

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

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

Издается с 1961 г.

№ 7

Москва 2017

ОБЩИЙ РАЗДЕЛ

УДК 004.9 : [001.8 : 002]

П.А. Калачихин

Паттерны конструирования показателей научного цитирования*

Приводятся паттерны структуры базовых типов показателей научного цитирования, в том числе альтернативных показателей цитирования новейшего поколения. Паттерны учитывают тонкие особенности показателей цитирования. Паттерны индикаторов цитирования являются достаточно универсальными для того, чтобы быть использованными в качестве шаблонов для автоматической генерации новых индикаторов цитирования с заданными свойствами при создании интеллектуальных модулей наукометрических информационных систем.

Ключевые слова: альтметрики, индексы, индикаторы, метрики, паттерны, рейтинги, цитирование, экспертные оценки

ВВЕДЕНИЕ

В отечественных и зарубежных научных исследованиях уделяется очень много внимания проблемам, связанным с научными публикациями и публикационной активностью. Однако в них превалирует инте-

рес к различным этическим аномалиям: плагиату и борьбе с ним, вопросам неправомерного цитирования, недобросовестному кооперированию во взаимном цитировании среди отдельных научных сообществ. Между тем, вопрос о соответствии существующих данных по научной литературе с реальной динамикой числа научных исследований конкретно нигде не рассматривается. Более того, различные наукометрические базы данных содержат разнообразные статистические сведения о динамике в области научных

* Работа выполнена в рамках исследования по теме 0003-2015-0008 Госзадания ВИНТИ РАН и при поддержке Российского фонда фундаментальных исследований (проект РФФИ № 17-07-00256)

публикаций. Поэтому представляется актуальным глубже рассмотреть структуру показателей цитирования, которая в значительной мере определяет, насколько адекватной будет оценка динамики числа научных публикаций при помощи того или иного показателя цитирования.

Цель настоящего исследования состоит в том, чтобы на основе наукометрических методов и методов анализа данных повысить качество управления фундаментальными научными исследованиями, а задача исследования заключается в изучении показателей, стимулирующих публикационную активность исследователей в научно-образовательных организациях, не обязательно только лишь в отечественных.

Изучение показателей научного цитирования на уровне методик невозможно без использования математического языка, на котором показатели цитирования могут быть формализованы, так как такой язык наиболее краток, точен и универсален. В нашем исследовании была применена особая математическая нотация, вобравшая в себя черты дескриптивного языка теории множеств и структур данных, которые в дискретной математике задаются при помощи типовых синтаксических конструкций.

Таким образом, в настоящем исследовании используются формальные методы, родственные таким классическим инструментам проведения исследований как анализ, моделирование и абстрагирование. При этом все результаты исследования носят рекомендательный характер, допуская альтернативную точку зрения и возможность критики в ходе обсуждения. Были предприняты усилия для того чтобы сделать результаты исследования максимально конструктивными, потому что эти результаты подчинены более масштабной прикладной задаче – построению наукометрических информационных систем нового поколения, базирующихся на интеллектуальных алгоритмах.

ОБЗОР И КЛАССИФИКАЦИЯ ПОКАЗАТЕЛЕЙ ЦИТИРОВАНИЯ

Цитатные метрики широко распространились за последние годы. Быстрое развитие этого направления свидетельствует о том, что наилучший подход должен учитывать оценку публикаций, исследователей, организаций и научных изданий. Таким образом, использование цитатных метрик может иметь серьезные последствия, которые затронут научное сообщество, и поэтому они являются интересными не только лишь для экспертов в наукометрии.

Количество цитирований публикаций стремиться к тому, чтобы иметь очень широкое распределение по множеству публикаций, собравших мало цитирований, и малое распределение по нескольким публикациям, получившим большое количество цитирований. Эти списки цитирований могут быть получены благодаря предпочтениям читателей, согласно которым части публикаций удастся привлечь большое количество цитирований. Цитирования всегда говорят о том, что публикация вызвала интерес, но точные причины, почему именно эта публикация вызвала интерес, могут варьироваться для различных публикаций.

В идеале, каждый исследователь хочет, чтобы индикаторы цитирования измерили его деятельность монотонным способом: чем выше метрика, тем «лучше» публикация. Цитирования, полученные единственной публикацией, могут зависеть от множества факторов вне ее достоинств. Они могут включать научную область, которой публикация принадлежит (у различных областей могут быть различные нормы количества публикаций и цитирования, и, таким образом, различные «плотности» цитирования), продолжительность периода времени (с момента публикации), тип документа (реферат, обзор, обычная статья, передовая статья), и покрытие баз данных, где цитирования учитываются. Ошибочно делать преждевременный вывод о том, что показатели цитирования настолько запутанны, что они должны быть отклонены в пользу оценок экспертов, компетентных в данной области [1].

Сложный индикатор – сумма стандартизированных и приведенных к результатам одинакового типа индикаторов цитирований. Для каждой стандартизации высшая оценка дается исследователю с самым высоким значением по соответствующему индикатору. Достоинство сложных индикаторов заключается в том, что они могут помочь ослабить различия между областями, потому что один тип индикатора дает преимущество для некоторых областей, но другой тип индикатора может препятствовать такому преимуществу. Однако сравнения между различными областями должны осуществляться осторожно. Даже не известно, каково идеальное пропорциональное представление исследователей из различных областей среди тех, которые являются ведущими исследователями.

Простые индикаторы цитирований – с одной стороны – могут иметь преимущества в том, что они прозрачнее, чем сложные индикаторы. С другой стороны, сложный индикатор требует более сложных вычислений, и поэтому воздействия различных процитированных публикаций труднее различить. В некоторых случаях различия между сложными и несложными индикаторами не совсем очевидны. Однако сложный индикатор может быть более информативным, чем любой из его элементов, если он охватывает преимущества этих элементов и не сокращает их. Это является причиной, по которой сложный индикатор выигрывает у односоставных в способности распознавать всеми признанных крупных исследователей [2].

По состоянию на текущий момент широко известны индикаторный подход к оценке научной деятельности и подход, основанный на экспертных оценках. Последний полагается на субъективную оценку качества исследования, но при этом он сильно зависит от компетентности экспертов, а также требует значительных ресурсных затрат. Индикаторный подход подразумевает количественные оценки величин, свойственных исследователям и научно-исследовательским организациям. Расчеты индикаторов должны производиться на основании материалов публикаций. Таким образом, индикаторами не только охватываются, но и количественно

измеряются различные стороны научно-исследовательской деятельности.

Индикаторы научных цитирований принято разделять на три подвида, основанные:

- на подсчете количества публикаций;
- на подсчете количества цитирований;
- на подсчете как количества публикаций, так и количества цитирований [3].

Академические публикации представляются как узлы в сети, а цитирования – как направленные ссылки между ними. Дополнительные функции, относящиеся к оценке влияния, сохраняются как атрибуты узла. Они включают год публикации, заголовок, сведения об исследователях и домен. Подсчет так называемого «*Eigenfactor score*» – метрики влияния каждой публикации, которая учитывает количество и качество цитирований, вычисленных через весь набор данных, – также сохраняется как атрибут для каждого узла [4].

Данные о цитированиях, представленные в числовом формате, позволяют оценить влияние результатов, полученных в ходе выполнения научно-исследовательских работ, на современную научную мысль, а также пригодны для измерения полезности результатов по отношению к деятельности других исследователей. Данные о цитированиях, взятые вне контекста, не могут служить информационной базой для оценки качества публикаций, поэтому подобную информацию необходимо облачать в форму индикаторов, показывающих вероятность, с которой некоторая заданная публикация может оказаться значимой. Сказанное справедливо прежде всего по отношению к показателю «*импакт-фактора исследователя*», который рассчитывается на базе первично оцениваемых показателей импакт-факторов журналов [5].

Конструирование и сопоставительный анализ показателей цитирования открывают новые горизонты перед развитием научно-исследовательской деятельности, предоставляя возможность управлять стратегией и ходом этой деятельности, а также ставить оценки, сравнивая полученные результаты с результатами других научных исследований. Показатели цитирования не всегда могут объективно учитывать специфические особенности научной деятельности, поэтому их рекомендовано использовать в связке с экспертными оценками. Будучи подвижными по своей природе, показатели цитирования все время находятся в состоянии динамического изменения, поэтому полученные с их помощью оценки могут быть объективными только в течение ограниченного времени [6].

В процессе практического использования экспертных оценок выявлен целый ряд «узких мест», к числу которых следует отнести долгий период ожидания окончательных экспертных оценок; зависимость от информационной базы; зависимость от предметной области; неопределенность положительности или отрицательности знака оценки. Прочим показателям цитирования свойственны такие отрицательные стороны, как неуместное использование импакт-факторов изданий для оценки качества размещенных в них публикаций; тенденция оценивать многогранную научную деятельность исследователей по неполному набору критериев;

«накручиваемость» индикаторов цитирования всевозможными способами [7].

Отличительные особенности научного цитирования по отношению к цитированию в других сферах издательской деятельности заключаются в активном использовании методов квалиметрии, которые позволяют разрабатывать показатели цитирования «нового поколения» с необычайными полезными свойствами и интересными характеристиками. Так на стыке нескольких дисциплинарных направлений возникла *вебометрия*, в центре внимания которой располагаются так называемые «*альтернативные метрики*» или просто «*альтметрики*». Альтметрики исследуют содержание социальных сетей, чтобы обеспечить альтернативу или улучшение оценок факторов влияния журналов и анализ отношения числа кликов к числу показов, измеряя влияние и значение научных публикаций [8].

Альтернативные метрики в настоящее время – одна из самых популярных тем в наукометрических исследованиях. Существуют две важные причины популярности этой темы. С одной стороны, доступность данных на платформах социальных сетей обеспечила легкий источник для статистических исследований, который с нетерпением был принят научным сообществом. С другой стороны, успех этого нового источника данных соответствует стремлению научной политики к оценке широкого влияния научных исследований. В настоящее время такое влияние изучается при помощи научных метрик, однако, возможно, что подобные исследования могут быть выполнены при помощи альтернативных метрик [9].

Альтметрики могут измерять влияние на уровне журнальной публикации, столь же наблюдаемое посредством применения инструментов для социального общения, а также воздействующее, как это может быть измерено, на другие значительные результаты исследований. Новые метрики предоставляют возможность обнаруживать идеи, которые невозможно было ранее получить, и они подсчитываются скорее по сравнению с традиционными метриками, которые полагаются на подсчет цитирований. Альтметрики обеспечивают широкие и более глубокие идеи о влиянии академических публикаций, исследователей, подразделений организаций, университетов вне досягаемости традиционных метрик. В долгосрочной перспективе альтметрики смогут, возможно, пережить импакт-факторы и другие традиционные, основанные на цитированиях метрики, но это, скорее всего, зависит от существенного сдвига в способах, которыми академические выводы формализуются [8].

Цитирования из онлайн-изданий могут быть автоматически собраны посредством веб-поиска, и поэтому они способны оказывать влияние на области, основанные на конференциях, несмотря на то, что они слишком редки, чтобы быть собранными в целях проведения подобного исследования. Web-цитирование может быть полезным как доказательство существенного влияния всех типов публикаций и должно включать отбор комментариев для того чтобы им стать релевантными. Напротив, цитирования из научных блогов могут являться доказательством более широкого академического интереса к опубликован-

ной работе, но, поскольку при этом их трудно систематизировать, то они не могут быть практическими источниками для доказательств. Тем не менее, они могут быть особенно полезными, чтобы изучать небольшие области исследований, которые генерируют большие объемы информации [10].

Эффективность и продуктивность деятельности научно-исследовательских организаций и отдельных исследователей недостижимо полностью выразить через набор неких «универсальных» показателей цитирования, чтобы получилась база для формирования рейтинговых оценок. В таком случае, необходимо будет принимать во внимание как количественные, так и качественные показатели слагаемых успеха научно-исследовательской деятельности, уровень научно-исследовательских и опытно-конструкторских разработок, а также их внутреннюю значимость для научного сообщества и социальное влияние на широкую аудиторию. Индикаторы цитирования являются всего лишь инструментами для обеспечения поддержки принятия решений экспертами, и поэтому на практике им отводится второстепенная роль [3].

Итак, каждый показатель цитирования может быть отнесен к одному из следующих типов:

- индексы (T_{ind});
- импакт-факторы (T_{if});
- экспертные оценки (T_{exp});
- альтметрики (T_{alt});
- рейтинги (T_{rat}).

Границы типов T_{ind} , T_{if} , T_{exp} , T_{alt} , T_{rat} простираются достаточно широко, включая как классические показатели, так и их многочисленные модификации, а также более оригинальные, но менее известные показатели с нестандартными методиками оценки. Для полноты картины к ним следует добавить смешанный тип, который подразумевает комбинирование первичных показателей, принадлежащих к разным типам, при расчете составного показателя $T_{mix} = T_{ind} \cup T_{if} \cup T_{exp} \cup T_{alt} \cup T_{rat}$.

МЕТОДИКИ РАСЧЕТА ИНДИКАТОРОВ ЦИТИРОВАНИЯ

На сегодняшний день некоторые исследователи утверждают, что им известно никак не меньше ста индикаторов цитирования. При подготовке настоящего исследования были рассмотрены и проанализированы несколько десятков таких индексов, признанных в научном сообществе.

Среднее количество цитирований p всех публикаций в каждой области рассчитывается по формуле:

$$p = \frac{1}{n} \sum_{i=1}^n c_i, \quad (1)$$

где c_i обозначает количество цитирований публикации i ; n – количество публикаций.

Количество цитирования публикации i может быть разделено на среднее количество цитирова-

ний в области публикации i , приведенных к нормализованному показателю цитирования – NCS_i для публикации i :

$$NCS_i = \frac{c_i}{\rho}. \quad (2)$$

Среднее количество R – связанных процитированных ссылок – всех публикаций вычисляется для каждой области и в течение каждого года с момента выхода публикации:

$$R = \frac{1}{N} \sum_{i=1}^N r_i, \quad (3)$$

где r_i – количество связанных ссылок в желаемом справочном окне публикации i .

$CSNCR_i$ – новый нормализованный индикатор на публикационном уровне.

Количество цитирований центральной публикации можно разделить на ожидаемое количество цитирований:

$$CSNCR_i = \frac{c_i}{e_i}, \quad (4)$$

где c_i – цитаты полученной публикации;

e_i – ожидаемое количество цитирований на основе среднего количества связанных процитированных ссылок по всем публикациям i в том же самом году публикации и в той же области.

Когда анализируется n публикаций, среднее число всех значений $CSNCR_i$ используется в качестве независимого от размера индикатора $MCSNCR$:

$$MCSNCR = \frac{1}{n} \sum_{i=1}^n CSNCR_i. \quad (5)$$

Альтернативно, сумма по всем значениям $CSNCR_i$ может использоваться в качестве зависимого от размера индикатора, $TCSNCR$:

$$TCSNCR = \sum_{i=1}^n CSNCR_i = n \cdot MCSNCR. \quad (6)$$

При объединении формул (4) и (5) $MCSNCR$ может быть переписан следующим образом [11]:

$$MCSNCR = \frac{1}{n} \sum_{i=1}^n \frac{c_i}{e_i}. \quad (7)$$

Индекс, известный как CPT , был специально разработан для того чтобы выявлять ключевые фронты исследований. CPT – отношение среднего воздействия цитирований фронта исследования к возрасту (возникновению) его публикаций. Чем выше это число, тем исследовательский фронт более «горячий» или данная тема более эффективна (продуктивна):

$$CPT = \frac{C}{P \cdot T}, \quad (8)$$

где C – количество цитирований публикаций, т.е. суммарное число публикаций, цитирующих основные публикации; P – количество основных публикаций; T – период цитирований публикаций, кото-

рый рассчитывается как количество лет, прошедших до издания данной публикации после выхода наиболее ранней процитированной в данной публикации работы [12].

Существует многообразие вариаций индекса Хирша. Методики расчета многих «хиршеподобных» индикаторов удобнее записывать не на языке исчисления предикатов и даже не при помощи математических формул, а в словесном виде, воспринимаемом человеком, но не являющемся доступным для распознавания компьютером.

Так называемый «обобщенный» h_a -индекс набора публикаций равен h_a в том случае, когда каждая публикация из этого набора процитирована как минимум $a \cdot h$ число раз, а все остальные публикации набрали менее, чем $a \cdot h$ цитирований.

Другая вариация индекса Хирша – g -индекс исследователя – равен максимальному числу g , такому, что у данного исследователя имеется g публикаций со средней цитируемостью по всему их набору в расчете на одну публикацию не менее g раз. Наличие g публикаций со средней цитируемостью g или большим количеством раз говорит о том, что в сумме эти публикации получили не менее чем g^2 цитирований.

Для устранения расхождений, которые дают публикации с большими количествами цитирований, удобно использовать a_h -индекс, рассчитываемый как отношение количества всех цитирований $N_{c,tot}$ исследователя к квадрату его индекса Хирша:

$$a_h = \frac{N_{c,tot}}{h^2}. \quad (9)$$

Похожий на предыдущий показатель, m_h -индекс позволяет учитывать продолжительность научного стажа исследователя. Он равен частному от деления индекса Хирша на количество лет научной карьеры с момента выхода самой первой публикации [13].

Следующий показатель, который принято называть индексом значимости публикаций IPR , рассчитывается в соответствии со следующей формулой:

$$IPR = \sum_{i=1}^m (n_i^2 \cdot imp_i), \quad (10)$$

где n_i – это количество публикаций в i -ом издании, наделенном импакт-фактором imp_i .

Другой показатель, называемый индексом качества опубликованных работ $IQPW$, вычисляется следующим образом [14]:

$$IQPW = \sum_{i=1}^m n_i \cdot imp_i^2, \quad (11)$$

где n_i – это количество публикаций в периодическом издании, обладающем импакт-фактором imp_i .

При наличии высоких значений индекса Хирша сильно проявляется его «инерционность», так как со временем становится все сложнее и сложнее его уве-

личить. В таком случае, для контроля над продуктивностью научно-исследовательской деятельности и для прогнозирования результативности будущих исследований предпочтительно применять более рациональные модификации индекса Хирша, такие как Sh -индекс и h_{rat} -индекс. Целые части этих индикаторов равны индексу Хирша, рассчитываемому по стандартной методике. Дробные части обоих индикаторов показывают, насколько исследователь готов к переходу на следующее значение индекса Хирша.

В Sh -индексе дробная часть рассчитывается как доля выполнения плана по публикациям для получения следующего значения индекса Хирша. Дробная часть Δ рассчитывается по формуле:

$$\Delta = \frac{\sum_{j=1,h} r_j}{h+1}, \quad (12)$$

где $r_j = \begin{cases} 1, & \text{если } r_i > h \\ 0, & \text{иначе} \end{cases}$ – индикатор принадлежности

публикации ядру Хирша.

В h_{rat} -индексе интервал $[h, h+1]$ разбивается на $2h+1$ равных интервалов в соответствии с минимально необходимым количеством цитирований для получения следующего значения индекса Хирша. $2h+1$ цитирований необходимы при наихудших начальных условиях, когда $c_j = h$, $j=1, \dots, h$ и $c_{h+1} = 0$. Для вычисления h_{rat} к h прибавляем Δ , т.е. долю выполнения плана цитирований. Дробная часть ядра рассчитывается следующим образом [15]:

$$\Delta = 1 - \frac{(h+1-c_{h+1}) + \sum_{j=1,h} (1-r_j)}{2h+1}. \quad (13)$$

i -индекс используется для оценки публикационной активности научной организации наряду с h -индексом: научная организация имеет индекс i , если не менее i учёных имеют h -индекс не менее i . Таким образом, величина i -индекса зависит от количества наиболее высокочитируемых исследователей, работающих в данной организации; i -индекс будет низким там, где высокий h -индекс обеспечивают публикации одного-двух сотрудников [16].

h_2 -индекс учитывает так называемый «вес» самых цитируемых публикаций у конкретного исследователя. Этот индекс определяется как наибольшее натуральное число h_2 такое, что каждая из h_2 наиболее цитируемых публикаций процитирована не менее, чем h_2^2 раз.

e -индекс применяется для подсчета цитирований cit_j в ядре Хирша h публикаций. Методика вычисления e -индекса имеет следующий вид:

$$e^2 = \sum_{j=1}^h cit_j - h^2, \quad (14)$$

m -индекс является медианой количества цитирований h публикаций, входящих в ядро Хирша данного исследователя. Этот показатель является вариацией a -индекса, с его помощью может быть

осуществлена попытка подсчета распределений количества цитирований публикаций, входящих в ядро Хирша.

***a*-индекс** рассчитывается как среднее количество ссылок cit_j на публикации, которые принадлежат ядру Хирша h :

$$a = \frac{1}{h} \sum_{j=1}^h cit_j. \quad (15)$$

***m*-квотиент** используется в качестве инструмента для соизмерения исследователей, обладающих различной продолжительностью профессиональной деятельности. Этот показатель выражается через отношение h -индекса к количеству лет у стажа научно-исследовательской деятельности:

$$m_q = \frac{h}{y}. \quad (16)$$

***a_r*-индекс**, в свою очередь, является модификацией a -индекса, которая имеет следующий вид:

$$a_r = \sqrt{\sum_{j=1}^h \frac{cit_j}{a_j}}, \quad (17)$$

где h – классический индекс Хирша; cit_j – количество цитирований j -ой публикации; a_j – время, прошедшее с момента публикации.

***h_w*-индекс** используется для измерения производительности, оценка которой принимает вид:

$$h_w = \sqrt{\sum_{j=1}^{r_0} cit_j}, \quad (18)$$

где cit_j – число цитирований j -ой публикации; r_0 – максимальный ранг; k – порядковый номер публикации в ядре Хирша, для которой справедливо $r_w(k) \geq cit_k$.

Подчиненный индикатор $r_w(k)$ определяется на основании сравнения взвешенного ранга публикации и количества цитирований этой публикации:

$$r_w(k) = \frac{\sum_{j=1}^k cit_j}{h}. \quad (19)$$

Кроме того, в дополнение ко всему сказанному, следует также упомянуть ***h_g*-индекс** и ***q²*-индекс**, называемые накопительными индексами, которые рассчитываются следующим образом [17]:

$$h_g = \sqrt{h \cdot g}, \quad (20)$$

и

$$q^2 = \sqrt{h \cdot m}, \quad (21)$$

где h -индекс рассматривается как количественная характеристика, а g - и m -индексы содержат оценки качественных характеристик публикационной деятельности исследователя.

Таким образом, были рассмотрены основные примеры (1) – (21) методик оценки индексов научно-го цитирования, относящихся к типу T_{ind} .

ПАТТЕРНЫ ПОКАЗАТЕЛЕЙ ЦИТИРОВАНИЯ

Паттерн (от англ. «*pattern*») – это повторяемая архитектурная конструкция, применяемая в рамках некоторого часто возникающего контекста. Обычно паттерн не является законченным образцом, это лишь пример решения задачи, который можно использовать в различных ситуациях. Алгоритмы по своей сути являются паттернами вычисления, так как решают вычислительные задачи. При этом следует выделить так называемые **рекурсивные паттерны**, возникающие в том случае, когда компоненты паттерна вкладываются сами в себя.

Большинство исследований имеют предварительный характер и теоретическую природу, основываясь при этом на потенциальных эффектах вместо того, чтобы документировать фактические последствия. Большие эмпирические исследования часто ограничиваются определенными аспектами и редко затрагивают более общие изменения в научно-исследовательской деятельности и производстве. Ряд механизмов оценки, такие как использование национальных индикаторов для распределения ресурсов, также появились относительно недавно. Требуется немало времени для того, чтобы полностью оценить возникающие при этих действиях эффекты. Эта задача является довольно сложной, если не фактически невозможной, так как здесь нужно отделить влияние выбранных методов оценки от эффектов, полученных другими социальными факторами. Все эти трудности должны стимулировать, а не препятствовать дальнейшим междисциплинарным, смешанным методам обнаружения эффектов от оценки исследований и оценочных метрик [18].

Для решения проблемы построения УИНЦ (универсального индекса научного цитирования) должен быть получен паттерн, при помощи которого может быть сформирован паттерн данного показателя. Такой паттерн должен быть формализован, для этой цели подходят объектно-ориентированные средства визуализации – шаблоны (англ. «*templates*») классов или онтологические нотации в свободной форме.

CiTO (*citation typing ontology*, «онтология цитирования») является одной из таких онтологий, служащей для описания природы ссылочных цитирований в журнальных публикациях и других научных работах, относится также к другим публикациям и веб-ресурсам, а также учитывает описание публикаций в виде семантической сети. Здесь цитирования описываются с точки зрения фактических и риторических отношений между цитирующими и процитированными публикациями, а также глобальными частотами цитирований в тексте каждой процитированной публикации и природой самих процитированных работ, включая текущее состояние публикаций и полученные ими экспертные оценки [19].

Информационно-аналитические системы имеют дело с большим количеством наукометрических данных. Основная онтологическая единица – это отношение, субъектом которой является исследователь, а

объектом – результат деятельности этого исследователя. Разумеется, каждый объект может быть связан такими отношениями с несколькими субъектами. У некоторых объектов есть родительские объекты. Физически каждому типу связи исследователя и объекта соответствует таблица в реляционной базе данных. Каждый объект относится к своему классу. Все классы организованы в соответствующее их отношениям дерево классов. Некоторые классы, как правило «верхние», соответствуют отдельным таблицам в реляционной базе данных, а их подклассы определяются на основе правил. Возможно и обратное, а именно объединение некоторых физических классов в один класс онтологии.

У объекта могут также быть некоторые свойства, типами которых, как правило, являются либо действительные числа, либо булевы значения. Свойства могут как храниться непосредственно в таблице, содержащей сам объект, так и вычисляться на основе более сложных запросов. Формула состоит из множества строк, каждая из которых определяет некоторое множество объектов и задаёт функцию, сопоставляющую подходящему объекту некоторое действительное число [20].

Описание паттернов создания показателей цитирования требует унифицировать все известные на сегодняшний день унификаторы при помощи схем со спецификациями или таблиц с битовыми шкалами. Вся сложность заключается в том, что, как уже говорилось ранее, методику расчета для произвольного индекса цитирования не всегда можно записать в формульном виде.

Для каждого из типов показателей цитирования должен быть задан собственный паттерн, точнее, класс паттернов, так как паттерны одного и того же типа T_x показателя цитирования x могут быть реализованы разными способами.

Классы паттернов задаются при помощи примитивных типов данных и производных от них типов, а также свойств и методов. Формализация общей структуры семейства $\mathcal{P} = \{P_{ind}, P_{if}, P_{exp}, P_{alt}, P_{rat}, P_{mix}\}$ паттернов показателей цитирования представлена на рисунке.

Для формального описания паттернов нам понадобилась собственная нотация, основанная на языке теоретико-множественных конструкций. Итак, паттерн P_{ind} для индексов цитирования T_{ind} включает в себя следующие элементы:

- компоненты (в том числе рекурсивные, т.е. компоненты из компонент);
- целевые функции (деление, умножение и т.д.);
- функции свертки (сумма, произведение, среднее взвешенное и т.д.);
- унарные операторы (возведение в степень и т.д.);
- коэффициенты, параметры и константы.

Паттерн P_{ind} для индекса цитирования может быть записан в следующем формате:

$$P_{ind} \equiv k \cdot F \left[G \left(\{ R_i(x_i) \} \right) \right], \quad (22)$$

где k – числовой коэффициент, $k \in \mathbb{R}^+$; F – функция свертки, $F \in \{\sqrt{}, \Sigma, \min(\max), \delta(\text{сигнатура})\}$; G – целевая функция, $G \in \{-, \setminus\}$; R_i – унарный оператор, $R_i(x) \in \{x^{n_i}\}$ и $\{R_i(x_i)\} \neq \emptyset$, $n_i \in \mathbb{Q}$; x_i – индекс цитирования, рекурсивно задаваемый тем же самым паттерном P_{ind} , или «элементарный» показатель.

В некоторых вырожденных, т.е. упрощенных, случаях для некоторых из индикаторов, относящихся к T_{ind} , может быть справедливо, что $k=1$, или $\{R_i(x_i)\} = R(x)$, или $R_i(x_i) = x_i$ и т.д.

С нашей точки зрения, импакт-фактор является универсальным весовым коэффициентом, который может использоваться при расчете индексов T_{ind} и возможно некоторых других типов показателей научного цитирования.

Структуру импакт-факторного показателя в самом общем виде возможно записать при помощи паттерна P_{if} :

$$P_{if} \equiv \{P, R\}, \quad (23)$$

где P – набор параметров (количество лет с некоторого момента и т.п.); R – набор ограничений (сужение до некоторого класса объектов и т.п.).

Структуру показателя научного цитирования, рассчитываемого методом экспертных оценок и поэтому называемых «экспертными» показателями типа T_{exp} , возможно формализовать при помощи следующего паттерна P_{exp} :

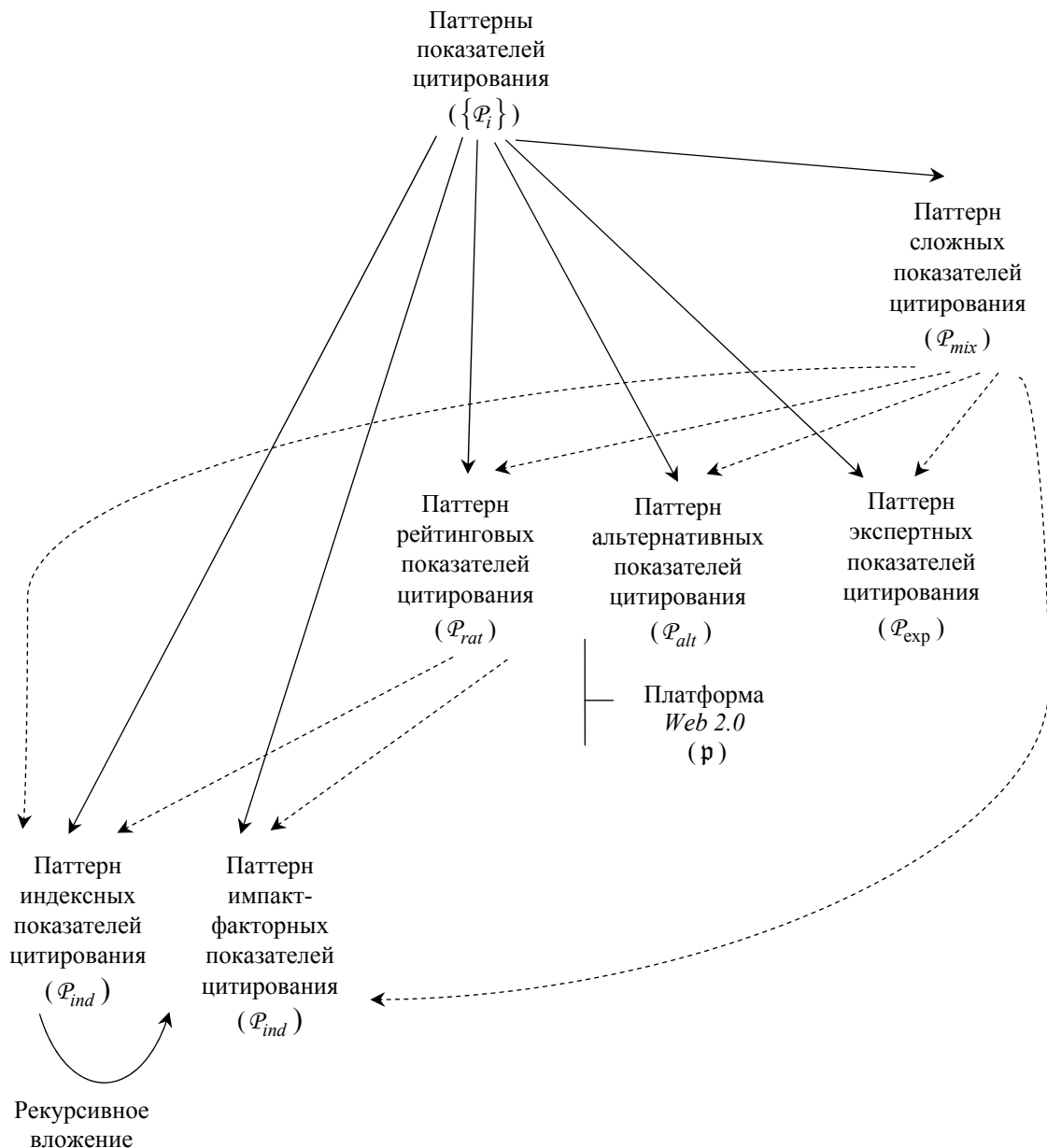
$$P_{exp} \equiv k \cdot \left\| A \left[E_{ij} = \left\{ E_i, x_j = (l_{x_j} | v_{x_j}) \right\} \right] \right\|, \quad (24)$$

где k – некоторый числовой коэффициент, $k = 10^n$; $\| \cdot \|$ – некоторая унарная операция (нормирования, приведения); A – функция агрегирования (среднее взвешенное и т.п.); E_{ij} – экспертная оценки j -го x_j показателя i -м экспертом E_i ; l_{x_j} – качественная шкала оценки показателя x_j ; v_{x_j} – числовая шкала оценки показателя x_j .

Все альтметрики T_{alt} являются платформо-ориентированными показателями. Другими словами, они являются «зеркальным» отражением индексных T_{ind} и импакт-факторных T_{if} показателей цитирования:

$$T_{alt} = (\{T_{ind}, T_{if}\} | p), \quad (25)$$

где $|p$ означает привязку к некоторой платформе p . Такая платформа должна поддерживать технологию Web 2.0.



Внутренняя структура семейства $\mathcal{P}$ паттернов показателей цитирования

Таким образом, паттерн $\mathcal{P}_{alt}$ структуры альтернативного показателя, относящегося к типу T_{alt} , идентичен индикаторному $\mathcal{P}_{ind}$ и импакт-факторному $\mathcal{P}_{ind}$ паттернам или их сочетанию:

$$(\mathcal{P}_{alt} \subset \{\mathcal{P}_{ind}, \mathcal{P}_{if}\}) \vee (\mathcal{P}_{alt} = \mathcal{P}_{ind} \star \mathcal{P}_{if}), \quad (26)$$

где под операцией $\star$ понимается такое сочетание паттернов, которое подразумевает подстановку на место компонент одного из паттернов показателей цитирования, имеющих структуру другого паттерна.

Вообще говоря, рейтинговые показатели T_{rat} являются некоторым подмножеством составных показателей T_{mix} цитирования, т.е. $T_{rat} \subset T_{mix}$, так как многие из них формируются за счет «слияния» более элементарных составляющих. Но при этом любой рейтинг допускает ранжирование, т.е. отображение в

натуральный ряд $\mathbb{N}$. Некоторые рейтинги ограничены на некотором интервале с максимальным значением sup и минимальным значением inf . Кроме того, некоторые рейтинги рассчитываются с шагом ± 1 и являются целыми числами. Для оценки рейтингов используются модели, которые, как правило, являются гибкими.

Показатели научного цитирования смешанного типа T_{mix} являются комбинацией остальных типов T_{ind} , T_{if} , T_{exp} , T_{alt} , T_{rat} :

$$T_{mix} = \{T_{ind}, T_{if}, T_{exp}, T_{alt}, T_{rat}\}. \quad (27)$$

Конкретный составной показатель цитирования может включать в себя компоненты не только по полному набору из пяти типов, но и выборочно по некоторому подмножеству S из множества указанных пяти типов.

В таком случае, «смешанный» паттерн $\mathcal{P}_{mix}$ задается следующим образом:

$$\mathcal{P}_{mix} = \prod_{i \in S} \mathcal{P}_i, \quad (28)$$

где $\mathcal{P}_i$ – i -й паттерн, а i пробегает множество $S \subset \{T_{ind}, T_{if}, T_{exp}, T_{alt}, T_{rat}\}$ и при этом $S \neq \emptyset$.

Ассоциативная операция «умножения» в (28) задается аналогично операции $\star$ из (26): как подстановка показателей цитирования на место компонент базового показателя с произвольным уровнем вложения.

В заключение следует сказать, что сконструировать новую метрику «эффективности индикаторов цитирования», принадлежащих к перечисленным типам, конечно же, возможно. Однако, если оценивать эту метрику экспертными методами, на выходе будет получено общее заключение комиссии экспертов. В другом случае, если для этой же цели использовать индикаторный метод, то конечный результат связан с влиянием особенностей базы данных и других «технических» моментов. Таким образом, вопрос о поиске универсального показателя научного цитирования остается открытым на повестке сегодняшнего дня хотя бы потому, что до сих пор не согласован критерий, по которому было бы возможно сравнивать (хотя бы попарно) показатели цитирования и определять, какой из них «хуже», а какой «лучше».

ЗАКЛЮЧЕНИЕ

В настоящем исследовании был произведен краткий обзор и дана типологическая классификация показателей научного цитирования. Новизна предложенной классификации заключается в том, что в ней формально заданы границы типов известных наукометрических показателей. Кроме того, для целого ряда показателей научного цитирования, известных как «индексы» цитирования, были приведены методики их оценки. При этом удалось показать, что, несмотря на определение индексов цитирования, при помощи языка исчисления предикатов и словесных формулировок, получивших распространение, методика оценки любого индекса цитирования может быть записана в виде формулы, строго подчиняющейся конечному и детерминированному алгоритму. Но главным результатом нашего исследования следует считать формализацию семейства паттернов, по шаблонам которых устроены существующие, а также могут создаваться новые показатели цитирования. Этот результат является попыткой продолжить дискурс о создании онтологий предметных областей, связанных с научным цитированием, который ведется нашими западными коллегами.

На основании проведенного исследования будет уместно сделать несколько рекомендаций, касающихся стратегии обращения с наукометрическими показателями. Во-первых, нужно крайне внимательно относиться к «тонким» параметрам при выборе того или иного показателя научного цитирования, в том числе числовых констант и источников данных, в зависимости от контекста, в котором он будет использован. Во-вторых, при выборе показателей цитирования нужно задействовать не только «логиче-

ское», но и «системное» мышление, рассуждая в категориях проектирования информационных систем. В-третьих, было бы полезно как можно активнее экспериментировать с показателями цитирования, «скрещивая» их разным образом на пути к следующему за альтметриками поколению показателей цитирования, контуры которого сейчас еще выглядят туманными.

В дальнейшей перспективе привлекательной представляется возможность использования паттернов наукометрических индикаторов при разработке так называемого «генератора индикаторов цитирования», который будет служить интеллектуальным модулем наукометрической информационной системы нового образца, только если уместно будет говорить об этом заранее. В таком случае, при помощи использования паттернов в качестве шаблонов станет возможно конструировать новые индикаторы цитирования с заданными наперед свойствами исходя из анализа «узких мест» в тех или иных наукометрических системах и предъявляемых к ним требований в виде спецификаций.

СПИСОК ЛИТЕРАТУРЫ

1. Ioannidis J.P.A., Boyack K., Wouters P.F. Citation Metrics: a primer on how (not) to normalize // PLoS Biol. – 2016. – Vol. 14, №. 9. – P. e1002542.
2. Ioannidis J.P.A., Klavans R., Boyack K.W. Multiple citation indicators and their composite across scientific disciplines // PLoS Biol. – 2016. – Vol. 14, №. 7. – P. e1002501.
3. Кабакова Е.А. Использование наукометрических показателей при оценке научной деятельности // Вопросы территориального развития. – 2014. – № 8 (18). – С. 1-10.
4. Portenoy J., Hullman J., West J.D. Leveraging Citation Networks to Visualize Scholarly Influence Over Time // arXiv preprint arXiv:1611.07135. – 2016.
5. Маршакова-Шайкевич И.В. Роль библиометрии в оценке исследовательской активности науки // Управление большими системами: сборник трудов. – 2013. – №. 44. – С. 210-251.
6. Третьякова О.В., Кабакова Е.А. Возможности и перспективы использования индексов цитирования в оценке результатов деятельности научного учреждения // Экономические и социальные перемены: факты, тенденции, прогноз. – 2013. – №. 6 (30). – С. 189-200.
7. Бусыгина Т.В. Альтметрия как комплекс новых инструментов для оценки продуктов научной деятельности // Идеи и идеалы. – 2016. – Т. 2, №. 2. – С. 79-87.
8. Galligan F., Dyas-Correia S. Altmetrics: rethinking the way we measure // Serials review. – 2013. – Vol. 39, №. 1. – P. 56-61.
9. Bornmann L. Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics // Scientometrics. – 2015. – Vol. 103, №. 3. – P. 1123-1144.
10. Thelwall M., Kousha K. Web indicators for research evaluation. Part 1: Citations and links to aca-

- demic articles from the Web // El profesional de la información. – 2015. – Vol. 24, №. 5. – P. 587-606.
11. Bornmann L., Haunschild R. Citation score normalized by cited references (CSNCR): The introduction of a new citation impact indicator // Journal of Informetrics. – 2016. – Vol. 10, №. 3. – P. 875-887.
 12. King C. Research fronts 2016. The hottest areas in science. – URL: <http://stateofinnovation.com/research-fronts-2016-the-hottest-areas-in-science>.
 13. Акоев М.А., Маркусова В.А., Москалева О.В., Писляков В.В. Руководство по наукометрии: индикаторы развития науки и технологии: [монография] / под ред. М. А. Акоева. – Екатеринбург: Изд-во Урал. ун-та, 2014. – 250 с.
 14. Алескеров Ф.Т., Катаева Е.В., Писляков В.В., Якуба В.И. Оценка вклада научных работников методом порогового агрегирования // Управление большими системами: сборник трудов. – 2013. – № 44. – С. 172-189.
 15. Штовба С.Д., Штовба Е.В. Обзор наукометрических показателей для оценки публикационной деятельности ученого // Управление большими системами: сборник трудов. – 2013. – №. 44. – С. 262-279.
 16. Земсков А.И. О некоторых библиометрических индексах // Научные и технические библиотеки. – 2016. – № 8. – С. 18-28.
 17. Цыганов А.В. Краткое описание наукометрических показателей, основанных на цитируемости // Управление большими системами: сборник трудов. – 2013. – №. 44. – С. 248-261.
 18. Rijcke S., Wouters P.F., Rushforth A.D., Franssen T.P., Hammarfelt B. Evaluation practices and effects of indicator use – a literature review // Research Evaluation. – 2016. – Vol. 25, №. 2. – P. 161-169.
 19. Shotton D. CiTO, the citation typing ontology // Journal of biomedical semantics. – 2010. – Vol. 1, №. 1. – P. S6.
 20. Афонин С.А., Козицын А.С., Шачнев Д.А. Использование онтологического представления структуры реляционной базы для агрегации наукометрических данных // Научный сервис в сети Интернет. – 2016. – С. 58-63.

Материал поступил в редакцию 14.04.17.

Сведения об авторе

КАЛАЧИХИН Павел Андреевич – кандидат экономических наук, старший научный сотрудник ВИНТИ РАН, Москва
e-mail: pakalachikhin@viniti.ru

Надежность предсказания по аналогии

Исследуется вопрос выбора достаточного числа запусков спаривающей цепи Маркова, чтобы с высоким уровнем доверия были порождены такие гипотезы, что в любой пример с высокой вероятностью доопределения, представленный на прогноз, вложится хотя бы одна порожденная гипотеза. Применяемая техника, основана на методе повторной выборки В.Н. Вапника и А.Я. Червоненкиса.

Ключевые слова: сходство, спаривающая цепь Маркова, нижнее полупространство, повторная выборка

ВВЕДЕНИЕ

Новый вероятностно-комбинаторный подход к интеллектуальному анализу данных был предложен автором настоящей публикации [1] на основе результатов, полученных группой отечественных ученых под руководством проф. В.К. Финна [2, 3], которые сформулировали и исследовали ДСМ-метод автоматического порождения гипотез. Так как некоторые составляющие заимствованы из теории формальных понятий [4], подход был назван вероятностно-комбинаторным формальным методом, сокращенно ВКФ-метод.

ВКФ-метод, подобно ДСМ-методу, соединяет вместе три когнитивные процедуры (индукция, абдукция и аналогия), основанные на операции сходства [5]. Главной когнитивной процедурой ВКФ-метода, как и в ДСМ-методе, является индуктивное обобщение обучающих примеров. Но, в отличие от ДСМ-метода, где вычисляются все ВКФ-гипотезы (они же ДСМ-гипотезы), наш подход применяет спаривающую цепь Маркова для порождения некоторого заранее заданного числа N вероятностно порожденных ВКФ-кандидатов. Возникает вопрос: имеется ли какое-либо указание, как выбрать это число N ?

В настоящей работе мы получим ответ на этот вопрос, воспользовавшись методом повторной выборки, впервые примененным В.Н. Вапником и А.Я. Червоненкисом [6] в задаче об оценке вероятности равномерной сходимости частот к вероятностям по классу подмножеств, имеющих конечную емкость (современные авторы называют эту характеристику размерностью Вапника-Червоненкиса).

ВСПОМОГАТЕЛЬНЫЕ ОПРЕДЕЛЕНИЯ И ФАКТЫ

Сходство является бинарной операцией на множестве X , объемлющем множество объектов, т. е. представляет собой отображение $\cap: X \times X \rightarrow X$.

Элементы множества X мы будем называть *фрагментами*. Для независимости результата нахождения сходства нескольких объектов операция сходства должна удовлетворять аксиомам *нижней полурешетки*. Для выражения тривиальности сходства добавляется специальный *пустой фрагмент* $\emptyset$ со свойством наименьшего элемента.

Важнейшим примером для нас будет нижняя полурешетка, состоящая из битовых строк фиксированной длины с побитовым умножением в качестве операции сходства. Пустым фрагментом будет являться строка, состоящая из одних нулей. Каждый бит может быть отождествлен с бинарным признаком, тогда битовая строка соответствует множеству признаков, в которых встречаются единицы. При этом операция сходства соответствует пересечению множеств признаков, а пустой фрагмент – пустому множеству признаков. Ясно, что в этом примере строка из одних единиц будет соответствовать наибольшему элементу.

Важность этого примера, с одной стороны, объясняется тем, что операция побитового умножения допускает эффективную реализацию на современных ЭВМ. Существуют специальные классы объектов (например, *dynamic_bitset* в C++), реализующие удобное оперирование битовыми строками.

С другой стороны, теорема Рудольфа Вилле [4] утверждает, что эта конструкция, если к ней добавить наибольший элемент, позволяет построить любую конечную решетку из нижней полурешетки битовых строк с операцией побитового умножения.

Собирая вместе битовые строки, представляющие объекты, мы получаем прямоугольную таблицу I , которую мы будем называть *формальным контекстом*. Формальный контекст задает бинарное отношение между элементами множества O , которые мы называем *именами объектов* (или просто *объектами*), и элементами множества F , которые мы называем *при-*

знаками. Если в строчке, соответствующей объекту $o \in O$, и столбце, соответствующим фрагменту $f \in F$, стоит единица, то мы говорим, что объект o обладает признаком f , и обозначаем это через oIf . В противном случае, говорим, что объект o не имеет признака f .

Для подмножества $A \subseteq O$ объектов его сходством называется подмножество

$$A' = \{f \in F : \forall o \in A [oIf]\} \subseteq F.$$

Полагаем $\mathcal{O}' = F$. Это определение совпадает с последовательным вычислением побитового умножения строк, соответствующих отобранному во множестве A объектов.

Для подмножества $B \subseteq F$ признаков его сходством называется подмножество

$$B' = \{o \in O : \forall f \in B [oIf]\} \subseteq O.$$

Полагаем $\mathcal{O}' = O$.

Определение 1. Пару $\langle A, B \rangle$ назовем **ВКФ-кандидатом**, если $A = B' \subseteq O$ и $B = A' \subseteq F$.

Обычно такие пары называют *формальными понятиями*, но мы предпочитаем сменить название, так как оригинальное название подвергается обоснованной критике со стороны специалистов по философии и искусственному интеллекту.

Легко проверить, что множество всех **ВКФ-кандидатов** (для фиксированного формального контекста) образует решетку.

Теорема Р.Вилле [4] гласит: **Любая конечная решетка изоморфна решетке всех ВКФ-кандидатов для подходяще выбранного формального контекста.**

Введем ключевое понятие для последующего изложения.

Определение 2. Операция *закрывай-по-одному-вниз* на ВКФ-кандидате $\langle A, B \rangle$ и объекте $o \in O$ порождает пару

$$CbODown(\langle A, B \rangle, o) = \langle (A \cup \{o\})'', (A \cup \{o\})' \rangle.$$

Операция *закрывай-по-одному-вверх* на ВКФ-кандидате $\langle A, B \rangle$ и признаке $f \in F$ порождает пару

$$CbOUp(\langle A, B \rangle, f) = \langle (B \cup \{f\})', (B \cup \{f\})'' \rangle.$$

Операция $CbODown$ соответствует шагу алгоритма «закрывай-по-одному», который был предложен С.О. Кузнецовым [7] для вычисления всех ВКФ-кандидатов перебором сверху-вниз. Следующая лемма доказывается прямыми вычислениями.

Лемма 1. Для любого ВКФ-кандидата $\langle A, B \rangle$ и любого объекта $o \in O$ пара $CbODown(\langle A, B \rangle, o)$ является ВКФ-кандидатом. Для любого ВКФ-кандидата $\langle A, B \rangle$ и любого признака $f \in F$ пара $CbOUp(\langle A, B \rangle, f)$ является ВКФ-кандидатом.

Определение 3. Порядок на ВКФ-кандидатах: $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$, если $B_1 \subseteq B_2$.

Это двойственное (с точки зрения теории формальных понятий) определение приводится в настоящем виде для согласованности с традицией отечественной школы.

Лемма 2. Для всякой упорядоченной пары ВКФ-кандидатов $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ и любого объекта $o \in O$ имеем

$$CbODown(\langle A_1, B_1 \rangle, o) \leq CbODown(\langle A_2, B_2 \rangle, o).$$

Для любой упорядоченной пары ВКФ-кандидатов $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ и любого признака $f \in F$ выполнено

$$CbOUp(\langle A_1, B_1 \rangle, f) \leq CbOUp(\langle A_2, B_2 \rangle, f).$$

ОБЩАЯ СХЕМА ВКФ-МЕТОДА

В ВКФ-методе используются следующие когнитивные процедуры:

- 1) индуктивное обобщение данных (используя цепь Маркова для порождения вероятностным образом заранее предписанного числа гипотез);
- 2) абдуктивное уточнение и принятие гипотез (порождая дополнительные гипотезы для объяснения исходных примеров);
- 3) предсказание целевого свойства у новых объектов по аналогии с обучающими примерами.

Эти процедуры являются модификацией соответствующих процедур ДСМ-метода. Основное отличие – использование вероятностного алгоритма поиска сходств (ВКФ-кандидатов) с использованием спаривающей цепи Маркова:

Data: множество обучающих (+)-примеров; внешние функции $CbOUp(\cdot)$ и $CbODown(\cdot)$ операций «закрывай-по-одному»

Result: ВКФ-кандидат $\langle A, B \rangle$

$O := (+)$ -примеры, $F :=$ признаки; $I \subseteq O \times F$ – формальный контекст для (+)-примеров;

$R := O \cup F$; $Min := \langle O, O' \rangle$; $Max := \langle F', F \rangle$;

while ($Min \neq Max$) **do**

 Выбираем случайный элемент $r \in R$;

if ($r \in O$) **then**

$Min := CbODown(Min, r)$;

$Max := CbODown(Max, r)$;

end

else

$Min := CbOUp(Min, r)$;

$Max := CbOUp(Max, r)$;

end

end

Алгоритм 1. Спаривающая цепь Маркова

Заметим, что состоянием спаривающей цепи Маркова является упорядоченная пара ВКФ-объектов $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$. Первоначально меньший ВКФ-кандидат совпадает с наименьшим

ВКФ-кандидатом $Min := \langle O, O' \rangle$, а больший – с наибольшим $Max := \langle F', F \rangle$.

В цикле к обоим ВКФ-кандидатам применяется одна и та же операция $CbODown$ с выбранным объектом, или $CbOUp$ с выбранным признаком. Процесс останавливается, когда меньший ВКФ-кандидат совпадет в большем. Тогда этот общий ВКФ-кандидат и выдается Алгоритмом 1.

Следующая теорема из статьи автора [3] объясняет, откуда здесь возникает цепь Маркова.

Теорема 1. Алгоритм 1 соответствует цепи Маркова.

Определение 4. Состояние вида $\langle A, B \rangle \leq \langle A, B \rangle$ спаривающей цепи Маркова для совпадающей пары ВКФ-кандидатов называется *эргодическим*. Состояние вида $\langle A_1, B_1 \rangle < \langle A_2, B_2 \rangle$ называется *невозвратным*.

Теперь мы имеем классическую теорему о сходимости с вероятностью единица:

Теорема 2. Вероятность того, что состояние $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ спаривающей цепи Маркова окажется невозвратным, стремится к нулю, когда $t \rightarrow \infty$.

В статье [8] приводится доказательство этого результата, но это – классический результат в теории цепей Маркова [9], поэтому здесь мы не будем приводить его доказательство.

Индуктивное обобщение обучающих примеров осуществляется следующей процедурой:

Data: множество обучающих (+)- и (-)-примеров; число N порождаемых ВКФ-гипотез

Result: выборка S ВКФ-гипотез объема N

$O := (+)$ -примеры, $F :=$ признаки; $I \subseteq O \times F$ – формальный контекст для (+)-примеров;

$C := (-)$ -примеры, $S := \emptyset$; $i := 0$;

while ($i < N$) **do**

Породить ВКФ-кандидата $\langle A, B \rangle$ с помощью цепи Маркова;

$hasObstacle := false$;

for ($c \in C$) **do**

if ($B \subseteq c'$) **then**

$hasObstacle := true$;

end

end

if ($hasObstacle = false$) **then**

$S := S \cup \{\langle A, B \rangle\}$; $i := i + 1$;

end

end

Алгоритм 2. Процедура индуктивного обобщения

Проверка условия $B \subseteq c'$ в Алгоритме 2 означает, что фрагмент B ВКФ-кандидата $\langle A, B \rangle$ вкладывается в фрагмент (множество признаков) контр-примера c .

Любое такое вложение означает, что ВКФ-кандидат нарушает условие «запрета на контр-пример». Если ВКФ-кандидат преодолевает все такие проверки, то он становится *ВКФ-гипотезой* (о причине наличия целевого свойства).

Вторая когнитивная процедура ВКФ-метода – алгоритм абдуктивного уточнения множества гипотез (подробнее см. в статье [8]).

Третьей когнитивной процедурой является предсказание целевого свойства по аналогии с обучающими примерами:

Data: расширенная выборка S^+ ВКФ-гипотез, файл (τ)-примеров

Result: предсказанные свойства (τ)-примеров

$X := (\tau)$ -примеры;

for ($o \in X$) **do**

$PredictPositively(o) := false$;

for ($\langle A, B \rangle \in S^+$) **do**

if ($B \subseteq o'$) **then**

$PredictPositively(o) := true$;

end

end

end

Алгоритм 3. Процедура предсказания по аналогии

Эта процедура пытается найти вложение хотя бы одного фрагмента $B \subseteq o'$, соответствующего хотя бы одной из порожденных (индукцией и абдукцией) ВКФ-гипотез $\langle A, B \rangle$, в каждый (τ)-пример o . Если такое вложение случается, то для этого (τ)-примера предсказывается наличие целевого свойства по аналогии с родителями $A \subseteq O$ ВКФ-гипотезы $\langle A, B \rangle$, чей фрагмент B вложился. Иначе предсказывается отсутствие целевого свойства у этого (τ)-примера o .

Для выбора числа N запусков спаривающей цепи Маркова (Алгоритма 1) внутри Алгоритма 2 полезна оценка достаточного числа ВКФ-гипотез для надежного предсказания по аналогии, составляющая главный результат настоящей работы.

ОСНОВНОЙ РЕЗУЛЬТАТ

Определение 5. Объект o , описываемый фрагментом $\{o\}' \subseteq F$ (множеством признаков), *предсказывается положительным* с помощью ВКФ-гипотезы $\langle A, B \rangle$, если $B \subseteq \{o\}'$. Если число признаков равно $n = |F|$, то фрагмент B нужно рассматривать как вершину n -мерного гиперкуба $B \in \{0, 1\}^n$.

Определение 6. Нижнее полупространство $H^\downarrow(o)$, определяемое объектом o с фрагментом $\{o\}' \subseteq F$, задается линейным неравенством

$$x_{j_1} + \dots + x_{j_k} < \frac{1}{2},$$

где $F \setminus \{o'\} = \{f_{j_1}, \dots, f_{j_k}\}$. Допускается также вырожденное нижнее полупространство $0 < \frac{1}{2}$, соответствующее $\{o'\}' = F$, и совпадающее со всем $\{0, 1\}^n$.

Класс нижних полупространств, определяемых объектами, обозначим через $Sub^\downarrow$.

Лемма 3. *Пример о предсказываемся положительным тогда и только тогда, когда в любом его нижнем полупространстве содержится фрагмент хотя одной ВКФ-гипотезы.*

Доказательство. По Определению 1 условие, что ВКФ-гипотеза $\langle A, B \rangle$ предсказывает объект o положительным, эквивалентно $B \subseteq \{o\}'$, т. е. условию $B \cap (F \setminus \{o'\}) = \emptyset$. Но в обозначении Определения 2 это означает, что $\forall i [f_{j_i} \notin B]$. Последнее эквивалентно условию $0 = f_{j_1} + \dots + f_{j_k} < \frac{1}{2}$.

Определение 7. Зафиксируем $\varepsilon > 0$ – точность предсказания. Объект o назовем ε -важным, если суммарная вероятность появления таких ВКФ-гипотез $\langle A, B \rangle$, что $B \in H^\downarrow(o)$ будет больше ε . Семейство ВКФ-гипотез назовем ε -сетью, если для каждого ε -важного объекта найдется хотя бы одна ВКФ-гипотеза из этого семейства, которая предскажет этот объект положительно.

Определение 8. Максимальное число подмножеств, откалываемых нижними полупространствами $H \in Sub^\downarrow$ у множества мощности l , задает функцию роста $m^{Sub^\downarrow}(l)$.

Лемма 4. Для $l \geq n$ верно $m^{Sub^\downarrow}(l) = 2^n$.

Доказательство. Легко проверить, что $Sub^\downarrow$ откалывает любое из 2^n подмножеств у множества $\{(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)\}$. Однако $Sub^\downarrow$ содержит 2^n нижних пространств.

Применим метод повторной выборки Вапника–Червоненкиса [6]. В следующей лемме $|S_2 \cap H|$ понимается с учетом кратности, так как элементы повторной выборки S_2 могут повторяться.

Лемма 5. Для любого ε при $l > \frac{2}{\varepsilon}$ для независимых случайных выборок S_1 и S_2 ВКФ-гипотез объемов l имеем оценку

$$\mathbf{P}^l \left\{ S_1 : \exists H \in Sub^\downarrow [S_1 \cap H = \emptyset], PH > \varepsilon \right\} \leq 2 \cdot \mathbf{P}^{2l} \left\{ S_1 S_2 : \exists H \in Sub^\downarrow \left[S_1 \cap H = \emptyset, |S_2 \cap H| > \frac{\varepsilon \cdot l}{2} \right] \right\}.$$

Доказательство. Для выборки S_1 рассмотрим нижнее полупространство $H \in Sub^\downarrow$, удовлетворяющее условиям $S_1 \cap H = \emptyset$ и $PH > \varepsilon$. Из неравенства треугольника следует, что

$$\mathbf{P}^l \left\{ S_2 : l \cdot PH - |S_2 \cap H| \leq \frac{\varepsilon \cdot l}{2} \right\} \leq \mathbf{P}^l \left\{ S_2 : |S_2 \cap H| > \frac{\varepsilon \cdot l}{2} \right\}.$$

Покажем, что при $l > \frac{2}{\varepsilon}$ выполняется

$$\begin{aligned} \mathbf{P}^l \left\{ S_2 : l \cdot PH - |S_2 \cap H| \leq \frac{\varepsilon \cdot l}{2} \right\} &= \\ &= \mathbf{P}^l \left\{ S_2 : |S_2 \cap H| \geq l \cdot PH - \frac{\varepsilon \cdot l}{2} \right\} \geq \frac{1}{2}. \end{aligned}$$

Это – вероятность для биномиальной случайной величины $|S_2 \cap H|$ быть не меньше своего математического ожидания $l \cdot PH$ за вычетом $\frac{\varepsilon \cdot l}{2} > 1$ (при $l > \frac{2}{\varepsilon}$). Известно, что медиана биномиального распределения отличается от его среднего меньше, чем на единицу. Это и доказывает нужное нам неравенство.

Из-за независимости выборок S_1 и S_2 имеем

$$\begin{aligned} \mathbf{P}^l \{ S_1 : S_1 \cap H = \emptyset, PH > \varepsilon \} \cdot \mathbf{P}^l \left\{ S_2 : |S_2 \cap H| > \frac{\varepsilon \cdot l}{2} \right\} &= \\ = \mathbf{P}^{2l} \left\{ S_1 S_2 : S_1 \cap H = \emptyset, |S_2 \cap H| > \frac{\varepsilon \cdot l}{2} \right\}. \end{aligned}$$

Поэтому

$$\begin{aligned} \frac{1}{2} \cdot \mathbf{P}^l \{ S_1 : S_1 \cap H = \emptyset, PH > \varepsilon \} &\leq \\ \leq \mathbf{P}^l \left\{ S_2 : l \cdot PH - |S_2 \cap H| > \frac{\varepsilon \cdot l}{2} \right\} \cdot &\mathbf{P}^l \{ S_1 : S_1 \cap H = \emptyset, PH > \varepsilon \} \\ \leq \mathbf{P}^{2l} \left\{ S_1 S_2 : S_1 \cap H = \emptyset, |S_2 \cap H| > \frac{\varepsilon \cdot l}{2} \right\}. \end{aligned}$$

Объединяя события в левой и в правой частях по $H \in Sub^\downarrow$, удовлетворяющим условиям $S_1 \cap H = \emptyset$ и $PH > \varepsilon$, получаем утверждение леммы.

Лемма 6. Для любого ε для двух независимых случайных выборок S_1 и S_2 ВКФ-гипотез объемов l имеем оценку:

$$\begin{aligned} \mathbf{P}^{2l} \{ S_1 S_2 : \exists H \in Sub^\downarrow [S_1 \cap H = \emptyset], [S_2 \cap H] > \varepsilon \cdot l \} &\leq \\ \leq m^{Sub^\downarrow}(2l) \cdot 2^{-\varepsilon \cdot l}. \end{aligned}$$

Доказательство. Зададим отображение g выборки $S_1 S_2$ в мультимножество над $\{0, 1\}^n$, фиксирующее состав выборки. Вероятность $\mathbf{P}^{2l}$ на $(\{0, 1\}^n)$ индуцирует вероятностную меру $g(\mathbf{P})$ на составах выборки объема $2l$. Из-за независимости и одинаковой распределенности ВКФ-гипотез в разных прогонах спаривающей цепи Маркова вероятности на выборках одного состава – одинаковы. Фиксируем состав выборки v . Для конкретного нижнего полупространства $H \in Sub^\downarrow$ условная вероятность оценивается через гипергеометрическое распределение:

$$\begin{aligned} \mathbf{P}\{S_1 S_2 : S_1 \cap H = \emptyset, |S_2 \cap H| > \varepsilon \cdot l \mid g(S_1 S_2) = v\} &\leq \\ &\leq \frac{\binom{l}{\varepsilon \cdot l}}{\binom{2l}{\varepsilon \cdot l}} = \frac{l! (2l - \varepsilon \cdot l)!}{(l - \varepsilon \cdot l)! (2l)!} = \\ &= \frac{l \cdot (l-1) \cdot \dots \cdot (l - \varepsilon \cdot l + 1)}{2l \cdot (2l-1) \cdot \dots \cdot (2l - \varepsilon \cdot l + 1)} \leq 2^{-\varepsilon l}. \end{aligned}$$

Тогда условная вероятность на выборках заданного состава

$$\mathbf{P} \left\{ \begin{aligned} &S_1 S_2 : \exists H \in \text{Sub}^l [S_1 \cap H = \emptyset], \\ &[S_2 \cap H] > \varepsilon \cdot l \mid g(S_1 S_2) = v \end{aligned} \right\} \leq m^{\text{Sub}^l} (2l) \cdot 2^{-\varepsilon l}.$$

Правая часть не зависит от состава v , поэтому, интегрируя по мере $g(\mathbf{P})$, получаем утверждение леммы.

Теорема 3. Для n признаков и любых $\varepsilon > 0$ и $1 > \delta > 0$ достаточно породить

$$N \geq \frac{2 \cdot (n+1) - 2 \cdot \log_2 \delta}{\varepsilon}$$

ВКФ-гипотез, чтобы вероятностью $\geq 1 - \delta$ все ε -важные объекты могли быть предсказаны положительно.

Доказательство. По леммам 5 и 6 для $N > \frac{2}{\varepsilon}$

имеем оценку

$$\begin{aligned} \mathbf{P}^N \left\{ S \subseteq \{0,1\}^n : \exists H \in \text{Sub}^l [S \cap H = \emptyset, \mathbf{P}H > \varepsilon] \right\} &\leq \\ &\leq 2 \cdot m^{\text{Sub}^l} (2N) \cdot 2^{-\frac{\varepsilon \cdot N}{2}}. \end{aligned}$$

Теперь по лемме 4 остается решить неравенство $2 \cdot 2^n \cdot 2^{-\frac{\varepsilon \cdot N}{2}} \leq \delta$ относительно N , чтобы получить утверждение теоремы.

ЗАКЛЮЧЕНИЕ

В настоящей работе получена оценка необходимого числа N запусков спаривающей цепи Маркова, чтобы с высоким уровнем доверия ($\geq (1 - \delta)$) были порождены такие ВКФ-гипотезы, что в любой ε -важный объект, представленный на прогноз, вложится хотя бы одна из порожденных ВКФ-гипотез. Техника, которую мы применяем, основана на методе повторной выборки В.Н. Вапника и А.Я. Червоненкиса [6], однако наши результаты точнее тех, которые могут быть получены прямым сведением к теореме о равномерной сходимости частот к вероятностям.

* * *

Автор благодарит проф. В.К. Финна за внимание к работе, проф. Е.М. Бениаминова за полезные обсуждения и своих коллег по лаборатории 35 ФИЦ ИУ РАН за поддержку и полезные дискуссии.

СПИСОК ЛИТЕРАТУРЫ

1. Vinogradov D.V. VKF-method of hypotheses generation // Communications in Computer and Information Science. – 2014. – Vol. 436. – P. 237-248
2. Аншаков О.М., Скворцов Д.П., Финн В.К. Логические средства экспертных систем типа ДСМ // Семиотика и информатика. – 1986. – Вып. 28. – С. 65–102
3. ДСМ-метод автоматического порождения гипотез: Логические и эпистемологические основания / ред. В.К. Финн, О.М. Аншаков. – М.: URSS, 2009. – 432 с.
4. Ganter Bernard, Rudolf Wille. Formal Concept Analysis. Transl. from German. – Berlin: Springer-Verlag, 1999. – 284 p.
5. Гусакова С.М., Финн В.К. Сходства и правдоподобный вывод // Известия АН СССР. Сер. «Техническая кибернетика». – 1987. – № 5. – С. 42–63
6. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов (статистические проблемы обучения). – М.: Наука, Гл. ред. физ.-мат. лит., 1974. – 416 с.
7. Кузнецов С.О. Быстрый алгоритм построения всех пересечений объектов из нижней полурешетки // Научная и техническая информация. Сер. 2. – 1993. – № 1. – С. 17–20.
8. Виноградов Д.В. Вероятностное порождения гипотез в ДСМ-методе с помощью простейших цепей Маркова // Научная и техническая информация. Сер. 2. – 2012. – № 9. – С. 20–27; Vinogradov D.W. Random Generation of Hypotheses in the JSM Method using Simple Markov Chains // Automatic Documentation and Mathematical Linguistics. – 2012. – Vol. 46, № 5. – P. 221-228.
9. Кемени Дж., Снелл Дж., Кнепп А. Счетные цепи Маркова / пер. с англ. – М.: Наука, Гл. ред. физ.-мат. лит., 1987. – 416 с.

Материал поступил в редакцию 04.05.17.

Сведения об авторе

ВИНОГРАДОВ Дмитрий Вячеславович – кандидат физико-математических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН и Российского государственного гуманитарного университета, Москва e-mail: vinogradov.d.w@gmail.com

С.С. Терещенко

Информационно-аналитические аспекты процесса реформирования вузов и университетских сетей (по материалам диссертационных исследований)

Представлен обзор диссертаций по математическим, программным, алгоритмическим и техническим средствам разработки и внедрения систем информатизации и автоматизации управления в вузах и университетах России, по формализации средств накопления знаний и определения индивидуальной траектории обучения, а также по оценке учебных заведений и знаний студентов. Приводятся примеры использования веб-технологий, экспертно-аналитических, многоуровневых аналитических и административных систем. Показана преемственность в применении методов и технологий Государственной системы НТИ для проектирования современных информационно-сетевых технологий на новом техническом уровне (избирательное распространение информации, ретроспективный поиск, создание БД для ситуационного анализа сети сайтов международного уровня на основе методов лингвистического масштабирования. Обсуждаются технологии оперативного обслуживания типа сигнальной информации, дайджестов, досье из современных сетевых приложений с учетом индивидуальных классификационных и тезаурусных предпочтений экспертов, их групп или аналитических экспертных служб.

Ключевые слова: система автоматизации управления в университете, программно-технологические средства, государственная система НТИ, лингвистическое масштабирование, мониторинг, рейтингование, научно-образовательная сфера

ВВЕДЕНИЕ

В Российской Федерации и в других странах постсоветского пространства развиваются процессы реструктуризации научной и образовательной сфер, активно влияющих на сопутствующие интеллектуальные процессы в социальной, экономической, производственно-технологической деятельности. Значительные изменения происходят в сети академических институтов, университетских и вузовских структур, их промышленных образованиях и предприятиях оборонного комплекса. Это радикально меняет информационную инфраструктуру в рамках всего государства, что отражается на использовании информационно-аналитических ресурсов в управлении государством, а также на разработке стратегий инновационного развития, ресурсосбережения, антикризисного и ситуационного управления, программ опережающей подготовки профессионалов новой формации. В условиях сложной международной обстановки в 2017 г. президент Российской Федерации наметил следующие направления развития страны: сбережение человеческого капитала, трансформирование его в новые социальные, научно-образовательные, промышленно-технологические и оборонные проекты, продолжение инновационного развития. Это требует критической оценки сохранившихся информационных ресурсов сети научно-технических библиотек, центров НТИ отраслей и регионов, университетских информационно-аналитических структур, а

также экспертно-аналитического потенциала академий и институтов оборонного комплекса.

ВИНИТИ РАН как головной орган государственной системы научно-технической информации в своей программе перспективного развития [1] консолидирует интеллектуальные ресурсы и расширяет сеть информационно-аналитического сотрудничества с академическими и университетскими коллективами, поддерживает информационно-аналитические, экспертные, референтские, педагогические и исследовательские контакты с соответствующими кафедрами и факультетами ведущих вузов и университетских кластеров, особое внимание уделяет реализации федеральной программы «Развитие научного потенциала высшей школы», изучению научных достижений по новым направлениям фундаментальной и прикладной информатики и их внедрению в университетскую практику, а также информатизации всех подразделений высшей школы – педагогических, научно-исследовательских, инновационных, административных, включая сетевое взаимодействие с центрами НТИ.

ПРОГРАММНО-ТЕХНОЛОГИЧЕСКИЕ РЕШЕНИЯ

По материалам докторской диссертации А.М. Гудова «Программно-технологический комплекс для развития информационной среды образовательного учреждения на основе системы электронного документооборота», защищенной в Кемеровском госу-

дарственном университете [2], разработаны и внедрены 20 информационных систем документального и фактографического типов, используемых в административном управлении университета, депозитарии электронных образовательных ресурсов «Деканат» и «Учебные процессы», а также аналитическая составляющая научных исследований и конференций. Все эти системы объединены интерфейсом пользователя в пакете «Kemsu WEB», реализующем веб технологии, системы электронного документооборота, комплексы справочников и систем регламентации доступа.

Справочники и прикладные базы данных обеспечиваются многоуровневой службой вызова, использующей технологии системы управления базами данных. Общее хранилище для всех типов систем включает следующие компоненты: WEB-приложения, программы аналитических отчетов, запросно-ориентированную структуру аналитических данных для постоянных запросов (база гиперкубов данных). Загрузка и выгрузка данных осуществляется XML-файлами с описанием структур данных (XML-язык описания метаданных документов).

Разработчик этой универсальной университетской среды использует самые современные общесистемные решения интегральных информационных систем, подобные хорошо известным при создании Государственной системы НТИ и системы «АССИСТЕНТ», а также принцип интегральности, т.е. одноразовый ввод данных и многократное их использование для разных задач управления и справочно-информационного обслуживания, включая режим избирательного распространения информации. В наше время это обеспечивается одним из процессов WEB-технологий.

В исследовании А.М. Гудова формируется единая информационная среда из следующих компонентов: организационные структуры информатизации университета (нескольких вузов), средства связи и коммуникации, центры высокопроизводительных вычислений, центры коллективного пользования, функциональные подсистемы, информационные ресурсы. Накопление информации носит разный временной характер в обработке, хранении, уточнении и в аналитической обработке документов. Автор предлагает универсальную единую информационную среду как прототип автоматизированных систем управления, информационно-поисковых систем (документальных и фактографических), баз и банков данных, систем управления базами данных с возможностью комплексования и интерпретации ГРИД-технологий в различных масштабах (в пределах одного университета, нескольких вузов, центров информации, научно-технических библиотек и центров НТИ федерального или территориального подчинения) с учетом принципов программирования на основе современных стандартов открытых систем (*Open Systems*) и сервисно-ориентированной архитектуры (*Service-Oriented Architecture*).

В рассматриваемой диссертации использованы такие методы исследований, как системный анализ, теория оптимизации, элементы теории нейронных сетей, методы извлечения данных из текстов и создания онтологий, методы функционального и объектно-ориентированного проектирования, новые техно-

логии разработки и тестирования программных систем. Особое внимание уделено математическим, логическим, семиотическим и лингвистическим моделям и взаимодействию информационных процессов на базе специализированных вычислительных комплексов и современных телекоммуникационных систем в активной интернет-среде.

Технологии интегральной системы ВИНТИ РАН «АССИСТЕНТ» объединили информационно-аналитические ресурсы федеральных (всесоюзных), отраслевых, территориальных и республиканских центров НТИ в рамках Государственной автоматизированной системы НТИ (ГАСНТИ). Это позволило создать информационно-аналитический базис для администраций субъектов Российской Федерации – основу формирования сети региональных информационно-аналитических служб. Администрация Кемеровской области поддержала проект «Создание областного реестра информационных ресурсов, баз данных научно-технической информации, информации учебного назначения и электронных средств обучения (2006 год)», а Министерство образования и науки РФ предложило проект «Создание типового информационно-вычислительного портала вуза» в рамках целевой программы «Развитие научного потенциала высшей школы».

Эти проекты с учетом новой концепции развития ВИНТИ РАН, будут способствовать активизации объединения сохранившихся фрагментов созданной ГАСНТИ, которые восстанавливаются в отраслях, субъектах РФ, странах СНГ. Эффективные условия для этого определяются государственными стандартами и нормативами: ГОСТ Р 7.0.10-2010 (ISO 15836-2003) Онтологическая модель единой информационной среды – модель объектов онтологии – «Дублинское ядро»; ГОСТ Р 3898-2010 Системы электронного документооборота; ГОСТ Р 51141-98 Делопроизводство и архивное дело. Термины и определения; ГОСТ Р ИСО 9000-2001 Системы менеджмента качества. Основные положения и словарь, а также законом РФ № 24-ФЗ от 20.02.1995 «Об информации, информатизации и защите информации» и законом РФ от 10.01.2002 № 1-ФЗ «Об электронно-цифровой подписи».

Для активной работы информационного комплекса кроме требований к сервис-ориентированной, процессно-ориентированной, гетерогенно-информационно-ориентированной архитектуре, нужно особо выделить требования к построению распределенной корпоративной информационной системы – Технологический подход – Сервисная шина предприятия, которая включает описание процессов, XML-сообщений (XML-файлы), сервисов, каталогов сервисов. Профиль открытых систем «Единая информационная система» базируется на профиле взаимодействия открытых систем (BOC) России, тип распределенных сетей связи и передачи данных по этим сетям – на эталонной модели OSI/RM, программное обеспечение сервера приложений – на *Apache Tomcat* WEB-сервера, программное обеспечение баз данных, хранилищ данных – на объектно-реляционной СУБД *ORACLE 10G R2*. Этот проект выбран вместо двух СУБД – реляционной и объектной, бизнес-логика реализо-

вана на языках *PL/SQL*, языке *AVA* и внедрен в пяти университетах России.

Другой вариант из коллекции архитектурных подходов к созданию системных программ разрабатывается в Южно-Уральском государственном университете в Челябинске. Он описан в кандидатской диссертации М.И. Нестерова «Управление развитием информационно-аналитической системы современного вуза на основе архитектурного подхода» [3], где практически показана разработка интегрированной информационно-коммуникационной среды университета на основе баз, банков и хранилищ данных. Исследованы возможности объектно-ориентированного, процессного и архитектурного подходов.

Основой архитектурного подхода является концепция структуры университета или предприятия, включающей функциональные подразделения, информационные технологии (ИТ), управление архитектурным процессом. Разработчик выбрал методику *TOGAF (The Open Group Architecture Framework)* и архитектуру федеральной организации (*FEA*), объединяющие бизнес-архитектуру, организационную структуру, информационные потоки, технологии сбора и использования информации, ее хранение. Этот подход использован при создании информационно-аналитической системы в управлении развитием вуза с выделением следующих подсистем: административно-управленческой, образовательной, научно-исследовательской, информационно-технологической, хозяйственной. Автоматизации подлежат процессы приема абитуриентов, учебная деятельность, практики студентов, стимулирование работы преподавателей, организация НИР и ОКР, контроль выполнения плановых показателей, установленных учредителем в течение учебного года, а электронизации – научно-техническая библиотека, электронный документооборот, бизнес-процессы внедренческой деятельности.

Архитектура приложений обеспечивает управление данными и их использование в прикладных ИПС. Технологическая архитектура содержит аппаратные и программные средства согласно операционным требованиям. Информационно-аналитические процессы обеспечивают работу сети личных кабинетов по административному регламенту. Аналитика включает статистический анализ, прогноз показателей работы университета в соответствии с требованиями Минобрнауки РФ. Особое внимание в диссертации уделено издательской деятельности по показателю «Количество статей в научной периодике, индексируемой иностранными и российскими организациями (*Web of Science*, РИНЦ)». Автор предлагает планировать подготовку научных публикаций.

Техническое оснащение Южно-Уральского государственного университета включает: серверное оборудование, сетевые приложения (проводные и беспроводные – *Wi-Fi*), рабочие станции, программное обеспечение; информационная система университета и его 10 филиалов работает в сетях с поддержкой протокола *http/https*; пользователь применяет операционные системы (*Windows, Linux, MacOS*) и любой веб-браузер (*Internet Explorer, Google Chrome* и др.), ему обеспечивается доступ к системе даже из-за рубежа. Электронный документооборот реализуется на

пакетах программ (*MS Office, Libre Office*). В Информационно-аналитической системе зарегистрировано 4 тыс. пользователей. Разработка может быть рекомендована крупным вузам и университетам. К сожалению, Южно-Уральский государственный университет не имеет пока совместной технологии по аналитической прогнозной работе с Челябинским ЦНТИ и аналитической службой администрации Челябинской области.

В диссертации О.А. Бубаревой «Модель, алгоритмы и программное обеспечение интеграции данных информационных систем на основе онтологий (на примере вуза)» [4] рассмотрены модели, алгоритмы и программное обеспечение интеграции данных информационных систем вуза на основе онтологий и компьютерных сетей. Предлагаемая автором единая онтология верхнего уровня понятий устраняет проблемы неоднородности семантического расширения предметных областей в деятельности научно-образовательных структур и, используя вычислительный метод оценки семантической близости концептов двух онтологий на основе генетического алгоритма, не требует корректировки моделей данных и программ при работе с неоднородными онтологическими спецификациями. Такой подход позволил разработать и внедрить гетерогенную информационно-аналитическую систему вуза для управления учебным процессом, планирования и финансового обеспечения, а также подсистемы «Успеваемость студентов», «Аудиторный фонд», «Расписание». На программы этих подсистем получены авторские свидетельства Федеральной службы по интеллектуальной собственности.

Концептуальная модель системы интеграции данных включает следующие модули: администрирования, интеграции данных, изменения онтологических описаний, разрешения конфликтов, пользовательского интерфейса, формирования шаблонов запросов, изменения онтологических описаний, сравнения и отображения онтологий, репозитория метаданных с онтологией предметных областей. Хранилище данных реализовано на *OLAP*-структуре в программной среде – *ORACLE Irg* (репозиторий на *Oracle Spatial*).

Когнитивный и онтологический подходы, использованные при разработке информационно-аналитической системы для поддержки процессов управления кафедрой вуза в Пензенском государственном университете, описаны в кандидатской диссертации А.А. Акимова «Информационно-аналитическая система для поддержки процессов управления кафедрой вуза» [5]. Автор отмечает «островную» автоматизацию в создании информационно-аналитической инфраструктуры для научно-образовательной сферы, предлагает интеграцию хранилищ данных (ХД), многомерные базы данных (МБД), оперативный анализ данных (ОАД, *OLAP*), интеллектуальный анализ данных (*Data Mining*, ИАД), онтологическую модель информационной системы анализа деятельности кафедры вуза, включающую множество концептов предметной области, инфраструктуры, управления, а также методику построения когнитивных карт, которые описывают деятельность кафедр и ранжируются экспертами. В диссертации выделены 11 когнитивных карт: оценка

деятельности кафедр, учебно-методическая работа, повышение квалификации преподавателей, организационно-методическая работа, подготовка научно-педагогических кадров, список рейтинговой оценки, квалификация профессорско-преподавательского состава, научно-исследовательская работа, владение информационными технологиями, инновационная деятельность, поощрения. Для прогнозирования автор использовал сценарный анализ – сочетание неформального логико-эвристического подхода при оценке вероятных сценариев развития кафедральных ситуаций и показателей, их сопровождающих.

Комплекс информационно-аналитической системы включает следующие модули: ввода, обработки и хранения, формирования отчетности, мониторинга, администрирования. Формат документов – Microsoft Word, Microsoft Excel. Средства разработки системы: СУБД *Microsoft SQL Server*, язык программирования C++, фреймворк *ASP.NET MVC*, спецификация формирования требований к системе – FURPS+. Сетевой протокол – *HTTP/HTTPS* (На стороне клиента необходимо наличие веб-браузера, поддерживающего *JavaScript*), серверная часть – веб-сервер *Internet Information Server (IIS)*, сервер баз данных – *Microsoft SQL Server 2008* с поддержкой *Analysis Services*. Аппаратная часть определяется в зависимости от общей загрузки.

Еще одна диссертация по развитию онтологического метода проектирования представлена Воронежской государственной лесотехнической академией: Л.М. Манучарян «Разработка средств повышения релевантности ИПС на базе онтологий» [6], в которой исследуются проблемы искусственного интеллекта и семантические вычисления в области связанных данных с использованием предметно-структурированной информации. Автором разработаны технологии извлечения информации из глобальной сети. Результаты поиска представляются в виде онтологий для практического использования в массиве персональной информации с устранением неоднозначности при поиске фамилий и имен.

Примененные методы повышают релевантность базы онтологий в АИПС при извлечении и семантическом представлении структурированной информации из неструктурированного текста. Для извлечения информации в диссертации исследованы: дерево грамматического разбора, графы зависимостей, онтологии с экземплярами, процессы валидации (проверки) и структуризации. Извлечение информации происходит без учета предметной области. Разработанная технология может быть реализована в виде веб-приложения.

Система извлечения информации работает с предметно-специфическими сведениями из необработанного текста в виде сложных (или несложных) связей. Используемые алгоритмы выявляют ключевые сущности и связи на основе фундаментальных знаний, они позволяют сопоставлять извлеченные фрагменты с базовыми знаниями согласно тестируемой предметной области, представляя результаты в виде RDF-графов. При разборе извлеченных предложений связи в словах сравниваются с библиотекой типизированных зависимостей (в стэнфордской

библиотеке грамматического разбора). Для каждого предложения автор рекомендует формировать дерево грамматического разбора и граф зависимостей (*Stanford parser*-технология). Разработанная Л.А. Манучарян модель поиска персональной информации основана на текстовой кластеризации по тематическим признакам и отличается шестью идентификационными признаками (название Web-страницы, URL-адрес документа, метаданные, фрагменты, контекстное предложение, совокупность слов). Модель использована в АИПС *Arm-Personfinder* компании *SceonDev*.

МОНИТОРИНГ РЕЙТИНГОВАНИЯ НАУЧНО-ОБРАЗОВАТЕЛЬНОЙ СФЕРЫ

Оценка приоритетов в науке и в образовании становится существенным аргументом при их инновационном развитии. Естественно, информатизация и аналитическое обеспечение этого процесса требуют развития и дополнительных ресурсов.

Диссертационное исследование, Я.П. Бондарева «Разработка моделей и алгоритмов оценки и управления положением в рейтинге субъектов и объектов высшего образования [7]» выполнено в рамках темы НИР Воронежского института высоких технологий «Разработка технологии создания многоуровневых веб-ориентированных систем нового поколения». Автор предлагает рассматривать самооценку вуза и его развитие с минимальными затратами для восходящего продвижения в сетевом рейтинге. Он создает единую информационную среду института, веб-ориентированную с электронным документооборотом и связями между всеми субъектами и объектами, а также студенто-ориентированные учебный процесс и образовательную систему, и анализирует учебный контент и индивидуальные траектории обучения, рейтинговые технологии, интеграцию мотивационной составляющей, учебные и научные достижения, социальные взаимосвязи, модуль поддержки принятия решений, гибкую архитектуру системы. Разработанные рейтинговые оценки студентов, преподавателей, зав. кафедрой, декана факультета предлагаются в баллах (от «0» до «100» баллов), устанавливаются причинно-следственные связи неуспеваемости обучающихся. Для изменения положения в рейтинге автор использует алгоритмы интеллектуальной поддержки, многокритериальную задачу дискретного программирования с булевыми переменными методом ветвей и границ, многомерную задачу о «Ранце», предлагает оптимизацию образовательной, научно-исследовательской, международной, финансово-экономической, организационно-инфраструктурной деятельности, а также автоматизированный вариант системы рейтингования (электронный документооборот и веб-технологии).

Информатика и аналитика рейтингования научно-образовательной сферы включает оценку участников этих процессов на всех стадиях обучения – «школа – вуз – исследовательская структура». Рассмотрим один из подходов оценивания абитуриентов вуза на примере Сибирского государственного индустриального университета, представленный в диссертации Н.Б. Бабичевой «Методы и модели рейтингового

оценивания абитуриентов вуза» [8]. Автор отмечает, что при отборе абитуриентов средний балл по ЕГЭ не отражает многих способностей будущих студентов, и необходимо специально оценивать довузовские знания такими дополнительными экспертными нормативными моделями для технических вузов, как модель оценки с лингвистическими переменными для разных специальностей и направлений обучения («важно», «почти важно», «не важно»); комплексная методика, объединяющая балльный метод, метод с весовыми коэффициентами и элементами нечеткой логики.

В разработанной АИС «Абитуриент» предлагаемая методика позволила повысить качество сдачи первой сессии по профильным предметам на 10-15%. В этой методике учитывались американская, британская, скандинавская и европейская континентальная модели перехода от школы к вузу, что способствует повышению вероятности успешного освоения программ высшего профессионального образования, творческого участия в исследовательских и конструкторско-технологических разработках. Рейтинговый расчет предлагается в баллах с использованием весовых коэффициентов и лингвистических переменных (трудно переводимых в числовую форму). Рейтинговая оценка по комбинированной методике корректируется по новым правилам приема абитуриентов и с учетом изменения и развития дисциплин вуза «Прикладная информатика», «Информатика и программирование». В АИС «Абитуриент» активно использован интернет-сервер университета по предварительной работе с будущими абитуриентами. Предлагаемая методика рейтингового оценивания может использоваться и для гуманитарных вузов.

Следующая диссертация по рейтингованию рассматривает развитие интеллектуальных обучающих систем, электронных систем обучения, концепцию виртуальных учебных сред и систем управления обучением. Исследование «Разработка веб-ориентированной интеллектуальной системы представления и проверки знаний на основе концептуальной модели предметной области» [9] проведено О.Ю. Лазаревой в Московском госуниверситете печати им. И.Ф. Федорова. Автор описывает предметную область «дидактической единицы», их совокупностью – «Учебная дисциплина», «Компетенция». Предметная область представлена следующей семантической сетью: (Дидактическая единица) → (Тесты, Лекции, Лабораторные работы) → (Дисциплины) → (Компетенции) → (Профиль обучения). Когнитивная карта (модель знаний экспертов) – это совокупность оцененных экспертами отношений (шкала весов). Модель учащегося – это совокупность усвоенных дидактических единиц. Динамическая оверлейная модель учащегося оценивает успешно усвоенные знания. Теория нечетких множеств фиксирует пробелы в знаниях, что фиксируется операторной моделью дидактической единицы в терминах языка логического программирования *SWI-Prolog* в интеграции с языком *PHP*. Когнитивные карты визуализируются с использованием JavaScript-библиотеки *D3*. Эта интеллектуально-обучающая система внедрена в «Электронном

учебном центре – вуз» и представляет собой совокупность веб-сайтов.

При разработке системы использованы следующие языки программирования и разметки, библиотек и систем: *HTML*, *CSS*, *PHP*, *XML*, *MySQL*, база данных включает данные о пользователях, информацию о курсах и тестах, результаты тестирования. База знаний – совокупность когнитивных карт дисциплин, алгоритмы работы с картами. Модель предметной области в виде семантической сети с отношением «предыдущие - последующие» становится когнитивной картой, если этим отношениям заданы веса, определяющие важность одного концепта после изучения другого концепта. Таким образом, веб-ориентированная интеллектуальная обучающая система совмещает возможности систем управления обучением и интеллектуальных обучающих систем.

Результаты исследований в Московском государственном университете прикладной биотехнологии свидетельствуют о непредсказуемости системы образования, а рейтинговая или статистическая оценка часто не учитывает динамику ее состояния, поведения и взаимодействия активных элементов (преподавателей и студентов) при заданных и изменяющихся условиях управления]. В диссертации Е.А. Назайкина «Мультиагентное имитационное моделирование образовательного процесса накопления знаний» [10] поэтапное взаимодействие обучаемых и обучающих предлагается оптимизировать с учетом информационных, психофизиологических и социальных факторов. Когнитивное состояние студентов характеризуют интеллект по *IQ*, внимание, способность к восприятию, уровень априорных знаний, коэффициент эффективности переработки информации, скорость восприятия информации, текущий уровень знаний, а эмоциональное состояние – психотип личности, степень удовлетворенности обучением, эмоциональная реакция. Персональные характеристики – это сознательность, трудолюбие, утомляемость, социальное состояние – индекс социометрического статуса, уровень взаимодействия. На процесс обучения возникает эмоциональная реакция – удовольствие, ожидание, безразличие, отрицание, надежда, интерес, разочарование.

Исходное состояние студента автор описывает уравнением регрессии от исходных параметров и рассчитывает с помощью корреляционного и регрессионного анализа. Эффективность процесса накопления знаний он определяет отношением воспринятой информации к представленной. Более точное значение дает алгоритм обучения искусственной нейронной сети для значимых факторов с использованием многослойного персептрона. Преподаватели оцениваются по таким когнитивным параметрам, как объем знаний, уровень умений, степень владения предметной областью. Взаимодействие «студент - преподаватель» в диссертации описывается моделью в терминах объектно-ориентированного языка – *Simplex-MDL (Model Description Language)* с декларированием переменных состояния, сенсорных связей в процессе обучения в виде алгебраических и дифференциальных уравнений.

Процесс накопления знаний реализуется студентами на лекциях, практических занятиях или самостоятельно. Связь между студентами и преподавателями можно описать мультиагентной моделью (успеваемость, эффективность, запрос о помощи, оценка эмоциональной реакции). Рассмотренная в диссертации мультиагентная имитационная модель образовательного процесса может быть реализована в универсальной имитационной системе *SimplexMDL* и языке описания эксперимента.

Требованиям федерального государственного образовательного стандарта третьего поколения (ФГОС ВПО 3) посвящена работа О.В. Махныткиной «Моделирование и оптимизация индивидуальной траектории обучения студента (ИТОС)» [11], выполненная в Алтайском государственном университете. Решаемая автором проблема – сложная, слабо формализованная, большой размерности, должна учитывать индивидуальные особенности студента, рынок труда, сложности процесса обучения и неоднородность шкал измерений результатов обучения, иерархическую их структуру. Предлагаемая гибридная экспертная система учитывает компетентность выпускника, участие в научных работах, его индивидуальные особенности. По сравнению с ФГОС ВПО 2-го поколения понятие компетенции расширяется – это динамическая совокупность знаний, умений, навыков, способностей, личных качеств после завершения обучения («компетентность» трактуется как обобщающее понятие по отношению к понятию «компетенция», и оба термина – как носящие надпредметный, междисциплинарный характер).

Автор выделяет интегральную компетентность, обобщающую компетенцию, уровень промежуточных компетенций, а также общекультурную и профессиональную компетентность. Совокупность дисциплин, разделов учебного плана, тематика НИР определяются тематической областью формирования частной компетенции. Оценка уровня формирования компетентности проводится экспертной системой и аналитической подсистемой. Научно-исследовательские работы проводятся в каждом семестре. Индивидуальная траектория обучения включает упорядоченный набор дисциплин и тематику НИР для конкретного студента. Для получения качества обучения выше заданного общего критерия используется алгоритм оптимальности Р. Беллмана.

Автоматизированная информационная система для построения индивидуальных траекторий обучения студентов (АИС ИТОС) и их информационно-аналитического обеспечения реализует сбор, ввод, учет и хранение всей информации – это учебный план, учебные курсы и программы, сведения об успеваемости студентов, о результатах НИР, учебный график, методические материалы. Учитываются предложения будущих работодателей. Все это приводит в итоге к повышению уровня компетентности выпускника, ориентированного на конкретную рабочую должность.

Проблеме информатизации при оценке научно-исследовательской сферы в вузах, которым необходимо кроме педагогической деятельности, активизировать фундаментальные и прикладные исследования, посвящена диссертация «Поддержка принятия

решений при управлении научно-исследовательской работой студентов вуза» [12], выполненная Е.В. Измайловой в Пермском национальном исследовательском политехническом университете. Автор отмечает, что социальные системы – наука, образование, воспитание, трудно оценивать экономическими критериями – доходом, издержками, рентабельностью. Чаще вузы оцениваются по качеству подготовки студентов. В программы профессиональной ориентации выпускников включается участие в научно-исследовательской работе. Показано, что существующие системы автоматизации деятельности вузов – «ПАРУС-ВУЗ», «1С : Университет», «Галактика» не оценивают исследовательские направления, их значимость при проверке и контроле. Автор разработала Автоматизированную систему управления научно-исследовательской работой студентов (НИРС) на основе многоагентной технологии, включающей структурные подразделения, множество задач, образовательную среду.

Многоагентная система относится к самоорганизующимся системам и объединяет следующих агентов: стратегического (администрация), решающего центрального (декан, зав. кафедрой), специалистов, программного. При различных сценариях исследовательской работы для всех ее участников автор предлагает формировать критерии самооценки. Веса критериев предварительно были определены по результатам всероссийского опроса, в котором участвовали 14 докторов и 19 кандидатов наук. Оценивалась важность 92 показателей, не учитываемых в официальной оценке. Предлагаемая автоматизированная информационная система научно-исследовательской работы студентов (АИС НИРС) территориально распределена по нескольким структурам Пермского Университета, для нее использованы возможности Интернета, а для программной реализации выбран язык программирования *PHP* и система управления базами данных – СУБД *MySQL*. Рассмотренная АИС упростила подготовку ежегодных отчетов и участие в рейтинговании внутри университета.

Последняя в нашем обзоре диссертация А.А. Артамонова «Трёхуровневая система агентного поиска и обработки научно-технической информации» [13] имеет существенное значение для служб НТИ отраслевого, территориального, ведомственного и всероссийского уровней в теоретическом и имиджевом отношениях. Эта разработка Национального исследовательского ядерного университета «МИФИ» объединила все лучшее, накопленное в свое время во Всесоюзном институте межотраслевой информации, ЦНИИТЭИ Электроника, Информэлектро и других организациях Государственной системы НТИ. Автором были рассмотрены режимы ИРИ, СИ, подготовки ежегодных итогов науки и техники. А.А. Артамонов предлагает трехуровневую систему агентного поиска и обработки НТИ, но в режиме *Web*-технологий и экспертного анализа, что расширяет функциональное назначение «Пользовательского Агента».

В диссертации дается определение «Агента» по уровням разработанной системы: агент-почтальон (*P*) –

«программный робот», взаимодействующий с информационными ресурсами Интернета по заданному расписанию и собирающий информацию в текстовую БД; агент курьер (*K*) – агент, взаимодействующий с агентом почтальоном и текстовой БД, систематизирующий информацию по соответствующим тематикам; агент-референт (*R*) работает с пользователем и агентом-курьером при подготовке пертинентной информации из баз данных системы.

Система формирует следующие аналитические продукты: на первом уровне – дайджесты, досье на объекты интереса и семантические сети; на втором уровне – тематические тезаурусы; на третьем уровне – персональные БД с рубрикацией информации по этапам разработки изделия, по областям приложения знаний и технологий, а также персональные словари тематических областей, пользователей, объектов, стадий жизненного цикла.

Одним из источников информации для диссертации был Отдел оборонных исследований агентства *DARPA*, США, другим – Семантическая сеть по лазерным протехнологиям, построенная на материалах международной организации *Laserlab-Europe*, сайт которой содержит более 10 тыс. страниц. Автор проводит ситуационный экспресс-анализ информационных ресурсов, описание и анализ ситуации в исследованиях, программных средств, инструментальных технологий, экспертов, интерпретирует результаты анализа, проводившегося с использованием 9 тыс. ключевых слов по частоте их использования.

По материалам диссертации разработан учебный курс по теме «Информационно-аналитические исследования с использованием интеллектуальных агентных систем» при подготовке многоязычных дайджестов семантических сетей, динамических досье, аналитических отчетов. Автором получены 4 свидетельства о государственной регистрации баз данных. Метод лингвистического масштабирования агентной системы создан на основе международного стандарта *TMX 1. 4b – Specification*. Система внедрена в Научно-техническом институте межотраслевой информации (НТИМИ) и в Центре плазменных и лазерных технологий Научно-исследовательского ядерного университета МИФИ.

ЗАКЛЮЧЕНИЕ

Обзор недавних диссертационных исследований показывает, что реформирование университетов и других высших учебных заведений осуществляется за счет автоматизации и охватывает все большее число видов их деятельности – учебный процесс, научно-исследовательскую работу преподавателей и студентов, информационное, программное, техническое, экономическое и даже хозяйственное обеспечение. При этом сохраняются основные принципы и многие методы, разработанные при создании Государственной системы НТИ в ее центральных и отраслевых организациях с учетом многократно возросших возможностей информационно-коммуникационной технологии.

В отличие от статей в научных журналах, где обычно рассматриваются частные вопросы реформи-

рования и автоматизации учебных заведений, в диссертациях эти проблемы исследуются комплексно, а результаты этих исследований редко обобщаются в реферативных и других вторичных изданиях.

СПИСОК ЛИТЕРАТУРЫ

1. Биктимиров М.Р., Сюнтюренко О.В., Гиляревский Р.С. Новая концептуальная основа развития информационной деятельности ВИНТИ РАН // Научно-техническая информация. Сер.1. – 2016. – № 1. – С. 3-7; Biktimirov M.R., Gilyarevskii R.S., Syuntyurenko O.V. A New Conceptual Basis for the Development of the Information Activities of the All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences // Scientific and Technical Information Processing. – 2016. – Vol. 43, № 1. – P. 1-7.
2. Гудов А.М. Программно-технологический комплекс для развития информационной среды образовательного учреждения на основе системы электронного документооборота: автореф. дисс. ... докт. техн. наук. – Новосибирск: Сиб. гос. ун-т телекомму. и информат., 2013 – 36 с.
3. Нестеров М.И. Управление развитием информационно-аналитической системы современного вуза на основе архитектурного подхода: автореф. дисс. ... канд. техн. наук. – Челябинск: Южно-Уральский гос. ун-т, 2013. – 23 с.
4. Бубарева О.А. Модель, алгоритмы и программное обеспечение интеграции данных информационных систем на основе онтологий (на примере вуза): автореф. дисс.... канд. техн. наук. – Бийск: Алтайский гос. техн. ун-т им. И.И. Ползунова, 2014. – 16 с.
5. Акимов А.А. Информационно-аналитическая система для поддержки процессов управления кафедрой ВУЗа: автореф. дисс.... канд. техн. наук. – Пенза, Пензенский гос. ун-т – 2012. – 17 с.
6. Манучарян Л.А. Разработка средств повышения релевантности ИПС на базе онтологий: автореф. дисс.... канд. техн. наук / Воронежская гос. лесотехн. акад. – Тамбов, 2012. – 16 с.
7. Бондарев Я.П. Разработка моделей и алгоритмов оценки и управления положением в рейтинге субъектов и объектов высшего образования: автореф. дисс. канд. техн. наук. – Воронеж: Воронежский ин-т высоких технологий, 2013. – 16 с.
8. Бабичева Н.Б. Методы и модели рейтингового оценивания абитуриентов вуза: автореф.... дисс. канд. техн. наук. – Новокузнецк: Сиб. гос. индустриальный ун-т, 2013. – 19 с.
9. Лазарева О.Ю. Разработка веб-ориентированной интеллектуальной системы проверки знаний на основе концептуальной модели предметной области: автореф.... дисс. канд. техн. наук. – М.: Моск. гос. ун-т печати им. И. Федорова, 2015 – 24 с.

10. Назойкин Е.А. Мультиагентное имитационное моделирование образовательного процесса накопления знаний: автореф.... дисс. канд. техн. наук – М.: Моск. гос. ун-т прикладной биотехнологии, 2011 – 23 с.
11. Махныткина О.В. Моделирование и оптимизация индивидуальной траектории обучения студента: автореф.... дисс. канд. техн. наук. – Новосибирск: Сиб. гос. ун-т телекоммун. и информат., 2013 – 23 с.
12. Измайлова Е.В. Поддержка принятия решений при управлении научно-исследовательской работой студентов вузов: автореф.... дисс. канд. техн. наук. – М., 2012 – 19 с.
13. Артамонов А.А. Трёхуровневая система агентного поиска и обработки научно-технической информации: автореф.... дисс. канд. техн. наук. – М.: Национальный исследовательский ун-т «МИФИ», 2017. – 21с.

Материал поступил в редакцию 07.04.17

Сведения об авторе

ТЕРЕЩЕНКО Сергей Сергеевич – доктор технических наук, ведущий научный сотрудник ВИНТИ РАН, Москва.
e-mail: nti@viniti.ru

Мультилингвистическая среда информационно-образовательного взаимодействия

Представлена структура процесса обучения при индивидуальной языковой подготовке специалистов. На основе концепции сетевой образовательной среды рассмотрена реализация базовых компонентов мультилингвистической среды информационно-образовательного взаимодействия. Основу построения информационно-терминологического базиса такой среды составляют частотные словари, по которым определяется приоритетность изучения тех или иных терминов. Показано, что использование таких словарей, полученных при анализе языкового материала, качественно улучшает процесс обучения иностранной лексике.

Ключевые слова: адаптивное обучение, образовательная среда, языковая подготовка, мультилингвистическая технология

ВВЕДЕНИЕ

Внедрение автоматизированных систем обучения, активное применение новых информационных технологий в образовательных учреждениях открывают широкие возможности для интенсификации и индивидуализации учебного процесса [1]. В последнее время в связи с развитием компьютерных технологий широкое распространение получили новые формы образования – системы индивидуального обучения. Применение компьютера позволяет сделать процесс получения знаний интересным, увлекательным, высокооперативным. Повышение качества обучения достигается за счет индивидуальной работы учащегося с учебным материалом.

Систему обучения представляют следующие взаимодействующие элементы в:

- программы обучения, разработанные с учетом реальных потребностей в повышении уровня знаний пользователей;
- подразделения, занимающиеся выявлением потребностей в обучении, управлением и оценкой эффективности и качества обучения;
- сеть учреждений и подразделений, непосредственно занимающихся обучением, независимо от их организационно-правовых форм, типов и видов, реализующих принятые программы обучения.

Использование автоматизированных обучающих систем непосредственно для обучения связано с выполнением компьютером следующих взаимосвязан-

ных функций: управление учебной деятельностью; хранение и выдача учебной информации; моделирование лабораторных экспериментов, явлений, ситуаций, закономерностей и т.д.; анализ сообщений и ответов обучаемых; регистрация, хранение и обработка результатов учебной деятельности.

Мультилингвистическая адаптивно-обучающая технология [2] применяется для создания систем обучения, способствующих интенсивному накоплению иностранного профессионально-ориентированного словарного запаса у специалистов и студентов, сталкивающихся с иностранной лексикой в процессе работы с иностранной литературой или сотрудничества с иностранными коллегами.

В центре современной мультилингвистической адаптивно-обучающей технологии лежат два взаимосвязанных компонента – информационно-терминологический базис (ИТБ) и технология обучения [3].

МУЛЬТИЛИНГВИСТИЧЕСКАЯ СРЕДА ИНФОРМАЦИОННО-ОБРАЗОВАТЕЛЬНОГО ВЗАИМОДЕЙСТВИЯ

Анализ, представленный в [4], дает основание сделать вывод, что мультилингвистическая адаптивно-обучающая технология соответствует (в зависимости от реальной реализации в виде компьютерной системы) ряду возможных процессов обучения (ПО). Например, структура, отражающая ситуацию одновременного обучения группы учащихся с использованием одного мультилингвистического информаци-

онно-терминологического базиса, представлена на рис. 1. Частная реализация такой структуры – система одномерной языковой подготовки (англо-русский, англо-немецкий информационно-терминологический базис и т.д.).

Отметим, что эта структура (см. рис. 1) приводит к повышению производительности обучения только при тщательном подборе группы обучаемых. Имея в реальной системе разных учеников и определяя цель как обязательное обучение всех, мы будем достигать лишь производительности самого «медленно обучающегося» ученика [5, 6].

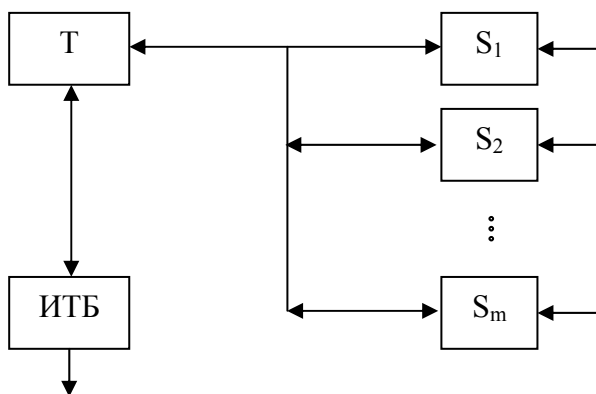


Рис. 1. Одновременное обучение группы учащихся с использованием одного информационно-терминологического базиса (ИТБ): Т – преподаватель (в общем случае, виртуальный); $S_1, S_2, \dots, S_m$ – обучаемые.

Одновременному обучению группы с использованием многих (набора разных) информационно-терминологических базисов, т.е. каждый обучаемый самостоятельно (или по рекомендации преподавателя) выбирает свой базис ($ИТБ_1, ИТБ_2, \dots, ИТБ_k$), соответствует структурная схема, представленная на рис. 2. К этой структуре добавляются множественные связи компонентов, и она является более гибкой. Сокращается не только время обучения, но и повышается его

эффективность, так как присутствуют элементы индивидуального подхода к ученику и имеется возможность адаптации информационно-терминологического базиса к требованиям обучаемого.

Если в предыдущей структуре (см. рис. 1) установить множественные связи преподавателя с учениками (в этом случае нагрузка на преподавателя еще больше возрастет), то мы получим процесс обучения, который в [5, 7] назван индивидуальным обучением в группе, когда преподаватель контролирует обучение, т. е. решает, какую следующую порцию знаний и умений ученику надо получить. Преимущество данной схемы может быть реализовано только в случае, если такой компонент как преподаватель будет дублироваться. Это повысит пропускную способность этого компонента и позволит использовать достоинства как индивидуального, так и группового обучения. При этом необязательно предполагать, что преподаватели (обучающие алгоритмы) – одинаковы. Имеется возможность построить адаптивные обучающие алгоритмы, настраиваемые в каждом конкретном случае (на каждом сеансе обучения) на конкретного обучаемого. Этот подход соответствует структуре с множеством преподавателей, реализующих различные алгоритмы обучения или адаптирующихся к обучаемому, позволяя, тем самым, сократить время обучения слабого ученика и повысить эффективность обучения сильного.

Назовем множество информационных связей структурной схемы процесса обучения мультилингвистической средой информационно-образовательного взаимодействия. Следует отметить, что, в отличие от традиционного процесса обучения, который при наличии множества предметных дисциплин становится многомерным, процесс обучения, реализуемый с помощью мультилингвистической адаптивно-обучающей технологии, остается одномерным процессом при наличии многомерного информационно-терминологического базиса, который может быть реализован как одномерный, если установлено взаимно однозначное соответствие между разноязычными компонентами $ИТБ_1, ИТБ_2, \dots, ИТБ_k$.

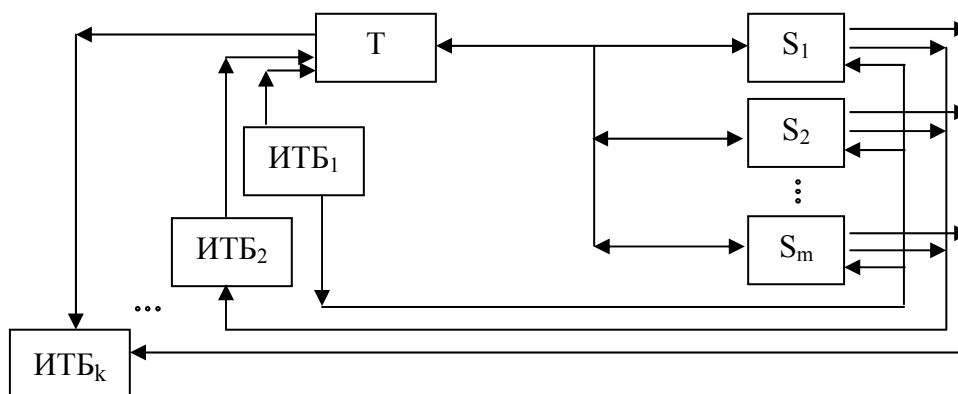


Рис. 2. Одновременное обучение группы с использованием многих информационно-терминологических базисов (ИТБ): Т – преподаватель; $S_1, S_2, \dots, S_m$ – обучаемые.

Множество элементов структуры процесса обучения, информационно связанных друг с другом посредством среды образовательного взаимодействия [4], как правило, образуют информационно-образовательную сеть (в нашем случае – мультилингвистическую).

Очевидно, что реализация мультилингвистической информационно-образовательной сети в традиционном обучении – практически невыполнимая задача. Автоматизация процесса обучения для данной структуры требует, в первую очередь, автоматизации формирования информационно-терминологического базиса (по запросам обучаемых, вплоть до динамического его формирования), а также компьютерной поддержки технологии адаптивного обучения в интерактивном режиме.

РЕЛЯЦИОННАЯ МОДЕЛЬ ИНФОРМАЦИОННО-ТЕРМИНОЛОГИЧЕСКОГО БАЗИСА

Формирование информационно-терминологического базиса мультилингвистической адаптивно-обучающей технологии (МЛ-технологии) позволяет использовать для его информационно-логической организации современные структурные методологии и базовые конструкции диаграмм Варнье–Орра [4].

Информационно-терминологический базис мультилингвистической технологии строится на основе результатов анализа языкового материала. Под языковым материалом здесь следует понимать некоторое множество текстов интересующей разработчика предметной области изучаемого языка. Объем языко-

вого материала может варьироваться в зависимости от средств анализа, наличия оригинальных текстов и необходимого количества терминов.

Основу построения информационно-терминологического базиса составляют частотные словари, по которым определяется приоритетность изучения тех или иных терминов. Использование таких словарей, полученных при анализе языкового материала, качественно улучшает процесс обучения иностранной лексике [2, 5].

Согласно разработанной нами реляционной модели информационно-терминологического базиса, все его элементы по своим частотным характеристикам разделяются на две группы: лексемы, имеющие наибольшую частоту (основные лексемы), и лексемы, частота которых не превышает некоторого числового порога. Из них впоследствии формируются реляционные ряды, каждый из которых однозначно соответствует одной из основных лексем. При этом структура мультилингвистического компонента (МЛ-компонента), отражающего основную лексему, дополняется соответствующими переходными вероятностями.

Построение информационно-терминологического базиса согласно реляционной модели (рис. 3) позволяет применять реляционную методику обучения, делающую акцент не только на терминологию, но и на ассоциативные зависимости между терминами. Таким образом решается задача наведения ассоциативных полей внутри изучаемого языка, и благодаря этому увеличивается эффективность процесса обучения в целом.

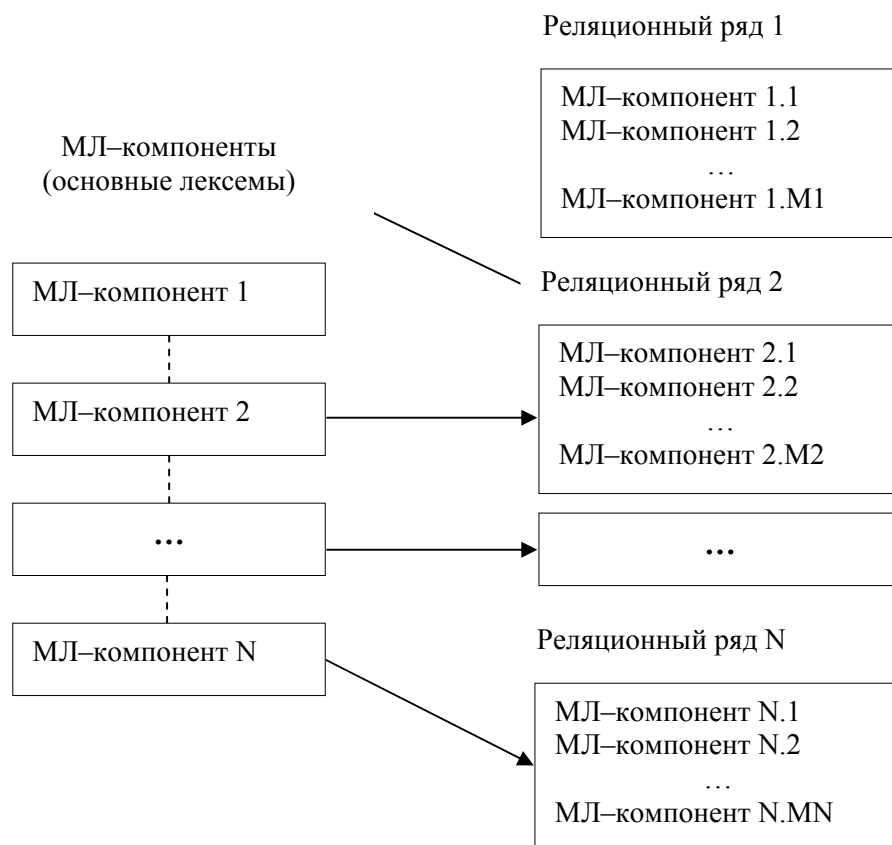


Рис.3. Реляционная модель структуры информационно-терминологического базиса:
МЛ – мультилингвистическая адаптивно-обучающая технология.

Применение реляционной модели хорошо сочетается с другими методами оптимизации структуры информационно-терминологического базиса, в частности, с методами оптимального разбиения на блоки и модули.

Человеческую память можно представить в виде огромного количества ячеек. Множество ячеек (K_N) для слов родного языка можно обозначить как:

$$K_N = \{k_1, \dots, k_N\}, \quad (1)$$

где каждая из ячеек k содержит слово родного языка.

При изучении первого иностранного языка (например, английского) в результате пополнения словарного запаса в памяти генерируется новое множество ячеек для запомненных слов английского языка

$$K_E = \{k_1, \dots, k_E\}. \quad (2)$$

В дальнейшем у обучаемого при актуализации i -го слова на одном из языков возникают ассоциации со значением этого слова на другом языке:

$$A_i^{NE}, i = 1, \dots, E. \quad (3)$$

Ассоциация – это связь между психическими элементами, возникшая в результате опыта и обуславливающая при актуализации одного элемента связи проявление и другого [2]. Степень связи этих элементов может быть различной в зависимости от скорости восприятия информации и забывания, времени забывания слова после его последнего заучивания и индивидуальных особенностей памяти обучаемого.

При изучении второго иностранного языка (например, немецкого) через определенный промежуток времени в памяти возникают все новые ячейки для запомненных немецких слов, множество которых обозначим:

$$K_G = \{k_1, \dots, k_G\}. \quad (4)$$

Множество K_G частично перекрывает множество запомненных английских слов K_E , так как число элементов множества K_E в зависимости от времени после последнего заучивания уменьшается. Выражая объем памяти как функцию времени, можно проследить зависимость – даже отлично выученные иностранные слова со временем забываются. Новые заученные немецкие слова занимают бывшие «английские» ячейки, которые к тому моменту становятся пустыми. При этом необходимо отметить, что ячейки, зарезервированные для слов родного языка, остаются неизменными. Механизмом, лежащим в основе этой закономерности, выступает перевод информации из кратковременной памяти в долговременную.

У обучаемого возникают также новые ассоциации (A) между словами родного и изучаемого немецкого языка:

$$A_i^{NG}, i = 1, \dots, E. \quad (5)$$

Степень связи этих языков может варьироваться вследствие различных скоростей восприятия и забывания слов, времени их забывания после последнего заучивания вплоть до перекрывания ассоциаций с английскими значениями слов.

Например, при попытке вспомнить английское, либо немецкое значение слова, осуществляется поиск двух разных ячеек из обоих множеств K_G и K_E . Поиск ячеек довольно сложен вследствие того, что ассоциации со словами немецкого языка, который изучался позднее, гораздо сильнее, и могут затмить ассоциации с английскими словами. Происходит своеобразный конфликт ассоциаций, и «победа» одной из них (не обязательно требуемой в данный момент) на этом уровне зависит от индивидуальных особенностей памяти человека.

Мультилингвистическая адаптивно-обучающая технология [2, 6] предполагает изучение терминологической лексики иностранного языка с учетом связи элементов этой лексики с элементами терминологии ранее изучавшихся языков. Генерация ассоциативного поля вокруг запоминаемых терминов путем использования ассоциаций между элементами различных иностранных терминологий на практике достигается предоставлением терминов-подсказок на выбранном пользователем языке и дает возможность интенсивно пополнять словарный запас. При этом значения термина на всех изучаемых иностранных языках хранятся в одной и той же ячейке памяти обучаемого, причем элемент терминологии изучаемого языка фиксируется в ячейке с элементом ранее изучавшегося языка. Так как этот процесс возможен в двух или более направлениях (при количестве иностранных языков $N \geq 2$), то происходит заучивание терминов на всех изучаемых языках. Возникают множественные ассоциации в зависимости от числа иностранных языков. Например, при изучении немецкой лексики при поддержке терминами английской и русской лексики возникают троичные ассоциации. Соответственно, процесс обращения к множеству ячеек происходит один раз, и ассоциации не конфликтуют.

Решение задачи формирования информационно-терминологического базиса МЛ-технологии, согласно новым принципам построения его структуры, было реализовано в программной системе TuMLasv.1.0 [6].

АДАПТИВНО-ОБУЧАЮЩИЙ ПОДХОД К ОСВОЕНИЮ ИНОСТРАННОЙ ТЕРМИНОЛОГИЧЕСКОЙ ЛЕКСИКИ НА ОСНОВЕ МУЛЬТИЛИНГВИСТИЧЕСКОГО ЧАСТОТНОГО СЛОВАРЯ

Как было упомянуто, главное требование к современным компьютерным обучающим системам – обеспечивать максимальную степень индивидуализации процесса обучения, т.е. адаптировать его к каждому конкретному ученику.

Адаптация в обучающей системе – это процесс изменения параметров и структуры модели объекта (обучаемого) и обучающих воздействий на основе текущей информации, получаемой в ходе

обучения для достижения оптимального состояния объекта при его начальной неопределенности и изменяющейся среде.

Ядром мультилингвистической адаптивно-обучающей технологии является алгоритм обучения, работающий на основе адаптивной модели обучения Л.А. Расстригина, учитывающей индивидуальные факторы процессов запоминания и забывания каждого конкретного обучаемого для мультилингвистического представления информации [8].

Отношения между учеником и учителем рассматриваются как отношения между объектом управления и управляющим устройством, что позволит использовать методы теории управления. Так как построение точной модели сложного объекта практически невозможно, необходимо построить приближенную модель и адаптировать ее параметры для обеспечения адекватности реальному объекту (обучаемому). Именно этот подход, реализуемый для адаптивно-обучающей технологии (МЛ-технологии), мы и рассматриваем. Основным прикладным аспектом технологии является решение практической задачи изучения терминологической лексики иностранных языков и интенсивного накопления профессионально-ориентированного терминологического базиса обучаемого, так как именно с этими проблемами наиболее часто сталкиваются специалисты, читающие литературу и слушающие лекции на иностранных языках.

Применение мультилингвистического подхода позволяет генерировать ассоциативное терминологическое поле вокруг запоминаемых терминов и модифицировать адаптивную модель обучаемого с учетом ассоциативных параметров при коррекции скорости забывания и вероятностей незнания элементов обучающей информации.

Информационно-терминологический базис мультилингвистической адаптивно-обучающей технологии, как было упомянуто, представлен в виде электронных частотных словарей, последующее разбиение которых по частотам используется при обучении [5, 6].

Создание двуязычного частотного словаря связано с определенными трудностями – качественного и количественного характера: нелегко решить, какие именно тексты следует считать достаточно хорошо представляющими литературный язык. Не менее сложным является и вопрос о достаточном для такого словаря общем объеме текстов. Словари, базирующиеся на выборках меньшего объема, недостаточно представительны, чтобы по ним можно было судить о лексике языка в целом.

Частотный словарь представляет ту выделившуюся за последнее время область статистической лексикографии, которую можно назвать семантической. Статистическая лексикография занимается вопросами классификации, составления и применения частотных словарей. Частотные словари отличаются от словарей всех других типов тем, что они, во-первых, указывают при входных единицах частоты употребления этих единиц в конкретном массиве текстов и, во-вторых, включает только те единицы, которые встретились в этом массиве.

Частотные словари различаются по ряду признаков, важнейшими из которых являются язык, содер-

жание и объем текстов, послуживших материалом для словаря определенного естественного языка, его слова и словоформы.

Мы провели статистическое исследование и описали тексты по информатике и системному анализу объемом около 30000 слов, в результате чего был составлен список объемом около 2000 слов, адекватность которых была установлена для всех трех языков.

Первый англо-немецко-русский словарь по информатике и системному анализу, изданный авторами в 2004 г., состоял из алфавитного англо-немецко-русского словаря, включающего словарные статьи. Каждая статья со всеми относящимися к ней материалами составляла отдельную единицу словаря и была сформирована по гнездовому принципу. Глубина гнезда определялась показателем распространенности слова в исследуемых подязыках.

В этом словаре нашли отражение некоторые важные при обучении иностранному языку и его использовании качественные и количественные аспекты употребительной лексики по системному анализу и информатике на английском, немецком и русском языках, полученные в результате статистического анализа и описания текстов, которые будут способствовать организации рационального усвоения лексики и накоплению словарного запаса.

Далее был разработан англо-немецко-русский частотный словарь по системному анализу в электронном машиностроении и авиакосмической отрасли [9], дополненный многоязычной терминологией в количестве 300 специальных терминов и адаптированный для обучения посредством методики построения внутриязыковых ассоциативных полей – лексически связанных компонентов (ЛСК-методики [6]). Для примера представим блок лексически связанных компонентов словаря, его структура изменилась и теперь он состоит из двух равноправных блоков: блока лексически связанных компонентов и блока справочно-учебного назначения.

Первый блок предназначен для обучения согласно ЛСК-методике и содержит лексически связанные компоненты, представленные в виде двухуровневых структур данных, где на верхнем уровне находятся основные лексемы, выделенные полужирным шрифтом, на нижнем – связанные лексемы и сочетания.

Последовательность подачи материала в ЛСК-методике такова:

- 1) основная лексема → перевод, подсказка на иностранном языке;
- 2) связанная лексема → перевод, подсказка на иностранном языке;
- 3) лексическое сочетание основной и связанной лексем → перевод сочетания, но не лексем по отдельности);
- 4) переход к следующей связанной лексеме;
- 5) переход к следующему лексически связанному компоненту.

Особенность такого обучения заключается в формировании ассоциативных связей внутри изучаемой терминологии, причем большинство связей приходится на основные лексемы, специально отобранные как наиболее значимые для изучения.

БЛОК ЛЕКСИЧЕСКИ СВЯЗАННЫХ КОМПОНЕНТОВ

1 accuracy,47	Genauigkeit, 35 f (Femininum)	Точность,139
2 adequate,31	Adäquate, 16	Адекватный,32
3 adequate accuracy,3	geforderte Genauigkeit,1 f	Требуемая точность, 59
4 available,16	Verfügbar,2	Доступный, 16
5 available accuracy,2	VerfügbareGenauigkeit, 2 f	Доступная точность,16
6 degree,23	Grad, 19 m (Maskulinum)	Степень, 67, Порядок,11
7 degreeof accuracy,7	Genauigkeitsgrad, 9 m	Степень точности, 21
8 extreme,10	Extrem, 5	Крайний,10
9 extreme accuracy,6	Extremgenauigkeit, 2 f	Предельная точность,10
10 order, 27	Ordnung, 28 m	Степень,17, Порядок,6
11 orderofaccuracy, 16	Genauigkeitsordnung, 9 f	Степень точности,21
12 relative , 20	Relativ, 20	Относительный,20
13 Relative accuracy,4	relativeGenauigkeit, 4 f	Относительная точность,9
14 action,164	Handlung, 90 f	Действие,172
15 elementary, 27	elementar, 30	Элементарный,27
16 elementaryaction, 14	Elementare Handlung,31 f	Элементарное действие,38
17 integral, 32	Integral, 22 n (Neutrum)	Интеграл,23
18 integralaction, 26	Integralhandlung, 23 f	Интегральное действие,17
19 out, 13	Ausgang, 13 m	Выход,28
20 outofaction, 11	AusserBetrieb, 8 m	Неработающий,14

БЛОК СПРАВОЧНО-УЧЕБНОГО НАЗНАЧЕНИЯ

560 a prioridata, 3	Apriori - Daten, 3 pl	Априорные данные, 3
561 abbreviate, 8	abkürzen, 3	Сокращать, 11
562 abbreviate a notation, 3	Bezeichnungabkürzen, 2 f	Сокращать обозначение, 3
563 ability, 4	Fähigkeit, 4 f	Способность, 8
564 above? 32	über, 100	Выше,8, свыше, 19
548 absolute, 131	absolut, 53	Абсолютный, 17
565 absoluteaccuracy, 9	absoluteGenauigkeit, 4 f	Абсолютная точность,48
566 absoluteaddress, 2	absoluteAdresse, 12 f	Истинный адрес, 16
567 absolutelay, 4	absoluteVerzögerung, 7 f	Абсолютная задержка, 12
568 absolute error,18	absoluterFehler, 10 m	Абсолютная ошибка, 17
569 absoluteinequality, 4	absoluteUngleichung, 2 f	Безусловное неравенство, 11, абсолютное неравенство, 3
570 absolutemaximum, 24	absolutesMaximum, 9 n	абсолютный максимум, 17

Второй блок содержит всю терминологию, включая лексемы, вошедшие в блок лексически связанных компонентов, которые выделены полужирным курсивом и формально представляют собой ссылки на соответствующие элементы первого блока, и предназначен для использования в качестве справочного материала при переводе иностранных текстов, а также при обучении.

Словарь организован как англо-немецко-русский, однако он может быть использован и как двуязычный, так как в нем имеются указатели немецких и русских слов, составленные в алфавитном порядке.

УКАЗАТЕЛЬ НЕМЕЦКИХ ТЕРМИНОВ

- 1487 abbilden, 2
- 479 Abbildung, 6 f**
- 1358 Abfrageastatur, 3 f
- 539 Abfragezeit, 2 f**
- 862 abgekürzt, 3
- 863 abgekürzteBezeichnung, 3 f
- 975 abhängiger, 5
- 976 abhängigeUnterbrechung, 2 f
- 977abhängige Variable, 2 f
- 974 Abhängigkeit, 6 f

562 abkürzen, 3
505 Ablassen, 14 n
1176 Ablaufdiagramm, 2 n
1174 Ablaufplan, 2 n
422 Ablesen, 6 n

Нумерация позволяет легко найти необходимое слово в блоке лексически связанных компонентов (первом блоке) и в терминологическом блоке (втором блоке). Слова, выделенные полужирным курсивом, являются ссылками на соответствующие элементы первого блока.

ЗАКЛЮЧЕНИЕ

Основываясь на концепции сетевой образовательной структуры и опираясь на частные формализованные представления учебного процесса при языковой подготовке, нами выполнен анализ возможных структур мультилингвистической среды информационно-образовательного взаимодействия. Это позволило сделать вывод о том, что мультилингвистическая адаптивно-обучающая технология соответствует (в зависимости от реальной реализации в виде компьютерной системы) ряду возможных процессов обучения. При этом рассмотрена проблема учета ассоциативных характеристик лексем при формировании информационно-терминологического обеспечения, которое позволяет генерировать ассоциативное поле вокруг запоминаемых терминов с использованием ассоциаций между элементами различных иностранных терминологий, что дает возможность интенсивно пополнять словарный запас на практике. Программная система TuMLasv.1.0 обеспечивает информационную поддержку мультилингвистической технологии.

Разработанный и дополненный многоязычной терминологией англо-немецко-русский частотный словарь по системному анализу в электронном машиностроении и авиакосмической отрасли адаптирован для обучения посредством методики построения внутриязыковых ассоциативных полей – лексически связанных компонентов. Это обеспечено структурным разделением словаря на два равноправных блока: блок лексически связанных компонентов и блок справочно-учебного назначения.

Электронный словарь, являясь информационно-терминологическим обеспечением мультилингвистической адаптивно-обучающей технологии, позволяет автоматизировать процесс изучения терминологической лексики иностранного языка с учетом связи ее элементов с элементами терминологии ранее изучавшихся языков и генерации ассоциативного поля вокруг запоминаемых терминов.

* * *

Авторы выражают признательность профессору Института германистики Рурского университета в г. Бохуме (RUB) доктору Х.-Р. Флуку и его коллегам за содействие и помощь в подготовке и коррекции немецко-русской части словаря. Это сотрудничество стало возможно при финансовой поддержке Германской службы академических обменов (DAAD) и Научно-исследовательского фонда аэропорта во Франкфурте-на-Майне

(Flughafen-Frankfurt-Main-Stiftung), а также в рамках совместной программы «Михаил Ломоносов» Министерства образования и науки РФ и DAAD.

СПИСОК ЛИТЕРАТУРЫ

1. Монахов М.Ю. Информационная образовательная сеть // Информационные технологии. – 2001. – № 7. – С. 36-47.
2. Ковалев И.В., Карасева М.В., Лесков В.О. Компоненты информационной поддержки мультилингвистической адаптивно-обучающей технологии // Системы управления и информационные технологии. – 2009. – Т. 35, № 1.3. – С. 360-363.
3. Ковалев И.В., Карасева М.В., Суздальева Е.А. Системные аспекты организации и применения мультилингвистической адаптивно-обучающей технологии // Образовательные технологии и общество. – 2002. – Т. 5, № 2. – С. 198-212.
4. Kovalev I.V., Kovaleva T.A., Karasyova M.V., Ezhemanskay S.N. System aspects of multilingual adaptive-training technology organization and usage // Proceedings of Modeling and Simulation, MS'2004 AMSE International Conference on Modelling and Simulation, MS'2004. – Lyon-Villeurbanne, 2004.
5. Ковалев И.В., Зеленков П.В., Яркова С.А., Шевчук С.Ф. Оптимизация обработки данных в распределенных образовательных средах // Программные продукты и системы. – 2007. – № 3. – С. 28.
6. Ковалев И.В., Карасева М.В., Лесков В.О. Алгоритмизация процедур включения связанных лексем в структуру информационно-терминологического базиса // Программные продукты и системы. – 2009. – № 4. – С. 28.
7. Ковалев И.В., Карасева М.В., Лесков В.О. Информационно терминологический базис как совокупность лексически связанных компонентов // Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева. – 2009. – № 1-1 (22). – С. 54-56.
8. Растринин Л.А., Эренштейн М.Х. Адаптивное обучение с моделью обучаемого. – Рига: Зинатне, 1988. – 160 с.
9. Ковалев И.В., Карасева М.В. Англо-немецко-русский частотный словарь по системному анализу в электронном машиностроении и авиакосмической отрасли. – Красноярск: Сиб. гос. аэрокосмич. ун-т., 2013. – 216 с.

Материал поступил в редакцию 06.04.17.

Сведения об авторах

КОВАЛЕВ Игорь Владимирович – доктор технических наук, профессор, ректор Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева, г. Красноярск
e-mail: rector@sibsau.ru

КОВАЛЕВ Дмитрий Игоревич – аспирант Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева, г. Красноярск
e-mail: kleniks@yandex.ru

КАРАСЕВА Маргарита Владимировна – кандидат технических наук, доцент, доцент кафедры «Системного анализа и исследования операций» Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева, г. Красноярск
e-mail: karaseva-margarita@rambler.ru

ПЕРШАКОВА Кристина Константиновна – младший научный сотрудник научно-исследовательского управления Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева, г. Красноярск
e-mail: kleniks@yandex.ru

ТУЕВА Евгения Вячеславовна – аспирант Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева, г. Красноярск
e-mail: kleniks@yandex.ru

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 001.4 : 81'373.4

Ю.В. Сложеникина, В.С. Звягинцев

Термины-эпонимы: *PRO ET CONTRA*

Показаны достоинства терминов-эпонимов: связь с культурно-историческими коннотациями, мемориальная, деонтологическая функции, функции индивидуализации и глобализации. Недостатки терминов-эпонимов: существенные признаки явления не фиксируются в форме терминологического знака, частотность омонимии, разное содержание терминов с предельно близкой формой, наличие орфографических / орфоэпических вариантов, необходимость введения в структуру термина дифференцирующих компонентов, неблагозвучность, несоответствие психологическим особенностям речевой деятельности. Терминоведы настоятельно рекомендуют по возможности сокращать количество отыменных наименований, заменяя их дублетными классификационными терминами.

Ключевые слова: термин, терминология, эпоним, нормализация, кодификация терминологии

С точки зрения мотивированности значения терминов выделяют два основных типа специальных единиц. В форме мотивированных терминов отражены необходимые и достаточные признаки обозначаемого понятия, что позволяет понимать значение термина, не прибегая к его дефиниции (*держатель, переходник, соединительное устройство, адаптер* – в технической терминологии). Немотивированными называются термины, в звуковых комплексах которых не содержится указание на существенные признаки понятия, другими словами, это единицы, буквальное значение которых не распознается. Однако между двумя противоположными разновидностями единиц в профессионально ориентированной лексике находится тип частично мотивированных терминов. Их основную часть составляют так называемые эпонимы (от греч. *εponυμος* из *ερι* – «после» + *ονομα* – «имя» – дающий свое имя) – лицо, от имени которого произведено название. Эпонимом называется также языковая единица, созданная по модели: *закон Бартоломе, закон Бругмана, правило Гавлика, закон Ваккгернагеля, гипотеза Сепира-Уорфа, Женевская школа* (в лингвистике).

Мы относим эпонимы к категории частично мотивированных терминов по двум причинам: название содержит указание на второстепенный признак понятия, эпонимические термины могут быть ложноориентированными.

Считать подобные термины лишь частично мотивированными позволяют два обстоятельства. Суть первого в том, что название содержит указание на

второстепенный, неклассификационный признак терминируемого объекта (лицо, имеющее отношение к его созданию, изучению). Эпонимы, как правило, появляются не в результате научного обобщения, а на основе эмпирической деятельности. Кроме того, эпонимические термины иногда могут быть ложноориентированными. Например, классическое описание клинической картины *капилляротоксикоза* составили Уильямс (1808), Бейтмен (1819) и Латур (1828), однако болезнь называется *синдромом Шендлейна-Геноха*, хотя Шендлейн описал её только в 1832, а Генох – в 1874 г.

Эпонимические термины связаны с культурно-историческими коннотациями [1, с. 469-471]. С прагматических позиций, эпонимы выполняют мемориальную функцию, хотя по большей части она реализуется в ограниченном кругу специалистов в тех или иных областях науки и техники. Например, доктор медицинских наук, член Российского научного общества истории медицины В. Остроглазов пишет, что психиатрия XXI в. столкнулась с явлением «потерявших себя молодых людей», обращающихся в полицию с вопросом: «Скажите мне, кто я?». В психиатрическом Центре им. Сербского эту проблему определили «биографической амнезией». В. Остроглазов считает такой подход «наивным» и отмечает, что «в обширном семействе амнезий нет адекватного термина». Поэтому в 2003 г. он предложил латинский неологизм, более точно отражающий загадочный феномен – *autopersonamnesia* – забвение собственной личности. Однако спустя 10 лет ученый-медик на-

стаивает на создании эпонима для аутоперсонамнезии – синдром Эминеску. По мнению В. Остроглазова, румынско-молдавско-украинский поэт Эминеску – поэтический гений XIX в., «в жизни которого мотив абсолютного отрицания и желание полного забвения собственной личности пронизывает все его творчество и самую жизнь, оканчивающуюся угасанием самосознания и жизни в психиатрической клинике. ...архетипический мотив полного небытия, аннигиляции, нирваны, забвения себя и забвения собственной личности в человеческом мире никем не был выражен с такой поэтической силой». Имя поэта как бы «очеловечивает» сухой научный язык психопатологии, а в широком смысле вводит его в социально-культурное пространство языка в целом [2, с. 91-96].

Очевидно, что еще большую историко-научную и социально-культурную ценность приобретают эпонимы-интернационализмы, выполняющие в привычном употреблении функцию, с одной стороны индивидуализации, а с другой – глобализации. Например, спинозизм (англ. *spinozism*, нем. *Spinozismus*) от фамилии голландского философа Бенедикта Спинозы (1632-1677); бенедиктинец (англ. *benedictine*, нем. *Benediktiner*) от имени Св. Бенедикта Нурсийского; кириллица (англ. *cyrillic*, нем. *Kyrilliza*) от имени славянского просветителя Кирилла (827-869); пасквиль (англ. *pasquil*, нем. *Pasquill*) от имени римского башмачника XV в. Пасквино; филиппика (англ. *philippic*, нем. *Philippika*) от имени македонского царя Филиппа II (около 382-336 до н.э.); эпикурейство (англ. *epicureanism*, нем. *Epikureismus*) от имени древнегреческого философа Эпикура (342/341 до н.э.-271/270 до н.э.) и др. [3, с. 185-186].

В некоторых профессиональных сферах наличие терминов-эпонимов может объясняться деонтологическими причинами, например, введение в специальную номинацию медицины антропонима позволяет избежать травмирующих ситуаций при постановке клинического диагноза: вместо *тяжелый дерматоз неясной этиологии с острым течением и летальным исходом – синдром Лайелла*, вместо *форма старческого слабоумия – синдром Вернике*.

Эпонимические термины представлены практически во всех терминологиях, вследствие чего они достаточно часто становились предметом научного исследования: в языке авиации [4], в терминосистеме теплоэнергетики [5], в экономической терминологии [6], в медицине [7], в инженерной психологии [8], в орнитологии [9], в специальном языке альпинизма [10], в издательском деле [11], в дискурсе моды [12], в кулинарии [13] и др. Это вызывает необходимость рассмотреть их достоинства и недостатки в специальных семантических полях в аспекте лексической вариантности.

Известно, что мотивированность термина часто бывает неполной, потому что любой объект, тем более научный, имеет множество разных свойств. И именуемый волен положить в основу номинации существенный, с его точки зрения, признак – потенциально каждый научный феномен может иметь несколько синонимичных наименований (специальный предмет один, а именуемых понятий, стоящих за данным предметом, может быть несколько). Эпони-

мические термины возникают прежде всего при описании сложных, неоднозначных явлений в тех областях, в которых практическая деятельность часто опережает научные исследования и классификации, и поэтому существенные признаки явления еще не фиксируются в форме терминологического знака. Возникает номинативный «кризис», выходом из которого может стать создание эпонимической единицы.

Например, современное состояние медицинской науки не всегда позволяет точно выявить этиологию заболевания и отнести определенную патологию к конкретной нозологической форме. В таких случаях может быть поставлен диагноз, в наименование которого входит частично мотивированный эпонимический термин, включающий в свой состав имя исследователя, описавшего данное заболевание. Так, *синдромом Гудпасчера* названо редкое аллергическое заболевание, при котором у больного по неизвестным причинам вырабатываются антитела против тканей почек и легких. Возможно, эти тела и являются непосредственной причиной болезни, однако пока данный факт научно не доказан, этот вид болезни будет именоваться эпонимическим термином.

Все же подчеркнем, что основным вектором в становлении системы терминов, адекватно отражающей систему понятий любой науки, должен быть процесс создания квалификационных наименований [14].

Требование квалификационной номинации обусловлено рядом существенных недостатков эпонимических терминов. Первым из них отметим омонимию, которая в отличие от общезыковых формальных совпадений, не разрешается в специальном контексте. Например, термин *тракт Павлова* восходит к эпониму *Владимир Алексеевич Павлов* (отечественный хирург, специалист в области анатомии центральной нервной системы). Омонимичный ряд терминов: *учение Павлова*, *Павлова кишечная фистула* – в качестве производного имеет эпоним *Иван Петрович Павлов* (русский физиолог, лауреат Нобелевской премии). Имя собственное в составе специальной единицы вызывает прежде всего ассоциации с прецедентными именами, поэтому только биографическая справка может соотнести номинацию и запечатленное ею историческое лицо. Так как существенная часть имен, зафиксированных эпонимами, известна преимущественно узкому кругу специалистов, носители непрецедентных имен остаются в «ущемленном положении». Нейтрализации такого вида омонимии могли бы способствовать специальные эпонимические отсылки в словарях. Однако среди лексикографических медицинских изданий только 20% словарей фиксируют необходимые данные как часть словарной статьи.

Как негативное качество эпонимов признаем разное содержание терминов с предельно близкой формой, например, *Арнольда связка* – «связка, соединяющая шиловидный отросток лучевой кости с ладьевидной костью» и *Арнольда связки* – «связки, укрепляющие сесамовидные кости стопы». Аналогично, *Келликера клетка* – «клетка молекулярного

слоя коры мозжечка» и *Келликера клетки* – «мужские зародышевые клетки» и подобное.

Антропонимичные термины иноязычного происхождения могут иметь орфографические / орфоэпические варианты типа *Литтла болезнь* – *Литтля болезнь*, *Грехема Стилла шум* – *Грэма Стилла шум*, *Гофбауэра клетки* – *Хофбауэра клетки*, *Алькока канал* – *Олькока канал*, *Манроу симптом* – *Манро симптом* и др., что нередко затрудняет их идентификацию.

Можно отметить еще на одно нежелательное качество эпонимических терминов: в многокомпонентных наименованиях имя собственное называет видовой признак, зачастую не являющийся существенным в определении концепта. Неточность значения влечет за собой необходимость введения в структуру термина дифференцирующих компонентов, выраженных цифровой или адъективной частью: *Раубера вена (1)*, *Раубера вена (2)*; *Жданова рентгеноконтрастная масса (белая)*, *Жданова рентгеноконтрастная масса (красная)*. Введение дополнительных компонентов усложняет структуру термина и увеличивает его протяженность, нарушая требование краткости.

С произносительной точки зрения, термины-эпонимы не всегда соответствуют требованию благозвучности, например, *вирус Тхоттапалам*, *вирус О' Ньонг-Ньонг*, *вирус Пахйоки* и т. п.

Укажем также, что эпонимические термины не отвечают целому ряду психологических особенностей человека, его памяти: они не соответствуют принципам упорядоченности языковых единиц и доступности извлечения языковой информации. Пользование ими не основывается на семантических ассоциациях, более того, такая лексема даже не всегда указывает на тематическое поле, к которому она принадлежит. Специальная лексика достаточно условна, но все-таки эта условность должна основываться на естественных связях слов и именуемых с их помощью объектов. Эпонимические термины своей внешней формой не актуализируют мыслительно-вербальные связи, они плохо запоминаются, с трудом извлекаются из памяти, потому что, с точки зрения плана выражения, они не системны и не сигнализируют собой о том или ином гнезде таксономически связанных терминов.

Недостатки терминов-эпонимов служат веским основанием, чтобы считать эпонимию наиболее простым, но все же не лучшим путем преодоления терминологических трудностей. Поэтому терминоведы настоятельно рекомендуют, по возможности, сокращать количество отыменных наименований, заменяя их дублетными классификационными терминами.

Рассмотрим конкретный пример вариантной замены антропонимичного термина квалификационным (в медицинской терминологии). Врачом Альпортом было описано заболевание почек, которое и называли по имени его первого исследователя *синдромом Альпорта*. Со временем, когда удалось выявить этиологию заболевания, вызываемого дефектным геном в X-хромосоме, болезнь получила иное наименование – *наследственный нефрит*. Новое название не только определяет его принадлежность к

тематической группе «нефрология», не только квалифицирует саму патологию как воспалительный процесс, но и содержит указание на врожденный (не приобретенный) характер болезни. Кстати, и само родовое понятие *нефрит* (от греч. *nephros* – «почка») пришло на смену эпонимическому термину *Брайтова болезнь*, восходящему к имени английского врача З. Брайта, впервые установившего связи между отдельными клиническими проявлениями нефрологических болезней и механизмами их возникновения.

В переименованиях, подобных описанному выше, происходит перераспределение сем, составляющих основу номинации. В отыменном названии роль интегральной семы временно приобретает периферийный компонент, характеризующий не сами существенные свойства понятия, а экстралингвистические (антропонимические) признаки его становления и изучения. Впоследствии, по мере накопления эмпирического материала, происходит переоценка семантических черт именуемого понятия, которая имеет следствием перегруппировку интегральных и дифференциальных сем. В результате вторичная номинация строится уже с учетом классификационных, родовых признаков исследуемого явления.

Эпонимический термин можно считать микрономинацией, при которой первичное наименование еще не определяет статус термина в специальном языке и не обобщает семантические свойства нового научного объекта и схожих понятий. Поэтому эпонимический термин необязательно должен закрепиться в языке науки и, скорее всего, будет заменен более подходящим. Это положение основывается на ряде статистических и фактических данных. Потребности языка медицины заставляют использовать любую возможность замены эпонимических терминов единицами других типов. Так, в процессе унификации анатомической терминологии около 1300 эпонимических единиц были заменены синонимами-апеллятивами: вместо *Евстахиева труба* – *слуховая труба*, вместо *капсула Глиссона* – *околососудистая фиброзная капсула*, вместо *Морганиево отверстие* – *слепое отверстие языка*. Иногда предпочтение отдается экспрессивной единице – семантическому деривату, например, вместо термина *Галлера ноготь* – *птичья шпора*; аналогично, вместо *Мальгигиева сеть* – *чудесная сеть* (ЭСМТ, Т. 3, с. 395, с. 279). Терминология анатомии человека в соответствии с принятой в настоящее время Парижской международной номенклатурой сохранила эпонимические термины в качестве дублетов только в тех случаях, когда они служат основой для образования производных клинических терминов. Так, в качестве варианта словосочетания *верхнечелюстная пазуха* используется термин *Гайморова пазуха*, сохранение которого объясняется не только высокой степенью внедренности, но и наличием клинических терминов, в которых элемент *гаймор-* означает «относящийся к верхнечелюстной пазухе»: *гайморография*, *гаймороскоп*, *гаймороскопия*, *гайморотомия*, *гайморит* и др.

Таким образом, можно утверждать, что термины-эпонимы не отвечают многим принципам построения

оптимальных терминополь (требованию системности, оптимальной протяженности термина, отсутствия синонимии и многозначности и др.). Поэтому они зачастую заменяются эквивалентными вариантами, учитывающими в словопроизводстве классификационные и ассоциативные признаки термилируемых обозначений.

СПИСОК ЛИТЕРАТУРЫ

1. Сложеникина Ю.В., Растягаев А.В. Философский диалог об именах: Платон, Сумароков, современное состояние // *Онтология проектирования*. – 2015. – Т. 5, № 4 (18). – С. 463-471.
2. Остроглазов В. Синдром Эминеску. Обоснование нового эпонима // *Независимый психиатрический журнал*. – 2013. – № 1. – С. 91-96.
3. Какзанова Е.М. Интернациональные эпонимы антропонимического содержания // *Язык и культура в эпоху глобализации: сб. научн. тр. Второй Международной научн. конф.: в 2-х ч. / ред.кол.: Е.В. Белоглазова и др.* – М., 2015. – С. 182-190.
4. Асмукович И.В. Эпонимы в составе английских авиационных терминов // *Вісник Житомирського державного університету імені Івана Франка*. – 2011. – № 58. – С. 80-83.
5. Тулев Е.С., Беззатеева Э.Г. Эпонимы в терминосистеме теплоэнергетики // *Язык науки и техники в современном мире: материалы IV Международной научн.-практ. конф.* – М., 2015. – С. 198-199.
6. Калугина Ю.Е. Структурно-семантические характеристики эпонимов экономического подъязыка науки // *Вестник Московского государственного областного университета. Серия: Лингвистика*. – 2010. – № 3. – С. 74-79.
7. Извекова Т.Ф., Грищенко Е.В., Пуртов А.С. Эпонимы в медицинской терминологии // *Медицина и образование в Сибири*. – 2014. – № 3. – С. 18.
8. Клёстер А.М. Особенности эпонимов в немецкой терминологии инженерной психологии // *Известия Самарского научного центра РАН. Социальные, гуманитарные, медико-биологические науки*. – 2014. – Т. 16, № 2-2. – С. 416-419.
9. Курашкина Н.А. Роль компонентов топонимов и эпонимов в составе научных названий птиц // *Языки в диалоге культур: материалы Международной научн.-практ. конф., посвященной 75-летию д-ра филол. наук, профессора Р.З. Мурсова*. – М., 2014. – С. 216-231.
10. Бобырева Н.Н. Эпоним в специальном языке альпинизма // *Ученые записки национального общества прикладной лингвистики*. – 2015. – № 1 (9). – С. 61-67.
11. Масейко К.С. Термины-эпонимы издательского дела в английском языке // *Современная наука. Новые перспективы: сб. научных докладов*. – М., 2014. – С. 116-119.
12. Какзанова Е.М. Дискурс моды и его воплощение в интернациональных эпонимах // *Коммуникация в современном поликультурном мире: прагматика лингвистического знака / отв. ред. Т.А. Барановская*. – М., 2015. – С. 189-196.
13. Бобкова Н.С. Эпонимы в номинации французских блюд // *Древняя и Новая Романия*. – 2014. – Т. 14, № 2. – С. 17-25.
14. Сложеникина Ю.В. Вторичная терминологическая номинация: опора на классификационные признаки // *Научно-техническая информация. Сер. 2*. – 2006. – № 1. – С. 32-36.

Материал поступил в редакцию 20.03.17.

Сведения об авторах

СЛОЖЕНИКИНА Юлия Владимировна – доктор филологических наук, доцент, декан филологического факультета Московского городского педагогического университета (Самарский филиал), профессор кафедры филологии и массовых коммуникаций, г. Самара
e-mail: goldword@mail.ru

ЗВЯГИНЦЕВ Вячеслав Сергеевич – аспирант кафедры филологии и массовых коммуникаций Московского городского педагогического университета (Самарский филиал)
e-mail: zvs063@mail.ru

ИНФОРМАЦИОННОЕ ПИСЬМО И ПРИГЛАШЕНИЕ
МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ К 65-ЛЕТИЮ ВИНТИ РАН
«ИНФОРМАЦИЯ В СОВРЕМЕННОМ МИРЕ»
Москва, 25-26 октября 2017 г.

подробная информация на сайте: <http://www.viniti.ru>

Главный организатор:

Всероссийский институт научной и технической информации
Российской академии наук (ВИНИТИ РАН)

Соорганизаторы:

Российская академия наук
Федеральное агентство научных организаций
Российский фонд фундаментальных исследований
Министерство образования и науки РФ

Проблемно-тематическое направление конференции: современный издательский процесс, интеллектуальная собственность, научные библиотеки, информационное обеспечение научной и инновационной деятельности, информационные технологии для научной и библиотечной отрасли, информационная безопасность, международное сотрудничество и информационный обмен, инфометрия, классификации, стандартизация, образование для отрасли, экономика информации

Основные вопросы, предлагаемые к обсуждению:

- Популяризация научных знаний: Новые модели распространения научной информации
- Редакционно-издательская деятельность в цифровой среде: продукты и сервисы
- Издательские стандарты и технологии
- Перспективы развития книжного дела. Проекты и программы
- Взаимодействие цифровых и печатных ресурсов в научно-технической библиотеке
- Информационно-библиотечное обслуживание: сервисный подход
- Управление данными и навигация в современной научной библиотеке
- Научные библиотечные консорциумы – основные подписчики на научную литературу
- Перспективы развития национальных систем научно-технической информации
- Государственные проекты и программы поддержки информационного обеспечения научно-образовательной деятельности
- Тенденции развития региональных аналитических центров
- Информационное обеспечение экспертной деятельности. Использование информационно-аналитических систем для управления наукой и образованием
- Формальные и неформальные каналы развития современных научных коммуникаций

- Современные агрегаторы научной литературы открытого доступа как источник научной информации
- Машинная обработка данных и аналитические исследования: Приоритеты и сотрудничество
- Использование специальных сервисов компании CrossRef для идентификации научных публикаций
- Роль поисковых систем в современном издательском процессе
- Защита данных от несанкционированного использования. Маркеры безопасности. Политика безопасности открытых систем
- Вопросы достоверности и доверенности при обработке информационного потока
- Межгосударственный обмен научно-технической информацией на евразийском пространстве
- Информационное взаимодействие в рамках СНГ
- Международное партнерство при хранении и обработке больших массивов данных
- Современное состояние систем классификации знаний как инструмента индексирования и поиска данных по перспективным направлениям науки и критическим технологиям
- Современные библиометрические методы определения научных лидеров: Новые математические модели
- Анализ читательской аудитории научной литературы путем вебметрического анализа
- Подготовка специалистов в сфере научно-информационной деятельности
- Мастер-класс по работе с классификационными системами (УДК, ГРНТИ)
- Информация как источник цифрового капитала и фактор социальных изменений
- Информационная деятельность как фактор национальной экономики
- Новейшие бизнес-модели для публикаций открытого и закрытого доступа

На конференции планируются доклады представителей ведущих информационных центров и научно-технических библиотек России, СНГ и дальнего зарубежья.

В рамках юбилейной конференции состоится научно-практический семинар по классификационным системам «Перспективные направления научных исследований и критические технологии в классификационных системах». Предполагается проведение специализированных обучающих мероприятий по УДК индексированию. Запланировано заседание методического совета пользователей ГРНТИ и УДК. Участники конференции получают свидетельства о повышении квалификации.

Материалы конференции будут опубликованы в сборнике Трудов и на CD-ROM, основные – в сборнике **«Научно-техническая информация»**.

Доклады

Принимаются оригинальные работы, имеющие научное и прикладное значение, соответствующие тематическим направлениям конференции и НЕ ОПУБЛИКОВАННЫЕ ГДЕ-ЛИБО РАНЕЕ.

Предлагаемый доклад должен отвечать следующим требованиям:

1. Необходимо указать название доклада, фамилию, имя, отчество (полностью) авторов/соавторов, название организации, город, страну, выделить автора, который будет представлять доклад.
2. Необходимо наличие аннотации, раскрывающей содержание доклада. Размер аннотации - не более 850 знаков (включая пробелы).
3. Доклады принимаются только в электронной форме; тексты – в формате MS Word; схемы, диаграммы, фотографии, сканированные виды экранов и т. п. - в формате JPG. Объем доклада вместе с аннотацией, рисунками, приложениями и т.п. не более 10 страниц формата А4.
4. Доклад необходимо выслать по электронной почте до 11 сентября 2017 г. в адрес оргкомитета: conf@viniti.ru

Доклады, не соответствующие вышеуказанным требованиям,
НЕ РАССМАТРИВАЮТСЯ.

Программный комитет оставляет за собой право определять статус доклада (пленарный доклад, доклад, стендовый доклад), включать принятые доклады в те или иные секции.

Время для выступления: пленарные доклады – 15–20 мин., доклады на отдельных мероприятиях – до 10 мин. Доклады включаются в Труды на основании решения экспертов оргкомитета.

Контакты: 125190, Москва, ул. Усиевича, 20, ВИНТИ РАН

Телефоны: 8 (499) 152 61 13, 8 (499) 155 42 52, 8 (499) 151 02 61. Факс 8 (499) 943 00 60

Интернет-сайт: <http://www.viniti.ru> Эл. почта: conf@viniti.ru

Центр (Отдел) научно-информационного обслуживания (ЦНИО) ВИНИТИ РАН

Информационные услуги, предоставляемые ЦНИО ВИНИТИ РАН:

- проведение тематического поиска и консультации поисковых экспертов;
- подготовка списков научной литературы;
- подбор, копирование полнотекстовых материалов из первоисточников на бумажном носителе и в электронном виде;
- библиометрическая оценка публикационной активности исследователей и научных организаций с использованием российских и зарубежных баз данных;
- информационное обеспечение информационно-аналитической деятельности по подготовке и предоставлению аналитических обзоров и других научных материалов.

ВИНИТИ РАН располагает следующими информационными ресурсами:

- фондом НТЛ, включающим более 2,5 млн. отечественных и иностранных журналов, книг, депонированных рукописей, авторефератов диссертаций и другой научной литературы, ретроспектива – с 1991 года;
- базами данных и Интернет-ресурсами: БД ВИНИТИ (разработка ВИНИТИ), БД SCOPUS, БД Questel (патенты) и другими реферативными ресурсами;
- полнотекстовыми электронными ресурсами (статьи, патенты, материалы конференций).

Ознакомиться с информацией о доступных полнотекстовых и реферативных ресурсах можно на сайте ВИНИТИ www.viniti.ru

К услугам пользователей – **Электронный Каталог ВИНИТИ** <http://catalog.viniti.ru>
и **служба электронной доставки документов.**

Осуществляется платное информационное обслуживание по разовым заказам и на договорной основе с предоставлением всех необходимых финансовых документов.

Проводится индивидуальное обслуживание пользователей в читальном зале ЦНИО ВИНИТИ.

Обращаться в ЦНИО ВИНИТИ:

- адрес: 125190, Россия, г. Москва, ул. Усиевича, 20;
- телефоны: 8(499) 155 -42 -43, 8(499) 155 -42 -17;
- эл. почта cnio@viniti.ru, fdk@viniti.ru;
- факс 8(499) 930 -60 -00 (для ЦНИО).

База данных (БД) ВИНИТИ РАН

Федеральная база отечественных и зарубежных публикаций по естественным, точным и техническим наукам, генерируется с 1981 г., обновляется ежемесячно, пополнение составляет около 1 млн документов в год. Тематическое наполнение соответствует реферативному журналу ВИНИТИ. Для поиска одновременно по всем или нескольким тематическим фрагментам генерируется единая Политематическая БД.

БД ВИНИТИ РАН в сети INTERNET

Сервер ВИНИТИ - <http://www.viniti.ru> – обеспечивает on-line доступ к Базе данных ВИНИТИ РАН круглосуточно без выходных.

На основе БД ВИНИТИ РАН предоставляются следующие услуги:

- **Диалоговый поиск** научно-технической информации **в режиме on-line**;
- **Демо-версия**, позволяющая ознакомиться с основными функциями поисковой системы, составом данных, формами представления документов и получить навыки работы с системой;
- **Поисковые эксперты ВИНИТИ** выполняют тематический поиск по разовым или постоянным запросам, а также окажут **консультационные услуги**.

БД ВИНИТИ РАН на CD-ROM

Любые наборы тематических фрагментов БД ВИНИТИ или их разделов за любой период с 1981 г., а также **проблемно-ориентированные выборки** из БД ВИНИТИ по актуальным направлениям научных исследований могут быть предоставлены на договорной основе **в поисковой системе (ИПС) "Сокол"**, работающей под управлением Microsoft Windows и обеспечивающей следующие возможности:

- **Чтение** документов в режиме последовательного просмотра или выборочно по оглавлению за весь период заказанной ретроспективы
- **Поиск** документов по автору, заглавию, источнику, ключевым словам или словосочетаниям, реферату, рубрикам, году издания, стране, языку и т.д. (всего более 20 признаков)
- **Словарь** системы поможет правильно подобрать термины для поиска и выбрать глубину их усечения.
- Для **уточнения поиска** можно дополнительно использовать год издания документа, язык текста документа, рубрики, шифры тематических разделов БД.
- Выполненные **запросы можно сохранять** для их последующего использования и/или редактирования.

125190, г. Москва, ул. Усиевича, 20, БД ВИНИТИ РАН.

Отдел взаимодействия с потребителями – (499) 155-45-25, (499) 152-58-81

E-mail: csbd@viniti.ru, sales@viniti.ru

WWW: <http://www.viniti.ru>