экстракции именных групп, т.е. когда список кандидатов уже достаточно хорошо отфильтрован и представляет собой ненормализованные именные группы, дает возможность использовать упрощенный, ориентированный только на именные группы алгоритм нормализации. Спецификой подхода является вычисление  $n$ -грамм ( $n=1,2,3,4$ )<sup>1</sup> из *исходного текста* без предварительной обработки и последующая фильтрация кандидатов посредством *удаления  $n$ -грамм*, которые *не являются* грамматически корректными.

База лингвистических знаний модели включает в себя знания об ограничениях на структуру именных групп в конкретном языке, специальным образом отсортированные лексиконы словоформ патентной лексики (стоп-листы)<sup>2</sup>, статические и динамические правила фильтрации. Знания об ограничениях на структуру именных групп – это знания о том, какие части речи *не могут стоять* в начале, середине и конце таких групп. Например, для английского и русского языка эти знания во многом одинаковы: именная группа не может начинаться с предлога, союза и личной формы глагола (1); в ее середине не могут быть наречие и личная форма глагола (2); в конце недопустимы предлог, союз, прилагательное, наречие и личная форма глагола (3). Следующим компонентом базы знаний являются стоп-листы (1), (2) и (3), представляющие собой комбинации списков словоформ, отсортированных по их принадлежности к определенным частям речи.

Статические правила фильтрации отражают знания о грамматических ограничениях именной фразы в конкретном языке и выполняются над множеством вычисленных  $n$ -грамм путем сравнения их компонентов с соответствующими стоп-листами. Примеры правил приведены ниже:

Правило 1: Если первое слово в  $n$ -грамме совпадает со словом из (1), удалите  $n$ -грамму

Правило 2: Если среднее слово в  $n$ -грамме совпадает со словом из (2), удалите  $n$ -грамму

Правило 3: Если последнее слово в  $n$ -грамме совпадает со словом из (3), удалите  $n$ -грамму

Динамические правила фильтрации основаны на вычисляемых в процессе обработки текста статистических параметрах кандидатов в именные группы.

Экстракция многокомпонентных именных фраз начинается с простого вычисления  $n$ -грамм необработанного текста. При этом  $0 < n < 5$ . Затем применяются статические правила фильтрации и происходит формирование первого множества кандидатов. Преимуществами такой фильтрации являются простота реализации и исключение из обработки этапа предварительного тэгирования  $n$ -грамм и/или использования шаблонов именных фраз, что требует значи-

тельных лингвистических ресурсов. Каждое статическое правило, взятое отдельно, может «пропустить» некоторое количество «плохих» кандидатов, но последовательное применение правил к словам одной и той же  $n$ -граммы, как правило, компенсирует недостатки покрываемости стоп-листов. «Плохой» кандидат, пропущенный одним правилом, будет удален другим.

На следующем этапе экстракции именных групп из первого множества кандидатов создается матрица расширения, где для каждой «уцелевшей» 1-граммы формируется частотный список содержащих ее более длинных кандидатов. Матрица позволяет определить, «имеют ли право» кандидаты в именные группы, являющиеся частью более длинных кандидатов, быть включенными в окончательный результат. Для принятия решения по этому вопросу вводится критерий «Уникальность» ( $U$ ), который определяется как разность между частотой  $n$ -граммы и суммой частот ее  $(n + 1)$ -граммных расширений. Низкое значение  $U$  ( $U = 0$  или  $U < 0$ ) показывает, что кандидат вряд ли используется в тексте индивидуально и должен быть отброшен.

Второй список кандидатов подвергается дополнительной фильтрации повторным применением статических правил. Затем выполняется нормализация (приведение всех именных групп к единственному числу), которая осуществляется по сильно упрощенному алгоритму.

## ВЫЧИСЛЕНИЕ РЕЛЕВАНТНОСТИ ИМЕННЫХ ГРУПП

Выделение ключевых именных групп из списка всех извлеченных на первом этапе работы алгоритма лексем этого типа осуществляется на основе вычисления для каждой из них вектора релевантности  $R = (F, n, N, T, U, M, D, d, C, B)$ , где

$R$  – релевантность именных групп;

$F$  – их частота;

$n$  – длина (1–4 компонента);

$N$  – средняя частота именных групп определенной длины;

$T$  – количество самых частотных слов документа, содержащихся в именных группах. Наиболее частотными считаются 30% самых частотных из 1-грамм, полученных на первом этапе экстракции; этот прием достаточно часто используется при вычислении весов ключевых слов, см., например, [15];

$M$  – сумма частот слов, входящих в именную группу;

$U$  – уникальность;

$D$  – количество предложений в тексте;

$d$  – количество предложений, в которых данная именная группа встречается хотя бы один раз;

$C$  – вхождение именной группы в ФОРМУЛУ ИЗОБРЕТЕНИЯ (этот параметр показывает, что такая группа встречается в ФОРМУЛЕ ИЗОБРЕТЕНИЯ хотя бы один раз);

$B$  – вхождение именной группы в ПРЕАМБУЛУ (этот параметр показывает, что группа встречается в ПРЕАМБУЛЕ хотя бы один раз).

<sup>1</sup> 4-х граммные модели считаются наиболее пригодными для практических приложений, поскольку вычисление более длинных  $n$ -грамм на больших текстах может вызвать безостановочную работу алгоритма.

<sup>2</sup> Не стоп-слова из служебной лексики.

Значения вектора релевантности рассчитываются по эмпирически заданной формуле как комбинация математических операций (+, -, \*, / и *log*) над значениями параметров и любыми целыми коэффициентами. Формула вычисления вектора релевантности может содержать все или отдельные параметры. Значения параметров *F*, *n*, *N*, *T*, *U*, *M*, *D*, *d* вычисляются динамически в процессе экстракции именных групп, в то время как *C* и *B* имеют фиксированную произвольно заданную величину.

Фильтрация именных групп для включения их в список ключевых выполняется на основе определенного порогового значения, установленного для вектора *R*, который, в свою очередь, может быть задан в виде определенного процентного отношения наиболее релевантных именных групп к общему числу всех именных групп или в виде фиксированного числа, например, 10% или 7 наиболее релевантных. При этом ключевые именные группы можно вычислять как на всем тексте патента, так и на его отдельных частях, например, только на основе текста ФОРМУЛЫ ИЗОБРЕТЕНИЯ, даже если он небольшой, поскольку ни алгоритм экстракции, ни вектор релевантности не зависят от длины текста.

Практика показывает, что экстракция ключевых именных групп из патентного документа осуществляется наиболее корректно, если задавать не одну, а четыре формулы релевантности: отдельно для однокомпонентных и многокомпонентных групп для разделов ОПИСАНИЯ и ФОРМУЛЫ ИЗОБРЕТЕНИЯ.

Поскольку вектор релевантности вычисляется с учетом принадлежности именной группы к одному из разделов патентного документа – ОПИСАНИЕ, ФОРМУЛА ИЗОБРЕТЕНИЯ и ПРЕАМБУЛА (параметры *C* и *B*), а также с учетом количества предложений в тексте и вхождений в них этой группы (параметры *D* и *d*), в процесс экстракции включены процедуры сегментации патентного документа, что на определенных этапах может создавать проблемы.

Сегментация патента на ОПИСАНИЕ и ФОРМУЛУ ИЗОБРЕТЕНИЯ не вызывает ошибок благодаря стандартной структуре патента. ФОРМУЛА ИЗОБРЕТЕНИЯ англоязычного патента всегда начинается с сигнальных фраз «*We claim*», «*I claim*», «*claims*», «*what we claim is*» и т. д., которые можно исчислить. Однозвенная ФОРМУЛА ИЗОБРЕТЕНИЯ и пункты многозвенной формулы всегда формулируются в виде одного предложения, при этом в случае многозвенной ФОРМУЛЫ выделение отдельных пунктов также может быть выполнено с высокой надежностью на основе исчисляемых сигнальных фраз.

Сегментация ПРЕАМБУЛЫ в тексте ФОРМУЛЫ ИЗОБРЕТЕНИЯ более проблематична. Мы разработали специальную процедуру, включающую определенные лингвистические знания. Что касается сегментации части ОПИСАНИЕ на предложения, то наличие в ней специальных символов, запятых, точек, заглавных букв и т. д., часто не являющихся сигналами о начале или конце предложений, делают сегментацию предложений крайне ненадежной. Поэтому мы не включаем параметры *D* и *d* в вычисление векторов релевантности именных групп части ОПИСАНИЕ.

Таким образом, алгоритм экстракции ключевых многокомпонентных именных групп включает процедуры:

- ВЫЧИСЛЕНИЕ *N*-ГРАММ.
- ФИЛЬТРАЦИЯ именных групп.
- Первая фильтрация кандидатов именных групп на основе статических правил.
- Построение матрицы расширения.
- Вторая фильтрация именных групп-кандидатов на основе динамически вычисляемого параметра уникальности с помощью *U*- критерия.
- Третья фильтрация именных групп посредством повторного применения статических правил.
- ВЫЧИСЛЕНИЕ РЕЛЕВАНТНОСТИ именных групп до лемматизации.
- ЛЕММАТИЗАЦИЯ.
- ПЕРЕСЧЕТ РЕЛЕВАНТНОСТИ именных групп после лемматизации.
- ВЫВОД КЛЮЧЕВЫХ именных групп.

После лемматизации в случае появления одинаковых именных групп дубликаты убираются, а релевантность каждой группы пересчитывается с учетом значений релевантности всех ее лемматизированных текстовых форм. Лемматизированные списки ключевых именных групп ОПИСАНИЯ и ФОРМУЛЫ ИЗОБРЕТЕНИЯ затем объединяются, и значение релевантности для совпадающих групп пересчитывается еще раз. Полученный список ранжируется по релевантности и может быть выведен полностью или вырезан с определенным порогом.

Описанный алгоритм экстракции ключевых именных групп реализован в виде программных инструментов на языке *C++* для операционных сред *Windows* и *Linux*, запускаемых через программную строку. Программы экстракции могут быть адаптированы к формату базы патентов и автоматически обрабатывать и, например, индексировать ключевыми фразами поток вновь поступающих патентов. Время вычисления ключевых именных групп по разработанному методу зависит от многих факторов, включая нагрузку на сервер, скорость сети и размер входного текста. Размер патентов варьируется от нескольких килобайт до 1,5 мегабайт. На обычном персональном компьютере процесс обработки одного патента, как правило, занимает доли секунды. XML-файл объемом 8 мегабайт, содержащий 150 патентов, обрабатывается менее чем за полторы минуты.

## МЕЖЪЯЗЫКОВАЯ АДАПТАЦИЯ МОДЕЛИ ЭКСТРАКЦИИ

Описанная методология экстракции именных групп может быть легко перенесена на новые языки, предметные области и адаптирована к извлечению других типов лексических групп, например, глагольных. Поскольку алгоритмы экстракции и вычисления векторов релевантности универсальны, возможность межъязыкового переноса определяется возможностью адаптации лингвистических знаний модели, что в связи с ограниченностью этих знаний представляется реальным. Необходимо знать ограничения, накладываемые на структуру интересующей пользователя лексической группы, что позволит создавать новые стоп-листы и лемматизатор. Лемматизатор –

это задача разработчиков, притом, что первые две из указанных задач (выявление ограничений на структуру лексических групп определенных типов и построение стоп-листов) могут быть решены пользователем. Например, учитывая, что ограничения, накладываемые на структуры английской, русской и немецкой именных групп практически одинаковы, при переносе экстрактора с английского на русский или немецкий языки достаточно английские слова в стоп-листах заменить на русские или немецкие, соответственно, принадлежащие к тем же самым частям речи, что и английские. При переносе на русский язык для упрощения процедуры нормализации используется эвристический алгоритм, который идентифицирует различные текстовые формы именных групп, принадлежащие к одной лексеме (например, текстовые формы «хороший стол», «хорошего стола», «хорошими столами» и т.д. принадлежат к одной

лексеме «хороший стол») и выбирает в качестве основной самую частотную из них. В табл. 1 показана адаптация англоязычного экстрактора к извлечению многокомпонентных именных групп, ранжированных по релевантности, из патентов на русском языке. Приведены результаты экстракции из текста одного русскоязычного патента, а также из короткого текста первого пункта многозвенной формулы изобретения этого патента: «Турбинная установка, содержащая: облопачивание, включающее криволинейные лопатки, внутренний конец каждой из которых заделан в полости, открытой с одной стороны; и генератор, расположенный в полости и соединенный с облопачиванием, причем каждая криволинейная лопатка имеет динамическую структуру и выполнена с возможностью увеличения в размерах под действием гидростатического давления и с возможностью сжатия под действием противодействия».

Таблица 1

**Фрагмент списка наиболее релевантных ключевых именных групп отечественного патента 2502890 и полный список именных групп первого пункта формулы изобретения этого патента.**

Отечественный патент 2502890	Первая формула патента 2502890
[88] турбинные установки	[10] Турбинная установка
[49] вал генератора	[8] криволинейная лопатка
[46] генератор	[6] внутренний конец
[46] облопачиванию	[5] действием гидростатического давления
[34] криволинейные лопатки	[5] облопачивание
[34] полость	[4] возможностью сжатия
[31] угол наклона	[4] возможностью увеличения
[30] турбине	[4] действием противодействия
[26] лопаток облопачивания	[4] динамическую структуру
[23] газ	[4] полости
[23] направлению потока текучей среды	[2] генератор
[22] внутренняя полость	[2] размерах
[22] зона протекания текучей среды	[2] стороны

Таблица 2

**Ключевые именные группы, извлеченные из патентных документов на английском, французском и немецком языках по одной методике**

English: US patent 7271765	French: WO/2003/012682	German: EP0317764
[2437] wireless location system	[[1829] système de gestion	[558] oberen Endbereich des Schafte
[1323] wireless transmitter	[1692] base de données topologique	[441] Halteelement
[1291] location processor	[1262] mémoire de graphe topologique	[172] hinteren Schaftteil
[1279] location record	[1171] modèle de graphe topologique	[143] Schieberichtung
[1258] location data signal	[1119] moyens de gestion	[104] Befestigungsanordnung
[1227] location estimation	[1102] table de type DIP	[91] Klemmeinrichtungen
[1200] location calculation	[1087] gestion de données	[88] angeordneten Halteelement
[1137] centralized database system	[1082] graphe représentatif de l'hypergraphe	[87] Halteelement angeformter Ansatz
[1070] location information	[1079] système de gestion coopérant	[84] Halteelement bei Skischuhen Verwendung
[1053] location processing	[1069] type de rôle	[83] umgreifenden Halteelement
[1051] location datum	[1061] type de donnée d'interface	[78] Schaft angeordneten Lageranordnungen
[1036] signal collection system		
[1012] wireless transmission		
[1000] location coordinate		

При переносе модели с английского на языки, отличающиеся накладываемыми ограничениями на структуры лексических групп, следует иначе комбинировать списки словоформ нового языка, предварительно отсортированные по частям речи<sup>3</sup>. Так, при переносе английского экстрактора на французский язык для извлечения именной терминологии необходимо учитывать, например, различный порядок наречий–прилагательных–существительных в английском и французском языках: *rigorously\_Adv constant\_Adj speeds\_N = vitesses\_N riouusement\_Adv constants\_Adj*. Результатом учета этой разницы будет включение в стоп-лист (1) наречий и прилагательных, поскольку они запрещены в начале французской именной группы, и, наоборот исключения из стоп-листа (3) прилагательных, поскольку они в конце французской именной группы разрешены. Мы провели пилотные эксперименты по переносу английского экстрактора ключевых именных групп на немецкий и французский языки. Предварительные результаты (табл. 2) доказывают перспективность нашего подхода.

Аналогичным образом решается задача настройки экстрактора на извлечение новой лексической группы в рамках одного языка, которая сводится к созданию новых комбинаций отсортированных по частям списков словоформ. Например, если необходимо извлечь глагольные группы, то стоп-лист (1) должен включать существительные и прилагательные, поскольку они не могут стоять в начале глагольной группы и т. д.

Для повышения качества работы новых вариантов экстракторов инструменты должны «натренировываться» для каждой предметной области. Это сводится к чистке и увеличению покрываемости стоп-листов, что можно сделать непосредственно через пользовательский интерфейс.

## ИНТЕРФЕЙС

Стартовая страница интерфейса экстрактора ключевых слов, настроенного на работу с англоязычными патентами показана на рисунке. Программная оболочка экстрактора универсальна и может быть наполнена знаниями для экстракции на различных языках.

По умолчанию в качестве конечного результата экстрактор выдает список одно-, двух-, трех- и четырехсловных именных групп с показателями их релевантности в обработанном тексте.

Результаты экстракции показаны в центральном окне правой панели и автоматически сохраняются в папке программы. Цифры в квадратных скобках указывают релевантность, вычисленную по формулам, заданным в соответствующих полях левой панели. Знак «\*» означает, что именная группа была найдена в ФОРМУЛЕ ИЗОБРЕТЕНИЯ и ОПИСАНИИ. Эта функция интерфейса может использоваться для проверки правильности составления текста патента: в соответствии с патентным законодательством ФОРМУЛА ИЗОБРЕТЕНИЯ не должна включать термины, не упомянутые в ОПИСАНИИ. Под окном

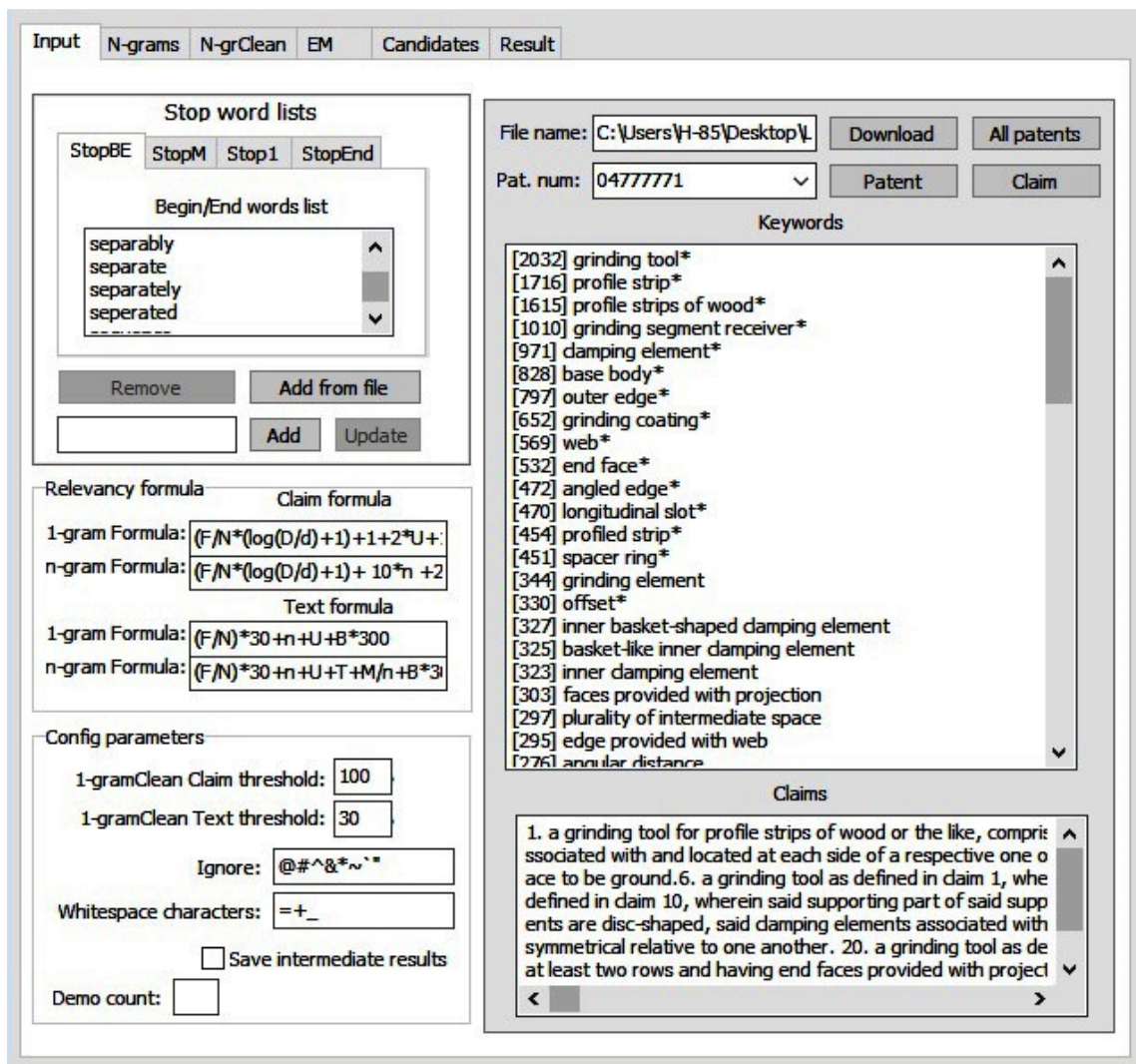
результатов для контроля расположено окно, визуализирующее текст формулы изобретения. Кнопки над окном ключевых слов позволяют загружать папки с большим количеством патентов и обрабатывать а) все патенты одновременно; б) один из патентов; в) только ФОРМУЛУ ИЗОБРЕТЕНИЯ определенного патента. Все результаты экстракции автоматически сохраняются в отдельных файлах с именами соответствующих патентов.

В верхней части левой панели находится встроенный интерфейс доступа к основному блоку лингвистических знаний – стоп-листам, которые визуализированы в верхнем окне с закладками *StopBE*, *StopM*, *StopI*, *StopEnd*. Закладки открывают для просмотра каждый из листов с возможностью их редактирования либо в пословном режиме с помощью кнопок «Remove» и «Add», блочно или целиком с помощью кнопки «Add from file». Пословное редактирование полезно, если, например, нужно исключить из результатов экстракции такие общие слова, как «system», «device» и т. д., поскольку иначе они будут «загрязнять», список ключевых именных групп из-за высокой частотности. Достаточно добавить эти слова в лист «*StopI*», который их «убьет».

Стоп-листы могут создаваться/редактироваться отдельно без интерфейса в текстовых файлах и затем загружаться в экстрактор через интерфейс. Закладки в верхней части интерфейса открывают страницы с промежуточными результатами процесса экстракции: списками «сырых» *n*-грамм с частотными показателями (закладка *N-grams*), результатами первой фильтрации кандидатов (*N-grClean*), интерактивной матрицей расширений (*EM*), результатами второй фильтрации (*Candidates*) и результатами третьей фильтрации, что необходимо для контроля корректности обработки при редактировании стоп-листов. На левой панели ниже окна администрирования знаний расположены интерактивные окна для задания параметров экстракции, что дает пользователю широкий спектр возможностей для экспериментов с пороговыми значениями и комбинациями параметров вектора релевантности.

Показанная на рисунке реализация интерфейса принимает на вход патенты в форматах их хранения в определенных базах, что может быть удобно, например, для консалтинговых компаний со своей базой патентов. Другая реализация интерфейса позволяет загружать папки или отдельные патенты в *txt*, *html* и *xml* форматах, обрабатывать их отдельно или (с помощью соответствующей кнопки интерфейса), предварительно объединив тексты в папках в один корпус и вычисляя для него ключевые слова. Последнее более удобно для индивидуального администрирования знаний, так как дает больше информации при их контроле и имеет больше настроек для сортировки и форматов выдачи результатов. И, наконец, третья «легкая» версия интерфейса реализована с фиксированными знаниями и настройками. Все варианты интерфейсов разработаны для операционной среды *Windows*.

<sup>3</sup> Это можно сделать на основе предварительного корпусного анализа.



Пользовательский интерфейс экстрактора ключевых именных групп с доступом к модулю адаптации знаний. Знания настроены на англоязычные патенты.

## ЗАКЛЮЧЕНИЕ

Разработанная методика экстракции, сочетающая статистические приемы обработки и неглубокие лингвистические знания, может давать хорошие результаты не только для языков с бедной морфологией, но и для высоко флективных языков, таких, как русский. Предложенная процедура экстракции обеспечивает лингвистически корректные результаты, не теряя низкочастотные лексемы. Адаптируемость экстрактора к различным языкам, предметным областям и типам лексических групп допускает его использование для обработки любых типов информации. В частности, он широко используется преподавателями, аспирантами и студентами Южно-уральского государственного университета при проведении лингвистических исследований.

## СПИСОК ЛИТЕРАТУРЫ

1. Liu Z., Huang W., Zheng Y., Sun M. Automatic keyphrase extraction via topic decomposition // Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. – Cambridge (MA), 2010. – P. 366–376.
2. Sag I.A., Baldwin T., Bond F., Copeland A., Flickinger D. Multiword expressions: A Pain in the neck for NLP // Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing CICLing '02. – London, UK, 2002. –P. 1–15.
3. Witschel Hans Friedrich. Terminology extraction and automatic indexing – comparison and qualitative evaluation of methods. (2005). – URL: <http://wortschatz.uni-leipzig.de/~fwitschel/papers/TKEIndexing.pdf>
4. Patent Keyword Extraction Algorithm Based on Distributed Representation for Patent Classification. – URL: <https://www.mdpi.com/1099-4300/20/2/104/pdf>
5. Kim, J., Choi, J., Park, S., Jang, D. Patent keyword extraction for sustainable technology management. – 2018. – URL: <https://www.mdpi.com/2071-1050/10/4/1287/htm>
6. Suzuki S., Takatsuka H. Extraction of keywords of novelties from patent claims. – URL: <https://www.mdpi.com/2071-1050/10/4/1287/htm>



- <https://pdfs.semanticscholar.org/05c7/e6cb9adbabd0e4bab919b60be507cf7a994a.pdf>
7. Li Y.-R., Wang L.-H., Hong Ch.-F. Extracting the significant-rare keywords for patent analysis // Expert Systems with Applications. – 2009. – Vol. 36, Issue 3, Part 1. – P. 5200–5204.
  8. Шабанов Д.В., Коробкин Д.М., Фоменков С.А., Колесников С.Г. Метод извлечения описаний технических функций из патентных текстов // Известия Волгоградского государственного технического университета. – 2018. – № 5(215). – С. 68–76.
  9. Fattori, M., Pedrazzi G., Turra R. Text mining applied to patent mapping: a practical business case // World Patent Information. – 2003. – Vol. 25. – P. 335–342.
  10. Браславский П., Соколов Е. Сравнение четырех методов автоматического извлечения двухсловных терминов из текста. – 2006. – URL: <http://www.dialog-21.ru/digests/dialog2006/materials/html/Braslavski.htm>
  11. Гайнанов Р.Р. Разработка системы автоматического извлечения ключевых слов из текста. – URL: [http://conf59.mipt.ru/static/reports\\_pdf/1640.pdf](http://conf59.mipt.ru/static/reports_pdf/1640.pdf)
  12. Salton G. Automatic text processing. – London: Addison-Wesley, 1988.
  13. Liu Y., Ciliax B.J., Borges K., Dasigi V., Ram A., Navathe S.B., Dingleline R. Comparison of two schemes for automatic keyword extraction from MEDLINE for functional gene clustering. – URL: <https://ieeexplore.ieee.org/abstract/document/1332452>
  14. Jones K.S. A Statistical interpretation of term specificity and its application in retrieval // Journal of Documentation. – 2004. – Vol. 60, № 5. – P. 493–502.
  15. Matsuo Y. Keyword extraction from a single document using word co-occurrence statistical information // International Journal on Artificial Intelligence Tools. – 2004. – Vol.13, № 1. – P. 157–169.
  16. Girish K.P. Keyword extraction from a single document using centrality measures. pattern recognition and machine intelligence. – New-York: Springer, 2007. – P. 503–510.
  17. Seretan V., Wehrli E. Multilingual collocation extraction with a syntactic parser // Language Resources and Evaluation. – 2009. – № 43(1). – P. 71–85.
  18. Piao S. L., Rayson P., Archer D., McEney T. Comparing and combining a semantic tagger and a statistical tool for MWE extraction // Computer Speech & Language. – 2005. – Vol. 19, Issue 4. – P. 378 – 397.
  19. Гринева М., Гринева М. Анализ текстовых документов для извлечения тематически сгруппированных ключевых терминов // Труды ИСП РАН. – 2009. – Т. 16. – С. 155–165.
  20. Litvak M. Graph-based keyword extraction for single-document summarization – URL: <https://dl.acm.org/citation.cfm?id=1613178>
  21. Turney P.D. Learning algorithms for keyphrase extraction // Information Retrieval. – 2000. – Vol. 2, № 4. – P. 303–336.
  22. Sarkar K., Nasipuri M., Ghose S. A New approach to keyphrase extraction using neural networks // International Journal of Computer Science Issues. – 2010. – Vol. 7, № 2. – P.16–25.
  23. Barker K., Cornacchia N. Using noun phrase heads to extract document keyphrases // Advances in Artificial Intelligence. – 2000. – Vol. 1822. – P. 40–52.
  24. Vadas D. Statistical Parsing of Noun Phrase Structure. – URL: [http://sydney.edu.au/engineering/it/~dvadas1/papers/vadas09\\_phd\\_thesis.Pdf](http://sydney.edu.au/engineering/it/~dvadas1/papers/vadas09_phd_thesis.Pdf)

*Материал поступил в редакцию 04.02.19*

#### **Сведения об авторе**

**ШЕРЕМЕТЬЕВА Светлана Олеговна** – доктор филологических наук, доцент Южно-уральского государственного университета, профессор кафедры лингвистики и перевода Института лингвистики и международных коммуникаций, директор Научно-образовательного центра лингво-инновационных технологий, г. Челябинск  
e-mail: linklana@yahoo.com

Г.И. Кустова

## Относительные наречия и относительные прилагательные по данным Национального корпуса русского языка\*

*Рассматриваются наречия на -о и -и (типа административно, коммерчески, планоно, туристически, финансово, химически и т.п.), которые не упоминаются в грамматиках как особый класс, а в статье названы относительными, так как они образованы от относительных отсубстантивных прилагательных. Показано сходство таких наречий по семантике и функциям с соответствующими относительными прилагательными и их отличие от других наречий на -о и -и – качественных (ср. наглядно, дружески и под.). Приведены данные о контекстах употреблений относительных наречий в Национальном корпусе русского языка.*

**Ключевые слова:** наречие, относительное прилагательное, актант, валентность аспекта, валентность сферы, обстоятельство образа действия

### КЛАССИФИКАЦИИ ПРИЛАГАТЕЛЬНЫХ И НАРЕЧИЙ

Наречия\* с субстантивными корнями, которые рассматриваются нами в настоящей статье (ср. административно, коммерчески, политически, туристически, финансово, химически и т.п.), морфологически образуются от относительных отсубстантивных прилагательных, связанных, в свою очередь, с существительными, и естественным образом наследуют от этих прилагательных какую-то часть их свойств и признаков.

Классификации прилагательных и наречий устроены по-разному. Прилагательные принято делить на качественные и относительные. Качественные прилагательные характеризуются целым рядом признаков: они имеют степени сравнения, диминутивные и аугментативные суффиксы, сочетаются с показателями степени, образуют наречия на -о и т.д. Базовые свойства качественных прилагательных обусловлены их семантикой – они обозначают градуируемые признаки, значение которых моделируется шкалой признака [1]. Относительные прилагательные выделяются «отрицательно» – как не обладающие перечисленными свойствами, что тоже обусловлено их семантикой – а именно, неградуируемостью.

Прототипическими относительными прилагательными считаются отсубстантивные (ср. железный, автомобильный, городской, детский), однако в широком смысле относительными можно считать лю-

бые прилагательные, которые обозначают неградуируемый признак, – в том числе и образованные от глаголов и причастий (стиральный, моющий, выдвигной), от наречий (вчерашний, здешний), от числительных (тройной), – поскольку они не являются качественными.

Хотя многие наречия образованы от прилагательных, их не принято делить на качественные и относительные, и сам термин «относительные наречия» не является употребительным (термин «качественные наречия» встречается). Классификация наречий не симметрична классификации прилагательных и построена по другому принципу. Наречия делятся на характеризующие и обстоятельственные, а характеризующие, в свою очередь, – на наречия образа действия (быстро, решительно, дружески, по-медвежьи) и количественные. В [2, § 1650] значение образа действия определяется как значение качественности, но не поясняется, что следует понимать под качественностью (если исходить из аналогии с качественными прилагательными, то это градуируемый признак, ср.: *очень быстро* [перемещался], *довольно решительно* [сказал], *вполне дружески* [относился]).

В школьной классификации наречия образа действия отвечают на вопрос «как?». Но если под качественностью понимать градуируемость, то не все наречия образа действия качественные: кроме качественных наречий на -о и -и в эту группу входят также наречия, образованные от падежных форм и предложных оборотов (боком, бегом, столбом, наизнанку, наотмашь, вприпрыжку, повзводно, напрямую), которые не сочетаются с показателями степени (ср. \**весьма боком*).

Помимо этого, с помощью суффиксов -о и -и образуются наречия от относительных прилагательных,

\* Исследование выполнено в рамках проекта РФФИ № 17-29-09154 офи\_м «Динамика языковой системы: корпусное исследование синхронной вариативности и диахронических изменений в текстах разных типов». Языковые примеры извлечены из Национального корпуса русского языка (НКРЯ). – URL: [www.ruscorpora.ru](http://www.ruscorpora.ru)

которые и интересуют нас и которые мы тоже будем называть относительными: *аварийно, административно, амбулаторно, анатомически, денежно, коммерчески, композиционно, лекарственно, медицински, музыкально, научно, планово, промышленно, профессионально, социально, сюжетно, территориально, технически, туристически, физически, финансово, химически, экономически, юридически* и т.п. Таких наречий, как свидетельствуют данные Национального корпуса русского языка (далее – НКРЯ или Корпус), немало, и не все они отражаются даже в электронных словарях, не говоря уже о бумажных. Эти наречия употребительны, в основном, в книжной и устной литературной речи. Но если про относительные прилагательные всегда имеется особый раздел в грамматиках и учебных пособиях, то об относительных наречиях в грамматических описаниях вообще ничего не говорится, и они остаются за пределами классификаций (ср. [2, 3]; есть лишь упоминание в специальной работе [4]). Между тем возникают вопросы: к какому разряду они относятся? Что обозначают? и т.д. Технически эти наречия отвечают на вопрос «как?» и формально должны были бы быть причислены к наречиям образа действия. Однако не вполне ясно, какой смысл можно придать термину «образ действия» в случае наречий с субстантивным корнем. Несомненно одно – эти наречия естественно сравнивать с относительными прилагательными. Поэтому мы начнем с прилагательных, а затем перейдем к наречиям.

Сразу оговоримся, что многие наречия, соотносительные с отсубстантивными прилагательными, имеют качественные значения, например, *железно пообещал, очень профессионально выполнил прыжок, анатомически точное изображение, жили сытно и денежно*. Эти значения в настоящей статье не приводятся. Мы будем рассматривать только те контексты, в которых реализованы относительные (неградуируемые) значения.

**ЗАМЕЧАНИЕ.** Отметим, что от глагольных прилагательных типа *стиральный, выдвижной, подъемный*, которые тоже обозначают неградуируемый признак, т.е. являются относительными, наречия не образуются, и это не случайно.

## ОТНОСИТЕЛЬНЫЕ ПРИЛАГАТЕЛЬНЫЕ

Относительные прилагательные присоединяются к существительным. Возникшее в результате присоединения сочетание AS ([nomen] adiectivum + [nomen] substantivum) представляет собой свернутую (редуцированную) предикацию и имеет внутреннюю семантическую формулу, которую слушающий должен реконструировать. Эта формула зависит от семантики обоих элементов сочетания и обязательно содержит внутренний предикат, который может быть «наведенным», а может извлекаться из семантики существительного. Наведенный предикат встречается как в контексте натуральных классов (ср. *рыба* – само существительное не содержит внутреннего предиката): *речная рыба* = ‘водится в реке’, так и в контексте существительных, имеющих внутренний

предикат: например, *сапоги* содержат внутренний предикат обуви (‘носить’), но *итальянские сапоги* – это не ‘сапоги, которые носят в Италии’, а ‘сапоги, которые производятся в Италии (и импортируются из Италии)’.

Формула сочетания AS может строиться на основе предиката, связанного с существительным. Существительные, которые содержат внутренний предикат, – это, например, функциональные названия предметов (*нож* – ‘резать’, *клей* – ‘клеить’) и функциональные названия лиц (*врач* – ‘лечить’). Для краткости будем называть их артефактами и агентами соответственно. Кроме того, есть событийные (абстрактные) существительные, ср. *приглашение, строительство, проверка, праздник*, которые являются «опредмеченными» предикатами и обозначают ситуации. Эти существительные часто называют номинализациями, однако этот термин предполагает их производность от глагола. Но очень многие существительные либо вообще не имеют соотносительного глагола (ср. *аудит, кризис, стресс*), либо семантически разошлись с ним (*доступ, успех*). Поэтому вместо термина «номинализации» будем использовать термин «ситуативы».

Одна из главных функций относительных субстантивных прилагательных, как неоднократно отмечали исследователи, см. [5–9], – это уточнение, конкретизация, они обозначают разновидность чего-либо: тип предмета (*хлебный / рыбный / кухонный / охотничий нож; посудный / платяной / книжный шкаф*); специализацию деятельности лица (*оперный / эстрадный певец*); тип ситуации (*пассажирские / грузовые перевозки*).

Относительные прилагательные при артефактах конкретизируют, уточняют предназначение артефакта (основная функция артефакта заключена в его внутреннем предикате, см. [10]). Конкретизация может выражаться через:

- актант: объект, ср. *рыбный нож* – ‘резать рыбу’; *обоинный клей* – ‘клеить обои’, или субъект, ср. *детская обувь* – ‘носят дети’;
- функциональное пространство / помещение / учреждение (где используется, применяется артефакт): *кухонный нож; садовые ножницы; больничный халат*;
- функциональную сферу: *авиационный / медицинский клей*.

Объект внутреннего предиката в двух последних случаях не меняется: садовые ножницы не режут сад, а авиационный клей не клеит авиацию, т.е. эти прилагательные не являются актантами внутреннего предиката, но косвенно указывают на него. Например, *кухонный нож* используется при приготовлении еды.

Относительные прилагательные при агентах тоже конкретизируют род деятельности субъекта через:

- объект, ср. *зубной / глазной врач* – ‘лечит зубы / глаза’;

**ПРИМЕЧАНИЕ.** Для агентов такая модель не очень характерна, ср. <sup>1</sup>*книжный читатель* (читает книги), <sup>2</sup>*рыбный продавец* (продает рыбу), <sup>3</sup>*лисий охотник* (охотится на лис).

- функциональное пространство / помещение / учреждение, ср. *фабричный рабочий, заводская администрация, больничный персонал, банковские служащие*;

- функциональную сферу: *финансовый работник, коммерческий директор, промышленный / сельскохозяйственный рабочий, медицинский персонал*.

Ситуативы представляют для нас наибольший интерес, поскольку они ближе всего к глаголам (а с глаголами сочетаются наречия). Что обозначают прилагательные при ситуативе? В принципе, это те же три значения: актант, функциональное пространство и сфера.

Относительные прилагательные от функциональных существительных, обозначающие субъекта ситуатива, используются ограниченно и обычно представляют собой устойчивые сочетания: *прокурорская проверка* – вид проверки, *читательская конференция* – вид конференции (бывает *научная, практическая*); *милиейская облава*; *журналистское расследование*; *депутатский запрос*. При этом с другими существительными эти прилагательные не употребляются. Так, в НКРЯ не встретились сочетания <sup>?</sup>*прокурорский вызов* ('прокурор вызвал кого-л.'), <sup>?</sup>*прокурорский прием* ('прокурор принимает посетителей'). Сочетание *прокурорское заявление* (в значении 'прокурор заявил') встретилось в газетном корпусе 1 раз, а в основном – ни разу. При этом это сочетание звучит более разговорно, чем *заявление прокурора* – в отличие от сочетания *прокурорская проверка*, которое относится к официальной речи.

Для обозначения объекта относительные прилагательные тоже употребляются очень ограниченно: *жилищное / гаражное / офисное строительство; пассажирские / грузовые перевозки*. При этом в Корпусе не встречаются сочетания <sup>?</sup>*лесная рубка*, <sup>?</sup>*лесные заготовки*, <sup>?</sup>*болотное осушение*, <sup>?</sup>*ковровая чистка*, <sup>?</sup>*мостовой взрыв*; <sup>?</sup>*книжная покупка*, <sup>?</sup>*книжное чтение*. В XIX в., по данным НКРЯ, использовалось сочетание *хлебная уборка*, которое в современных текстах не встретилось. Прилагательное может соответствовать и другим типам актантов: *автобусная поездка* – 'на автобусе'.

ПРИМЕЧАНИЕ. С прилагательными от названий функциональных пространств и помещений ситуативы тоже сочетаются очень ограниченно, встречаются лишь отдельные сочетания, ср. *садовые работы* (в саду), *городское благоустройство*, *фабричное производство*, но не <sup>?</sup>*больничный ремонт* (ремонт больницы), <sup>?</sup>*кухонное приготовление* (на кухне), <sup>?</sup>*магазинные продажи* (в магазине).

Кроме того, к ситуативам, конечно, могут присоединяться обстоятельства, – например, обстоятельства места, ср. *речная прогулка*, и времени, ср. *вечерняя прогулка*. Но эти случаи мы не рассматриваем.

Можно было бы отметить, что неупотребительность сочетаний типа <sup>?</sup>*лесная рубка* или <sup>?</sup>*книжная покупка* – вопрос узуса, никакого семантического запрета здесь нет. Тем не менее существуют, по-видимому, и более содержательные объяснения этого ограничения.

Основной способ выражения собственно актантного отношения – падежные формы: *рубка леса, покупка книг*. Прилагательное не является в чистом виде выражением актанта, а, в силу грамматической специфики, остается характеристикой референта существительного. Сочетание AS, хотя и не является фразеологизмом, имеет номенклатурный, в каком-то смысле терминологический характер – является устойчивым наименованием разновидности объекта или ситуации в, так сказать, канцелярском языке. Так, *прокурорская проверка* – официальное название данного типа ситуации (вид проверки). И если бы в канцелярском языке потребовалось противопоставить покупки книг покупателям продуктов, то номинации *книжные* и *продуктовые покупки* звучали бы вполне стандартно.

При этом прослеживается следующая тенденция: наибольшие ограничения обнаруживаются для физических ситуаций и физических (предметных) актантов (типа <sup>?</sup>*ковровая чистка*). Чем более абстрактно прилагательное по семантике, тем свободнее образуются сочетания AS. При этом прилагательные в таких сочетаниях обозначают сферу, к которой относится ситуация. Например, в НКРЯ встретилось 20 сочетаний *врачебное обследование* (где ситуация характеризуется через субъекта) и около 800 сочетаний *медицинское обследование* (где ситуация характеризуется через сферу). Если сравнить, например, употребление в Корпусе прилагательных *денежный* (с более конкретной семантикой) и *финансовый* (с более абстрактной), обнаруживаются следующие соотношения (дается суммарное число вхождений по основному и газетному корпусам): *денежные злоупотребления* = 2 vs. *финансовые злоупотребления* = 105; *денежное решение* = 1 vs. *финансовое решение* = 81.

ПРИМЕЧАНИЕ. Количество сочетаний *денежные требования* (112) превосходит количество сочетаний *финансовые требования* (85), т.к. *денежные требования* – это юридический термин, который фигурирует в юридических документах, и даже в свободном обсуждении юридических конфликтов авторы текстов предпочитают использовать это терминологическое сочетание.

Таким образом, в сочетаниях с ситуативами у относительных прилагательных наиболее широко представлено значение сферы – т.е. относительные прилагательные используются для конкретизации типа ситуации через отнесение её к той или иной сфере жизни или деятельности.

## ОТНОСИТЕЛЬНЫЕ НАРЕЧИЯ

Теперь обратимся к относительным наречиям.

Относительные наречия с субстантивными корнями образуются от относительных отсубстантивных прилагательных и имеют сходную семантику, но другие грамматические свойства, в первую очередь – сочетаемость. Если рассматривать наречия в целом, то разные группы наречий сочетаются с разными частями речи: качественные наречия – с глаголами (*быстро бежал, аккуратно сложил*); количественно-степенные – с глаголами, прилагательными, наречиями, предикативами (*очень устал; очень холодный*;

довольно холодно ответил; здесь не так уж холодно); компаративами (*намного выше*); отдельные наречия сочетаются даже с существительными (*пальто нараспаишу*). Субстантивные наречия сочетаются с глаголами, ср. *коммерчески осваивать* (в том числе с причастными формами, ср. *фабрично упакован*), и прилагательными, ср. *финансово независимый*.

ПРИМЕЧАНИЕ. Особый тип употребления относительных наречий – когда они относятся не к конкретному слову, а ко всей предикации, ср.: *Корнуэл представляет собой непросветленную умом волю (силу). Композиционно он противоположен Эдмунду* [М. А. Чехов. О технике актера (1945)]. В русской грамматической традиции в таких случаях используется термин «детерминант», введенный Н.Ю. Шведовой [11, 12]). В литературе наречия с таким типом употребления называют также сентенциальными, наречиями с плавающей сферой действия [13–17]. Однако это отдельная большая тема, которую мы рассмотрим в другой работе.

Но если сравнивать относительные наречия и относительные прилагательные на уровне лексической сочетаемости, то она часто совпадает. Существуют пары соотносительных сочетаний «прилагательное + существительное» (AS) и «наречие + глагол» (AdvV), где второе можно представить как результат своего рода транспозиции первого, ср.: *финансовая поддержка [искусства] → финансово поддерживать [искусство]*. Это, конечно, не означает, что для каждого наречного сочетания имеется пара (некоторые существительные, как уже говорилось, не соотносятся с глаголами, ср. *финансовый кризис / крах*) или что наречные сочетания AdvV реально являются трансформациями адъективных AS. Тем не менее это дает основания сопоставить, так сказать, однокоренные сочетания AS и AdvV. Данные НКРЯ показывают, что если существуют оба сочетания, то адъективно-субстантивное – существенно более употребительно.

Так, выражение *плановое ведение [хозяйства]* встретилось в Корпусе 6 раз, а *[хозяйство будет] вестись плано* – 1 раз (здесь и далее приводится суммарное количество вхождений в основной и газетный корпус). На запрос «*техническое обслуживание*» НКРЯ выдает около 1500 вхождений, а на «*технически обслуживать*» – 2 вхождения (*Его организация по договору должна обслуживать узлы учета коммунальных ресурсов — технически обслуживать, своевременно менять фильтры, следить за правильностью показаний* [«Новгородские ведомости», 2013]; *На сегодняшний день ДКК просто не в состоянии технически обслуживать сделки на ММВБ* [РБК Daily, 2005]); на запрос «*техническое оснащение*» – больше 500 вхождений, а на «*технически оснастить*» – меньше 100 (ср.: *Необходимо поддержать государственное бесплатное здравоохранение, технически оснастить участковых врачей, особенно в сельской местности* [Труд-7, 2000.12.02]).

Конечно, не все сочетания AdvV имеют пару AS, но все относительные наречия являются производными от прилагательных. И квантитативные показатели говорят о том, что относительные наречия в целом существен-

но менее употребительны, чем относительные прилагательные. Например, *технически* встретилось в НКРЯ чуть больше 3700 раз, *технический* – более 60000; *медицински* – 51, *медицинский* – 46500.

Основные значения относительных прилагательных (в том числе в контексте ситуативов, которые наиболее близки к глаголам) – актант, функциональное пространство и сфера. Есть ли соответствия этой сочетаемости у наречий?

Хотя относительные прилагательные не являются прямым выражением актантов, они, как уже говорилось, соотносятся с актантами, ср.: *прокурорская проверка* – ‘прокурор проверил’; *пассажирские перевозки* – ‘перевозить пассажиров’, *автомобильные перевозки* – ‘перевозить автомобилями, на автомобилях’.

Относительные наречия обычно не соответствуют падежным формам – ни субъекту: *прокурорски проверит*, ни прямому объекту: *пассажирски перевозить*; ни инструменту: *автомобильно перевозить*.

ПРИМЕЧАНИЕ. Заметим, что многие из этих наречий употребляются в качественных значениях: – *Ты почему не позвонил?* – *прокурорски спросила жена* [Эльдар Рязанов, Эмиль Брагинский. Тихие омуты (1998)]

Запрет выражения наречиями основных участников ситуации можно объяснить семантически: наречие – это дополнительная, необязательная характеристика ситуации, поэтому странно превращать основных участников в такие дополнительные характеристики (а именно это происходит при их выражении наречиями).

Что касается таких участников, как инструмент и средство (стандартно выражаемых творительным падежом), то в этом случае нет противоречия между семантикой актанта и наречия: инструмент или средство, вообще говоря, входят в способ действия и соотносятся с обстоятельством образа действия. Показательно, что творительный падеж часто переходит в наречия образа действия: *боком, бегом*. Таким образом, здесь ограничения касаются, скорее всего, характера семантики самого наречия. Наречия с «предметной» семантикой в функции обстоятельства образа действия малоупотребительны. Есть лишь отдельные случаи, когда наречие можно перифразировать с помощью падежной формы: *Специальные классы звезд обозначаются буквенно в таком порядке: О, В, А, F, G, К, М — от очень горячих голубых звезд с поверхностной температурой 100 000° до красных с температурой в 3000 °* [И. А. Ефремов. Туманность Андромеды (1956)] – ‘обозначаются буквами’.

При этом наречия с более абстрактной семантикой довольно широко представлены в НКРЯ в значении образа действия: *Ведь это наше, понимаете, наше дело спорить с ними — административно же это не решается!* [Дмитрий Быков. Орфография (2002)] – ‘с помощью административных методов’; *Сталин дипломатически, а Тельманы агитаторски требовали от социал-демократического правительства не впускать меня в Германию — надо думать, во имя интересов пролетарской революции*

[Л. Д. Троицкий. Моя жизнь (1929-1933)] – ‘дипломатическими / агитационными средствами’ (ср. также: **медицински** доказанная вменяемость; **научно** обоснованные меры; **юридически** установленный факт – ‘медицинскими / научными / юридическими методами’).

Наречия, соотносительные с функциональными пространствами, типа *фабрично*, также выступают в значении образа действия: *На кителе, телогрейке и бушлате уже был **пришит фабрично** мой номер: Я-815* [Анатолий Жигулин. Черные камни (1988)] – машинным способом, не вручную.

Однако для относительных наречий с субстантивными корнями семантика образа действия не является основной, и их употребления в такой функции, по данным НКРЯ, немногочисленны. Основное значение таких наречий – значение аспекта / сферы (что соответствует значению сферы у прилагательных). Оно реализуется как в контексте глаголов с абстрактной семантикой и нефизическими актантами, ср.: *Когда за сольный концерт мне предложат половину от гонорара «Квартета II» за корпоратив, буду считать, что **коммерчески** и **музыкально** я **состоялся*** [Известия, 2013.02.06], – так и в контексте прилагательных, ср.: **коммерчески** выгодный / (бес)перспективный / безнадежный / интересный / оправданный / привлекательный / рентабельный / эффективный продукт, проект, бизнес.

Аспект (или сфера) – это специфическая валентность интерпретационно-оценочных слов. В частности, эта валентность характерна для интерпретационно-оценочных прилагательных [18]. В силу абстрактности своей семантики такие прилагательные часто нуждаются в уточнении, в проекции в какую-то конкретную сферу. Эту функцию выполняет валентность аспекта, которая у прилагательных может выражаться предложной группой по Дат. п.: *картина интересна по композиции*. Но эту же функцию конкретизации могут выполнять и относительные наречия. Вот контексты прилагательного *беспомощный*: [человек] – **педагогически** беспомощные учителя; **грозные, но политически** беспомощные силовики; [продукт деятельности (часто – информационный)] – **лингвистически** беспомощная аргументация, **методологически** беспомощная книжка; [деятельность] – **технически** беспомощная игра. Эту валентность можно перифразировать с помощью оборотов «с точки зрения», «в плане»: **юридически** беспомощная формулировка – ‘с юридической точки зрения’; **последние стихи поэтически** беспомощны – ‘с точки зрения поэзии’.

Данные Корпуса показывают, что если валентность аспекта может выражаться и предложным оборотом, и наречием, то более широко она представлена наречиями: сочетаний *беспомощный по Дат. п.* (ср. *Вся эта отрасль литературы представлялась Лему реакционной по форме и беспомощной по содержанию* [Александр Генис. Три «Соляриса» (2002) // «Звезда», 2003]) в НКРЯ встретилось 4, а сочетаний с наречием (типа **юридически** беспомощный) – 29; *выгодный по Дат. п.* (ср. *Финская компания известна как ведущий производитель надежных, удобных, вы-*

*годных по цене сухих отделочных смесей* [«Строительство», 2003.09.29]) – 9 вхождений в Корпус; наречие на *-ски + выгодный* (**коммерчески / стратегически / тактически / экономически / энергетически** выгодный) – около 200 вхождений.

Таким образом, относительные наречия – и по семантике, и по функциям – в большей степени похожи не на другие наречия, а на относительные прилагательные, прежде всего – в значении аспекта / сферы. Одна из главных функций относительных субстантивных прилагательных, как уже было сказано, – своего рода уточнение, конкретизация. Аналогичную функцию выполняют относительные наречия, конкретизируя аспект рассмотрения ситуации. Главным образом, это относится к сочетаниям таких наречий с прилагательными, а также с абстрактными глаголами. При этом данные Национального корпуса русского языка показывают, что если взять одно и то же наречие, то его сочетаний с прилагательными окажется на порядок больше, чем с глаголами. Так, сочетаний «**коммерчески** + прилагательное» (*коммерчески* выгодный, *коммерчески* грамотный, *коммерчески* успешный, *коммерчески* эффективный и т.п.) в основном и газетном корпусе более 400 (при этом причастия употребляются в таких контекстах в значении прилагательных, ср. *коммерчески* оправданный, *коммерчески* ориентированный), а сочетаний «**коммерчески** + глагол» (ср. *коммерчески* окупился) – меньше 20.

Если глагол обозначает целенаправленную деятельность, наречие может обозначать способ действия (средство достижения цели), ср.: *научно доказать*.

Таким образом, решающую роль в интерпретации относительных наречий (как и относительных прилагательных) играет их собственная семантика и семантика контекста.

## СПИСОК ЛИТЕРАТУРЫ

1. Кустова Г.И. Прилагательные // Материалы к корпусной грамматике русского языка. Вып. 3. Части речи и лексико-грамматические классы. – СПб: Нестор-История, 2018. – С. 40–107.
2. Русская грамматика. В 2-х тт. / под ред. Н.Ю. Шведовой. Т. I. – М.: Наука, 1980.
3. Панков Ф.И. Опыт функционально-коммуникативного анализа русского наречия. – М.: МАКС Пресс, 2008. – 448 с.
4. Битехтина Г.А. Семантико-синтаксические разряды определительных наречий в современном русском языке и условия их функционирования: автореф. ... канд. филол. наук. – М., 1979. – 24 с.
5. Павлов В.М. О разрядах имен прилагательных в русском языке // Вопросы языкознания. – 1960. – № 2. – С. 65–70.
6. Павлов В.М. Качественность и субстанциальная семантика // Теория функциональной грамматики. Т. 5. Качественность. Количественность. – СПб: Наука, 1996. – С. 8–53.
7. Павлов В.М. Предметно-относительный атрибут как типологическая константа // Проблемы

- функциональной грамматики. Полевые структуры. – СПб: Наука, 2005. – С. 169–192.
8. Панов М.В. Позиционная морфология русского языка. – М.: Наука-ЯРК, 1999. – 275 с.
  9. Рахилина Е.В. Когнитивный анализ предметных имен: семантика и сочетаемость. – М.: Русские словари, 2000. – 416 с.
  10. Кустова Г.И. Скрытая грамматика русских атрибутивных конструкций // Изв. РАН. Сер. лит. и яз. – 2015. – № 5. – С. 5–21.
  11. Шведова Н.Ю. Детерминирующий объект и детерминирующее обстоятельство как самостоятельные распространители предложения // Вопросы языкознания. – 1964. – № 6. – С. 77–93.
  12. Шведова Н.Ю. Существуют ли все-таки детерминанты как самостоятельные распространители предложения // Вопросы языкознания. – 1968. – № 2. – С. 39–50.
  13. Рудницкая Е.Л. Словарная статья наречия *легкомысленно* // Семиотика и информатика. Вып. 32. – М., 1991. – С. 156–159.
  14. Рудницкая Е.Л. Некоторые сентенциальные наречия в русском языке. Семантика. Синтаксис. Лексикография // Вопросы языкознания. – 1994. – № 1. – С. 114–125.
  15. Филипенко М.В. Семантика наречий и адвербиальных выражений. – М.: Азбуковник, 2003. – 304 с.
  16. Greenbaum S. Studies in English adverbial usage. – London, 1969. – 262 p.
  17. Thomason R., Stalnaker R. A semantic theory of adverbs // Linguistic Inquiry. Cambridge (Mass.). – 1973. – Vol.4, № 2. – P. 195–220.
  18. Кустова Г.И. Семантические и конструктивные валентности прилагательных // Prace filologiczne (Warszawa). – 2018. – № 73. – P. 223–238.

*Материал поступил в редакцию 12.02.19.*

#### **Сведения об авторе**

**КУСТОВА Галина Ивановна** – доктор филологических наук, ведущий научный сотрудник Института русского языка им. В.В. Виноградова РАН, Москва  
e-mail: galinak03@gmail.com

## **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](mailto:contact@viniti.ru); [sales@viniti.ru](mailto:sales@viniti.ru)