

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

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

Издается с 1961 г.

№ 2

Москва 2017

ОБЩИЙ РАЗДЕЛ

УДК 002 : 001.89

П.А. Калачихин

Методика проведения наукометрической экспертизы результатов исследований*

Предлагается новая методика экспертизы результатов научных исследований, построенная на инструментарии нечетких вычислений и методе экспертных оценок. Многокомпонентная методика использует наукометрические показатели для оценки качества результатов выполнения работ научно-исследовательскими институтами и научными коллективами. Методика способна адаптироваться к различным областям науки благодаря возможности варьирования ряда параметров и весовых коэффициентов в соответствии со спецификой применения.

Ключевые слова: научно-исследовательская деятельность, научно-исследовательские и опытно-конструкторские работы, научные коллективы, научные организации, научный потенциал

ВВЕДЕНИЕ

В современных условиях при обилии ролевых типов научных организаций, начиная от «генераторов знаний» и заканчивая «контролерами», существует

объективная потребность в осуществлении систематического мониторинга за развитием отечественной науки. Такой контроль сопряжен с обработкой больших объемов документальной информации (*big data*), поступающей на выходе из процессов научно-исследовательской деятельности. При этом очень важно владеть инструментом для проведения экспертизы качества результатов выполненных научных ис-

* Работа выполнена в рамках исследования по теме 0003-2015-0008 государственного задания ВИНТИ РАН

следований и выявления наиболее эффективных и перспективных кандидатов-исполнителей государственных заданий, а также претендентов на распределение финансирования по грантам. В таком случае, актуальной задачей является создание методики для оценки качества результатов деятельности научно-исследовательских институтов и коллективов.

Главная цель настоящего исследования заключается в разработке упомянутой методики специально для отечественных научно-исследовательских институтов. Для достижения этой цели нам необходимо проанализировать имеющийся опыт и описать методологию, а после этого – перейти непосредственно к разработке математических методов.

При такой постановке задачи объектами исследования будут выступать научно-исследовательские институты и коллективы, качество результатов исследований которых нам необходимо научиться оценивать *апостериорно*, опираясь на наукометрические показатели. Предмет этого исследования связан с использованием экономико-математических моделей, в которых происходит *априорная* оценка различных подвидов инновационного потенциала применительно к ожидаемым результатам научно-исследовательских и опытно-конструкторских разработок.

Несмотря на систематизацию и универсализацию научных знаний, в некоторых междисциплинарных направлениях, в целом, отечественная наука со временем становится все более и более дифференцированной. Однако эта тенденция вовсе не препятствует нам сформулировать гипотезу, согласно которой *оценку качества результатов научных исследований (в т. ч. исследований в области наукометрии) вполне оправданно проводить при помощи наукометрических показателей.*

Следует сослаться на компонентный подход, который нами применяется при разработке методики экспертизы качества результатов выполнения научных исследований. В соответствии с этим подходом, показатель качества результатов научных исследований должен быть разбит на несколько компонент, которые, в свою очередь, также подлежат декомпозиции на более элементарные показатели.

ВАЖНОСТЬ ОПЫТА ОЦЕНКИ НАУЧНЫХ ИССЛЕДОВАНИЙ

Примечательно, что на тепловой карте науки проекта «Карта российской науки»^{**} нет такого направления, как «наукометрия». В связи с тем, что наукометрия возникла в середине XX века, такой факт вряд ли можно объяснить «молодостью» направления. Это обстоятельство скорее свидетельствует о том, что наукометрия находится вдали от «горячих точек» на тепловой карте науки, другими словами, это направление не развивается столь интенсивно, как могло бы (возможно, по причине того, что наукометрия не представляет коммерческого интереса для потенциальных инвесторов). Тем не менее, несмотря на то, что наукометрические исследования являются не настолько популярной темой в мировом научном

сообществе как многие другие области науки, для нас очень важно уметь грамотно оценивать результативность исследований внутри наукометрии, подводя итоги таких оценок с детализацией по отдельным научно-исследовательским институтам или коллективам.

О всеобщем признании высокой значимости и ответственности роли наукометрии в управлении отечественной наукой красноречиво свидетельствует тот факт, что существуют нормативные документы, определяющие и регулирующие оценку результативности научно-исследовательской деятельности. Эти документы были сформированы на основе результатов выполнения комплекса программных мероприятий в 2005–2008 гг. Кроме того, в нашей стране в 2009–2015 гг. был издан ряд документов, касающихся отечественных научно-исследовательских институтов [1].

Отечественные научные организации обладают большим опытом проведения наукометрических экспертиз результатов научных исследований как в рамках принятых стандартов, так и за этими рамками. Например, в ВИНТИ РАН было выполнено специальное исследование по наукометрическому анализу одного из ведущих отечественных научных журналов. Этот анализ позволил выявить многогранную сеть научных коммуникаций отечественных исследователей и влияние конкурсного финансирования на развитие научных исследований в одной из областей науки. Благодаря результатам этого исследования стало ясно, что общие тенденции мировых наукометрических исследований показывают, какую важную роль для науки играют деятельность национальных научных фондов, частные инвестиции в передовые научные направления и стратегия выбора приоритетов в научных исследованиях [2].

МЕТОДОЛОГИЯ ОЦЕНКИ НАУКОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Выделяют три основных преимущества наукометрического анализа по сравнению с остальными методами экспертизы:

- 1) охватываются все науки в целом;
- 2) имеется возможность применять различные методики анализа на материале международных наукометрических баз данных;
- 3) материалом служат уже опубликованные работы и процитированные публикации, что придает результатам экспертизы большую объективность.

При сопоставлении экспертных оценок с оценками на основе показателей по цитированию, была установлена корреляция между различными методиками оценки. Но до сих пор преобладало мнение, согласно которому показатели на основе цитирования предпочтительнее использовать для оценки качества научных исследований, так как они имеют более широкий диапазон охвата различных характеристик результатов научных исследований.

Альтернативным наукометрическим методом является контент-анализ – тематический анализ исходящих публикаций для выявления информационных потребностей научных сотрудников. Его результаты впоследствии могут использоваться для сбора и ана-

^{**} <https://www.mapofscience.ru>

лиза тематически идентичных международных массивов публикаций и отбора журналов с наибольшим числом релевантных статей.

Итак, можно выделить три основных класса наукометрических показателей:

- количество публикаций автора, коллектива, института;
- количество цитирований, полученных этими публикациями;
- соотношение количеств публикаций и цитирований [2].

Научно-исследовательская деятельность оценивается по следующему блоку критериев:

- актуальность и перспективность направлений научных исследований, реализуемых научной организацией;
- вовлеченность научной организации в национальное и мировое научно-образовательное сообщество;
- инфраструктура научной организации и ресурсная обеспеченность научных направлений;
- перечень государственных и международных премий, призов, наград, почетных званий, полученных научной организацией или отдельными ее работниками.

Внутри каждого из этих блоков показателей наиболее важными являются следующие удельные показатели в расчете на одного исследователя:

- цитируемость работников научной организации;
- общее число участников всех проведенных организацией мероприятий;
- количество положительных решений по заявкам на выдачу охранных документов или свидетельств о государственной регистрации объектов интеллектуальной собственности;
- затраты на исследования и разработки за счет средств, полученных из иностранных источников;
- количество работников научной организации, являющихся членами редакционных коллегий зарубежных и отечественных научных журналов;
- объем средств, поступивших по договорам с отечественными и зарубежными организациями реального сектора экономики на выполнение исследований, оказание научно-технических и иных услуг;
- число диссертаций, защищенных работниками в отчетном году [1].

На практике, в зависимости от числа и уровня публикаций в периодических изданиях, формируются «показатели результативности научной деятельности» и индивидуальные рейтинги научных работников. Показатели публикационной активности влияют на стимулирующие выплаты сотрудникам и учитываются научными фондами при распределении грантов на научные исследования.

При оценке научных достижений или эффективности научного труда учитывается преимущественно научно-исследовательский труд теоретического характера, в результате которого, собственно, и возникают научные публикации. Совершенствование оценки результативности научных работ должно идти по пути разработки показателей, учитывающих всю систему научно-исследовательской деятельности, а не только ее исследовательскую составляющую

(кроме финансовой, кадровой и ресурсной), которая в наше время является завершающим этапом, опирающимся на все предыдущие [4].

При написании статьи на определенную тему каждый автор принимает решение включать или не включать ту или иную ссылку в статью. Другими словами, авторам приходится «голосовать» за ту или иную ссылку. Авторы, пишущие статьи на определенную тему, выполняют коллективную работу по определению характерных ссылок по этой теме. Это может быть использовано для оптимального объединения индивидуальных решений авторов, согласно которому автору может быть присвоен некоторый весовой коэффициент, который тем весомее, чем выше компетенция автора [5].

Для оценки и сопоставления характеристик научно-исследовательской деятельности принято использовать следующие показатели:

- ◆ количество ученых в коллективе;
- ◆ доля ученых *данного коллектива* в общем количестве ученых института или организации;
- ◆ количество публикаций ученых *данного коллектива*;
- ◆ доля публикаций ученых *данного коллектива* в общем количестве публикаций института или организации;
- ◆ общее количество ученых, участвовавших в публикациях *данного коллектива*;
- ◆ доля привлеченных ученых, не являвшихся членами *данного коллектива*, в общем количестве ученых *данного коллектива*;
- ◆ количество участия других авторов в публикациях ученых *данного коллектива*;
- ◆ доля участия других авторов в публикациях ученых *данного коллектива* в общем количестве участия авторов во всем массиве публикаций;
- ◆ среднее количество публикаций в расчете на одного ученого *данного коллектива*;
- ◆ среднее количество авторов в публикациях, в которых участвовали ученые *данного коллектива*;
- ◆ средневзвешенный импакт-фактор массива публикаций, в которых участвовали ученые *данного коллектива*;
- ◆ средневзвешенный агрегированный импакт-фактор массива публикаций, в которых участвовали ученые *данного коллектива*;
- ◆ научный уровень массива публикаций, в которых участвовали ученые *данного коллектива* по отношению к среднемировому уровню;
- ◆ среднее количество ссылок на публикацию, в которой участвовали ученые *данного коллектива*;
- ◆ количество проектов, в публикациях по которым участвовали ученые *данного коллектива*, и по результатам которых были получены патенты;
- ◆ количество патентов, полученных по результатам тех проектов, в публикациях по которым участвовали ученые *данного коллектива* [6].

Основные элементы данных, содержащихся в массиве исходной информации, используемые для определения научного уровня публикации, представлены такими реквизитами, как название организации и места работы авторов; город, страна; название и индекс журнала; номер гранта или исследовательско-

го проекта, поддержанного научными фондами или агентствами; название фонда; название тематической категории Web of Science, к которой относится та или иная статья из данного массива.

Производными показателями от предложенных базовых показателей будут являться:

- ожидаемый отклик на подмассив статей данного массива, которые опубликованы в данном журнале и соответствуют данной категории Web of Science;
- ожидаемый отклик на подмассив всех статей из данного массива, которые соответствуют данной категории Web of Science;
- ожидаемый отклик на подмассив статей из данного массива, которые опубликованы в данном журнале и соответствуют различным предметным категориям Web of Science;
- суммарный ожидаемый отклик на все статьи из данного массива;
- средневзвешенный импакт-фактор подмассива статей из данного массива, которые соответствуют данной категории Web of Science;
- средневзвешенный импакт-фактор данного массива статей;
- научный уровень подмассива всех статей из данного массива, которые соответствуют данной конкретной категории Web of Science;
- ожидаемый агрегированный отклик на подмассив всех тех статей из данного массива, которые соответствуют данной категории Web of Science;
- суммарный ожидаемый агрегированный отклик на все статьи данного массива;
- суммарное число статей в данном массиве;
- средневзвешенный агрегированный импакт-фактор данного массива статей;
- обобщенный научный уровень данного массива публикаций [7].

Существует более сотни показателей для измерения авторского уровня при написании научно-исследовательских статей. Число индикаторов кажется слишком высоким, учитывая, что они рассчитываются по достаточно сильно различающимся методикам. С таким количеством потенциальных индикаторов и при таком широком охвате важно исследовать характеристики индикаторов авторского уровня, чтобы обеспечить их использование не только экспертами при помощи программных средств, но также и самими исследователями. Так, например, возможно оценивать исследователей, используя максимальный h -индекс, измеренный через различные наукометрические базы данных. Ранжирование исследователей осуществляется с использованием всех наукометрических баз данных вместе. Предполагается, что различия в оценках h -индексов наукометрических баз данных происходят из-за ошибок, но такие ошибки незначительны. Несмотря на то, что максимальный индекс Хирша обеспечивает лучшую оценку h -индекса, чем какая бы то ни была единственная наукометрическая база данных, тщательное изучение перекрытий цитат и публикаций между базами данных позволит дать наилучшую оценку [8].

Помимо всего прочего, одним из сдерживающих факторов динамичного развития отечественной науки является дефицит грамотной научной экспертизы. Адекватное использование наукометрии возможно в том случае, если наукометрические показатели будут применяться как частный инструмент в продуманной системе поиска и выбора научных экспертов, в организации оценок и конкурсных процедур именно экспертными методами. Коллективная оценка цитируемых специалистов формирует экспертный «пул», который может включать людей с уже любыми индексами цитирования, если коллеги сочли их достаточно компетентными [9]. Нам представляется наиболее естественным оценивать научно-исследовательские институты экспертно, с привлечением большого количества специалистов. Самый лучший международный опыт, который следовало бы перенять, это опыт «экспертных панелей» [10]. Вместо суммарного индекса цитирования в качестве критерия подходит число активно работающих научных коллективов, которые часто могут быть обнаружены по индексам цитирования лидеров таких коллективов [9].

МЕТОДИКА ОЦЕНКИ КАЧЕСТВА РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ

Методика оценки качества результатов научных исследований должна отвечать требованиям по таким критериям, как достоверность результатов расчетов, лаконичность формальной записи, открытость вычислительных процедур, прозрачность выполняемых расчетов, интегрируемость с другими методиками и моделями, алгоритмизируемость на существующих языках программирования. Последний пункт требований является одним из самых немаловажных, так как придает методике практическую ценность при разработке программного обеспечения для наукометрических информационных систем.

На первый взгляд, проще всего адаптировать или модифицировать существующие методики оценки эффективности результатов научных исследований применительно к области наукометрических исследований. Но все же этого не следует делать, потому что, применяя наукометрические методы к оценке результатов наукометрических исследований, нам нужно быть уверенными в самих же этих наукометрических методах.

Главным результатом научно-исследовательской деятельности вполне обоснованно можно считать научную публикацию, поэтому наукометрическая экспертиза должна основываться на наукометрическом анализе выполненных работ. Однако, в связи с тем, что обрабатывать такие большие массивы информации весьма сложно, необходимо делать некоторые выборки из общего числа публикаций. Объем выборки должен рассчитываться в зависимости от общего количества научно-исследовательских институтов и коллективов, суммарного числа публикаций и числа экспертов.

В результате всех вычислений на выходе должен получаться рейтинг $R_\alpha - 0..100$ для α -го научно-исследовательского института или коллектива, если

научно-исследовательский институт занимается многопрофильными исследованиями:

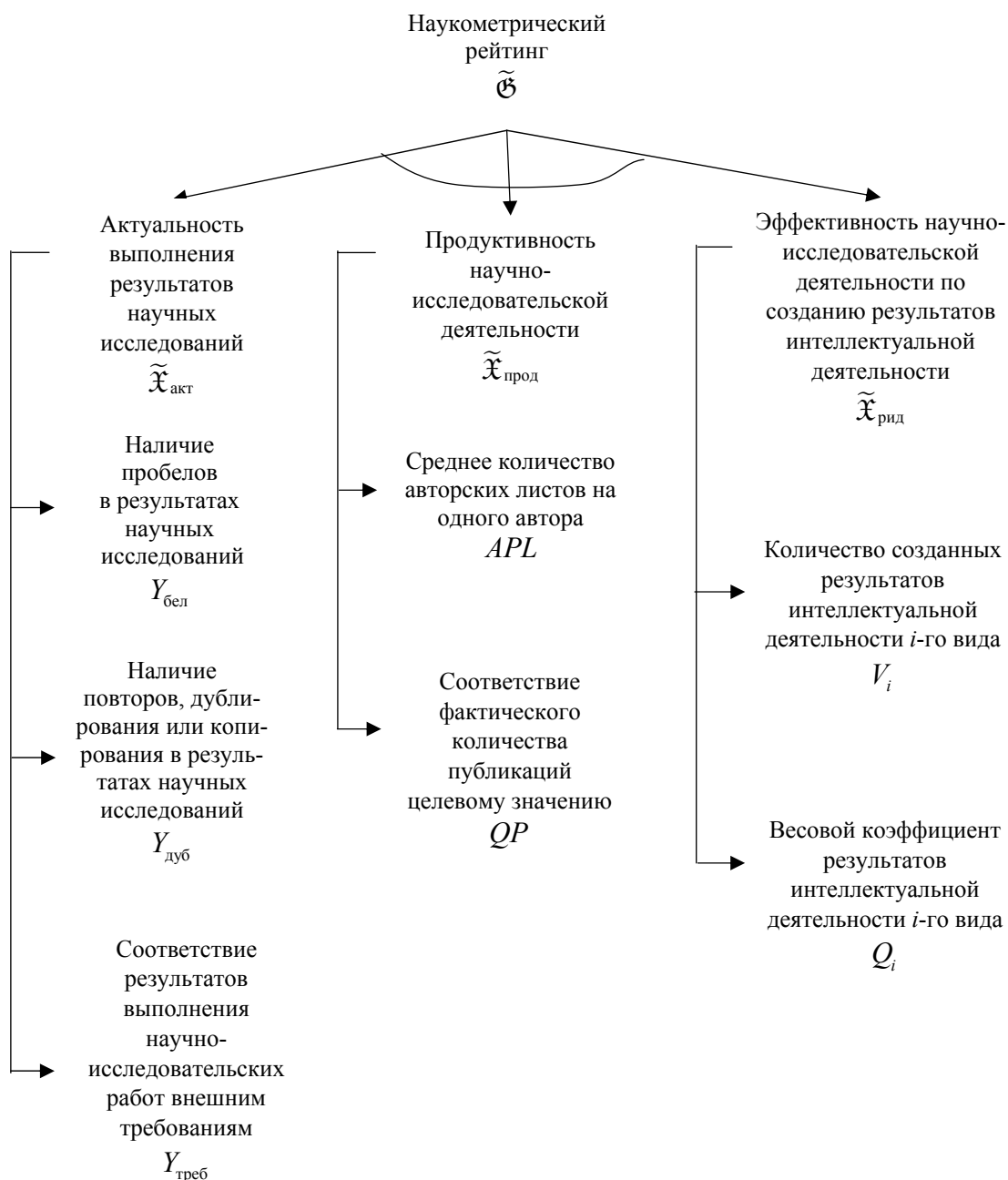
$$\mathcal{R}_\alpha \approx \mathcal{G}_\alpha \cdot CF_{\mathcal{G}_\alpha} = [\mathcal{G}_\alpha \cdot CF_{\mathcal{G}_\alpha}] \quad (1)$$

Этому должно предшествовать получение некоторой величины $\tilde{\mathcal{G}} = \{\mathcal{G}, CF_{\mathcal{G}}\}$, представленной в виде пары из обыкновенного числа $\mathcal{G} \in [0,1]$ и фактора уверенности $CF_{\mathcal{G}} \in [0,100]$.

Все компоненты показателя качества результатов научно-исследовательской деятельности на рисунке могут быть отнесены к результатам деятельности непосредственно в области наукометрии.

Как видно из рисунка, наукометрический показатель $\tilde{\mathcal{G}}$ качества результатов научно-исследовательской деятельности, выполненных научно-исследовательскими институтами или коллективами, формируется из следующих трех компонент:

- метрика $\tilde{\mathcal{X}}_{\text{акт}}$ актуальности результатов научно-исследовательской деятельности;
- индикатор $\tilde{\mathcal{X}}_{\text{прод}}$ публикационной продуктивности научно-исследовательской деятельности;
- индикатор $\tilde{\mathcal{X}}_{\text{рид}}$ эффективности научно-исследовательской деятельности по результатам интеллектуальной деятельности.



Дерево показателей для оценки качества результатов научных исследований

Информации из наукометрических баз данных для оценки показателей на рисунке, вероятнее всего, будет недостаточно, поэтому исходные данные должны извлекаться из отчетов о научно-исследовательской деятельности. Как уже указывалось, экспертная комиссия должна знакомиться с полными текстами научных работ, или, по крайней мере, делать это выборочно.

Агрегирование компонент $\tilde{\mathcal{X}}_{\text{акт}}$, $\tilde{\mathcal{X}}_{\text{прод}}$, $\tilde{\mathcal{X}}_{\text{рид}}$ в интегральный показатель $\tilde{\mathcal{G}}$ качества результатов научных исследований осуществляется по формуле:

$$\tilde{\mathcal{G}} = \tilde{\mathcal{X}}_{\text{акт}} \odot \tilde{\mathcal{X}}_{\text{прод}} \odot \tilde{\mathcal{X}}_{\text{рид}} \quad (2)$$

где $\odot$ – операция умножения нечетких пар (соответствий).

Компоненту $\mathcal{X}_{\text{акт}}$ следует рассчитывать как среднее геометрическое показателя $Y_{\text{бел}}$ наличия пробелов в результатах научных исследований; показателя $Y_{\text{дуб}}$ – наличия повторов, дублирования или копирования существующих результатов в полученных результатах научных исследований; показателя $Y_{\text{треб}}$ – соответствия результатов выполнения разработок внешним требованиям:

$$\mathcal{X}_{\text{акт}} = \sqrt[3]{Y_{\text{бел}} \cdot Y_{\text{дуб}} \cdot Y_{\text{треб}}} \quad (3)$$

Первичные показатели $Y_{\text{бел}}$, $Y_{\text{дуб}}$ и $Y_{\text{треб}}$ рассчитываются при помощи метода экспертных оценок:

$$Y_i = \sum_{j=1}^N (Ex_{ij}^* \cdot W_j), \quad (4)$$

где Ex_{ij}^* – приведенная оценка j -го эксперта i -й метрики; W_j – вес j -го эксперта; N – количество экспертов в составе экспертной комиссии; $i = \{\text{бел, дуб, треб}\}$.

Качественные шкалы одинаковой размерности для оценки наличия «белых пятен», определения научной новизны и соответствия тематике исследований результатов выполнения научно-исследовательских работ задаются по табл. 1.

Причем показатели экспертных оценок приводятся к нормализованному виду Ex_{ij}^* по формуле:

$$Ex_{ij}^* = \frac{Ex_{ij} - Ex_{\min i}}{Ex_{\max i} - Ex_{\min i}}, \quad (5)$$

где Ex_{ij} – балльная экспертная оценка j -го эксперта i -й метрики; $Ex_{\min i}$ – минимальное значение балльной метрики; $Ex_{\max i}$ – максимальное значение балльной метрики.

Весовые коэффициенты W_j формулы (4) рассчитываются в зависимости от компетентности экспертов, которую возможно определять разными способами, например, по одной из методик в [11]. При этом должно выполняться условие нормированности весов экспертов $\sum_{j=1}^N W_j = 1$, где N – количество экспертов в составе экспертной комиссии.

Компонента $\mathcal{X}_{\text{прод}}$ вычисляется на основании сравнения фактических наукометрических показателей продуктивности среднего количества $APL_{\text{факт}}$ авторских листов VAL , приходящихся на число авторов QA , и суммарного количества публикаций $QP_{\text{факт}}$, соответственно, с целевыми значениями $APL_{\text{цел}}$ и $QP_{\text{цел}}$ этих же показателей:

$$\begin{aligned} \mathcal{X}_{\text{прод}} &= \sqrt{APL \cdot QP} = \sqrt{\frac{APL_{\text{факт}}}{APL_{\text{цел}}} \cdot \frac{QP_{\text{факт}}}{QP_{\text{цел}}}} = \\ &= \sqrt{\frac{VAL}{QA} \cdot \frac{QP_{\text{факт}}}{QP_{\text{цел}}}} \end{aligned} \quad (6)$$

В формуле (6) учитываются только те авторы, которые подготовили (или приняли участие в подготовке), по крайней мере, одной публикации, т.е. журнальной статьи, авторской или коллективной монографии и т.п.

Таблица 1

Качественные шкалы метрик актуальности результатов выполнения научных исследований

Качественная шкала оценки метрики $Y_{\text{бел}}$	Качественная шкала оценки метрики $Y_{\text{дуб}}$	Качественная шкала оценки метрики $Y_{\text{треб}}$	Балл	Приведенное значение
1	2	3	4	5
Программа выполнения исследований реализована «в глубину» и «в ширину» крайне поверхностно и слабо	Исследование абсолютно не содержит никаких нетривиальных результатов	Результаты исследований абсолютно не согласуются с предъявленными к ним требованиями	1	0

Качественная шкала оценки метрики $Y_{\text{бел}}$	Качественная шкала оценки метрики $Y_{\text{дуб}}$	Качественная шкала оценки метрики $Y_{\text{треб}}$	Балл	Приведенное значение
1	2	3	4	5
Эпизодически исследованы только наиболее общие вопросы, проделана лишь вступительная работа в области решения обозначенных задач и проблем	Новизна большинства результатов исследования крайне низкая	Результаты исследований представляют дальние аналоги решения проблем и задач, прописанных в проектной документации	2	0,125
Результаты исследования охватывают выборочно область решения заявленных задач и проблем	Результаты исследования во многом повторяют полученные ранее, хотя имеются и оригинальные идеи	Результаты исследований плохо соответствуют плановой траектории выполнения исследований и нуждаются в уточнении	3	0,250
Получен ряд отдельных важных результатов без формирования целостной итоговой картины исследований в данной области	Результаты исследования представляют некоторую научную ценность и новизну	Некоторые результаты отвечают заявленным требованиям, но в целом полученные результаты требуют уточнения	4	0,375
План исследований реал- изован в целом удовле- творительно, но без ак- центирования наиболее значимых вопросов	Новизна результатов ис- следования позволяет го- ворить об определенном вкладе авторов в реше- ние круга задач по рас- сматриваемой проблема- тике	Результаты исследований удовлетворительно соответствуют плану научных исследований	5	0,500
Исследования прорабо- таны более чем удовле- творительно, однако, требуется доведение до завершенного состояния некоторых разработок	В исследовании имеются неординарные, позитив- ные результаты, подхо- дящие под критерии на- учной ценности	Многие результаты весьма точно отвечают на поставленные перед исследователями вопросы, однако, нуждаются в конкретизации	6	0,625
Результаты исследова- ния в значительной мере охватывают предметную область исследований, несмотря на имеющиеся отдельные пробелы	Почти все результаты исследования являются достаточно новыми и оригинальными	Результаты исследований по большой мере соответствуют критериям адекватности, хотя встречается неубедительные примеры	7	0,750
Степень разработанно- сти темы исследования превышает средний сложившийся научный уровень глубины заде- лов в смежных предмет- ных областях науки	В исследовании удалось получить ряд уникаль- ных результатов, имею- щих большое значение для науки	Результаты исследований так же точно соотносятся с требованиями, заявленными заказчиками исследований, как и работы прочих, наиболее передовых авторов, коллективов и организаций	8	0,875
Тематика исследований максимально тщательно и глубоко проработана	Исследование содержит прорывные результаты, позволяющие пересмотреть важные положения современных научных представлений	Результаты исследований полностью соответствуют исходным требованиям проектов, в рамках которых они выполняются	9	1,00

Целевые показатели $APL_{\text{цел}}$ и $QP_{\text{цел}}$ определяются экспертами по прогрессивным шкалам в зависимости от научного потенциала научно-исследовательских институтов. Научный потенциал научно-исследовательских институтов возможно оценивать согласно любому из подходов, изложенных в [12]. Научный потенциал также логично связывать с количеством сотрудников в научно-исследовательском институте или коллективе, а также с объемом денежного финансирования научно-исследовательского проекта, результаты выполнения которого выступают объектом экспертизы.

В таком случае для научно-исследовательских институтов и коллективов с большими значениями научного потенциала S должны быть заданы высокие целевые значения $APL_{\text{цел}}$ и $QP_{\text{цел}}$ по принципу «чем больше научный потенциал S , тем выше целевые значения»:

$$\uparrow S \rightarrow (\uparrow APL_{\text{цел}} \wedge \uparrow QP_{\text{цел}}). \quad (7)$$

Компонента $\mathfrak{X}_{\text{рид}}$ рассчитывается методом сравнения с некоторым целевым значением $Z_{\text{цел}}$ среднего взвешенного $Z_{\text{факт}}$ количества V_k созданных программ ($k=1$), баз данных ($k=2$), ноу-хау ($k=3$), изобретений ($k=4$) и т.д., взятых с весовыми коэффициентами U_k , определенными для каждого из типов результатов интеллектуальной деятельности и в обязательном порядке защищенных как объекты интеллектуальной собственности:

$$\mathfrak{X}_{\text{рид}} = \frac{Z_{\text{факт}}}{Z_{\text{цел}}} = \frac{\sum_k V_k \cdot U_k}{\sum_k U_k} \cdot \frac{1}{Z_{\text{цел}}}, \quad (8)$$

где суммирование ведется по всему набору задействованных видов результатов интеллектуальной деятельности, а весовые коэффициенты U_k определяются экспертами в зависимости от специфических особенностей научной области (например, наукометрии).

Целевое значение $Z_{\text{цел}}$ зависит от научного потенциала научно-исследовательского института, по аналогии с компонентой $\mathfrak{X}_{\text{прод}}$ в (6).

Компонента $\mathfrak{X}_{\text{акт}}$ наукометрического показателя $\mathfrak{E}$ качества результатов научно-исследовательской деятельности должна быть приведена к нечеткому виду. Приведение компоненты $\mathfrak{X}_{\text{акт}}$ актуальности результатов научных разработок осуществляется в соответствии с формулой:

$$CF_{\mathfrak{X}_{\text{акт}}} = 100 \cdot \left(1 - \max_j (Ex_{\text{акт}j}^*) + \min_j (Ex_{\text{акт}j}^*) \right), \quad (9)$$

где $CF_{\mathfrak{X}_{\text{акт}}}$ – коэффициент уверенности компоненты $\tilde{\mathfrak{X}}_{\text{акт}} = \langle \mathfrak{X}_{\text{акт}}, CF_{\mathfrak{X}_{\text{акт}}} \rangle$ актуальности результатов выполнения научно-исследовательских работ; $\max_j (Ex_{\text{акт}j}^*)$ и $\min_j (Ex_{\text{акт}j}^*)$ – соответственно, максимальная и минимальная приведенные экспертные оценки компоненты $\mathfrak{X}_{\text{акт}}$ (по всем экспертам в составе экспертной комиссии).

Преобразование к нечеткому виду остальных двух компонент происходит в зависимости от степени уверенности членов экспертной комиссии в исходной информации для расчетов. Такое решение обусловлено тем, что пара названных компонент вычисляется по алгебраическим формулам на основании данных, которые могут различаться по версиям тех или иных наукометрических источников (баз данных).

Для фактора уверенности $CF_{\mathfrak{X}_{\text{прод}}}$ компоненты $\tilde{\mathfrak{X}}_{\text{прод}} = \langle \mathfrak{X}_{\text{прод}}, CF_{\mathfrak{X}_{\text{прод}}} \rangle$ продуктивности выполнения научно-исследовательских работ будет справедливо:

$$CF_{\mathfrak{X}_{\text{прод}}} = \frac{CF_{VAL} \cdot CF_{QA} \cdot CF_{QP_{\text{факт}}}}{100^2}, \quad (10)$$

где CF_{VAL} – фактор уверенности, характеризующий достоверность значения показателя VAL количества опубликованных авторских листов; CF_{QA} – фактор уверенности, отражающий степень достоверности показателя QA количества авторов; $QP_{\text{факт}}$ – фактор уверенности, характеризующий достоверность значения показателя $QP_{\text{факт}}$ суммарного количества публикаций.

В свою очередь, для фактора уверенности $CF_{\mathfrak{X}_{\text{рид}}}$ компоненты $\tilde{\mathfrak{X}}_{\text{рид}} = \langle \mathfrak{X}_{\text{рид}}, CF_{\mathfrak{X}_{\text{рид}}} \rangle$ эффективности выполнения научно-исследовательских работ по результатам интеллектуальной деятельности будет справедливо:

$$CF_{\mathfrak{X}_{\text{рид}}} = \frac{\prod_{k=1}^n CF_{V_k}}{100^{n-1}}, \quad (11)$$

где CF_{V_k} – фактор уверенности, характеризующий достоверность того, что было создано именно V_k РИД k -го вида; n – количество различающихся видов результатов интеллектуальной деятельности, созданных в результате проведения научно-исследовательских работ.

Показатели факторов уверенности CF_{VAL} , CF_{QA} , $CF_{QP_{\text{факт}}}$, CF_{V_k} должны определяться коллегиальным решением экспертной комиссии. Для достижения цели следует использовать шкалы подстановки чисел вместо качественных признаков из табл. 2.

**Качественные шкалы оценки достоверности показателей продуктивности
и эффективности результатов научной деятельности**

Качественная оценка факторов уверенности $CF_{VAL}, CF_{QA}, CF_{QP_{факт}}, CF_{V_k}$	Количественная оценка факторов уверенности $CF_{VAL}, CF_{QA}, CF_{QP_{факт}}, CF_{V_k}$
«Нулевая уверенность»	0
«Крайне низкая достоверность»	10
«Очень низкая достоверность»	20
«Низкая достоверность»	30
«Сниженная достоверность»	40
«Удовлетворительная достоверность»	50
«Более чем удовлетворительная достоверность»	60
«Умеренная уверенность»	70
«Высокая уверенность»	80
«Очень высокая уверенность»	90
«Абсолютная уверенность»	100

По умолчанию, все источники данных следует считать надежными, поэтому процедура приведения к нечеткому виду эквивалентна присваиванию компонентам CF_{x_β} , где $\beta = \{\text{прод, рид}\}$, максимального константного значения:

$$CF_{x_\beta} = 100. \quad (12)$$

Операция умножения $\odot$ нечетких пар (соответствий) $\tilde{X}_A = \{X_A, CF_{X_A}\}$ и $\tilde{X}_B = \{X_B, CF_{X_B}\}$, осуществляется по правилу:

$$\tilde{X}_A \odot \tilde{X}_B = \left\{ X_A \cdot X_B, \frac{CF_{X_A} \cdot CF_{X_B}}{100} \right\}. \quad (13)$$

Уже после того, как рейтинги $\mathcal{R}_\alpha$ отдельных научно-исследовательских институтов (или коллективов в составе этих институтов) сформированы, возможно принять следующие меры по осуществлению следующих управляющих воздействий:

(I) расставить приоритеты финансирования в ближайшем будущем в зависимости от позиций научно-исследовательских институтов (а также коллективов) в ранжированном по показателю $\mathcal{R}$ списке;

(II) произвести отбор наиболее передовых научно-исследовательских институтов (а также коллективов), в зависимости от того, превышают ли их рейтинги $\mathcal{R}_\alpha$ некоторый заданный пороговый уровень $\mathcal{R}_{min}$ (например, $\mathcal{R}_{min} = 50$ или $\mathcal{R}_{min} = 75$).

Уровни качества результатов научных исследований задаются по регрессивному принципу «чем выше уровень качества $\mathcal{R}$, тем уже интервал $\Delta_{\mathcal{R}}$ его рейтинговой группы $\mathcal{R}$ »:

$$\uparrow \mathcal{R} \rightarrow \downarrow \Delta_{\mathcal{R}}. \quad (14)$$

Интервалы $\Delta_{\mathcal{R}}$ рейтинговых групп $\mathcal{R}$ рейтинга качества $\mathcal{R}$ результатов научных исследований задаются в соответствии с данными табл. 3.

Таблица 3

**Интервалы рейтинга качества результатов
научных исследований**

Наименования рейтинговых групп $\mathcal{R}$ в порядке возрастания	Интервалы $\Delta_{\mathcal{R}}$ рейтинга $\mathcal{R}$
«Низкий»	$0 < \mathcal{R} \leq 50$
«Средний»	$50 < \mathcal{R} \leq 75$
«Высокий»	$75 < \mathcal{R} \leq 88$
«Очень высокий»	$88 < \mathcal{R} \leq 100$

После получения всех рейтингов $\mathcal{R}_\alpha$ по α -м научно-исследовательским институтам и научно-исследовательским коллективам, их следует «нормировать» при помощи операции $\parallel$ таким образом, чтобы нормированные рейтинговые показатели $\parallel \mathcal{R}_\alpha \parallel$ в сумме по всем объектам наукометрической экспертизы давали максимально константное значение, т.е. 100:

$$\parallel \mathcal{R}_\alpha \parallel = 100 \cdot \frac{\mathcal{R}_\alpha}{\sum_{\alpha} \mathcal{R}_\alpha}. \quad (15)$$

Тогда к «ведущим» научно-исследовательским институтам (или коллективам) можно будет отнести тех исследователей, которые индивидуально набрали в сумме более 50 % от общего нормированного рейтинга.

ЗАКЛЮЧЕНИЕ

В настоящей статье была предложена новая методика проведения экспертизы результатов выполнения научных исследований, выполняемых целыми научно-исследовательскими институтами, а также их отдельными научными коллективами. Новая методика предназначена специально для оценки научно-исследовательской деятельности отечественных научных организаций в области наукометрии, но вполне может быть применена и к другим областям науки. На основании набора из трех компонент (актуальности, продуктивности и эффективности) в методике вычисляется интегральный показатель качества результатов научных исследований. После вычисления каждой из компонент качества результатов научных исследований происходит их преобразование к нечеткому виду. Первая из трех компонент получается на основании экспертных оценок, так как подразумевает оценку слабоформализуемых качественных признаков. Для получения остальных двух компонент используются обычные индикаторы, сравниваемые с целевыми значениями, а также функции вычисления среднего геометрического и средневзвешенного первичных показателей. Также нами предлагается простая рейтинговая модель, которая позволяет отбирать или поощрять отдельные организации и коллективы. Рассматриваемая рейтинговая модель строится на базе критерия сравнения рейтингов результатов научных исследований с некоторым пороговым значением.

На основе анализа предложенной методики можно сделать практические выводы о том, что в целях повышения качества результатов собственной деятельности научно-исследовательские институты и их научные коллективы должны стремиться выполнять планы исследовательских работ максимально четко, полно и детально, не упуская ничего существенного; избавляться от случайных или недобросовестных повторов, копирований или заимствований результатов, полученных другими исследователями; строго придерживаться требований к соблюдению тематических программ выполнения исследований; максимизировать количество опубликованных научных работ и объем опубликованных работ персонально по каждому из авторов в отдельности; получать в ходе исследований как можно больше конкретных и ценных результатов интеллектуальной деятельности, которые впоследствии будут возможно защитить авторскими правами. Используя предложенные рекомендации в качестве ориентиров, научно-исследовательским институтам и научным коллективам, возможно, удастся повысить качество результатов собственных исследований, а значит – открыть перед собой новые возможности для привлечения финансирования благодаря приобретению конкурентных преимуществ в борьбе за наиболее выгодные гранты и государственные задания.

СПИСОК ЛИТЕРАТУРЫ

1. Яковлева А.Ф. Об основных подходах к оценке результативности научных исследований в России // Измерение философии. Об основаниях и критериях оценки результативности философских и социо-

гуманитарных исследований / под. ред. А.В. Рубцова. – М., 2012. – С. 144-156.

2. Маркусова В.А., Рубвальтер Д.А., Либкинд А.Н. Оценки эффективности конкурсного финансирования: обзор библиометрических исследований // Научно-техническая информация. Сер. 1. – 2015. – № 5. – С. 23-29.
3. Гуреев В.Н., Мазов Н.А. Использование библиометрии для оценки значимости журналов в научных библиотеках (обзор) // Научно-техническая информация. Сер. 1. – 2015. – № 2. – С. 8-19.
4. Гиляревский Р.С. Публикационная активность как оценка научных достижений // Научно-техническая информация. Сер. 1. – 2014. – № 8. – С. 1-9.
5. Ежела В.В., Клименко С.В., Райков А.Н., Шарнин М.М. Семантический подход к оценке качества научных публикаций // Научно-техническая информация. Сер. 2. – 2015. – № 7. – С. 13-18.
6. Либкинд А.Н., Терехов А.И., Маркусова В.А., Рубвальтер Д.А., Либкинд И.А. Результаты выполнения конкурсных исследовательских проектов: библиометрия вклада различных групп ученых, организаций, городов, регионов и стран // Научно-техническая информация. Сер. 1. – 2015. – № 11. – С. 16-28.
7. Либкинд И.А. Определение научного уровня заданной совокупности публикаций // Научно-техническая информация. Сер. 2. – 2014. – № 8. – С. 1-9.
8. Wildgaard L., Schneider J.W., Larsen B. A review of the characteristics of 108 author-level bibliometric indicators. – URL: <https://arxiv.org/ftp/arxiv/papers/1408/1408.5700.pdf>.
9. Фейгельман М.В., Цирлина Г.А. Библиометрический азарт как следствие отсутствия научной экспертизы // Управление большими системами. – 2013. – № 44. – С. 332-345.
10. Соколов А.В. Форсайт: взгляд в будущее // Форсайт. – 2007. – № 1. – С. 8-15.
11. Найда А.А. Разработка методики количественного определения уровней компетентности экспертов // Межотраслевая информационная служба. – 2007. – № 1. – С. 19-29.
12. Мартынич С.А., Глухова Е.А., Галустова Л.Р. Требования к построению системы для оценки результативности и потенциала научной деятельности на уровне организации // Социальные аспекты здоровья населения. – 2013. – № 4. – С. 10.

Материал поступил в редакцию 06.12.16.

Сведения об авторах

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

Н.В. Харчевникова, В.Г. Блинова, Д.А. Добрынин, З.И. Жолдакова, В.И. Журко, Д.Ю. Федорцова

Интеллектуальная ДСМ-система по токсичности. Анализ данных по функциональной кумуляции химических соединений

Представлена структура интеллектуальной ДСМ-системы по токсикологии. Приведены результаты анализа данных по функциональной кумулятивной активности химических соединений, на основании которых предсказана степень активности неизученных веществ.

Ключевые слова: функциональная кумуляция, ДСМ-система, структурные гипотезы, степень активности

ВВЕДЕНИЕ

Понятие о функциональной кумуляции было разработано отечественными учеными в середине прошлого века. В отличие от материальной кумуляции, т.е. способности вещества накапливаться в организме при однократном, кратковременном и/или длительном поступлении, под функциональной кумуляцией в общем виде следует понимать «изменение токсичности ксенобиотиков во времени и, в частности, максимально достижимое увеличение токсичности (усиление действия) вещества в процессе острого, подострого и хронического воздействия» [1].

В соответствии с определением, функциональную кумуляцию изучают при разных длительностях и уровнях воздействия.

В зависимости от условий эксперимента степень кумуляции характеризуется разными методами и показателями, разработанными отечественными токсикологами [2]. При обосновании ПДК веществ в воде используется классификация, разработанная Г.Н. Красовским и модифицированная З.И. Жолдаковой [3], в которой класс кумуляции устанавливается по отношению среднесмертельной дозы острого опыта к пороговой дозе хронического эксперимента ($ЛД_{50}/ПД_{хр.}$). По величине отношения $ЛД_{50}/ПД_{хр.}$ вещества разделяются на 4 класса: первый – чрезвычайно опасные ($ЛД_{50}/ПД_{хр.} > 10^5$), второй – высоко опасные ($10^4 < ЛД_{50}/ПД_{хр.} < 10^5$), третий – умеренно опасные ($10^3 < ЛД_{50}/ПД_{хр.} < 10^4$) и четвертый – малоопасные ($ЛД_{50}/ПД_{хр.} < 10^3$).

Функциональная кумуляция – это сложный системный процесс, зависящий от дозы и длительности воздействия вещества и сопряженный с процессами адаптации и дезадаптации. Поэтому важной задачей является сопоставление материальной и функциональной кумуляции. Первым этапом должно быть

изучение зависимости способности к функциональной кумуляции от строения вещества. В отличие от материальной кумуляции, связь функциональной кумуляции со строением веществ ранее не изучалась.

Целью нашей работы является интеллектуальный анализ данных по функциональной кумуляции с использованием гетерогенной выборки веществ, принадлежащих к различным структурным рядам, и прогноз на этой основе степени способности к кумуляции неизученных веществ.

ОБЪЕКТЫ И МЕТОДЫ

Для интеллектуального анализа данных и прогноза биологической активности химических соединений на основе ДСМ-метода [4] создана экспериментальная система, которая включает следующие компоненты:

- модульный ДСМ-решатель, который состоит из генератора пересечений, генератора гипотез 1-го рода и генератора гипотез 2-го рода;
- автоматический кодировщик ФКСР, который в пакетном режиме преобразует химические структуры, заданные MOL файлами, в код ФКСР;
- функции определения сходства, которые реализуют операции пересечения, объединения и проверки включения над множествами кодов ФКСР;
- средства графической визуализации исходных данных и результатов эксперимента; интерфейсная оболочка, которая взаимодействует с пользователем и управляет процессом эксперимента.

Для проведения экспериментов по токсичности химических соединений создана новая версия проблемно ориентированного языка ФКСР [5] (фрагментарный код суперпозиций подструктур) в основу которого вошли новые центры, полученные из базы знаний ДСМ-экспериментов.

ДСМ-метод использовался для выделения структурных фрагментов молекул, ответственных за проявление функциональной кумуляции, а также для доказательства связи классов опасности со структурой веществ в пределах однородных рядов.

ДСМ-метод состоит из трех познавательных процедур: индукция – порождение причин эффектов на основе обнаруженных сходств фактов; аналогия – правдоподобные выводы, использующие наличие положительных и отрицательных причин в фактах с неопределенной оценкой, требующей уточнения, то есть прогноза данного свойства; абдукция – принятие гипотез для объяснения начального состояния базы фактов с помощью положительных и отрицательных гипотез. Объектами исследования являются структуры химических соединений, представленные в виде ФКСП-кодов.

Свойством в данном случае является биологическая активность химических веществ. Сходство – операция пересечения ФКСП-кодов веществ (т.е. числовых множеств дескрипторов), результатом которой являются фрагменты структур химического соединения. Применением правил правдоподобного вывода ДСМ-метод порождает положительные гипотезы, содержащие фрагменты структур, отвечающие за проявление биологической активности, и отрицательные гипотезы, наличие которых приводит к снижению либо отсутствию биологической активности. Далее выполняется процедура применения правил правдоподобного вывода для прогноза соединений, активность которых неизвестна. Сгенерированные ДСМ-методом структурные гипотезы имеют «родителей», т.е. ряд соединений, обладающих этим общим структурным фрагментом и относящихся к одному классу по активности. Активность неизученного соединения, в структуру молекулы которого входит такой структурный фрагмент, может быть предсказана. Т.е. родители определенной структурной гипотезы образуют структурный ряд, выделенный из гетерогенной обучающей выборки. Прогноз неизученных соединений возможен для тех из них, которые содержат в молекуле структуру, совпадающую с гипотезой, выделенной программой, что позволяет подбирать для прогноза соединения, относящиеся к вполне определенному структурному ряду.

КОМПЬЮТЕРНЫЕ ЭКСПЕРИМЕНТЫ И ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ.

В данной работе обучающая выборка (583 соединения) была автоматически выделена из базы данных WATERTOХ, содержащей значения показателей острой и хронической токсичности для более 2000 химических веществ. В выборку не были включены металлы, элементарноорганические соединения, а также соединения, для которых в базе нет информации по острой или хронической токсичности. Для каждого соединения выборки программой был определен класс опасности по величине $LD_{50}/PD_{хр.}$. Выборка содержала 43 соединения первого класса, 78 соединений второго класса, 206 соединений третьего класса и 256 соединений четвертого класса. При проведении компьютерных экспериментов соединения 1 и 2 классов были объединены в один класс. Была проведена серия компьютерных экспериментов с целью

выделить структурные фрагменты, определяющие отнесение соединения к одному из классов. В первом эксперименте соединения характеризовали двумя «свойствами»: первое свойство – быть либо чрезвычайно или высоко кумулятивным, либо умеренно кумулятивным [+ 0] и второе свойство – быть либо умеренно, либо малокумулятивным [0 -]. Здесь 0 – либо «+», либо «-».

О возможности прогноза кумулятивности с использованием сгенерированных гипотез судили по результатам скользящего контроля – каждое соединение из выборки последовательно делали тестовым – осуществлялся прогноз класса для этого соединения с использованием структурных гипотез, сгенерированных без участия этого соединения. Результат прогноза сравнивали с исходным классом опасности этого соединения. С последовательным применением двух вариантов метода – более строгого «с запретом на контрпримеры» и менее строгого «без запрета на контрпримеры» прогноз оказался возможным для около 70% соединений. Применение менее строгого метода дает возможность расширить число веществ, для которых возможен прогноз, однако возрастает и число ошибочных прогнозов, поэтому необходим анализ сгенерированных гипотез, с целью удаления неинформативных гипотез. По результатам эксперимента по двум свойствам, не удалось выделить гипотезы, однозначно определяющие отнесение соединения к классу умеренно опасных. Для большого числа соединений класс опасности был предсказан не полностью, было определено только одно свойство, что позволяло выделить структурные фрагменты, определяющие отнесение соединения к 1) высокоопасным либо умеренно опасным и 2) к умеренно либо малоопасным.

С учетом этой неопределенности точность, оцененная по скользящему контролю, составила 78%. В табл. 1 приведены гипотезы (структурные фрагменты), присутствие которых в молекулах соединений определяет их отнесение к классам высоко кумулятивных и мало кумулятивных. К высоко кумулятивным относятся соединения ряда этиленгликоля и лекарственных вещества – антибиотики, содержащие четырехчленный цикл, который образуют три атома углерода и атом азота, производные индола.

Необходимо отметить, что высокая кумуляция не всегда свидетельствует об опасности вещества, так как значение показателя может определяться не только низкой пороговой дозой хронического эксперимента, но и низкой острой токсичностью. Этот факт требует своего уточнения и объяснения.

В качестве структурных гипотез для малокумулятивных веществ выделены, в частности, пятичленные циклы, входящие в состав замещенных фталимидов, алициклические азотсодержащие структуры, третбутильная группа, структура бензолсульфокислоты. Если для неизученного в токсикологическом плане соединения предсказывается малая кумулятивность по критерию величины отношения $LD_{50}/PD_{хр.}$, это означает, что вещество не более опасно при хроническом воздействии, чем в остром опыте. Поэтому оно не должно попасть в список приоритетных веществ для изучения в очень трудоемком и длительном хроническом эксперименте.

Гипотезы (структурные фрагменты), определяющие отнесение соединения к классам высоко кумулятивных и мало кумулятивных соединений

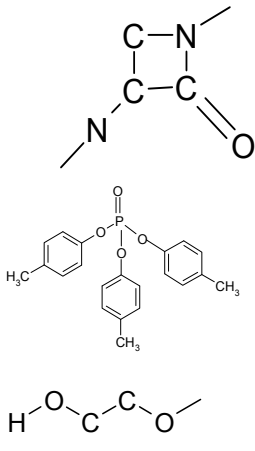
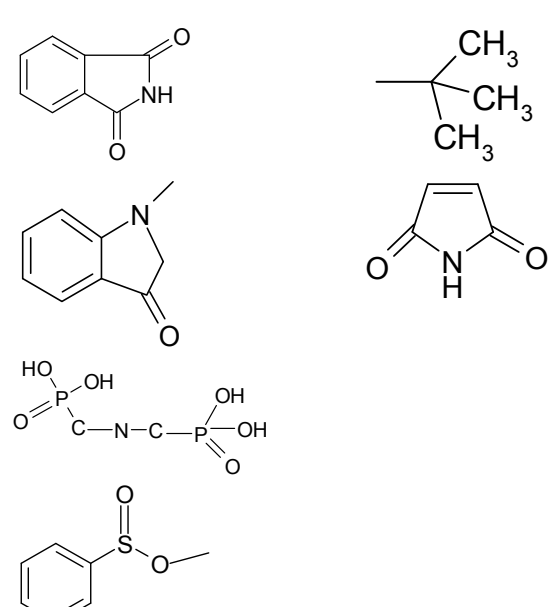
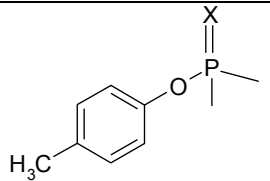
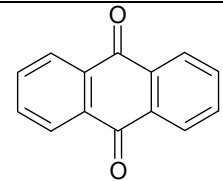
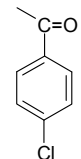
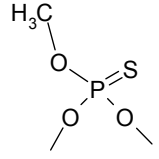
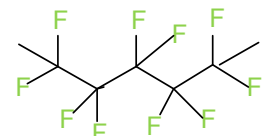
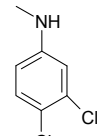
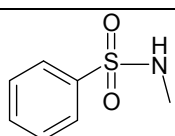
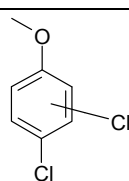
Гипотезы, определяющие принадлежность соединения к классу чрезвычайно и высоко опасных соединений по способности к кумуляции (+ +)	Гипотезы, определяющие принадлежность соединения к малоопасным по способности к кумуляции (- -)
	

Таблица 2

Гипотезы (структурные фрагменты), определяющие отнесение соединения к классам высоко опасных либо умеренно опасных соединений по способности к кумуляции и гипотезы, определяющие принадлежность соединения к умеренно опасным либо малоопасным по способности к кумуляции

Гипотезы, определяющие принадлежность соединения к классам высоко опасных либо умеренно опасных соединений по способности к кумуляции (+ 0). Здесь 0 – либо плюс, либо минус.	Гипотезы, определяющие принадлежность соединения к умеренно опасным либо малоопасным – по способности к кумуляции (0 -) Здесь 0 – либо плюс, либо минус
	
	
	
	
	S=P-OCH ₂ -

В табл. 2 приведены гипотезы (структурные фрагменты), определяющие отнесение соединения к классам высоко опасных либо умеренно опасных по способности к кумуляции и гипотезы, определяющие принадлежность соединения к умеренно опасным либо малоопасным.

Среди гипотез, определяющих принадлежность соединений к умеренно либо мало кумулятивным, выделены, например, структуры антрахинона, хлорзамещенные бензальдегиды и бензойные кислоты, хлорзамещенные анилины.

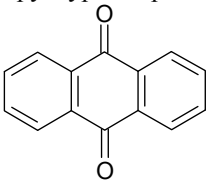
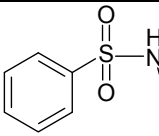
Для содержащихся в базе данных соединений, для которых не был проведен хронический эксперимент, с использованием ДСМ-метода был проведен прогноз. Так, согласно результатам прогноза, среди производных антрахинона соединения антрахинон 1-амино-2,4-дибромантрахинон и 1-хлорантрахинон относятся к 3 классу по кумулятивности, 1,2-дигидроксиантрахинон – к 4 классу. В результате эксперимента выделены несколько гипотез – структурных фрагментов, содержащих атом фосфора. В выборке содержалось 55 фосфорсодержащих соединений. Почти все эти соединения принадлежат к 3 либо 4 классам по кумулятивности. К 1 классу по кумулятивности принадлежат только два соединения, в состав которых входят фосфорсодержащие фрагменты

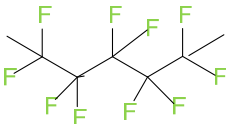
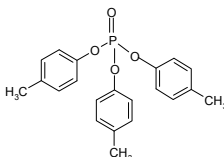
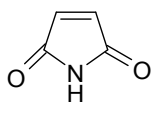
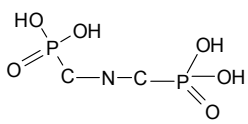
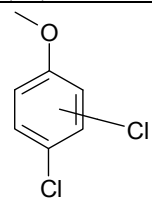
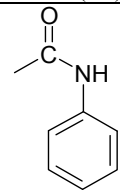
с метил-замещенным фенильным радикалом (табл. 1), а также соединение кислота ди(2-этилгексил)-дитиофосфорная. Остальные тио и дитиофосфорсодержащие соединения относятся к 3 и 4 классам. С использованием выделенных гипотез сделан прогноз для фосфорсодержащих соединений, для которых не был проведен хронический эксперимент. Соединение О,О-диметил-О-(3-метилтиофенил)тиофосфат (сульфидофос, байтекс) было отнесено к 3-ему классу, поскольку оно содержит фрагмент с метил-замещенным фенильным радикалом. Помимо этой «положительной» гипотезы по первому свойству, были выделены отрицательные гипотезы по второму свойству – структура метильного эфира тиофосфорной кислоты (табл. 2). Соединение О-этилдихлортиофосфат было отнесено к 4 классу на основании присутствия «отрицательной» гипотезы $S=P-OCH_2-$ (8 соединений «родителей»).

Была выделена гипотеза – структурный фрагмент, содержащий бензолсульфаниламидную группу (табл. 2). В базе данных присутствуют соединения, в состав которых входит эта группа, но нет результатов хронического эксперимента. Нами был осуществлен прогноз кумулятивности этих соединений. По результатам прогноза сульгин и сульфадимезин принадлежат к 3 классу по кумулятивности (табл. 3).

Таблица 3

Результаты прогноза класса опасности по способности к кумуляции

Гипотеза	Количество родителей	Результаты скользящего контроля		Прогнозируемые соединения	Результаты прогноза
		Количество соединений-родителей, правильно отнесенных к определенному классу	Количество соединений-родителей, правильно отнесенных к одному из двух соседних классов		
структура антрахинона  (0 -)	12	6	6	1-аминоантрахинон	(+ -)
				1-амино-2,4-дибромантрахинон	(+ -)
				1-хлорантрахинон	(+ -)
				1,5-дихлорантрахинон	(+ -)
				1,2-дигидроксиантрахинон	(- -)
 бензолсульфаниамиды (+ 0)	5	2	2	4-амино-N-(аминоимино-метил)бензолсульфонамид (сульгин)	(+ -)
				4-амино-N-(4,6-диметил-2-пиридинил)бензолсульфонамид (сульфадимезин)	(+ -)

Гипотеза	Количество родителей	Результаты скользящего контроля		Прогнозируемые соединения	Результаты прогноза
		Количество соединений-родителей, правильно отнесенных к определенному классу	Количество соединений-родителей, правильно отнесенных к одному из двух соседних классов		
 (+ -) перфторзамещенные алифатические соединения	5	5	-	1,1-дигидроперфторгептиловый спирт	(+ -)
				1,1-дигидроперфторгептилакрилат	(+ -)
 (+ +) метилфениловые эфиры фосфорной кислоты	3	3	-	три(триметилфенил)фосфат	(+ +)
 (- -) фталимиды	7	7	-	гексимид	(- -)
 (- -)	3	3	-	2-оксипропилен-1,3-диамино-N,N,N',N'-тетраметилфосфоновая кислота	(- -)
S=P-OCH ₂ - (0 -)	5	3	2	этилпаратион	(+ -)
 (0 -)	9	7	2	2,4,5- трихлорфеноксиэтил-α,α - трихлорпропионат (пентанат)	(+ -)
				2,4,5- трихлорфеноксиэтил-α,α - трихлорацетат (гексанат)	(- -)
 (0 -)	12	12	-	дифенилмочевина	(- -)
				фенилметилмочевина	(- -)

Также был сделан прогноз кумулятивности фталимида гексимида, перфторалифатических соединений и производных замещенной фосфорной кислоты. Необходимо отметить, что гипотезы, определяющие отнесение к определенному классу по кумулятивности обычно имеют ограниченное количество родителей (обычно два – три), тогда как гипотезы, определяющие отнесение соединения к одному из соседних классов, позволяют выделить довольно большой ряд структурно родственных соединений.

ЗАКЛЮЧЕНИЕ

В целом ДСМ метод позволил с удовлетворительной точностью выделить структурные фрагменты, наличие которых в молекуле ксенобиотика определяет высокую, умеренную и низкую способность вещества к функциональной кумуляции. Показана связь классов опасности по способности к кумуляции со структурой веществ в пределах однородных структурных рядов и возможность прогноза классов опасности по этому показателю для неизученных соединений.

Оценка степени способности к функциональной кумуляции позволяет в частности выделить вещества, для которых при предварительной оценке опасности можно ограничиться определением показателя острой токсичности и не проводить очень трудоемкий и длительный хронический эксперимент.

СПИСОК ЛИТЕРАТУРЫ

1. Штабский Б.М. Учение о кумуляции и его применение в профилактической токсикологии (Часть 1) // Актуальные проблемы транспортной медицины. – 2013. - № 2 (32). – С. 8-20.
2. Каган Ю.С., Красовский Г.Н., Штабский Б.М. Кумулятивные свойства химических соединений, их изучение и оценка. // Токсикометрия химических веществ, загрязняющих окружающую среду / под ред. А.А. Каспарова, И.В. Саноцкого. – М.: Центр международных проектов ГКНТ, 1986 г. – С. 104-133.
3. Методические указания МУ 2.1.5.720-98 “Обоснование гигиенических нормативов химических веществ в воде водных объектов хозяйственно-питьевого и культурно-бытового водопользования”. – М.: Федеральный центр Госсанэпиднадзора РФ. – 1999. – 55 с.
4. Финн В.К. Правдоподобные рассуждения в интеллектуальных системах типа ДСМ // Автоматическое порождение гипотез в интеллектуальных системах / под ред. В.К. Финна – М.: Книжный дом «ЛИБРОКОМ», 2009. – С. 10-49.
5. Ольшанский Д.Л., Добрынин Д.А., Блинова В.Г. Автоматическое кодирование химических соединений фрагментарным кодом супер-

позиций подструктур // Научно-техническая информация. Сер. 2 – 2015. – № 5. – С. 7-18.

6. Харчевникова Н.В., Блинова В.Г., Добрынин Д.А., Жолдакова З.И. Интеллектуальный анализ канцерогенности химических соединений с помощью стратегий метода, основанного на логике Джона Стюарта Милля // Научно-техническая информация. Сер. 2. – 2009. – №11. – С.18-23.
7. Харчевникова Н.В., Блинова В.Г., Добрынин Д.А., Максин М.В., Жолдакова З.И. Применение ДСМ метода и квантовомеханических расчетов для прогноза канцерогенности и хронической токсичности галогензамещенных алифатических углеводородов // Автоматическое порождение гипотез в интеллектуальных системах. / под ред. В.К. Финна – М.: Книжный дом «ЛИБРОКОМ», 2009. – С.351-378.

Материал поступил в редакцию 22.11.16.

Сведения об авторах

ХАРЧЕВНИКОВА Нина Вениаминовна – доктор биологических наук, ведущий научный сотрудник ФГБУ «НИИ экологии человека и гигиены окружающей среды им. А.Н.Сысина» Минздрава РФ, Москва
e-mail: kharchevnikova_n@mail.ru

БЛИНОВА Валентина Георгиевна – кандидат химических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН (ФИЦ ИУ РАН), Москва
e-mail: Blinova.V-Blinova@yandex.ru

ДОБРЫНИН Дмитрий Анатольевич – кандидат технических наук, старший научный сотрудник ФИЦ ИУ РАН
e-mail: minirobot@yandex.ru

ЖОЛДАКОВА Зоя Ильинична – доктор медицинских наук, профессор, ведущий научный сотрудник ФГБУ «НИИ экологии человека и гигиены окружающей среды им. А.Н. Сысина» Минздрава РФ
e-mail: brean38@mail.ru

ФЕДОРЦОВА Дарья Юрьевна – младший научный сотрудник ФГБУ «НИИ экологии человека и гигиены окружающей среды им. А.Н. Сысина» Минздрава РФ
e-mail: kirah1@yandex.ru

ЖУРКО Владимир Иванович – ведущий инженер ФГБУ «НИИ экологии человека и гигиены окружающей среды им. А.Н. Сысина» Минздрава РФ
e-mail: vladimir_zhurko@mail.ru

Предельная вероятность порождения случайного сходства при наличии контр-примеров

Ранее была найдена производящая функция возникновения случайного сходства с числом b примеров-родителей при t контр-примерах [1]. Здесь мы ограничиваемся случаем $b = 2$ и одинаковой вероятностью успеха p в испытаниях Бернулли для каждого атрибута каждого контр-примера и с вероятностью успеха p^2 для каждого атрибута в случайном сходстве. Если число n атрибутов устремить к бесконечности, вероятность успеха задать как $p = \sqrt{a/n}$ и рассмотреть $t = b \cdot \sqrt{n}$ контр-примеров, то вероятность возникновения случайного сходства, избегающего эти t контр-примеров, будет стремиться к $1 - e^{-a} - a \cdot e^{-a} \cdot [1 - e^{-b\sqrt{a}}]$.

Ключевые слова: испытания Бернулли, случайное сходство, контр-пример

ВВЕДЕНИЕ

Рассмотрим модель формирования множества контр-примеров и обучающей выборки, основанную на независимых последовательностях испытаний Бернулли. В этой работе число n обозначает количество несущественных (не входящих ни в какую реальную причину) атрибутов, которыми мы ограничиваемся. Для каждого контр-примера или элемента обучающей выборки образуем последовательность испытаний Бернулли с вероятностью успеха p , причем последовательности для разных объектов – независимы. Число t будет равно числу контр-примеров. Так как число родителей случайного сходства будет равно 2, Лемма 1 из работы [1] (воспроизведенная здесь) позволяет строить его как одну последовательность испытаний Бернулли с вероятностью успеха p^2 .

Теорема 1 из работы [1] оценивала вероятность того, что при наличии t контр-примеров возникнет случайное сходство, которое не устранилось ни одним из этих контр-примеров, по формуле $\sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot (1 - p^2 - p^{2+j})^n$. В настоящей статье мы докажем это равенство другим способом.

Основной результат настоящей статьи утверждает, что, если число n атрибутов устремить к бесконечности, вероятность успеха задать как $p = \sqrt{a/n}$ и рассмотреть $t = b \cdot \sqrt{n}$ контр-примеров, то вероятность возникновения случайного сходства, избегающего эти t контр-примеров, будет стремиться к $1 - e^{-a} - a \cdot e^{-a} \cdot [1 - e^{-b\sqrt{a}}]$.

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Определение 1. *Выборкой* обучающих примеров называется множество объектов O , каждый из которых описывается подмножеством множества атрибутов F . Отношение «объект $o \in O$ имеет атрибут $f \in F$ » задается как обычно для бинарных отношений в виде oEf .

Определение 2. *Пересечением* подмножества $A \subseteq O$ объектов назовем подмножество $A' = \{f \in F : \forall o \in A [oEf]\}$. Заметим, что $\emptyset' = F$. *Пересечением* подмножества $B \subseteq F$ атрибутов назовем подмножество $B' = \{o \in O : \forall f \in B [oEf]\}$. В этом случае $\emptyset' = O$.

Определение 3. *Сходством* называется всякая пара $\langle A, B \rangle$, состоящая из подмножества (называемого *списком родителей*) объектов $A \subseteq O$ и подмножества атрибутов $B \subseteq F$ (называемого *пересечением объектов*), удовлетворяющая условиям $A' = B$ и $B' = A$ (последнее условие называется *исчерпываемостью*). Дополнительно требуют $|A| \geq 2$.

Определение 4. *Последовательность испытаний Бернулли* – это распределение вероятностей на $\{0,1\}^n$ с $P(f_1 = \sigma_1, \dots, f_n = \sigma_n) = \prod_{j=1}^n p^{\sigma_j} \cdot (1-p)^{1-\sigma_j}$, где $\sigma_j \in \{0,1\}$ и $0 < p < 1$. Число $0 < p < 1$ называется вероятностью успеха $f_j = 1$ в j -ом испытании. В настоящей статье контр-примеры и элементы обу-

чающей выборки соответствуют независимым реализациям элементов этого вероятностного пространства.

Определение 5. Контр-пример $Z = \{j : f_j = 1\}$ для $(f_1, \dots, f_n) \in \{0, 1\}^n$ **устраняет** случайное сходство $X = \{j : a_j = 1\}$ для $(a_1, \dots, a_n) \in \{0, 1\}^n$, если $X \subseteq Z$.

ВСПОМОГАТЕЛЬНЫЕ УТВЕРЖДЕНИЯ

Лемма 1. Случайное сходство $X = \{j : a_j = 1\}$ порождается последовательностью испытаний Бернулли $(a_1, \dots, a_n) \in \{0, 1\}^n$ с вероятностью успеха p^2 .

Доказательство следует из условия независимости обучающих примеров и Определения 2.

Теорема 1. Вероятность того, что возникнет случайное сходство, которое не устранился ни одним из m контр-примеров, равна

$$\sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot (1 - p^2 - p^{2+j})^n.$$

Доказательство. Пусть $Z_1, \dots, Z_m$ – контр-примеры – независимые последовательности испытаний Бернулли с параметрами n, p . Пусть X – независимая от контр-примеров последовательность испытаний Бернулли с параметрами n, p^2 , соответствующая по Лемме 1 случайному сходству.

Зафиксируем $j = |X|$ – мощность случайного сходства. Тогда искомая вероятность равна $\sum_{l=0}^n \binom{n}{l} (p^2)^l (1 - p^2)^{n-l} (1 - p^l)^m$, так как $(1 - p^l)$ равна вероятности, что $X \not\subseteq Z_k$ для фиксированного контр-примера Z_k , поэтому $(1 - p^l)^m$ – вероятность $\forall k [X \not\subseteq Z_k]$, а $\binom{n}{l} (p^2)^l (1 - p^2)^{n-l}$ – вероятность порождения сходства мощности l .

Но

$$\begin{aligned} \sum_{l=0}^n \binom{n}{l} (p^2)^l (1 - p^2)^{n-l} (1 - p^l)^m &= \\ &= \sum_{l=0}^n \binom{n}{l} (p^2)^l (1 - p^2)^{n-l} \left(\sum_{j=0}^m \binom{m}{j} (-1)^j p^{lj} \right) = \\ &= \left(\sum_{j=0}^m \binom{m}{j} (-1)^j \right) \left(\sum_{l=0}^n \binom{n}{l} (p^{2+j})^l (1 - p^2)^{n-l} \right) = \\ &= \sum_{j=0}^m \binom{m}{j} (-1)^j (1 - p^2 + p^{2+j})^n. \end{aligned}$$

Замечание 1. Результат Теоремы 1 легко получить с помощью принципа «включение-исключение».

Лемма 2. Для любой константы C верно

$$\sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot C = 0.$$

Доказательство следует из бинома Ньютона

$$0 = (1 - 1)^m = \sum_{j=0}^m \binom{m}{j} \cdot (-1)^j.$$

Лемма 3. Для натурального $j > 0$ при $n \rightarrow \infty$ имеем разложение

$$\left(1 - \frac{a}{n} + \left(\frac{a}{n} \right)^{1+\frac{j}{2}} \right)^n - \left(1 - \frac{a}{n} \right)^n = a \cdot e^{-a} \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} + o \left(n^{-\frac{j}{2}} \right).$$

Доказательство.

$$\begin{aligned} \left(1 - \frac{a}{n} + \left(\frac{a}{n} \right)^{1+\frac{j}{2}} \right)^n - \left(1 - \frac{a}{n} \right)^n &= \left(1 - \frac{a}{n} \cdot \left[1 - \left(\frac{a}{n} \right)^{\frac{1}{2}} \right] + \frac{n \cdot (n-1)}{2} \cdot \left(\frac{a}{n} \right)^2 \cdot \right. \\ &\quad \left. \cdot \left[1 - 2 \cdot \left(\frac{a}{n} \right)^{\frac{1}{2}} + o \left(n^{-\frac{1}{2}} \right) \right] + \dots + \right. \\ &\quad \left. - \left(1 - \frac{a}{n} \right)^n = \left[\frac{n \cdot \dots \cdot (n-k+1)}{k!} \cdot (-1)^{k-1} \cdot \left(\frac{a}{n} \right)^k \cdot \right. \right. \\ &\quad \left. \cdot \left[1 + k \cdot \left(\frac{a}{n} \right)^{\frac{1}{2}} + o \left(n^{-\frac{1}{2}} \right) \right] + \frac{n!}{n!} \cdot (-1)^{n-1} \cdot \right. \\ &\quad \left. \cdot \left(\frac{a}{n} \right)^n \cdot \left[1 + n \cdot \left(\frac{a}{n} \right)^{\frac{1}{2}} + o \left(n^{-\frac{1}{2}} \right) \right] \right] - \\ &\quad - \left[1 - n \cdot \frac{a}{n} + \frac{n \cdot (n-1)}{2} \cdot \left(\frac{a}{n} \right)^2 + \dots + \frac{n \cdot \dots \cdot (n-k+1)}{k!} \cdot \right. \\ &\quad \left. \cdot (-1)^{k-1} \cdot \left(\frac{a}{n} \right)^k + \dots + \frac{n!}{n!} \cdot \left(\frac{a}{n} \right)^n \right] = \\ &= n \cdot \frac{a}{n} \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} - \frac{n \cdot (n-1)}{2} \cdot \left(\frac{a}{n} \right)^2 \cdot \left[2 \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} + o \left(n^{-\frac{1}{2}} \right) \right] + \dots + \\ &\quad + \frac{n \cdot \dots \cdot (n-k+1)}{k!} \cdot (-1)^{k-1} \cdot \left(\frac{a}{n} \right)^k \cdot \left[k \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} + o \left(n^{-\frac{1}{2}} \right) \right] + \dots + \\ &\quad + \frac{n!}{n!} \cdot (-1)^{n-1} \cdot \left(\frac{a}{n} \right)^n \cdot \left[n \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} + o \left(n^{-\frac{1}{2}} \right) \right] = a \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} - \\ &\quad - a^2 \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} + \dots + \frac{a^k}{(k-1)!} \cdot (-1)^{k-1} \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} + \dots + \\ &\quad + \frac{a^n}{(n-1)!} \cdot (-1)^{n-1} \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} + o \left(n^{-\frac{1}{2}} \right) = \\ &= a \cdot e^{-a} \cdot \left(\frac{a}{n} \right)^{\frac{j}{2}} + o \left(n^{-\frac{j}{2}} \right). \end{aligned}$$

Лемма 4. При $n \rightarrow \infty$ имеем разложение

$$1 - \left(1 - \frac{a}{n}\right)^n = 1 - e^{-a} + o(1).$$

Доказательство аналогично предыдущему, но проще и общеизвестно.

ОСНОВНОЙ РЕЗУЛЬТАТ

Теорема 2. При числе атрибутов $n \rightarrow \infty$ и вероятности появления атрибута у контр-примеров и обучающих примеров равной $p = \sqrt{a/n}$, вероятность возникновения случайного сходства, не устраненного никаким из контр-примеров, будет стремиться к $1 - e^{-a} - a \cdot e^{-a} \cdot [1 - e^{-b\sqrt{a}}]$.

Доказательство. По Теореме 1 нужно оценить

$$\begin{aligned} & \sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot (1 - p^2 + p^{2+j})^n = \\ & = \sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot \left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+\frac{j}{2}}\right)^n = \\ & = \sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot \left[\left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+\frac{j}{2}}\right)^n - \left(1 - \frac{a}{n}\right)^n\right] \end{aligned}$$

по Лемме 2.

$$\begin{aligned} & \sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot \left[\left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+\frac{j}{2}}\right)^n - \left(1 - \frac{a}{n}\right)^n\right] = \\ & = \left[1 - \left(1 - \frac{a}{n}\right)^n\right] + \\ & + \sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot \left[\left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+\frac{j}{2}}\right)^n - \left(1 - \frac{a}{n}\right)^n\right] = \\ & = [1 - e^{-a} + o(1)] + \\ & + \sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot \left[a \cdot e^{-a} \cdot \left(\frac{a}{n}\right)^{\frac{j}{2}} + o\left(n^{-\frac{1}{2}}\right)\right] \end{aligned}$$

выводится из Лемм 3 и 4. Оценка

$$\begin{aligned} & \sum_{j=1}^m \binom{m}{j} \cdot (-1)^j \cdot \left[a \cdot e^{-a} \cdot \left(\frac{a}{n}\right)^{\frac{j}{2}} + o\left(n^{-\frac{1}{2}}\right)\right] = \\ & = \sum_{j=1}^m \binom{m}{j} \cdot (-1)^j \cdot \left[a \cdot e^{-a} \cdot \left(\frac{b \cdot \sqrt{a}}{m}\right)^j + o(m^{-j})\right] = \\ & = a \cdot e^{-a} \cdot \left[\left(1 - \frac{b \cdot \sqrt{a}}{m}\right)^m - 1\right] + o(1) = \\ & = a \cdot e^{-a} \cdot [e^{-b\sqrt{a}} - 1] + o(1), \end{aligned}$$

где последнее равенство следует из Леммы 2 и завершает доказательство.

ЗАКЛЮЧЕНИЕ

В работе получены асимптотические формулы при $n \rightarrow \infty$ для вероятности возникновения случайного сходства, избегающего $m = b \cdot \sqrt{n}$ случайных контр-примеров, если наличие атрибута как в обучающем примере, так и в контр-примере возникает как испытание Бернулли с вероятностью успеха $p = \sqrt{a/n}$. Эту формулу можно использовать как оценку среднего числа «случайных гипотез», так как по обучающей выборке и множеству контр-примеров можно оценить вероятность появления каждого признака (как случайную величину Бернулли), а число признаков и контр-примеров известно.

Изложенные в работе результаты в равной степени применимы как к ДСМ-методу [2] автоматического порождения гипотез, так и к Анализ Формальных Понятий (АФП) (расширенному добавлением контр-примеров) [3], и к ВКФ-методу, развиваемому автором [4]. Более того, автор полагает, что именно из-за неизбежного появления случайных сходств исчерпывающая процедура вычисления всех сходств, лежащая в основе ДСМ-метода и АФП, приводит к переобучению. ВКФ-метод имеет более гибкие механизмы управления множеством порожденных гипотез и может ограничиваться из-за своего вероятностного характера более робастными моделями.

* * *

Автор благодарит профессора Е.М. Бениаминова и сотрудников лаборатории 35 Федерального исследовательского центра «Информатика и управление» РАН за полезные обсуждения.

СПИСОК ЛИТЕРАТУРЫ

1. Виноградов Д.В. Вероятность порождения случайного ДСМ-сходства при наличии контр-примеров // Научно-техническая информация. Сер. 2. – 2015. – № 3. – С. 1-5; Vinogradov D.V. The Probability of Encountering an Accidental DSM Similarity in the Presence // Automatic Documentation and Mathematical Linguistics. – 2015. – Vol. 49, №2. – P. 43-46.
2. Финн В.К. О машинно-ориентированной формализации правдоподобных рассуждений в стиле Ф. Бэкона – Д.С.Милля // Семиотика и информатика. – 1983. – Вып. 20. – С. 35-101.
3. Ganter B., Kuznetsov S.O. Formalizing Hypotheses with Concepts // Proc. 8th International Conference on Conceptual Structures ICCS 2001 / eds. G. Mineau, B. Ganter (Lecture Notes in Artificial Intelligence. Springer). – 2000. – Vol. 1867. – P. 342-356.
4. Vinogradov D.V. VKF-method of Hypotheses Generation // Analysis of Images, Social networks and Texts / eds. D.I. Ignatov et al. (Communications in Computer and Information Science. Springer). – 2014. – Vol. 436. – P. 237-248

Материал поступил в редакцию 24.11.16.

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

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 025.43.01 : 3

А.М. Галиева, А.В. Кириллович, Н.В. Лукашевич, О.А. Невзорова, Д.Ш. Сулейманов

Создание русско-татарского тезауруса по общественно-политической тематике: общие принципы и аспекты реализации*

Обсуждаются основные принципы и аспекты реализации нового лексического ресурса – русско-татарского тезауруса по общественно-политической тематике, разработанного на основе формата тезауруса РуТез. Двухязычный тезаурус имеет в целом концептуальную структуру тезауруса РуТез, однако допускаются структурные модификации, связанные с добавлением как единичных концептов на какой-либо ветви, так и добавление новых тематических ветвей, отражающих национальную специфику татарского языка. Обсуждаются основные способы отображения специфики лексико-семантической системы татарского языка при переводе, общая методология перевода названий концептов, формулируются основные требования к переводу имен концептов и их текстовых входов.

Ключевые слова: тезаурус, концепт, терминосистема, русский язык, татарский язык, общественно-политическая область

ВВЕДЕНИЕ

Разработка терминосистемы татарского языка с позиций системной организации лексики – актуальная проблема, пока не решенная для татарского языка в объеме, достаточном для применения в компьютерных технологиях.

В настоящее время применяют несколько способов моделирования терминосистем. Тезаурусное моделирование – один из главных подходов к проблеме. Любой тезаурус отражает логико-семантическую структуру терминологии, это позволяет использовать формальные языки для системного описания терминов, что в свою очередь обеспечивает применение тезаурусов для задач автоматической обработки текстов.

Исследования понятийной и терминологической систем русского языка выполнены в ряде проектов по разработке тезаурусных моделей лексики русского языка (проекты семантический словарь под ред. Н.Ю. Шведовой, Большой толковый словарь русских существительных под ред. Л.Г. Бабенко и др.). Предполагается использовать накопленный опыт в русском языке для формирования модели терминосистемы татарского языка.

В настоящей статье обсуждаются основные проектные решения, принятые при разработке русско-татарского тезауруса в общественно-политической области.

ОСОБЕННОСТИ ОБЩЕСТВЕННО-ПОЛИТИЧЕСКОЙ ОБЛАСТИ. ОБЩЕСТВЕННО-ПОЛИТИЧЕСКИЙ ТЕЗАУРУС

Общественно-политическая область – это широкая область современных общественных отношений, знания в которой входят в компетенцию образованных людей-неспециалистов [1, 2]. Эта область охватывает политику и международные отношения, экономику и финансы, промышленность, военные отношения, искусство, спорт и др.

Наибольшее проявление общественно-политической области находит в публикациях средств массовой информации общей направленности, которые включают обсуждения в рамках большого количества конкретных предметных областей, содержат большое количество терминов (*уголовное дело, инвестор, инфляция, федеральный закон*, и т.п.), и при этом направлены на непрофессионалов и понятны непрофессионалам. Отметим, что автоматическая обработка новостных сообщений относится к одному из самых востребованных направлений автоматической обработки текстов.

* Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 16-18-02074 «Разработка моделей связывания терминологии в разных языках (на материале русского и татарского языков)»).

Таким образом, существует совокупность терминов различных предметных областей, которые одновременно являются элементами лексической системы непрофессионалов. Существование такой терминологии не случайно. Непрофессионалы контактируют с профессионалами в повседневной жизни и должны владеть соответствующей терминологией.

Отметим, что среди известных информационно-поисковых тезаурусов некоторые можно отнести к тезаурусам в общественно-политической области, включая тезаурус Европейского союза EuroVoc, тезаурус ООН (UNBIS), тезаурус исследовательской службы конгресса США (LIV, 1994).

Для русского языка создан Общественно-политический тезаурус для автоматической обработки текстов [1, 2]. В настоящее время он частично опубликован, в составе тезауруса русского языка РуТез¹.

Создание тезауруса в общественно-политической области для некоторого языка имеет несколько важных преимуществ.

Во-первых, общественно-политическая область – это средняя зона между абстрактной лексикой, которая может встретиться в любой предметной области (Общий лексикон), часто очень многозначной и сложной для описания, и конкретными предметными областями (рис. 1).

Многозначность лексических единиц терминов в общественной области значительно ниже, чем в зоне Общего лексикона.

Во-вторых, общественно-политический тезаурус допускает расширение в любую профессиональную предметную область при возникновении соответствующей прикладной задачи. В-третьих, как ресурс его можно встраивать в различные приложения, связанные с обработкой текстов, например, мониторинг СМИ и социальных медиа, поиск и анализ законодательных актов и др.

Таким образом, создание тезауруса в общественно-политической области для некоторого языка является важной основой развития формализованного описания лексической системы и различных терминологических подсистем языка (рис. 2).

Именно так и произошло при построении системы тезаурусов для автоматической обработки текстов для русского языка: сначала появился Общественно-политический тезаурус, затем стали появляться тезаурусы в конкретных предметных областях, например, Авиа-Онтология [3] или Онтология по естественным наукам и технологиям – ОЕНТ [4], а также в результате дополнения надстройкой общего лексикона появился тезаурус русского языка – РуТез [1]. Поэтому важно начать создание тезаурусов для татарского языка именно с тезауруса в общественно-политической области.



Рис. 1. Положение общественно-политической области относительно специальных областей [1]

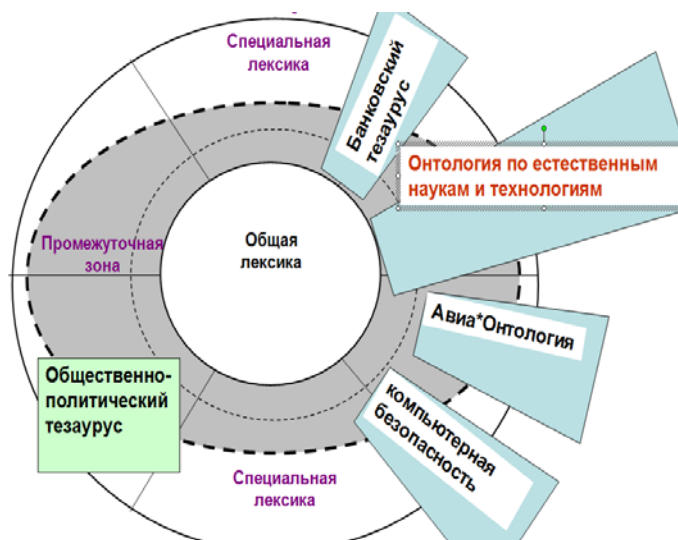


Рис. 2. Система тезаурусов для автоматической обработки текстов, созданных для русского языка по модели Общественно-политического тезауруса

¹ <http://www.labinform.ru/pub/ruthes/index.htm>

НАПРАВЛЕНИЯ ПРОЕКТИРОВАНИЯ ТЕЗАУРУСОВ

При рассмотрении близких работ можно выделить несколько направлений, в которых происходит проектирование тезаурусов. *Первое направление* связано с разработкой многоязычных информационно-поисковых тезаурусов, ориентированных на международную практику использования.

Необходимость создания многоязычных информационно-поисковых тезаурусов возникает у таких международных организаций, как парламент Европейского союза или библиотека ООН, для индексирования и организации поиска документов на разных языках. Создание таких ресурсов регулируется международными и национальными стандартами [5-7].

В ГОСТ 7.24-2007, созданном в соответствии с международным стандартом ISO 5964-1985, указывается, что:

- классы условной эквивалентности лексических единиц одноязычных версий должны соответствовать друг другу по объему выражаемых понятий;
- соответствующие друг другу дескрипторы отдельных одноязычных версий должны иметь, по возможности, максимально совпадающие парадигматические отношения (ГОСТ 7.24-2007, с. 2).

Также в стандарте рекомендуется выбрать в качестве базовой некоторую одноязычную версию и устанавливать эквивалентность дескрипторов других одноязычных версий по отношению к дескрипторам базовой версии.

Тезаурусы международных организаций обычно создаются в соответствии с вышеупомянутыми международными стандартами разработки информационно-поисковых тезаурусов.

Многоязычный тезаурус EuroVoc [8] был разработан специально для содержательной обработки и поиска документов по всем направлениям деятельности институтов Европейского Союза. Текущая версия тезауруса 4.4 опубликована для 23 языков государств, входящих в состав Европейского Союза, а также трех государств-кандидатов.

Структура всех языковых версий – одинакова. Для каждого языка дескриптор (preferred term) имеет соответствие с дескриптором на каждом из остальных языков. Таким образом, понятию тезауруса соответствуют эквивалентные выражения на всех языках, включенных в тезаурус.

При этом, помимо дескрипторов, в каждом языке у понятия могут быть дополнительные языковые соответствия (non preferred terms). В тезаурусе не устанавливаются никаких кросс-языковых соответствий между такими терминами, поэтому в каждом языковом варианте может быть отражена некоторая языковая специфика.

По таким же принципам устроены многоязычный тезаурус ООН UNBIS и тезаурус в области ядерной безопасности INIS. В обоих случаях в тезаурус сначала вводится дескриптор на английском языке, затем дескриптор снабжается единственным уникальным дескриптором на другом языке. Далее для каждого языка могут быть добавлены дополнительные варианты.

Многоязычный Тезаурус UNBIS [9], созданный Библиотекой им. Дага Хаммаршельда Департамента общественной информации ООН, содержит терминологию, используемую в качестве дескрипторов при анализе документов и других материалов, относящихся к программам и деятельности ООН. Будучи многоотраслевым, тезаурус отражает широкий круг вопросов, которыми занимается ООН; термины тезауруса по возможности точно и четко отражают специфику предметной области. В настоящее время тезаурус существует на всех официальных языках ООН: английском, арабском, испанском, китайском, русском и французском/

Международная система ядерной информации INIS – это автоматизированная библиографическая система в области мирного использования атомной энергии; образована государствами и международными организациями в рамках МАГАТЭ [10]. Для поиска информации в системе создан тезаурус INIS, содержащий 30 тысяч терминов, переведенных на восемь языков, включая русский.

Второе направление связано с моделированием лексических систем различных национальных языков и необходимостью создания мультязычной модели представления и обработки построенных лексических ресурсов. Распространенным подходом к развитию многоязычных лексических ресурсов является подход на основе тезаурусов формата Принстонского WordNet (*далее* – ворднеты). Принстонский WordNet является одним из наиболее известных лексических ресурсов в сфере компьютерной лингвистики и автоматической обработки текстов [11, 12]. Во многих странах были инициированы проекты создания ворднетов для национальных языков, а также возник вопрос о взаимосвязях между этими ворднетами [13-15].

Основными структурными единицами тезауруса WordNet являются синсеты – группы взаимозаменяемых слов. Использование синсетов делает ресурс зависимым от языка. Одному понятию реального мира может соответствовать несколько разных синсетов. Например, в отдельные синсеты выделяют разные части речи, новые и устаревшие термины, группы слов с различными грамматическими характеристиками (например, видовые формы глагола в русском языке) и т.д.

Существует несколько подходов к построению мультязычных ворднетов.

1. Пословный перевод. Новый ворднет создается на базе Принстонского WordNet посредством простого пословного перевода всех лексических входов [16, 17]. Преимущества такого подхода связаны с минимальными затратами ресурсов на разработку, основным недостатком является низкое качество полученного лексического ресурса.

2. Сохранение структурной модели. Новый ворднет создается на основе Принстонского WordNet и наследует его концептуальную структуру. При этом, состав синонимов, составляющих синсеты, варьируется для каждого конкретного языка [17-19]. Преимуществом такого подхода является совместимость с Принстонским WordNet, что обеспечивает легкость связывания ресурсов.

В проекте EuroWordNet [14, 15] одним из главных был вопрос о представлении многоязыковой структуры данных – нужно ли стремиться к разработке языково-независимой структуры, с которой необходимо сопоставить единицы каждого языка, или, может быть, нужно иметь единую систему синсетов – новая единица в иерархической сети может быть включена, если хотя бы один язык из рассматриваемых имеет лексему или устойчивый оборот с таким значением.

По принятому в проекте решению каждый ворднет должен сохранять специфику своего языка [14, 15]. При этом каждый ворднет должен содержать отсылки на значения Принстонского WordNet, что позволяет сравнивать ворднеты, обнаруживать непоследовательности в построении ворднетов и видеть различия в устройстве разных языковых систем (рис. 3).

Для установления связи между различными языками в проекте EuroWordNet, синсеты каждого ворднета имеют отсылку на так называемый межязыковой индекс (interlingual index – ILI), в качестве которого выбираются синсеты Принстонского WordNet. Индекс представляет собой неупорядоченный список синсетов с толкованиями для наиболее точного описания соответствия конкретных синсетов каждого языка и преодоления лексических пропусков, которые могут возникнуть в том или ином языке; предоставляется возможность использования нескольких разных отношений эквивалентности от синсетов конкретного языка к индексу ILI:

- EQ_SYNONYM: имеется прямое соответствие между синсетом языка и синсетом индекса;
- EQ_NEAR_SYNONYM: синсету соответствует несколько синсетов индекса,
- HAS_EQ_HYPERONYM: синсет является более специфичным, чем имеющиеся синсеты индекса,
- HAS_EQ_HYPONYM: синсет может быть связан только с более специфичными синсетами индекса.

Так, испанское слово *dedo*, соответствующее русскому слову *палец*, находится в отношении HAS_EQ_HYPONYM таким английским синсетам из индекса ILI как *toe* (палец ноги) и *finger* (палец руки).

В работе [1] описан формат представления концептуальных и лексических знаний, реализованный в проекте РуТез. На основе этой модели выполнен перевод на английский язык Общественно-политического тезауруса [2], представленного как лингвистическая онтология, т.е. основными единицами являются концепты, выявленные на основе существующих лексических единиц и выражений языка, каждый концепт имеет имя, к концепту приписан набор текстовых входов. Для построения английской версии такой лингвистической онтологии требуется существующий концепт онтологии снабдить однозначным и уникальным названием на английском языке и приписать к этому концепту англоязычные текстовые входы, на основе которых концепт будет распознаваться в тексте. Пример перевода концепта ЗДРАВООХРАНЕНИЕ представлен в табл. 1.

В [1] отмечается, что при переводе тезаурусов, лингвистических онтологий важно обеспечить эквивалентность всех переводов относительно отношений тезауруса (онтологии). Поскольку при переводе часто приходится использовать существующие двуязычные словари, то следует отметить, что при использовании переводов, необходимо проверять определения переводных эквивалентов в толковых словарях целевого языка, в связи с тем, что двуязычные словари часто содержат в качестве переводов слова и с более узким, и с более широким значением [1, 21].

Для разработки русско-татарского тезауруса использование формата РуТез является предпочтительным. Во-первых, этот ресурс описывает лексические входы (однословная и многословная терминология, соотносимая с соответствующими концептами) для русского языка, что дает готовый материал для построения двуязычного ресурса. Во-вторых, русский и татарский языки функционируют в общем геополитическом пространстве, что обуславливает взаимопроникновение лексики и общие черты ментальных пространств языков. Таким образом, в формате РуТез может быть обеспечен параллелизм концептуальных структур языковых пространств и отражены национально-специфические черты татарского языка.

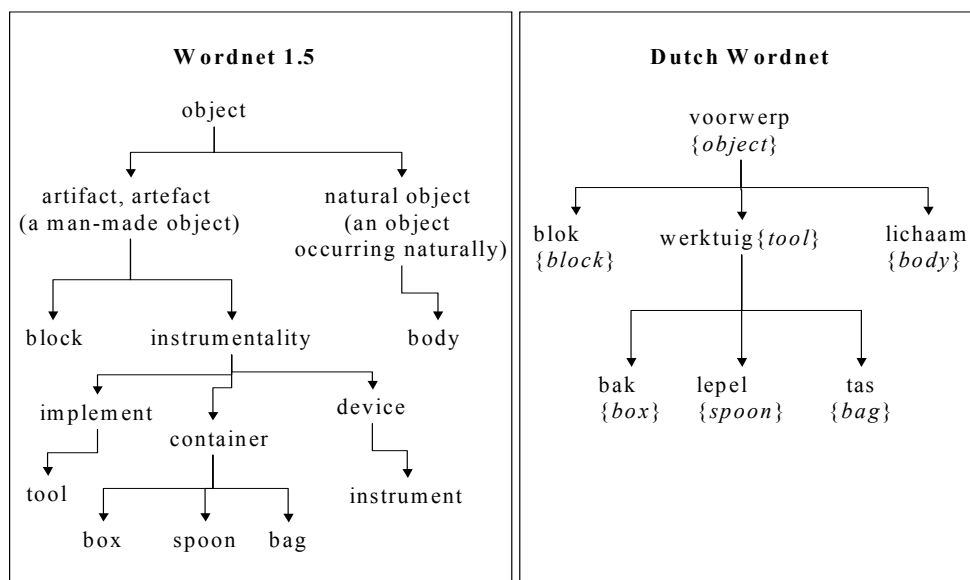


Рис.3. Различия в классификации объектов в англоязычном и голландском ворднетах [15]

Двуязычное представление концепта **ЗДРАВООХРАНЕНИЕ**

ЗДРАВООХРАНЕНИЕ	PUBLIC HEALTH
защита здоровья	<i>community health</i>
здравоохранительный	<i>health care</i>
здравоохранительные меры	<i>health care sector</i>
обеспечение здоровья	<i>health care system</i>
общественное здравоохранение	<i>health field</i>
оздоровление граждан	<i>health of population</i>
оздоровление населения	<i>health promotion</i>
охрана здоровья	<i>provision of health</i>
система здравоохранения	<i>public health</i>
	<i>public health field</i>

Современный подход к разработке лексических ресурсов должен обеспечивать возможность создания и публикации набора данных в пространстве открытых связанных данных LOD (Linked Open Data), т.е. структурированных данных, которые представлены в специальном формате для их совместного использования и установления взаимных связей. В настоящее время активно развивается ряд международных проектов в рамках пространства Linguistic Linked Open Data Cloud [22].

Публикация связанных данных требует стандартизации разрабатываемых ресурсов, что в свою очередь позволяет при связывании создавать сеть взаимно дополняемых лексических ресурсов. Связывание с центральным хранилищем терминологии способствует концептуальной совместимости. В [23] определены следующие основные преимущества связанных лингвистических данных: а) связывание через URIs, б) федерация, в) динамическое связывание между ресурсами, г) наличие богатой экосистемы форматов и технологий, например, формат RDF/OWL-представление WordNet [24], онтология для представления лексических ресурсов The Lexicon Model for Ontologies (Lemon) [25], RDF-представление многоязычного тезауруса BabelNet [26] и др. Актуальные решения по публикации тезаурусов русского языка в виде открытых связанных данных представлены в [27]. Публикация русско-татарского тезауруса в пространстве LLOD является одной из задач настоящего проекта.

ПРИНЦИПЫ МОДЕЛИРОВАНИЯ И МЕТОДОЛОГИЯ ПОСТРОЕНИЯ РУССКО-ТАТАРСКОГО ТЕЗАУРУСА

Определение принципов моделирования двуязычного ресурса предполагает установление: 1) формата словарной статьи; 2) объема и качества лексического материала; 3) методов выявления и разрешения конфликтной ситуации при переводе.

Моделирование русско-татарского тезауруса осуществляется в формате РуТез. При проектировании татарской компоненты в целом сохраняется концептуальная структура тезауруса РуТез, т.е. татарская компонента базируется на списке концептов тезауруса РуТез. Концептуальная структура тезауруса РуТез определяет направление (последовательность) разра-

ботки концептов в татарской части, при этом важным шагом является проверка параллельности разрабатываемых концептуальных структур. При проектировании татарской части сохраняется структура концептуальных связей тезауруса РуТез. Однако в случаях, когда семантические отношения между татарскими концептами не соответствуют семантическим отношениям между единицами тезауруса РуТез, эти отношения устанавливаются заново или корректируются.

Важной задачей при разработке татарской части двуязычного тезауруса является отображение специфики лексико-семантической системы татарского языка. Это достигается двумя основными способами.

1. Добавление новых концептов, отображающих не представленные или представленные неполно в тезаурусе РуТез тематические блоки, значимые для социокультурной жизни татарского общества (например: понятия, связанные с исламом; социальная иерархия восточных обществ; татарские этнокультурные реалии и т.п.).

2. Достижение концептов промежуточных уровней, не вербализованных в русском языковом мышлении, но выявляемых при сопоставлении с татарским языковым материалом (например, в русском языковом сознании вербализованы концепты ПАСЫНОК и ПАДЧЕРИЦА, но не представлен концепт, обозначающий ребенка одного из супругов вне зависимости от пола – такой концепт имеется в татарском языке).

Общая методология проектирования татарской части тезауруса включает следующие основные этапы.

На начальном этапе осуществляется перевод имен концептов и текстовых входов тезауруса РуТез на татарский язык. Источниками для перевода текстовых входов на татарский язык являются татарские тексты по соответствующей тематике, параллельные тексты, переводные и толковые словари, а также тематические общественно-политические словари [28–30].

Основные требования к переводу имен концептов можно обозначить следующим образом:

- ясность значения концепта из его обозначения;
- использование нейтральной и общеупотребительной лексики при обозначении концепта;
- при наличии вариантов (лексических и фразеологических дублетов) приоритет отдается обозначению, наиболее точно отражающему данный концепт, лаконичному и, если это не противоречит предыдущим ус-

ловиям, производится выбор в пользу наиболее частотного и официально используемого обозначения.

В идеале по имени концепта должно быть интуитивно понятно его содержание и объем.

В табл. 2 дано описание концепта БАЛАЧАК (ДЕТСТВО) в русско-татарском тезаурусе. Таблица дает общее представление о некоторых особенностях и сложностях перевода имен концептов тезауруса РуТез на татарский язык. Поскольку в татарском языке слово *яшь*, обозначающее возраст, многозначное, в скобках дается пояснение, указывающее на период жизни. Татарские переводные соответствия русского слова *младенчество* *нарасыйлык* и *сабылык* коннотативно окрашены, поэтому они малопригодны для языка деловых бумаг. Имя концепта НЕСОВЕРШЕННОЛЕТИЕ переводится не существительным, а именем действия.

Перевод текстовых входов предполагает, по возможности, максимальный охват существующих вариантов – параллельных наименований одного и того же явления (реалий), функционирующих в татарском языке.

Требование ясности и однозначности значения концепта, следующее из его обозначения, предполагает критический анализ терминов, используемых в настоящее время в татарском языке.

В ходе работы фиксируются и анализируются случаи, когда официально используемый термин является неудовлетворительным по ряду причин. Наиболее часто подобные термины возникают при калькировании русских терминов и терминологических сочетаний, не учитывающих особенностей лексической и грамматической системы татарского языка, что может создавать двусмысленность, крайне нежелательную для языка официальных документов.

В частности, переводчики портала “Официальный Татарстан” терминологическое сочетание *уголовное право* переводят как *жсинаять хокукы*. По нашему мнению, такой перевод является некорректным, так

как сочетание *жсинаять хокукы* может быть интерпретировано не только как ‘уголовное право’, но и как ‘право на преступление’ (ср.: *сайлау хокукы* ‘право избирать’, *яшәү хокукы* ‘право на жизнь’ и т.п.). При переводе имени концепта *уголовное право* мы предлагаем вариант объяснительного перевода *жсинаятькә каршы кәрәш хокукы*, а вариант *жсинаять хокукы* включается нами в число текстовых входов для названного концепта как реально используемая номинация.

Важной задачей при создании подобных тезаурусов является изменение состава концептов за счет внесения дополнений на различных уровнях. Поскольку концепты тезауруса РуТез сформированы на базе текстовых входов (слов и словосочетаний) на русском языке, тезаурус РуТез в конечном счете отражает русское языковое мышление. При сопоставлении с материалом татарского языка в ряде случаев выявляются языковые лакуны. В подобных случаях необходимо достроить тезаурус введением недостающих концептов, при этом возникает проблема изобретения новых обозначений для концептов (например, гиперонима для концептов *пасынок* и *падчерица*).

Другие изменения состава концептов возникают при добавлении новых концептов и текстовых входов целых тематических блоков (например, понятий, связанных с исламом, с особенностями татарской культуры), отсутствующих в тезаурусе РуТез.

Отдельной задачей является разработка списка текстовых входов для концептов тезауруса. Представление текстовых входов предполагает, по возможности, максимальный охват существующих вариантов – параллельных наименований. Поскольку языковая ситуация в Республике Татарстан находится в состоянии развития, для одних и тех же реалий, включая некоторых названия ведомств и государственных структур, могут использоваться параллельные номинации (см. табл.3).

Таблица 2

Концепт БАЛАЧАК (ДЕТСТВО)

Отношения	Имя концепта в русском языке	Имя концепта в татарском языке
	ДЕТСТВО	БАЛАЧАК
ВЫШЕ	ВОЗРАСТ	ЯШЬ (ЯШӘУ ЧОРЫ)
ЦЕЛОЕ	НЕСОВЕРШЕННОЛЕТНИЕ ДЕТИ	БАЛИГЪ БУЛМАГАН БАЛАЛАР
НИЖЕ	ДОШКОЛЬНЫЙ ВОЗРАСТ	МӨКТӘПКӘЧӘ ЯШЬ
НИЖЕ	МЛАДЕНЧЕСТВО	НАРАСЫЙЛЫК
АСЦ	НЕСОВЕРШЕННОЛЕТИЕ	БАЛИГЪ БУЛМАУ

Таблица 3

Названия организаций

Названия организаций	Названия организаций (перевод на татарский язык)
органы государственной безопасности	1) дәүләт куркынычсызлыгы органнары 2) дәүләт иминлеге органнары
Министерство земельных и имущественных отношений Республики Татарстан	1) Татарстан Республикасы жир һәм мөлкәт мөнәсәбәтләре министрлыгы; 2) Татарстан Республикасы жир һәм милек мөнәсәбәтләре министрлыгы; 3) Татарстан Республикасы жир һәм мөлкәт бәйләнешләре министрлыгы.

В подобных случаях в качестве имени концепта мы используем номинацию, наиболее ясно и точно обозначающую ту или иную реалию (явление), а в текстовых входах даем список всех используемых номинаций.

Под влиянием русского языка у многих татарских слов появляются новые значения - активно идет процесс семантического калькирования, эти новообразования также должны быть отражены в тезаурусе.

Работа над тезаурусом показывает, что в настоящее время продолжается активный процесс развития лексической системы татарского языка, при этом под влиянием русского языка у слов формируются новые значения (семантические кальки), а также активно продолжают процессы словообразовательного калькирования.

ПРОГРАММНАЯ ОБОЛОЧКА ДЛЯ ПОСТРОЕНИЯ ТЕЗАУРУСА

Разработка татарской компоненты тезауруса осуществляется на основе специальной программной оболочки (электронный тезаурус, см. рис. 4), разработанной в ходе данного проекта. Система электронного тезауруса реализована в формате веб-приложения в виде специального сайта и содержит открытую и закрытую части. Открытая часть используется для просмотра данных конечным пользователем, закрытая часть доступна только для авторизованных пользователей, обладающих полномочиями

по редактированию тезауруса. Основные функции в открытой части системы обеспечивают поиск и навигацию по концептам и текстовым входам тезауруса, просмотр концептов определенного домена. Редактор тезауруса обеспечивает функции добавления, правки и удаления текстовых входов и концептов, функции сравнения концептуальных структур русской и татарской частей тезауруса, поддерживает специальную разметку концептов, отражающую статусы завершенности и проблемности, а также имеет удобные средства визуализации с фильтрацией по различным параметрам.

ЗАКЛЮЧЕНИЕ

В настоящей статье изложены основные подходы и решения, полученные при проектировании нового лексического ресурса – русско-татарского тезауруса, разрабатываемого на основе формата тезауруса РуТез. Двухязычный тезаурус имеет в целом концептуальную структуру тезауруса РуТез, однако, в некоторых частных случаях допускаются структурные модификации, связанные с добавлением как единичных концептов на какой-либо ветви, так и с добавлением новых тематических ветвей, отражающих национальную специфику татарского языка. В настоящее время разрабатывается русско-татарский тезаурус по общественно-политической области, текущий размер которого составляет 5000 концептов.

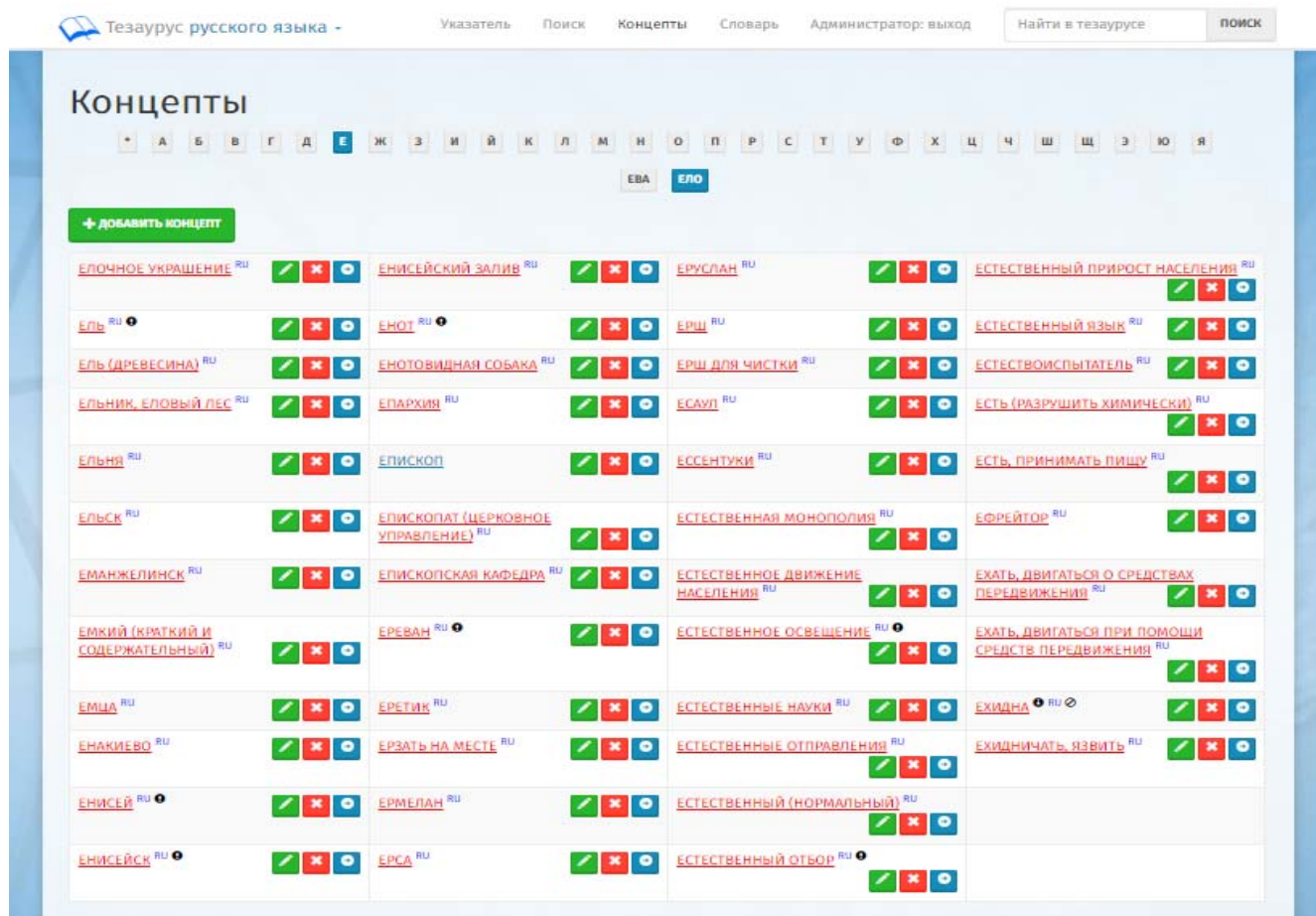


Рис. 4. Программная оболочка «Электронный тезаурус»

Построение татарской компоненты тезауруса позволяет уточнить (а также в определенной степени сформировать) новый пласт терминологии в общественно-политической сфере, исследовать влияние русского языка на процессы терминообразования в татарском языке.

СПИСОК ЛИТЕРАТУРЫ

1. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во МГУ, 2011. – 512 с.
2. Loukachevitch N., Dobrov B. The Sociopolitical Thesaurus as a resource for automatic document processing in Russian // Terminology. – 2015. – Vol. 21, № 2. – P. 237-262.
3. Лукашевич Н.В., Невзорова О.А. Авиантология: анализ современного состояния ресурса // Труды международной конференции Диалог-2004 "Компьютерная лингвистика и интеллектуальные технологии". – М.: Наука, 2004. – С. 424-430.
4. Добров Б.В., Лукашевич Н.В. Онтология по естественным наукам и технологиям ОЕНТ: структура, состав и современное состояние // Электронные библиотеки. – 2008. – Т. 11, № 1.
5. ISO 5964-1985. Documentation – Guidelines for the establishment and development of multilingual thesauri. – URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=12159 (дата обращения: 20.10.2016).
6. ISO 25964-1:2011. Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval. – URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=53657 (дата обращения: 20.10.2016).
7. ГОСТ 7.24-2007. Межгосударственный стандарт. Тезаурус информационно-поисковый, многоязычный. – М.: Стандартинформ, 2010.
8. EUROVOC Thesaurus. – URL: <http://eurovoc.europa.eu/drupal/?q=node/554> (дата обращения: 20.10.2016).
9. United Nations. UNBIS Thesaurus. English version. – New York: Dag Hammarskjöld Library of the United Nations, 1986.
10. INIS Thesaurus. English version. IAEA-INIS Reference Series. – 2016.
11. WordNet: An Electronic Lexical Database / ed. Fellbaum C. – Cambridge: MIT Press, 1998.
12. Miller G.A. WordNet: A lexical database for English // Communications of the ACM. – 1995. – Vol. 38, Issue 11. – P. 39-41.
13. Pianta E., Bentivogli L., Girardi C. Multiwordnet: Developing an aligned multilingual database // Proceedings of the First International Conference on Global WordNet. – Mysore, India: 2002. – P. 293-302.
14. EuroWordNet: A Multilingual Database with Lexical Semantic Network / Vossen P. (ed.). – Dordrecht: Kluwer, 1998.
15. Vossen P. Building wordnets. – 2005. – URL: <http://www.globalwordnet.org/gwa/BuildingWordnet.spt> (дата обращения: 20.10.2016).
16. Russian WordNet. – URL: <http://www.tendrow.org/> (дата обращения: 20.10.2016).
17. Сухоногов А. М., Яблонский С. А. Автоматизация построения англо-русского WordNet // Труды международной конференции «Диалог-2005». – Звенигород: 2005. – С. 46-51.
18. SKOS Simple Knowledge Organization System Reference, W3C Recommendation, 2009. – URL: <http://www.w3.org/TR/skos-reference/> (дата обращения: 20.10.2016).
19. Азарова И.В., Синопальникова А.А., Яворская М.В. Принципы построения wordnet-тезауруса RussNet // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог-2004 («Верхневолжский», 2-7 июня 2004 г.). – М.: 2004. – С. 542-547.
20. MultiWordNet. – 2004. – URL: <http://multiwordnet.fbk.eu/english/home.php> (дата обращения: 20.10.2016).
21. Гринев-Гриневич С.В. Терминоведение. – М.: Издательский центр «Академия», 2008. – 304 с.
22. Linguistic Linked Open Data Cloud (LLOD). – 2015. – URL: <http://linguistic-lod.org/lod-cloud> (дата обращения: 20.10.2016).
23. Chiarcos C., McCrae J., Cimiano P., Fellbaum C. Towards open data for linguistics: Linguistic linked data / A. Oltramari, Lu-Qin, P. Vossen, E. Hovy (eds.). New Trends of Research in Ontologies and Lexical Resources. – Heidelberg: Springer, 2013. – P. 7-25.
24. van Assem M., Gangemi A., Schreiber G. Conversion of WordNet to a standard RDF/OWL representation // Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06). – 2006. – URL: <http://www.cs.vu.nl/~guus/papers/Assem06a.pdf> (дата обращения: 20.10.2016).
25. Lemon – The Lexicon Model for Ontologies. – URL: <http://www.lemon-model.net/> (дата обращения: 20.10.2016).
26. Ehrmann M., Cecconi F., Vannella D., McCrae J.P., Cimiano P., Navigli R. Representing Multilingual Data as Linked Data: the Case of BabelNet 2.0. – LREC (2014). – 2014. – URL: http://wwwusers.di.uniroma1.it/~navigli/pubs/LREC_2014_Ehrmannetal.pdf (дата обращения: 20.10.2016).
27. Ustalov D. Russian Thesauri as Linked Open Data // Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue", Vol. 1. – Moscow: RGGU, 2015. – P. 616-625.
28. Амиров Ф.К. Русско-татарский юридический словарь. – Казань: Татар. кн. изд-во, 1996. – 103 с.
29. Низамов И.М. Краткий русско-татарский общественно-политический словарь / И.М.Низамов. – Казань: Татар. кн. изд-во, 1995. – 189 с.

30. Русско-татарский общественно-политический словарь / Ф.А. Ганиев, И.М. Низамов и др. под рук. Р.М. Харисова. – Казань: Татар, кн. изд-во, 1997. – 127 с.

Материал поступил в редакцию 08.11.16.

Сведения об авторах

ГАЛИЕВА Альфия Макаримовна – кандидат философских наук, доцент; ведущий научный сотрудник Института прикладной семиотики Академии наук Республики Татарстан, г. Казань
e-mail: amgalieva@gmail.com

КИРИЛЛОВИЧ Александр Витальевич – аспирант Высшей школы информационных технологий и информационных систем Казанского федерального университета
e-mail: alik.kirillovich@gmail.com

ЛУКАШЕВИЧ Наталья Валентиновна – кандидат физико-математических наук, ведущий научный сотрудник Московского государственного университета им. М.В. Ломоносова
e-mail: louk_nat@mail.ru

НЕВЗОРОВА Ольга Авенировна – кандидат технических наук, доцент, заместитель директора по научной работе Института прикладной семиотики Академии наук Республики; доцент кафедры информационных систем Казанского федерального университета Татарстан
e-mail: onevzoro@gmail.com

СУЛЕЙМАНОВ Джавдет Шевкетович – доктор технических наук, профессор, директор Института прикладной семиотики Академии наук Республики Татарстан; заведующий кафедрой информационных систем Казанского федерального университета
e-mail: dvdt.slt@gmail.com

СПРАВОЧНО-ИНФОРМАЦИОННЫЙ РАЗДЕЛ

УДК 069.9 (470) "2016" : 378

О.В. Сютнюрэнко

О национальной выставке-форуме «ВУЗПРОМЭКСПО – 2016»

Рассматриваются итоги работы в декабре 2016 г. выставки-форума, являющейся связующим звеном между отечественными разработчиками инновационных технологий и промышленностью. Результатам и перспективам отечественной прикладной науки и производства были посвящены доклады III Всероссийской научно-практической конференции «Исследования и разработки – 2016», проходившей в рамках форума. Показан тематический спектр докладов по направлению «Информационно-телекоммуникационные системы». Полная информация о «ВУЗПРОМЭКСПО–2016» представлена на сайте <http://vuzpromexpo.ru>.

Ключевые слова: технологические разработки, инновации, инжиниринг, промышленность, импортозамещение, модели партнерства, инвестиции, индустрия наносистем, энергоэффективность, информационно-телекоммуникационные системы, науки о жизни, микроэлектронные элементы.

В Центральном выставочном комплексе «Экспо-центр» на Красной Пресне 14-15 декабря 2016 г. прошла IV ежегодная национальная выставка-форум «ВУЗПРОМЭКСПО– 2016». Открывали выставку заместитель Председателя правительства Российской Федерации А.В. Дворкович и Министр образования и науки О.Ю. Васильева.

Выставка, организованная под патронажем Министерства образования и науки РФ, направлена на демонстрацию новейших отечественных разработок, выполненных в рамках кооперации промышленных предприятий, научно-исследовательских институтов и вузов страны.

Цель выставки – содействие взятому Правительством РФ курсу на инновационное развитие экономики России, в том числе продвижению на рынок отечественных научных разработок, которые по экономической эффективности и качеству сопоставимы с лучшими зарубежными аналогами.

Особенность выставки заключалась в широкомасштабной демонстрации научного потенциала высших учебных заведений РФ как генераторов новейших высокотехнологичных разработок. «Объем исследований и разработок, проводимых в стенах учебных лабораторий, увеличился за последнее десятилетие более чем в 5 раз. Численность исследователей достигла 47 тысяч человек. Самое важное, что доля молодых исследователей составляет 63%», – отметила О.Ю. Васильева.

Современный формат выставки позволял специалистам промышленных компаний обсуждать интересные их вопросы непосредственно с разработчиками технологий, а также делать заказы на дальнейшие исследования, актуальные для конкретного предприятия. Выставка-форум «ВУЗПРОМЭКСПО» является связующим звеном между отечественными разработчиками и промышленностью и, по существу, обрела статус общероссийского рынка инноваций и высоких технологий. Участниками выставки являлись государственные корпорации, вузы, научные организации, промышленные предприятия, малые инновационные предприятия, территориальные инновационные кластеры.

На площадке «ВУЗПРОМЭКСПО–2016» было представлено более 1200 технологических разработок, охватывающих практически все отрасли промышленности – от сельского хозяйства до космических систем. Выставку посетило более 7000 человек, среди которых руководители крупнейших компаний страны, а также ведущие эксперты – представители науки, средств массовой информации и гражданского общества.

Особое место в экспозиционной части «ВУЗПРОМЭКСПО» было отведено развитию инжиниринговых центров в России. Опыт и итоги работы инжиниринговых центров, созданных на базе вузов, освещался в деловой и конференционной программе.

Особое внимание специалистов привлек круглый стол «Взаимодействие компаний, реализующих про-

граммы инновационного развития, с образовательными организациями высшей школы», а также круглый стол «Повышение эффективности использования научно-образовательного потенциала высшей школы в интересах оборонно-промышленного комплекса России».

Одним из приоритетных направлений работы выставки была и остается поддержка талантливой молодежи. В рамках молодежной программы прошел финал III Всероссийского студенческого научно-технического фестиваля «ВУЗПРОМФЕСТ», цель которой – популяризация передовых инновационных технологий среди студентов технических вузов России, а также повышение уровня соревновательности и конкурентоспособности высших учебных заведений. Тема фестиваля в 2016 г. «Инновации, наука и техника в сельском хозяйстве».

В рамках выставки-форума проведено более 100 деловых мероприятий, которые были посвящены следующим темам:

- стратегии и программы инновационного развития России;
- импортозамещение промышленной продукции;
- инжиниринговая деятельность и промышленный дизайн;
- передовые производственные технологии;
- новые технологические платформы;
- технологическое прогнозирование;
- модели частно-государственного партнерства при реализации крупных проектов.

Пленарные и тематические заседания, круглые столы форума были ориентированы на выявление перспектив и конкретных механизмов взаимодействия науки, высшей школы и промышленности; обмен опытом по реализации высокотехнологичных проектов; переход к ряду практических решений – заключению соглашений о сотрудничестве, разработке планов финансирования совместных проектов.

Результатам и перспективам отечественной прикладной науки и производства были посвящены доклады III Всероссийской научно-практической конференции «Исследования и разработки – 2016», которая проводится ежегодно.

Основная задача конференции – демонстрация и обсуждение передовых научных и практических результатов прикладных научных исследований и экспериментальных разработок, в том числе междотраслевого характера, направленных на создание продукции и технологий для модернизации отраслей отечественной промышленности.

В рамках конференции были проведены тематические сессии по следующим направлениям:

- науки о жизни;
- индустрия наносистем;
- информационно-телекоммуникационные системы;
- рациональное природопользование;
- энергоэффективность, энергосбережение и ядерная энергетика;
- транспортные и космические системы.

Программный комитет конференции рассмотрел тезисы докладов, поданные исполнителями проектов Федеральной целевой программы «Исследования и

разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014-2020 годы» (ФЦП ИР) (1085 заявок) и по другим программам (71 заявки), и большинством голосов отобрал проекты, доклады по которым были включены в устную сессию работы конференции – всего 86 докладов. Полная информация доступна на сайте <http://vuzpromexpro.ru>. Были также организованы постерные сессии.

Определенный интерес представляют доклады по тематическому направлению «Информационно-телекоммуникационные системы»: «Исследование и разработка методов робастности алгоритмов автоматического распознавания русской слитной речи в условиях сложной акустической обстановки в режиме реального времени» (М.Л. Корневский – Санкт-Петербург, Национальный исследовательский университет (НИУ) информационных технологий, механики и оптики); «Создание опытного образца квантового устройства безопасности передачи данных» (Ю.В. Курочкин – ООО «Международный центр квантовой оптики и квантовых технологий»); «Система регистрации и декодирования биоэлектрической активности мозга и мышц человека» (В.Б. Казанцев – Нижегородский НИУ им. Н.И. Лобачевского); «Разработка системных компонентов инновационного роботизированного комплекса для реабилитации пациентов с нарушениями функций нижних конечностей вследствие травм и заболеваний головного и спинного мозга» (С.А. Минеев – Нижегородский НИУ им. Н.И. Лобачевского); «Разработка прототипа программного обеспечения распределенной высокопроизводительной интеллектуальной аналитической системы поддержки принятия клинических решений для многомасштабной эндоскопической диагностики и хирургии “Умный эндоскоп”» (В.В. Мизгулин – ООО «СИАМС»); «Разработка конструктивно-технологических решений по созданию семейства микроэлектронных элементов на структурах кремний на изоляторе (КНИ), обеспечивающего возможность создания высокотемпературных датчиков внешних воздействий различного функционального назначения» (В.Н. Мордкович – АО ПКК «Миландр»); «Исследование и разработка сенсорных узлов и универсальной цифровой платформы для построения самоорганизующихся и энергонезависимых беспроводных сенсорных сетей (умная пыль) для систем промышленной безопасности и экологического мониторинга» (А.М. Баранов – Московский авиационный институт); «Разработка многоэлементных неохлаждаемых КМОП-сенсоров теплового излучения на основе МЭМС-термопар для инфракрасных оптико-электронных систем технического зрения» (Е.А. Фетисов – Московский институт электронной техники); «Разработка конструктивно-технологических решений создания микросенсорного интеллектуального магниточувствительного интегрального датчика положения для контроля подвижных узлов современных интеллектуальных робототехнических комплексов на основе микросистемной техники» (Г.Н. Прокофьев – Зеленоградский нанотехнологический центр); «Анализ компетенций сотрудников с использованием технологий интеллектуальной обработки данных» (Н.С. Ники-

тинский – ООО «Центр разработки»); «Высокоточная магнитная градиентометрия и стереоскопическое распознавание лиц как основа уникальной системы скрытого дистанционного обнаружения носителей холодного и огнестрельного оружия в потоках людей высокой плотности» (доклад представлен ООО «Децима»); «Разработка технологии визуализации массивов неформализованных данных разнородной структуры, полученных путем автоматического анализа новостных потоков» (С.В. Знаменский – Институт программных систем им. А.К. Айламазяна РАН); «Разработка импульсных твердотельных генераторов миллиметрового и субмиллиметрового диапазонов волн на основе многослойных гетероструктур GaAs/AlGaAs» (В.А. Гергель – Институт радиотехники и электроники им. В.А. Котельникова РАН); «Кибернетическая модель и ее реализация на кластере универсальных и графических процессоров» (Е.Н. Надеждин – ГНИИ ИТТ «Информика»).

15 декабря в рамках выставки-форума прошел демонстрационный день проектов-выпускников Программы развития проектов Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014-2020 годы» (ФЦП ИР). Было представлено несколько десятков проектов, в первую очередь для потенциальных инвесторов и промышленных партнеров. Цель Программы развития ФЦП ИР – повышение экономической эффективности результатов прикладных научных исследований и экспериментальных разработок, финансируемых Минобрнауки РФ.

В постоянно растущую группу промышленных партнеров Программы развития ФЦП ИР входят, например ГК «Росатом», ПАО «Северсталь», НПО «Сатурн», АО «Наука и инновации», холдинг «Вертолеты России», фонд научной, научно-технической и инновационной деятельности «Энергия без границ» группы «Интер РАО ЕЭС», ГК «Мортон», независимая лаборатория «Инвитро» и другие. Подробная информация о Программе развития проектов ФЦП ИР и ее результатах представлена на сайте <https://acc.xpir.ru/about>.

14 и 15 декабря посетители «ВУЗПРОМЭКСПО–2016» смогли ознакомиться с проектом малого космического аппарата научного назначения «Аист 2Д», «старшие братья» которого были выведены на орбиту 28 апреля 2016 г. в ходе первой пусковой кампании с нового космодрома «Восточный».

Одним из экспонатов выставки-форума, вызвавших большой интерес, стали магнитные сплавы на основе редкоземельных металлов. Эту инновационную технологию разработали ученые Национального исследовательского технического университета «Московский институт стали и сплавов» совместно с АО «ВНИИХТ» и ГК «Скайград». В России такие материалы ранее не производились. Новая инновационная технология, которая была внедрена благодаря субсидии, выделенной Минобрнауки РФ в рамках 218 Постановления Правительства РФ о кооперации вузов и предприятий реального сектора экономики, в перспективе дает технологические возможности для получения магнитных материалов из техногенных отходов. Опытное производство уже запущено на ба-

зе ГК «АНТАРН» и к концу 2016 г. должна быть получена первая партия товарной продукции.

Инженеры Московского государственного университета дизайна и технологий (МГУДТ), создали новый металл-полимерный нанокомпозит – металлизированные полимерные нити. Уникальная технология включает две стадии производства: первая – предварительная химическая модификация поверхности полимерных нитей; вторая – металлизация полимерных нитей. Кроме того, новая технология обеспечивает изготовление термически плавких, легких, гибких и прочных электрических проводов для разных сфер деятельности. Металлизированная полимерная нить в 3-5 раз легче традиционных металлических проводов из-за низкой плотности полимеров, она абсолютно устойчива к воспламенению в условиях короткого замыкания. Разработку можно применять в легкой, нефтехимической и нефтедобывающей промышленности.

Ученые кафедры аналитической химии Химико-технологического института Уральского федерального университета им. Б.Н. Ельцина выпустили прибор «Эльсенс Эко», который позволяет определять содержание токсичных микропримесей в природных, питьевых и сточных водах, пищевых продуктах, продовольственном сырье, почвах и биологических объектах. В аналитическом приборе используются уникальные экологически безопасные и чувствительные сенсоры, позволяющие исключить из процесса анализа ртуть и ее растворимые соли, а также процедуры механической регенерации поверхности электродов, требующие вмешательства оператора. Кроме того, комплекс имеет программное и методическое обеспечение. Этот прибор уже появился на рынке и стоит 200 тыс. рублей, что примерно в 5 раз меньше, чем у зарубежных аналогов.

В рамках «ВУЗПРОМЭКСПО–2016» проходил конкурс инновационных проектов аэрокосмической отрасли. Организатор конкурса – Центральный аэрогидродинамический институт им. профессора Н.Е. Жуковского (ФГУП «ЦАГИ») – крупнейший государственный научный авиационный центр России. Конкурс ориентирован на поиск и поддержку инновационных проектов в авиакосмической области (в первую очередь со стороны промышленных и инвестиционных кампаний), основой которых являются результаты успешных научных и научно-технических исследований и разработок.

При поддержке отраслевого союза «Нейронет» в рамках «ВУЗПРОМЭКСПО–2016» проходил симпозиум «Нейронауки и нейротехнологии: фундаментальные аспекты и прикладные разработки». В ходе симпозиума обсуждались научно-технические вопросы современных исследований мозга с акцентом на разработку технологий моделирования нейронных сетей мозга и разработку технологий и устройств, реализующих когнитивные функции мозга, а также нейроинтерфейсов, расширяющих функциональные возможности человека. Заинтересованной аудиторией заслушаны доклады и выступления ведущих российских ученых в области нейронаук и нейротехнологий.

15 декабря в рамках деловой программы «ВУЗПРОМЭКСПО–2016» Минобрнауки РФ и НИУ

ВШЭ провели круглый стол на тему «Актуализация прогноза научно-технологического развития Российской Федерации на период до 2030 года: первые результаты в сфере авиастроения, фотоники, реабилитационной индустрии». Прогноз на период до 2030 г., подготовленный Минобрнауки РФ совместно с НИУ ВШЭ, включает обзор перспективных направлений, оказывающих наибольшее влияние на развитие отечественной науки, конкретизирует риски и определяет перспективные рынки, формирующие максимальный спрос на российские инновации.

Выставка-форум «ВУЗПРОМЭКСПО» уверенно заняла ведущую позицию специализированного мероприятия в России, посвященного демонстрации лучших практик в сфере внедрения инновационных разработок в отечественную промышленность. В экспозиционной части «ВУЗПРОМЭКСПО-2016» приняли участие 245 специалистов. Среди участников – 120 вузов, 69 предприятий реального сектора экономики, 12 научных институтов Российской академии наук, 32 технологические платформы, 30 инжиниринговых центров. Были представлены субъек-

ты Российской Федерации, а также программы, реализуемые Министерством образования и науки РФ. На 16000 кв.м. выставочной площадки демонстрировалось порядка 800 оригинальных разработок по актуальным направлениям развития промышленности Российской Федерации.

ВУЗПРОМЭКСПО – это крупнейшая федеральная многоотраслевая выставка технических и технологических достижений отечественной науки, уникальная площадка для взаимодействия российской науки, малого инновационного бизнеса, инвесторов и крупнейших предприятий страны.

Материал поступил в редакцию 19.12.16.

Сведения об авторе

СЮНТЮРЕНКО Олег Васильевич – доктор технических наук, профессор, ведущий научный сотрудник ВИНТИ РАН, Москва
e-mail: olegasu@mail.ru