МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ВИНИТИ РАН)

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г. **№ 10** **Москва 2023**

ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ

УДК [658.512:005.51/.53]:510.644.4

С.Г. Цапко, И.В. Цапко, Д.В. Тараканов

Оценка эффективности процессов проектирования сложных систем математическим аппаратом нечетких множеств

Представлен анализ методов оценки эффективности процессов проектировании сложных систем и предложены формальные показатели оценки их качества на основе математического аппарата нечетких множеств. За основу выбрана такая сложная система, как модель бизнес-процессов проектирования, испытаний и изготовления (создания) бортовой радиоэлектронной аппаратуры, в проектировании которой задействовано множество подразделений предприятия – от менеджеров, схемотехников, проектировщиков, до отдела закупки комплектующих, испытательных цехов, служб сопровождения.

Предложено считать все компоненты бизнес-процессов ресурсами, выделить из них материальные, трудовые, финансовые, временные, аппаратные и программные ресурсы, установить связь между ними и провести оценку весовых коэффициентов связей для поиска самой эффективной реализации процесса проектирования сложной системы с помощью лепестковой диаграммы относительных показателей качества. Для оценки эффективности каждого ресурса во времени и расчёта интегрального показателя эффективности бизнес-процессов жизненного цикла создания сложной системы использована Е-сетевая логико-динамическая модель выполнения проектно-конструкторских работ, которая позволяет учесть дина-

мические характеристики параллельных, взаимодействующих отделов, а также проектных этапов.

На основе аппарата нечётких множеств рассмотрен подход к оценке эффективности каждого этапа проекта создания сложной системы, что, в свою очередь, позволяет учесть неопределенность и нечеткость в оценке, а также определить важность каждого критерия на основе предпочтений инвестора. Это помогает принимать обоснованные решения о составлении инвестиционного портфеля и достижении желаемых финансовых целей.

Ключевые слова: *проектирование, сложная система, нечеткая логика, процесс, ресурс, показатели качества, лепестковая диаграмма, Е-сетевая модель, эффективность*

DOI: 10.36535/0548-0027-2023-10-1

ВВЕДЕНИЕ

Процесс проектирования сложных систем является многогранной итерационной задачей с множеством неизвестных, где каждый компонент процесса проектирования может оказать непредсказуемое воздействие на результат конечного продукта. В процессе проектирования сложных систем задействовано большое количество служб предприятия и специалистов разных профилей. На этапе эскизного проектирования сложной системы необходимо оценить возможности предприятия для обеспечения полного цикла создания изделия. Одним из важных факторов оценки является эффективность процессов проектирования и построение на ее основе структуры процессов взаимодействия подразделений предприятия.

Оценка эффективности процессов проектирования сложных систем – это одна из важнейших задач в области инженерии и науки о системах. Сегодня сложные системы играют ключевую роль в различных сферах деятельности, вопросы оценки их эффективности становятся все более актуальными. Проектирование сложных систем требует учета явлений неопределенности и размытости, которые сопутствуют процессу принятия решений в сложных и неопределенных средах.

Одним из подходов к анализу и оценке сложных систем является применение математического аппарата нечетких множеств. Нечеткая логика, основанная на теории нечетких множеств, позволяет моделировать неопределенность, присутствующую в реальных системах. Оценка эффективности процессов проектирования сложных систем с использованием нечетких множеств – это активно развивающаяся область исследований. Так, в работе [1] представлена концепция нечетких множеств, которая отличается от классических бинарных множеств, и определено понятие функции принадлежности, которая описывает степень принадлежности элементов к нечеткому множеству. Задавая нечеткое множество и его функцию принадлежности, можно моделировать неопределенность и нечеткую информацию. В книге [2] предложена более общая теория нечетких множеств и исследованы различные методы и подходы к нечеткому моделированию, а также рассмотрены операции над нечеткими множествами, такие как объединение, пересечение и дополнение, и введено понятие нечеткой логической связки для формулировки нечетких правил, кроме того описаны методы нечеткого вывода, такие как нечеткое заключение и нечеткое управление. Другие работы по математическому аппарату нечетких множеств – это [3], где авторы представили нечеткую инференцию на основе правил "если-тогда", и [4], в которой предложено использовать нечеткую инференцию на основе линейных моделей, что расширяет применение нечеткой логики, и методы для нечеткого моделирования и анализа.

В многочисленных исследованиях по разработке методов и моделей для оценки качества проектирования сложных систем с использованием нечетких множеств были описаны модели нечеткой оценки качества проектирования, основанные на принципе нечеткой свертки. Эти модели позволяют учитывать различные критерии и оценивать качество проекта на основе нечетких данных, учитывая неопределенность и неполноту информации. Эффективность процессов проектирования сложных систем рассмотрена работах [5–7], авторы которых подчеркивают необходимость анализа различных факторов, влияющих на эффективность проектирования, таких как стоимость, время, качество и риск, и обсуждают важность учета неопределенности и размытости в этих процессах, что обосновывает применение нечетких методов.

Работы [8, 9] представляют методы моделирования нечетких параметров и нечетких ограничений в процессе проектирования сложных систем, что позволяет учитывать различные уровни неопределенности и нечеткости в проектных решениях. Это особенно важно при работе с системами, где точные значения параметров не могут быть определены или являются субъективными. Авторы демонстрируют применение нечеткой логики для моделирования неопределенности и размытости в процессах проектирования и предлагают использовать нечеткие системы правил для описания связей между различными факторами и критериями эффективности. Описанные в [9] методы оптимизации процесса проектирования сложных систем с использованием нечетких множеств позволяют учитывать различные критерии оптимальности и нечеткие ограничения при выборе наилучших проектных решений. Это способствует повышению эффективности и качества процесса проектирования, принимая во внимание неопределенность и вариативность входных данных.

В работе [10] представлен такой подход к многокритериальной оценке эффективности системы, как анализ иерархий. Этот метод позволяет структуриро-

вать критерии оценки и определять их относительную значимость. Используя парные сравнения, эксперты могут присваивать каждому критерию числовые значения важности и устанавливать их взаимные веса. Нечеткий подход к анализу иерархий позволяет учитывать нечеткость и неопределенность в процессе принятия решений, что было использовано нами при решении задачи формального представления показателей качества.

Такие аналитические методы как линейное программирование, также нашли применение в многокритериальной оценке эффективности. С их помощью можно сформулировать математическую модель, учитывающую критерии и ограничения системы, и найти оптимальное решение, удовлетворяющее заданным критериям. Нечеткие методы линейного программирования расширяют возможности классического линейного программирования, рассматривая нечеткость и неопределенность в модели.

Графические методы, например, метод аналитической сети (ANP), предлагаемый в работе [10], позволяют моделировать сложные взаимосвязи между критериями и элементами системы. ANP представляет собой систему в виде сети, включающей критерии, элементы системы и связи между ними. Эксперты могут присваивать числовые значения взаимозависимостям между элементами и критериями, что позволяет определять важность каждого элемента и его вклад в общую эффективность системы. В настоящей статье мы использовали метод представления данных на основе лепестковой диаграммы с учетом нечеткого представления данных.

Нечеткие методы многокритериальной оценки предлагают более гибкое моделирование и анализ сложных систем. Они позволяют учитывать неопределенность и нечеткость при определении важности критериев, взаимосвязей между элементами системы, и более того, использовать нечеткие правила и экспертные знания в процессе оценки эффективности. Нечеткие методы нашли широкое применение в оценке эффективности процессов проектирования сложных систем, а также были использованы нами при построении функции принадлежности ресурсов сложной системы и вывода многофакторного интегрального показателя качества сложной системы. Это, в свою очередь, позволило учесть неопределенность и нечеткость при моделировании и анализе системы, что особенно полезно в случаях, когда доступная информация о системе является неполной или неоднозначной. Нечеткие методы предусматривают множество критериев и ограничений при оценке эффективности системы, что делает их гибкими и применимыми для различных областей и ситуаций.

В настоящей статье представлены результаты исследования процессов проектирования бортовой радиоэлектронной аппаратуры космических аппаратов для оценки их эффективности. Нахождение критериев оценки эффективности очень сложная и нетривиальная задача. Для этого мы использовали подход деления процесса на области ресурсов и оценки каждого ресурса в ходе проектной деятельности на основе аппарата нечетких множеств. Взаимное влияние ресурсов позволило нам выявить критические точки и определить степень использования каждого ресурса для повышения эффективности проектирования бортовой радиоэлектронной аппаратуры космических аппаратов. Результаты нашей работы были экстраполированы на другие области использования сложных систем.

АНАЛИЗ МЕТОДОВ ОЦЕНКИ ЭФФЕКТИВНОСТИ ПРОЦЕССОВ ПРОЕКТИРОВАНИЯ СЛОЖНЫХ СИСТЕМ

Для выбора метода оценки эффективности процессов проектирования сложных систем был проведен анализ существующих методов многокритериальной оценки систем с использованием нечетких множеств. Одним из распространенных методов является метод нечеткой аддитивной оценки (Fuzzy Additive Weighting, FAW) [11], основанный на присвоении весов каждому критерию и суммировании оценок, умноженных на соответствующие веса. Применение нечетких множеств позволяет учесть неопределенность и нечеткость при определении весов критериев и оценке их значимости. Данный метод был выбран нами как основополагающий для дальнейших исследований.

Другой метод, который можно использовать для оценки эффективности процессов проектирования сложных систем, – это метод нечеткой топсис-оценки (Fuzzy Technique for Order of Preference by Similarity to Ideal Solution, Fuzzy TOPSIS) [12]. В этом методе каждый альтернативный проект сравнивается с идеальным и анти-идеальным проектами по ряду критериев. Альтернативы ранжируются на основе их близости к идеальному и анти-идеальному решениям. Для учета нечеткости и неопределенности при оценке используются нечеткие множества. Однако данный метод не может быть применен для анализа сложных систем с неоднородной логикой, поэтому мы приняли его как альтернативный и не применимый в наших исследованиях.

Метод нечеткого анализа иерархий (Fuzzy Analytic Hierarchy Process, FAHP) [13] комбинирует идеи анализа иерархий и нечеткой логики, позволяя определять относительную значимость критериев и принимать решение на основе их весов. FAHP учитывает нечеткость и неопределенность в процессе принятия решений и может быть полезен при оценке эффективности процессов проектирования сложных систем. Некоторые идеи этого метода были нами использованы в комбинации с методом FAW, что дало возможность повысить точность результирующих прогнозов.

Метод нечеткой компромиссной оценки (Fuzzy Compromise Assessment, FCA) [14] основывается на поиске решения, предусматривающего нечеткие предпочтения и ограничения. Он позволяет находить оптимальные решения, удовлетворяющие требованиям по различным критериям, с учетом нечеткости и неопределенности, однако требует наличия четких критериев, что в данном случае затруднительно.

Альтернативой методу FCA является метод нечеткой взвешенной суммы (Fuzzy Weighted Sum, FWS) [15]. В этом методе каждый критерий умножается на его вес, а затем значения суммируются для получения итоговой оценки. Веса критериев могут быть определены на основе экспертных оценок или

других методов. Нечеткие множества используются для учета неопределенности и нечеткости при определении весов и оценке критериев. Этот метод был нами апробирован, но в связи со сложностью реализации отложен до рассмотрения в дальнейших научных исследованиях.

Нечеткие множества используются для формулирования нечетких правил и оценки выходных значений. Метод нечеткого принятия решений (Fuzzy Decision Making, FDM) [16] позволяет учесть неопределенность и нечеткость за счет принятия решений на основе нечетких правил, связывающих входные переменные с выходными.

При выборе метода оценки эффективности процессов проектирования сложных систем на основе практических изысканий мы получили симбиоз на базе рассмотренных методов, и все дальнейшие выкладки основываются на указанных теоретических наработках.

ФОРМАЛЬНОЕ ПРЕДСТАВЛЕНИЕ ПОКАЗАТЕЛЕЙ КАЧЕСТВА

Объектом исследования мы выбрали процесс проектирования сложных систем математическим аппаратом нечетких множеств и провели попытку оценки эффективности процесса проектирования на основе аппарата нечетких множеств. Практическим применением является процесс проектирования бортовой радиоэлектронной аппаратуры (РЭА) космического аппарата (КА), рассмотренный в [17]. Однако обобщенная оценка, приведенная в настоящей статье, возможна для различных областей использования сложных систем.

Производство бортовой РЭА представляет собой комплекс взаимосвязанных процессов, посредством которых из сырьевых ресурсов и электрорадиоизделий человеком создаются необходимые радиоэлектронные модули, предназначенные в дальнейшем для их испытаний или эксплуатации в составе космических аппаратов. Проектирование, а также технологии испытаний и производства, являются частями сложного процесса создания бортовой РЭА и не могут выполняться в отдельности, без учета взаимосвязей между собой и с другими этапами разработки. В итоге именно они определяют общие свойства изделий. Наличие параллельных и разнородных процессов требует разработки и развития оригинальных принципов и методов их оценки на основе методов динамического имитационного моделирования. При решении данной задачи один из центральных вопросов – это формирование показателей качества бизнес-процессов проектирования и испытаний бортовой РЭА космических аппаратов.

Предлагаемый нами подход построен на комплексной системе оценки показателей качества бизнес-процессов жизненного цикла (ЖЦ) проектирования, испытаний и изготовления (создания) бортовой РЭА и представляет собой иерархическую взаимосвязанную структуру, на нижнем уровне которой находятся относительные характеристики затрат ресурсов как каждого этапа создания бортовой РЭА, так и каждого подразделения, участвующего в отдельных этапах. При этом на одном этапе ЖЦ может быть задействовано несколько подразделений, а также любое подразделение может участвовать в нескольких этапах ЖЦ.

Следующий уровень структуры отражает эффективность не только какого-либо подэтапа ЖЦ создания бортовой РЭА, но и учитывает взаимосвязь между элементами и вклад каждого показателя в интегральную характеристику эффективности процессов проектирования, испытаний и изготовления РЭА космических аппаратов.

Конкретизация видов используемых ресурсов, необходимых для проектирования КА, дает следующую иерархическую структуру:

- верхний уровень – материальные R_1, трудовые R_2, финансовые R_3, временные R_4, аппаратные R_5 и программные R_6 ресурсы;
- средний уровень детализирует верхнюю страту: $R_{1,1}$, $R_{1,2}$ – аппаратные и программные средства, соответственно. При этом к аппаратным ресурсам можно отнести станки с ЧПУ, ЭВМ, коммуникационные системы и сети. Аналогично, программные ресурсы можно детализировать на: PDM, CAE, САПР 3D и т.п.

Для трудовых ресурсов необходимо определить квалификацию работника, его харизму, мотивацию, а также степень новизны заданий.

На рис. 1 представлена иерархическая структура ресурсов (среднего уровня иерархии). Предлагаемый нами подход позволяет шире использовать показатели качества в задачах оптимизации производственных процессов, оценки эффективности созданного изделия, а также интегрировать предлагаемую систему показателей с логико-динамической моделью процессов проектирования космических аппаратов.

Аккумулирующее влияние всех факторов можно вычислить с помощью правила дуальности. Мы предлагаем ввести следующие относительные показатели затрат ресурсов:

$$R_k = \frac{r_k}{R_{норм,k}}, \qquad (1)$$

где: r_k – фактическое потребление k-го ресурса,
$R_{норм,k}$ – нормативное потребление k-го ресурса.

Если $R_k = 1$, то затраты k-го ресурса соответствует нормативу. В формуле (1) при $R_k < 1$ текущие затраты меньше нормативного времени.

Для интегральной оценки качества проектируемой системы (с учетом значимости k-го ресурса) можно воспользоваться взвешенной характеристикой

$$R = \frac{\sum_{k=1}^{K} w_k R_k}{K}, \qquad (2)$$

где: w_k – весовой коэффициент значимости k – го относительного показателя затрат ресурсов;

K – общее количество показателей затрат ресурсов.

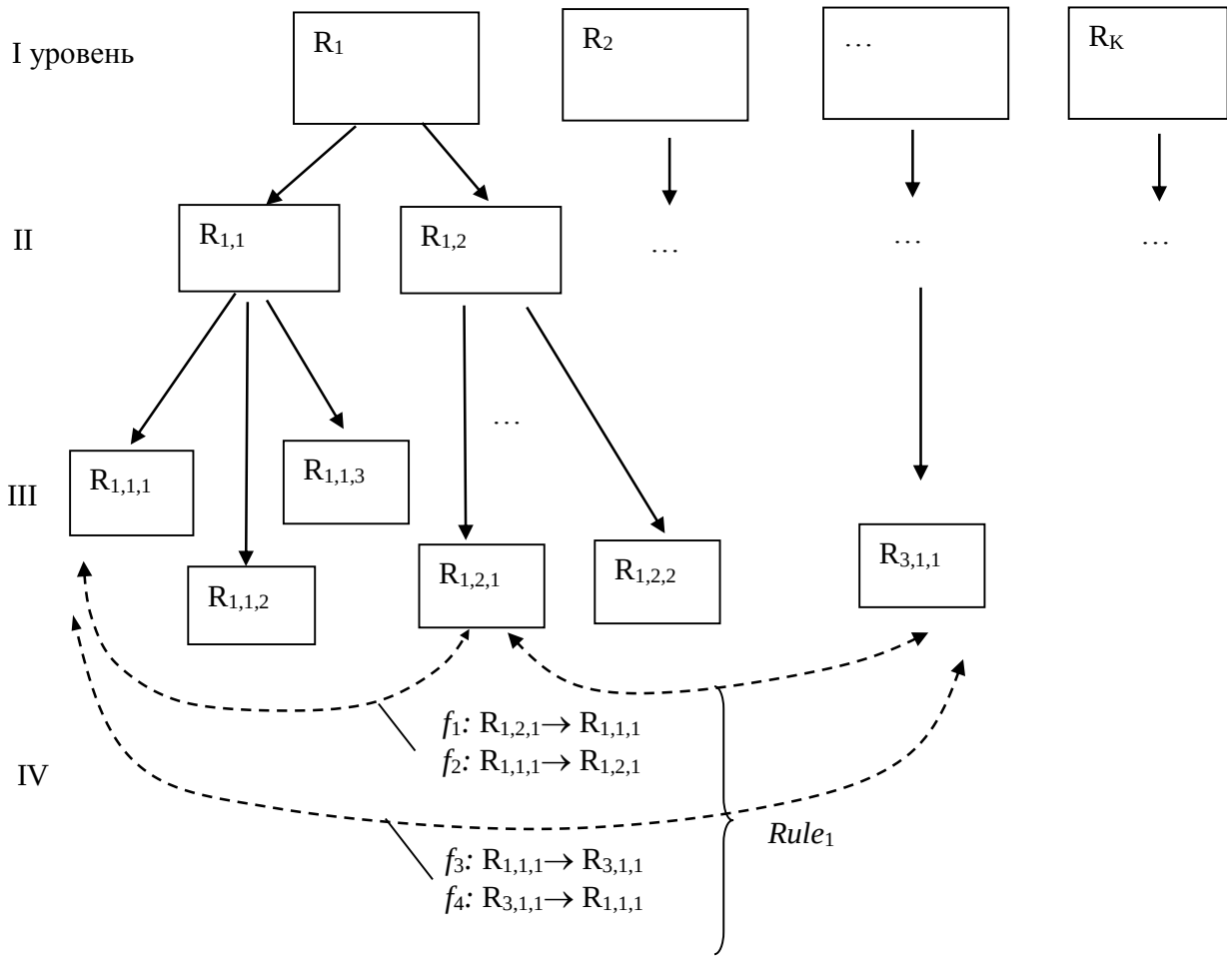

Рис. 1. Структура и влияние ресурсов между собой
(Пунктирными линиями отмечено влияние одного фактора на другой. Закон влияния одного фактора на другой вычисляется с помощью продукционных нечетких правил f_i: $R_j \rightarrow R_i$, где i-й ресурс влияет на значение ресурса j. Причем, взаимодействие может быть дуальным)

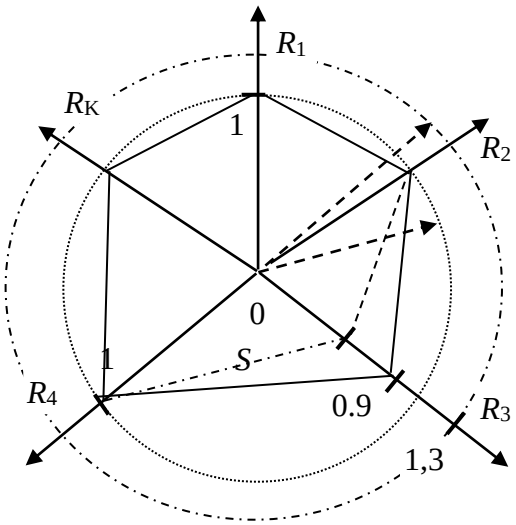

Рис. 2. Лепестковая диаграмма относительных показателей качества

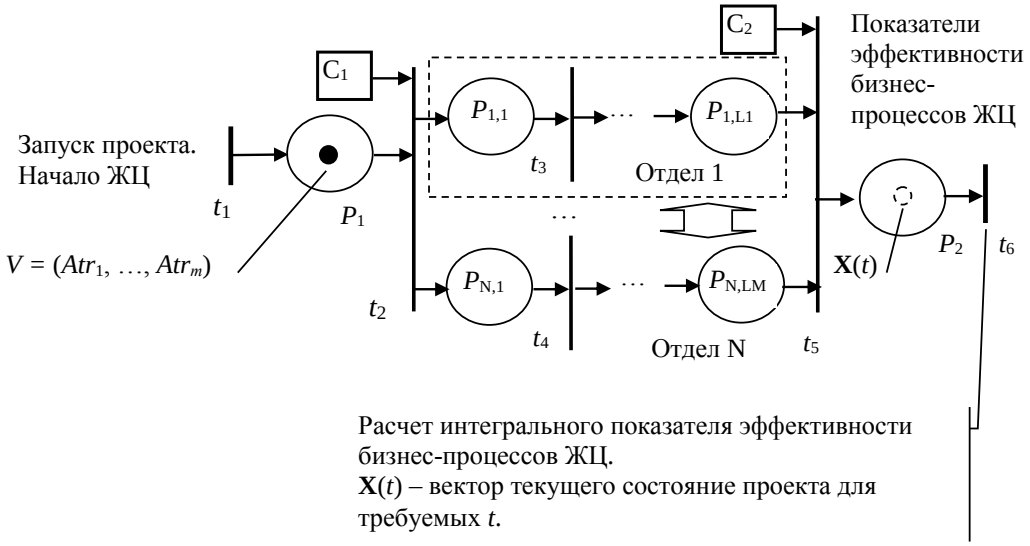

Рис. 3. Обобщенная Е-сетевая логико-динамическая модель выполнения проектно-конструкторских работ

Для визуального представления показателей качества проектирования РЭА космических аппаратов целесообразно построить лепестковую диаграмму (рис. 2), на которой по координатным осям откладываются относительные показатели качества R: чем меньше площадь S полученной фигуры, тем выше эффективность проектирования этапов космических аппаратов.

Представленные показатели качества характеризуются полнотой оценки используемого ресурса, простотой вычисления и наглядностью, но они не позволяют учитывать протекание взаимодействующих параллельных процессов при проектировании РЭА космических аппаратов, что необходимо для формирования рекомендаций эффективности взаимодействия команды разработчиков в процессе проектной деятельности.

С помощью модифицированных сетей Петри – *E*(evaluation)-сетей [17, 18] нами реализована логико-динамическая модель жизненного цикла проектно-конструкторской стадии создания проекта, представленная на рис. 3.

Е-сетевая иерархическая модель позволяет принять во внимание динамические характеристики с учетом параллельных, взаимодействующих между отделами, проектных этапов, что включает этапы (статус выполненных) работ в организации, отделе, лаборатории, а также используемые ресурсы. Результатом Е-сетевой модели являются оценки эффективности каждого ресурса во времени и расчёт интегрального показателя эффективности бизнес-процессов ЖЦ. Кроме того, эта модель отражает временную диаграмму работы организации с необходимой степенью детализации.

На рис. 3 срабатывание переходов t_1 и t_2 соответствует началу работы над проектом и распределению задач и ресурсов между отделами. Срабатывание переходов t_3 и t_4 соответствует выполнению определенного этапа работы. Е-сетевая модель учитывает параллельные процессы при создании проектируемой системы (иллюстрируют данные процессы переходы t_3, t_4). Кроме того, требуемая детализация модели позволяет учесть взаимодействия между отделами и смежными организациями. Время срабатывания Е-сетевых переходов происходит за конечное время Δt и определяется фактическим временем выполнения каждой операции (этапа). Позиции P_i отражают этапы проекта. Передача информации в Е-сетях осуществляется с помощью фишек V_i. Фишка сети V содержит в качестве атрибутов искомые ресурсы R_k. Компоненты вектора R_k включают в себя материальные ресурсы R_1, трудовые R_2, финансовые R_3, временные R_4, аппаратные R_5, программные R_6. Фишка – это динамический элемент Е-сетевой модели, который перемещается из одной позиции в другую в результате срабатывания переходов. Распределение фишек по позициям задает маркировку сети, которую можно трактовать как текущее состояние сети [17, 18]. При срабатывании перехода происходит (при необходимости) изменение атрибута согласно заданному закону.

Для учета взаимосвязей между этими факторами необходимо оценить затраты ресурсов с помощью продукционных нечетких правил R: $X \rightarrow Y$.

Перечень ресурсов определяется кортежем $\mathbf{R} = \{R_1, R_2, ..., R_K\}$, где $\mathbf{R}$ – это совокупность нечетких функций, которые определяют расход каждого ресурса. Каждая функция $f_R(x)$ показывает зависимость между этапом (шагом) и расходом ресурса. Затраты ресурсов на выполнение работы целесообразно представить в виде лингвистической переменной. Мы предлагаем ассоциировать с подмножеством материального ресурса R_1 – уровень сложности работы, $R_1 = \{ R_1^1 =$ «низкий», $R_1^2 =$ «средний», $R_1^3 =$ «высокий»$\}$. Для построения более точной модели можно использовать модальность высказывания, например, «очень» большие затраты и т.п.

Для примера в качестве трудового ресурса R_2 предлагаем нечеткую характеристику профессиональ-

ных навыков R_2 = {«низкие», «средние», «высокие»}. R_4= {низкая скорость выполнения задания», «средняя скорость выполнения задания», «высокая скорость выполнения задания»} – время на осуществление проекта. Пример графического представления функции принадлежности для *k*-го ресурса R_k представлен на рис. 4.

Интегральную характеристику, отражающую эффективность использования ресурсов, в данном случае можно выразить следующим образом:

$$\text{Rule}_i\text{: } if\ x_1 = R_1 \wedge x_2 = R_2 \wedge \ldots, then\ y_1 = B_1 \wedge y_2 = B_2, \quad (3)$$

что может быть представлено в виде нечеткой схемы логического вывода многофакторного интегрального показателя качества, представленного графически на рис. 5.

В качестве входных переменных мы используем массив относительных показателей материальных, финансовых, людских затрат. Так, на рис. 5 этот массив соответствует переменным x_i. На основании входных данных вычисляется степень истинности сужения каждой нечеткой (лингвистической переменной) $R_k(x_i)$. Далее на основании нечетких продукционных правил Rule(x_i, y_j) осуществляется активизация функций принадлежности $f_B(y)$ и процедура дефаззификации как вычисление центра массы. В итоге нечеткая система вывода сформирует результат в виде набора оценки эффективности $B(y)$. На рис. 6 показана детализация указанных принципов нечеткого логического вывода оценки эффективности процессов проектирования сложных систем.

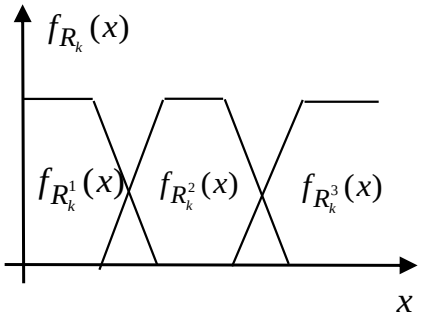

Рис. 4. Функции принадлежности для ресурсов:
$f_{R_k^1}(x)$ соответствует функции принадлежности с низким уровнем затрачиваемого *k*-го ресурса;
$f_{R_k^2}(x)$ и $f_{R_k^3}(x)$ соответствуют функции принадлежности со средним и высоким уровнями затрачиваемых ресурсов

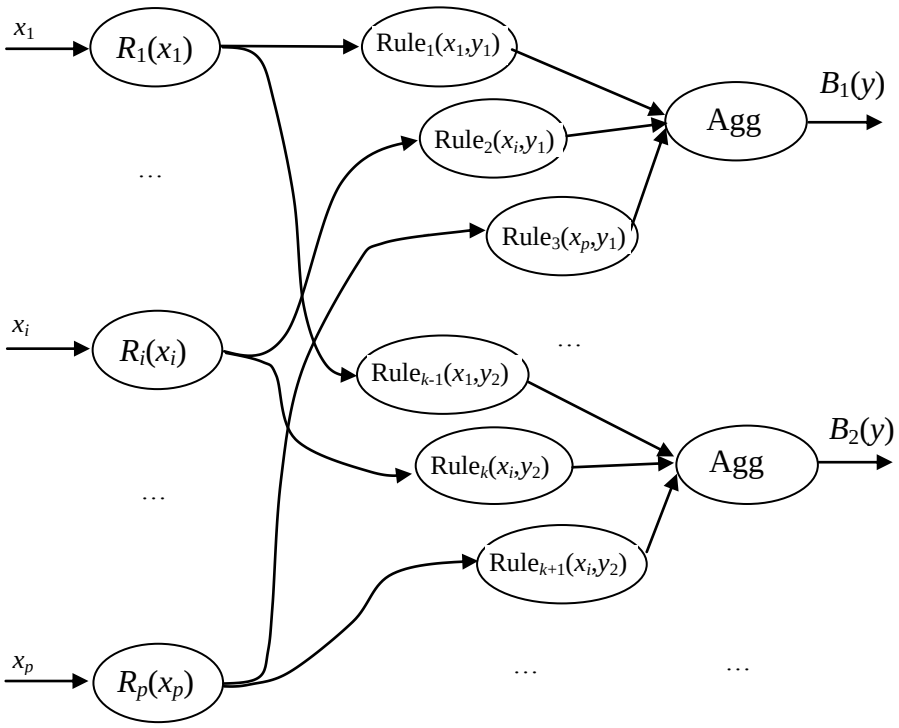

Рис. 5. Нечеткая схема логического вывода многофакторного интегрального показателя качества проектирования космических аппаратов

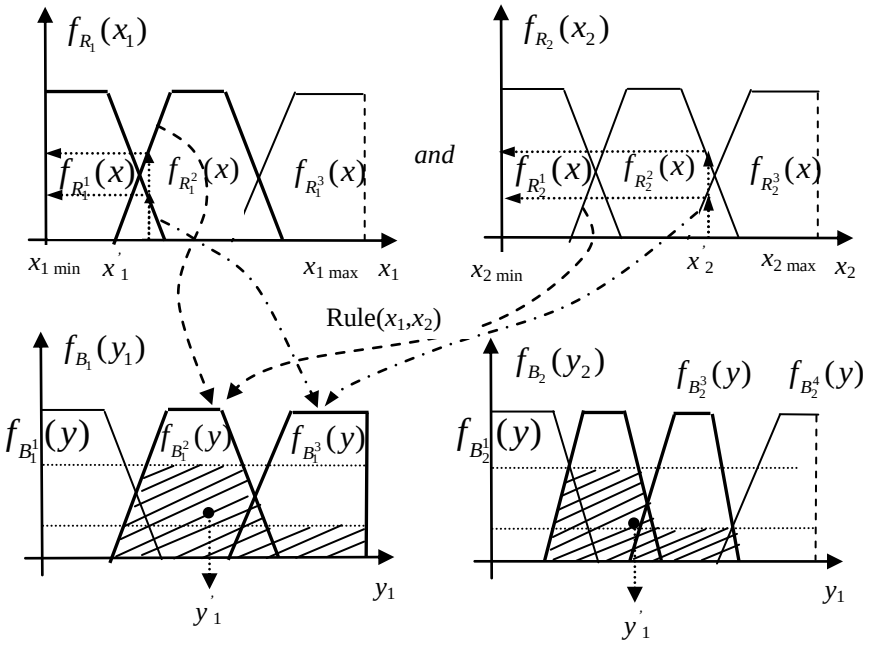

Рис. 6. Процедура вычисления эффективности проектирования космических аппаратов и систем (в качестве входных переменных представлены численные значения ресурсов x'_1 и $x'_{2)}$

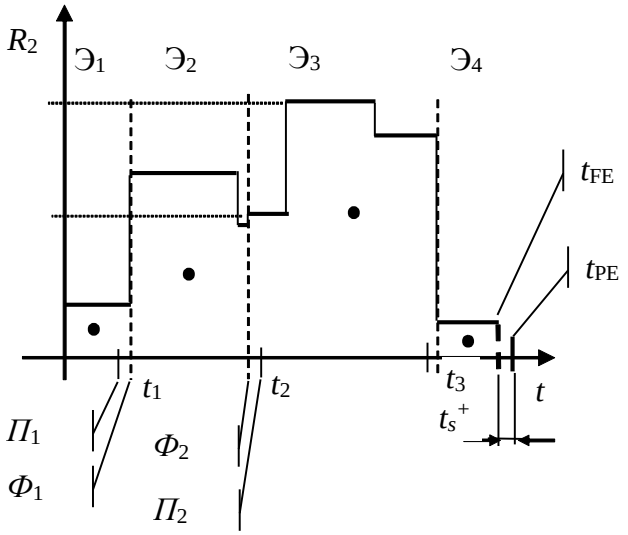

Рис. 7. Эффективность каждого этапа проекта
(П1 и Ф1 – плановое и фактическое выполнение первого этапа; переменные: t_{FE}, t_{PE} – фактическое и плановое окончание проекта; t_s^+, t_s^- – суммарное фактическое время выполнения проектирования и изготовление изделия, индекс «+» или «-» указывает на досрочное или несвоевременное выполнение проекта)

На основании функции принадлежности $f_{R_1}(x_1)$ и $f_{R_2}(x_2)$ вычисляется значение степени истинности суждения по каждому из ресурсов. Правило Rule(x_1, x_2) активизирует заданные оценки эффективности системы и по процедуре дефаззификации вычисляются численные характеристики оценок y'_1 и y'_2 по формуле вычисления центра массы:

$$y' = \frac{\int_{t1}^{t2} ty(t)dt}{\int_{t1}^{t2} y(t)dt}. \qquad (4)$$

В качестве иллюстрации на рис. 7 отражена эффективность трудовых ресурсов R_2. Процедура проектирования системы, согласно плану работ, включает в себя четыре этапа Э1– Э4.

Легко выявить относительное рассогласование между фактическим и планируемым временем выполнения каждого этапа проекта, а также относительное рассогласование времени начала и окончания работ. Эти характеристики позволяют адаптировать взаимодействие различных структурных подразделений и отдельных сотрудников для минимизации используемых ресурсов при проектировании и создании системы, а также для увеличения надежности готового изделия.

Предлагаемые нами модели и подходы позволяют повысить эффективность процесса проектирования и изготовления радиоэлектронной аппаратуры космических аппаратов, выполнить оценку эффективности бизнес-процесса, выявить «узкие» места и настроить реинжиниринг процесса до требуемых параметров эффективности.

ЗАКЛЮЧЕНИЕ

В настоящей статье описана попытка оптимизации бизнес-процессов проектирования сложных систем на основе оценки эффективности процессов проектирования математическим аппаратом нечетких множеств. Практическая реализация данного подхода возможна на основе аппарата раскрашенных сетей Петри. Для дальнейшей реализации мы предлагаем использовать бизнес-пакет AnyLogic. Эти исследования могут быть транспонированы и адаптированы для различных областей использования.

Нечеткие методы могут быть применены для оценки эффективности процессов проектирования сложных систем в различных областях, таких как промышленное производство, транспорт, энергетика и др. Одним из примеров применения нечетких методов многокритериальной оценки в проектировании сложных систем является оценка эффективности экологических проектов. В таких проектах следует учитывать различные экологические критерии, такие как загрязнение воздуха, выбросы водных веществ, энергопотребление и другие. Нечеткие методы позволяют моделировать нечеткость и неопределенность в оценке каждого критерия и определять важность каждого критерия с учетом экспертных знаний. Это позволяет выбирать наилучшие экологические решения и оценивать их эффективность.

Еще один пример – это применение нечетких методов в оценке качества продукции. При проектировании сложных систем, таких как автомобили, электроника или промышленное оборудование, необходимо учитывать множество критериев, включая надежность, производительность, стоимость и другие, определить важность каждого критерия и в дальнейшем принять решение на основе нечетких правил для улучшения качества продукции и удовлетворения запросов потребителей.

Нечеткие методы многокритериальной оценки могут быть применены и в задачах проектирования интеллектуальных систем управления транспортными сетями, включая газо- и нефтепроводы. В таких системах должны учитываться многие критерии, включая эффективность использования ресурсов, минимизацию задержек, улучшение качества, контроля целостности и многие другие. Нечеткие методы могут помочь моделировать неопределенность и нечеткость в оценке каждого критерия, а также определить веса критериев на основе предпочтений и экспертных знаний. Это позволяет разработать оптимальные стратегии управления транспортными сетями и повысить их эффективность.

Кроме всего указанного нечеткие методы многокритериальной оценки в области финансового планирования и портфельного управления позволяют при выборе инвестиционных проектов учитывать такие критерии, как доходность, риск, ликвидность и диверсификация. Нечеткие методы позволяют учесть неопределенность и нечеткость в оценке каждого критерия, а также определить важность каждого критерия на основе предпочтений инвестора. Это помогает принимать обоснованные решения о составлении инвестиционного портфеля и достижении желаемых финансовых целей.

Таким образом, применение нечетких методов многокритериальной оценки позволяет решать сложные задачи в различных областях, таких как экология, финансы, медицина, транспорт и другие. Они способствуют улучшению эффективности, качества и устойчивости проектируемых систем, а также помогают снижать риски и принимать обоснованные решения.

Важно отметить, что нечеткие методы многокритериальной оценки имеют свои преимущества и ограничения. Они требуют качественного описания нечеткости и экспертных знаний, а также подходящего выбора математических моделей и алгоритмов. Поэтому выбор метода должен основываться на конкретной задаче и ее особенностях.

Развитие нечетких методов многокритериальной оценки будет направлено на повышение точности и эффективности моделирования нечеткости и взаимосвязей между критериями, а также на учет новых видов данных и контекстов. Интеграция нечетких методов с другими интеллектуальными методами и технологиями, такими как искусственный интеллект и анализ данных, также представляет собой перспективное направление исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Fuzzy Sets and Decision Analysis. Editors H.J. Zimmermann, L.A. Zadeh, A.R. Gaines. – North Holland, New York, 1984
2. George J. Klir, Bo Yuan. Fuzzy sets and fuzzy logic: theory and applications. Prentice Hall; Upper Saddle River, NJ; 1995. – 574 p.
3. Демидова Г.Л., Лукичев Д.В. Регуляторы на основе нечеткой логики в системах управления техническими объектами. – Санкт-Петербург: Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2017. – 81 с.
4. Зиновьев И.П., Аникин И.В. Модель логического вывода на основе нечеткой линейной регрессии // Информатика, телекоммуникации и управление. – 2010. – № 5 (108). – URL: https://cyberleninka.ru/article/n/model-logicheskogo-vyvoda-na-osnove-nechetkoy-lineynoy-regressii (дата обращения: 11.05.2023).
5. Борисов В.В., Луферов В.С. Метод многомерного анализа и прогнозирования состояния сложных систем и процессов на основе нечетких когнитивных темпоральных моделей // Системы управления, связи и безопасности. – 2020. – № 2. – URL: https://cyberleninka.ru/article/n/metod-mnogomernogo-analiza-i-prognozirovaniya-sostoyaniya-slozhnyh-sistem-i-protsessov-na-osnove-nechetkih-kognitivnyh-temporalnyh (дата обращения: 11.05.2023).

6. Птускин А.С., Левнер Е., Жукова Ю.М. Многокритериальная модель определения наилучшей доступной технологии при нечетких исходных данных // Вестник МГТУ им. Н.Э. Баумана. Сер. Машиностроение. – 2016. – № 6(111). – URL: https://cyberleninka.ru/article/n/mnogokriterialnaya-model-opredeleniya-nailuchshey-dostupnoy-tehnologii-pri-nechetkih-ishodnyh-dannyh (дата обращения: 11.05.2023).
7. Pokushko M., Stupina A., Medina-Bulo I. et al. Applying the Data Envelopment Analysis method for evaluating the efficiency of the complex system operations in fuel and energy companies // JOP Conference Series: Metrological Support of Innovative Technologies, Krasnoyarsk, 04 марта 2020 года / Krasnoyarsk Science and Technology City Hall of the Russian Union of Scientific and Engineering Associations. – Krasnoyarsk, Russia: Institute of Physics and IOP Publishing Limited, 2020. – Vol. 1515. – P. 52039. – DOI 10.1088/1742-6596/1515/5/052039. – EDN ASRSOI.
8. Naderpajouh N., Afshar A., Nasirzadeh F. Fuzzy Approach to Project Delivery System Selection // 03rd International Project Management Conference, Symposia, Tehran. – 2007. DOI:10.13140/RG.2.1.1164.4326.
9. Полозюк О.Е. Оптимизация технологических процессов с использованием аппарата теории нечетких множеств // ГВУЗ «Приазовский государственный технический университет». – 2001. – № 11. – URL: https://cyberleninka.ru/article/n/optimizatsiya-tehnologicheskih-protsessov-s-ispolzovaniem-apparata-teorii-nechetkih-mnozhestv (дата обращения: 11.05.2023).
10. Saaty T.L. Decision making with the analytic hierarchy process // International Journal of Services Sciences. – 2008. – Vol. 1, № 1. – P. 83-98. DOI: 10.1504/IJSSCI.2008.017590.
11. Modarres M., Sadi-Nezhad S. Fuzzy Simple Additive Weighting Method by Preference Ratio // Intelligent Automation & Soft Computing. – 2005. – Vol. 11(4). – P. 235-244. DOI: 10.1080/10642907.2005.10642907.
12. Balwinder Sodhi, Prabhakar Tadinada. A Simplified Description of Fuzzy TOPSIS // Computer Science. Artificial Intelligence. *ArXiv*. – 2012. *n. pag*. DOI: 10.48550/arXiv.1205.5098.
13. Ali Emrouznejad, William Ho. Fuzzy Analytical Hierarchy Process. – New York: Chapman and Hall/CRC, 2017 – 430 p. DOI:10.1201/9781315369884.
14. Yang Kyoung-Mo, Kim Eung-Hee, Hwang, Suk-Hyung, Choi Sung-Hee. Conceptual analysis of fuzzy data using FCA // Proceedings of the 8th WSEAS International Conference on APPLIED COMPUTER SCIENCE (ACS'08). – 2008. P. 37-42.
15. Lu P., Jiang X. Fuzzy Bidirectional Weighted Sum for Face Recognition // The Open Automation and Control Systems Journal. – 2015. – Vol. 6. – P. 447-452.
16. Blanco-Mesa Fabio, Gil-Lafuente Anna, Merigo Jose M. Fuzzy decision making: A bibliometric-based review // Journal of Intelligent and Fuzzy Systems. – 2017. – Vol. 32. – P. 2033-2050. DOI: 10.3233/JIFS-161640.
17. Braginsky M., Tarakanov D., Tsapko S. E-network modelling of process industrial control systems in building computer simulators // 2016 International Siberian Conference on Control and Communications (SIBCON). – 2016. – P. 1-5. DOI: 10.1109/SIBCON.2016.7491659.
18. Braginsky M.Ya., Tarakanov D.V., Tsapko S.G. Hierarchical analytical and simulation modelling of human-machine systems with interference // Journal of Physics: Conference Series (Information Technologies in Business and Industry (ITBI-2016) : International Conference, 21–26 September 2016, Tomsk). – 2017. – Vol. 803 – P. 120-126.

Материал поступил в редакцию 31.07.23.

Сведения об авторах

ЦАПКО Сергей Геннадьевич – кандидат технических наук, доцент, доцент отделения информационных технологий Инженерной школы информационных технологий и робототехники Томского политехнического университета.
e-mail: tsapko@tpu.ru

ЦАПКО Ирина Валериевна – кандидат технических наук, доцент, доцент отделения информационных технологий Инженерной школы информационных технологий и робототехники Томского политехнического университета.
e-mail: tsiv@tpu.ru

ТАРАКАНОВ Дмитрий Викторович – кандидат технических наук, доцент, доцент кафедры автоматики и компьютерных систем Сургутского государственного университета
e-mail: sprtdv@mail.ru

УДК 004.451.5/.7.032.32/.34:519.872

В.И. Сумин, Е.Д. Грачев, Ю.Ю. Громов, В.М. Тютюнник

Математические модели определения времени обработки запросов на серверах информационных систем специального назначения

Для анализа функционирования серверов информационной системы специального назначения при различных режимах работы предложены математические модели, в основу формирования которых положены методы теории систем массового обслуживания. Эти модели позволяют анализировать функционирование серверов системы в двух режимах обработки запросов – синхронном и асинхронном – и оценивать время ожидания запросов, которые находятся в очереди.

Ключевые слова: *математические модели, сервер, информационные системы специального назначения, синхронная и асинхронная обработка запросов, системы массового обслуживания, функция распределения*

DOI: 10.36535/0548-0027-2023-10-2

ВВЕДЕНИЕ

В последнее время российские учреждения специального назначения начали создавать вычислительные сети большой размерности, обслуживающие рабочие места тысяч сотрудников по всей стране. При этом базы данных (БД) всех структурных подразделений каждого учреждения интегрируются в единое информационное пространство, что приводит к необходимости создания серверов для сложных информационных систем специального назначения (СИССН). Запросы от сотрудников таких систем, направленные на серверы структурных подразделений учреждений, часто приводят к их перегрузке и к многократному увеличению времени обработки запросов, а в результате – к снижению эффективности СИССН [1–3]. Существует ещё и проблема, для решения которой необходимо организовать управление рабочей нагрузкой серверов, обслуживающих БД, чтобы они эффективно участвовали в обработке запросов и учитывали особенности их индивидуальных характеристик [4–8]. В связи с этим целесообразно рассмотреть задачу определения времени обработки потока функционирующих в СИССН запросов к серверам.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ФУНКЦИОНИРОВАНИЯ СЕРВЕРА СИССН С СИНХРОННОЙ ОБРАБОТКОЙ ЗАПРОСОВ

Рассмотрим режим функционирования одного сервера сложной информационной системы специального назначения, когда сотрудник будет ожидать выполнения своего запроса только после того, как выполнен предыдущий запрос, находившийся в очереди. Такой режим наиболее распространён для web-технологий, которые используют серверы большинства СИССН.

Предложим вариант, когда СИССН имеет один сервер и N сотрудников, т. е. для сотрудников этой системы будет создаваться очередь запросов на сервер.

Условия функционирования серверов:

- каждый запрос сотрудника СИССН обслуживается на сервере в единой очереди, и этот запрос будет ждать обслуживания в порядке поступления на этот сервер;
- каждый сотрудник СИССН работает независимо, запросы сотрудников СИССН не теряются и смогут ждать обслуживания сколько нужно, а обслуживающий сервер СИССН не простаивает только в том случае, когда очереди запросов от сотрудников СИССН уже созданы;
- запросы сотрудников СИССН, которые уже обслужены, сразу покидают сервер и возвращаются на этот сервер, где стоят в очереди некоторое время, которое определяется случайной величиной, и только после этого снова поступают на этот сервер для обслуживания.

Будем считать, что вероятность численного значения времени обслуживания запросов определяется функцией распределения $G(t)$, а вероятность численного значения времени обслуживания запросов одинакова для запроса каждого сотрудника СИССН. Время обслуживания одного запроса считаем случайной величиной $\beta(k)$с заданной функцией распре-

деления $\beta(t,k)$, где: k – количество запросов, заблаговременно обслуживаемых на сервере; t – время обслуживания одного запроса. Каждая случайная величина имеет первый и второй моменты.

Под состоянием СИССН будем понимать количество запросов сотрудников этой системы к БД $m(0 \le m \le N-1)$ и количество запросов, которые обрабатываются на сервере $n(0 \le n \le K)$, где К – общее количество запросов. Таким образом, состояние системы задается парой (m,n). Будем рассматривать систему в момент окончания обработки сервером какого-либо запроса. Стационарную вероятность состояния СИССН обозначим как $p_{m,n}$. Поскольку система имеет конечное множество состояний, то существуют стационарные значения вероятностей состояний.

Приведём уравнения, которые показывают вероятности состояний СИССН:

$$p_{0,0} = p_{0,0}z_0(N-1,1) + p_{0,1}z_0(N-1,1)$$

$$p_{0,j} = p_{0,0}z_j(N-1,1) + \sum_{i=1}^{j+1} p_{0,i}z_{j-i+1}(N-i,i)$$

…….

$$p_{0,K-1} = p_{0,0}z_{K-1}(N-1,1) + \\ + \sum_{i=1}^{K} p_{0,i}z_{K-1-i+1}(N-i,i)$$

$$p_{0,K} = p_{0,0}z_K(N-1,1) + \sum_{i=1}^{K} p_{0,i}z_{K-i+1}(N-i,i) + \\ + p_{1,K}z_0(N-K,K)$$

……. (1)

$$p_{m,K} = p_{0,0}z_{K+m}(N-1,1) + \\ + \sum_{i=1}^{K} p_{0,i}z_{K-i+1+m}(N-i,i) + \\ + \sum_{i=1}^{m+1} p_{i,K}z_{m-i+1}(N-i-K,K)$$

…….

$$p_{N-K-1,K} = p_{0,0}z_{N-1}(N-1,1) + \\ + \sum_{i=1}^{K+1} p_{0,i}z_{N-i}(N-i,i) + p_{0,K}z_{N-K}(N-K,K) + \\ + \sum_{i=1}^{N-K-1} p_{i,K}z_{N-K-i}(N-K-i,K), j = \\ = 1,1,\dots,(K-1); m = 1,2,\dots(N-K-2),$$

где: $z_k(r,n)$ – вероятность того, что за время обслуживания запроса, поступившего на сервер от любого сотрудника СИССН, в очередь поступят k новых запросов при условии, что занято подготовкой запросов к БД r сотрудников ($r > k$) и на обслуживании уже находится n запросов к серверу планировщику.

Система линейных алгебраических уравнений (СЛАУ) (1) имеет N решений с нулевыми свободными членами. Если применить условие нормировки, то $\sum_{m=0}^{N-1}\sum_{n=1}^{K} p_{m,n} = 1$ при выполнении равенства $\sum_{k=0}^{r} z_k(r,n) = 1$ для всех $r = 1,2,\dots,(N-1)$. Такая СЛАУ решается любыми известными методами прикладной математики [2].

Для определения значения коэффициента $z_k(r,n)$ необходимо использовать допущение об однородности и независимости сотрудников СИССН, тогда:

$$z_k(r,n) = C_r^k q^k(r,n)\big(1-q(r,n)\big)^{r-k}, \quad (2)$$

где: $q(r,n)$ – вероятность того, что за время до окончания обслуживания запроса сотрудника на сервере, где обслуживается n запросов, закончится подготовка и будет сформирован и послан новый запрос одного из r сотрудников СИССН.

Учитывая допущение (2), получим:

$$z_k(r,n) = \int_0^{\infty} C_r^k G^k(t)\big(1-G^*(t)\big)^{r-k} dB(t,n), \quad (3)$$

где: $B(t,n)$ – функция распределения продолжительности обработки запроса, считая, что на сервере обрабатывается n запросов;

$G^*(t)$ – функция распределения продолжительности обработки запроса от начала обработки на этом сервере до его окончания:

$$G^*(t) = g_1^{-1}\int_0^t \big(1-G(\tau)\big)d\tau, \quad g_1 = \int_0^{\infty} dtG(t). \quad (4)$$

Если запросы обрабатываются в порядке поступления на сервер и его производительность обратно пропорциональна количеству одновременно обслуживаемых запросов n, то получим:

$$B(s,n) = \beta^n(s,1), \quad (5)$$

где: $\beta(s,n)$ – преобразование Лапласа-Стилтьеса (ПЛС) функции распределения $B(t,n)$;

$\beta(s,1)$ – ПЛС функции распределения только одного запроса.

Если сотрудник СИССН формирует разнотипный поток запросов, то:

$$B(t,n) = \sum_{j=1}^{M}\left(\frac{\sum_{i=1}^{N}\lambda_{ij}}{\sum_{j=1}^{M}\sum_{i=1}^{N}\lambda_{ij}} B_j(t)\right), \quad (6)$$

где: $B_j(t)$ – функция распределения времени, затрачиваемого на обработку запроса типа j;

М – множество однотипных запросов;

λ_{ij} – интенсивность потока запросов типа j от *i-го* сотрудника СИССН.

При нашем подходе $z_k(r,n)$ можно вычислять для всех видов функций распределения случайной величины с использованием численных методов.

Решение (1) позволяет определять стационарные вероятности состояний СИССН как систему массового обслуживания (СМО), а также её средние характеристики:

- время ожидания запросов сотрудников СИССН в очереди к серверу:

$$W = H + R, \quad (7)$$

где:

$$H = \sum_{m=0}^{N-n}\sum_{n=0}^{K}\left(p(m,n)\left\{m\left[\sum_{r=0}^{N-k}\sum_{k=0}^{K} b_1(k)p(r,k)\right]\right\}\right),$$

$$R = \sum_{m=0}^{N-n}\sum_{n=0}^{K}\left(p(m,n)\int_0^{\infty} tB^*(t,n)dt\right),$$

$$B^*(t,n) = b_1^{-1}(n)\int_0^{t}\left(1 - B(\tau,n)\right)d\tau,$$

$$b_1(n) = \int_0^{\infty} tdB(t,n);$$

- общее количество запросов на сервер, Q:

$$Q = \sum_{m=0}^{N-n}\sum_{n=0}^{K} mp(m,n). \quad (8)$$

Выражения (7) и (8) определяют основные характеристики функционирования сервера при синхронной обработке запросов.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ФУНКЦИОНИРОВАНИЯ СЕРВЕРА СИССН С РАЗДЕЛЕНИЕМ РЕСУРСОВ

Модель обработки запросов в асинхронном режиме будем строить с использованием методов теории системы массового обслуживания открытого типа с одним сервером, имеющим несколько входящих потоков. Будем полагать, что существует N сотрудников СИССН, которые через сервер посылают потоки запросов на БД с интенсивностями λ_{ij} (здесь $i = 1,2,...,N$ – индекс сотрудника, $j = 1,2, ..., M$ – индекс типа запроса). Обозначим поток запросов типа j через λ_j.

Время обработки запроса типа j является случайной величиной B_j с функцией распределения $B_j(t)$, с конечными первым и вторым моментами. Значение этой случайной величины определяет обработку только одного типа запроса. Время обработки запроса j-го типа вычисляется после обработки этого очередного запроса на сервере планировщике и равно β_j. При этом интенсивность обработки вновь поступившего запроса зависит от количества уже находящихся на сервере запросов таким образом, что каждый запрос получает равную долю от базовой интенсивности обработки запросов.

Так как все запросы находятся в общей очереди, то q_j – вероятность произвольно выбранного из очереди для обработки запроса типа j – определяется выражением (9):

$$q_j = \frac{\lambda_j}{\sum_{j=1}^{M}\lambda_j}. \quad (9)$$

Время обработки запроса j-го типа вычисляется после обработки этого очередного запроса на сервере. Если обрабатывается n запросов, то время обработки запросов равно $n\beta_j$. Состояние рассматриваемой СИССН определяется общим количеством запросов в момент времени t. Будем рассматривать состояние системы в то время, когда все запросы обработаны сервером. Если система находится в стационарном состоянии, тогда справедлива система уравнений, связывающая вероятности p_n нахождения системы СИССН в состоянии n:

$$\begin{aligned}
&p_0 = p_0 z_0(N-1) + p_1 z_0(1)\\
&p_1 = p_0 z_1(1) + p_1 z_1(1) + p_2 z_0(2)\\
&p_2 = p_0 z_2(1) + p_1 z_2(1) + p_2 z_1(2) + p_3 z_0(3)\\
&\ldots\ldots\\
&p_k = p_0 z_k(1) + p_1 z_k(1) + \sum_{m=2}^{k+1} p_m z_{k-m+1}(m)\\
&\ldots\ldots,
\end{aligned} \quad (10)$$

где: $z_k(m)$ – вероятность поступления k запросов от сотрудников СИССН за время до окончания обработки какого-либо запроса из m находящихся в системе СИССН, т. е. за интервал времени между двумя последовательными окончаниями обслуживании запросов.

Значение $z_k(m)$ определяется выражением:

$$z_k(m) = \int_0^{\infty}\frac{(\Lambda t)^k}{k!}e^{-\Lambda t}d\left[\sum_{r_1=1}^{m}\sum_{r_2=1}^{m}\ldots\sum_{r_m=1}^{m}\alpha_m(r_1,r_2,\ldots,r_M)\left(1-\prod_{j=1}^{M}\left(1-A_j(t)\right)^{r_j}\right)\right] = \int_0^{\infty}\frac{(\Lambda t)^k}{k!}e^{-\Lambda t}dR_m(t), \quad (11)$$

где:

r_j – количество запросов j-го типа, находящихся в СИССН, когда справедливо выражение $\sum_{j=1}^{M} r_j = m$; $\alpha_m(r_1, r_2, \ldots, r_N)$ – вероятность того, что среди m запросов от сотрудников СИССН находится r_1 запросов первого типа, r_2 запросов второго типа и т.д.;

$A_j(t)$ – функция распределения длительности времени до окончания обслуживания запроса типа j, находящегося в системе;

Λ – общая интенсивность поступления запросов в систему от всех сотрудников на запуск любых типов запросов $\Lambda = \sum_{j=1}^{M}\lambda_j = \sum_{i=1}^{N}\sum_{j=1}^{M}\lambda_{ij}$.

Считаем, что

$$B_j(t) = 1 - e^{-\lambda_j t}, \quad (12)$$

тогда

$$R_m(t) = 1 - \left(\sum_{j=1}^{M} q_j e^{\frac{-\beta_j}{m}}\right)^m. \qquad (13)$$

Введём в рассмотрение производящую функцию вероятностей $\pi(x) = \sum_{n=0}^{\infty} x^n p_n$

Преобразуя систему уравнений (10), получим:

$$\pi(x) = p_0 Z_1(x) + p_1 Z_1(x) + \\ + x p_2 Z_2(x) + \sum_{n=3}^{\infty} x^{n-1} p_n Z_n(x), \qquad (14)$$

где:

$$Z_n(x) = \sum_{i=0}^{\infty} x^i z_i(n) = \int_0^{\infty} \sum_{i=0}^{\infty} \frac{(\Lambda t)^i}{i!} x^i e^{-\Lambda t} dR_n(t) = \\ = \int_0^{\infty} e^{-t\Lambda(1-x)} dR_n(t) = p_n\big(\Lambda(1-x)\big), \qquad (15)$$

где: $p_n(.)$ – ПЛС функции распределения $R_n(t)$. Это выражение является достаточно общим, справедливым для различных видов функции $R_m(t)$.

Учитывая выражения (11) и (13), получим:

$$Z_n(x) = \int_0^{\infty} e^{-t\Lambda(1-x)} dR_n(t) = \\ = \int_0^{\infty} e^{-t\Lambda(1-x)} d\left(1 - \left(\sum_{j=1}^{M} q_j e^{\frac{-\beta_j t}{n}}\right)^n\right). \qquad (16)$$

Используя (14), можно вычислить $Z_n(x)$ и $\pi(x)$.

При $\beta_j = \beta$ выражение (16) (длительность обработки запросов для всех сотрудников СИССН является одинаковой) принимает вид:

$$Z_n(x) = \int_0^{\infty} e^{-t\Lambda(1-x)} dR_n(t) = \\ = \int_0^{\infty} e^{-t\Lambda(1-x)} d\left(1 - \left(\sum_{j=1}^{M} q_j e^{\frac{-\beta t}{n}}\right)^n\right) = \qquad (17)$$

$$= \int_0^{\infty} e^{-t\Lambda(1-x)} d\big(1 - e^{-\beta t}\big) = \frac{\beta}{\Lambda(1-x)+\beta} = \mathrm{Z(x)}.$$

Подставляя (17) в (14), получим выражение:

$$\pi(x) = \left(1 - \frac{\Lambda}{\beta}\right) \frac{\beta(1-x)}{\beta - x\Lambda(1-x) - x\beta} \qquad (18)$$

Из выражения (18) можно найти среднее количество запросов в системе СИССН:

$$R_1 = \pi'(x) = \frac{\Lambda}{\beta - \Lambda} \qquad (19)$$

Использование формулы Литтла (связывают среднее время пребывания запроса в очереди и среднее количество запросов в очереди) позволяет определять время ожидания запроса в очереди на сервер. При этом полученные значения времени обработки запросов позволяют управлять нагрузкой сервера СИССН в исследуемых режимах.

ЗАКЛЮЧЕНИЕ

В настоящей работе продемонстрирована возможность определения времени обработки потока запросов к серверам сложной информационной системы специального назначения (СИССН) на основе использования построенных математических моделей статического и динамического распределения запросов от сотрудников этой системы, причём, при различных режимах функционирования:

1) обрабатываются запросы от сотрудников по мере их поступления (синхронная обработка запросов);

2) новые запросы поступают без ожидания обработки предыдущих запросов (асинхронная обработка запросов или многозадачный режим).

Построенные математические модели опираются на использование аппарата теории массового обслуживания и позволяют оценивать время ожидания запросов от сотрудников СИССН в очереди к серверам, что даёт возможность осуществлять эффективное управление нагрузкой серверов при различных индивидуальных характеристиках.

СПИСОК ЛИТЕРАТУРЫ

1. Сумин В.И., Грачев Е.Д., Лукин М.А. Анализ методов управления нагрузкой серверов в распределённых информационных системах большой размерности // Вестник Воронежского института ФСИН России. – 2021. – № 3. – С. 116-124.
2. Сумин В.И., Грачев Е.Д. Математические модели для анализа алгоритмов управления нагрузкой серверов // Вестник Воронежского института ФСИН России. – 2023. – № 1. – С. 144-150.
3. Васкевич Д. Стратегии клиент/сервер. Руководство по выживанию для специалистов по реорганизации бизнеса. – Киев: Диалектика, 1996. – 384 с.
4. Мюллер С. Модернизация и ремонт серверов. – Москва: Диалектика/Вильямс, 2017. – 383 с.
5. Суханов В.В. Аналитическое обеспечение организации данных в распределённых информационных системах критического применения // Моделирование систем и процессов. – 2021. – Т. 14, № 3. – С. 60-67.
6. Суханов В.В., Ланкин О.В. Методика логического проектирования информационного обеспечения распределённых информационных систем критического применения // Моделирование систем и процессов. – 2021. – Т. 14, № 3. – С. 67-73.
7. Сумин В.И., Смоленцева Т.Е., Громов Ю.Ю., Тютюнник В.М. Анализ функционирования и структурная декомпозиция информационных систем специального назначения // Научно-техническая информация. Сер. 2. – 2021. – № 8. – С. 5-14. – DOI: 10.36535/0548-0027-2021-08-2; Sumin V.I., Smolentseva T.E., Gromov Yu.Yu., Tyutyunnik V.M. Analysis of the Functioning and Structural Decomposing of Special-Purpose Information Systems // Automatic Documentation and Mathematical Linguistics. – 2021. – Vol. 55, № 4. – P. 169-177. – DOI: 10.3103/S0005105521040087.

8. Сумин В.И., Громов Ю.Ю., Тютюнник В.М. Оптимизация функционирования информационных систем специального назначения // Научно-техническая информация. Сер. 2. – 2023. – № 5. – С. 1-6. – DOI: 10.36535/0548-0027-2023-05-1; Sumin V.I., Gromov Yu.Yu., Tyutyunnik V.M. Optimization of special-purpose information systems operation // Automatic Documentation and Mathematical Linguistics. – 2023. – Vol. 57, № 3. – P.135-139. – DOI: 10.3103/S0005105523030044.

Материал поступил в редакцию 21.07.23.

Сведения об авторах

СУМИН Виктор Иванович – доктор технических наук, профессор, профессор кафедры информационной безопасности телекоммуникационных систем Воронежского института ФСИН России
e-mail: viktorsumin51@yandex.ru

ГРАЧЕВ Егор Дмитриевич – старший инженер группы специальной связи и технической защиты информации службы по защите государственной тайны УФСИН России по Мурманской области
e-mail: grachev.egorka1994@gmail.com

ГРОМОВ Юрий Юрьевич – доктор технических наук, профессор, директор института автоматики и информационных технологий Тамбовского государственного технического университета (ТГТУ), профессор кафедры «Информационные системы и защита информации»
e-mail: gromovtambov@yandex.ru

ТЮТЮННИК Вячеслав Михайлович – доктор технических наук, профессор, генеральный директор Международного Информационного Нобелевского Центра (МИНЦ), профессор кафедры «Конструирование радиоэлектронных и микропроцессорных систем» ТГТУ, профессор кафедры библиотечно-информационных наук Московского государственного института культуры
e-mail: vmtyutyunnik@gmail.com

УДК 025.4:81’322

А.В. Мыльникова, В.А. Трусов, Л.А. Мыльников

Использование скелетных структур текстов для развития методов семантического поиска*

Рассматривается проблема формирования дескрипторов для сокращения объема выдачи данных, сокращения времени поиска в текстовых источниках информации с помощью таких новых факторов, как авторство, регион, эмоциональный окрас, популярность, категория текста без соответствующих меток, которые могут формировать дескрипторы. Предлагаемый подход позволяет использовать уникальные лексико-грамматические дистрибутивные закономерности, имеющиеся в текстах. Результаты исследования могут быть применены для определения авторства и типа текста.

Ключевые слова: *семантический поиск, анализ текста, скелетная структура текста, text mining, классификация, языковые данные, дистрибутивный анализ, дескриптор, спецификатор*

DOI: 10.36535/0548-0027-2023-10-3

ВВЕДЕНИЕ

Сложность использования современных поисковых машин, таких как *Yandex*, *Google*, *Yahoo*, *Bing*, *Ask*, заключается в отсутствии в них прикладных механизмов семантического поиска информации, позволяющих повышать релевантность выдаваемых данных и верифицировать их в заданной пользователем предметной области.

Большое количество исследований по развитию механизмов смыслового поиска информации проводится в библиотечно-библиографических системах и системах управления предприятиями. Существуют модели семантического поиска, основанные на различных методах. Так, в работе [1] предложена модель семантического поиска, основанная на онтологическом представлении документальной информации, позволяющая формировать образы новых смыслов предметной области, а в статье [2] представлена модель семантической обработки документальной информации, предназначенная для поиска в библиотечно-библиографических системах.

Следует отметить, что развитие и внедрение смысловых систем обработки информации осуществляется в корпоративных информационных системах предприятий. В качестве примера можно привести сервис компании *ABBYY* «*ABBYY Intelligent Search*», где используются подходы к обработке информации на принципах машинных грамматик [3]. Эти механизмы носят достаточно ограниченный характер в силу того, что предметная область смысловых систем имеет заранее известную иерархическую структуру с привязкой к ней документов и применение их в рамках распределённых информационных систем не позволяет в полной мере реализовать информационную потребность.

Необходимо обратить внимание на публикации, отражающие вопросы развития подходов к повышению результатов поиска и обработки документальной научно-технической информации на основе выявления смысловой структуры текстов [4], на основе формализации неструктурированной информации [5], на базе технологий искусственного интеллекта [6]. Концептуальный подход к семантической обработке научно-технической информации в распределенных информационных системах предложен в статье [7].

Развитие цифровой алгоритмизации позволило создавать механизмы смыслового поиска на основе различных методов сопоставления входящих запросов и исходящей информации. Для решения задач повышения эффективности смыслового поиска применяются механизмы машинного обучения, основанные на иерархических искусственных нейронных сетях [8], на рекуррентных нейронных сетях [9], на векторных моделях [10], на онтологических представлениях [11, 12], на системах классификации текстов [13], позволяющие качественно повышать результаты поисковых запросов и выстраивать эффективную систему ранжирования найденных документов. Все эти механизмы поиска и обработки информации либо носят универсальный характер – что не позволяет эффективно их использовать для решения задачи поиска научной информации, либо, наоборот, осуществляют поддержку конкретной узкой предметной области – и применять их для остальных предметных областей становится неэффективно.

* Исследование выполнено при финансовой поддержке Правительства Пермского края в рамках научного проекта № С-26/692.

В основе методологии разработки базы скелетных языковых структур лежит теория «смысл-текст» [14–16], структурно реализованная на примере анализа и классификации речевых актов деловой коммуникации [17]. Если следовать утверждению, что переход от смыслов к текстам и обратно может быть выполнен за счет моделирования языковых познаний говорящих, то это означает, что результаты эмпирического исследования позволят проводить границы языковых познаний говорящих [18], формализовав их в виде набора базовых языковых скелетных структур, лежащих в основе соответствующих актов деловой коммуникации.

Таким образом, с точки зрения внешней семантики, анализ речевых актов позволит получить ограниченный набор поверхностных языковых структур для представления определенного события, а также смоделировать вариативность семантического контекстуально- или ситуационно-обусловленного описания этого события в виде ситуаций и показать возможности стилистического окраса актуализируемых ситуаций.

Другие способы не смыслового анализа текста – это использование, во-первых, семантического дифференциала [19], который позволяет сопоставлять каждой части речи некоторый эмоциональный окрас по предзаданной группе шкал, и во-вторых, методов оценки интонационного или эмоционального окраса (фоносемантика, звукоцвет), которые могут применяться к отдельным частям речи.

Выявление эмоциональных особенностей позволяет рассматривать тексты с точки зрения соответствия их стилевых особенностей заданному типу текста (например, эссе, обзоры, рефераты, сочинения, газетные статьи и т.д. с учетом области знания и конкретным требований издания), интересу пользователей (успех или неудача у читателей или зрителей) и региональным или авторским особенностям.

Таким образом, используемую лексику можно определить как включенную в текст структуру, привносящую стилевые и эмоциональные черты. Однако в некоторых случаях этого оказывается недостаточно. Например, когда изобретаются новые слова или используется одно и то же слово в формах разных частей речи. При этом есть исследования, которые показывают, что структура организации текста оказывает влияние на качество перевода [20, 21] и позволяет делать выводы об авторстве и типе текста [22].

С учетом таких дополнительных характеристик можно повысить эффективность применяемых моделей и алгоритмов поиска и семантической обработки информации, содержащейся в распределенных информационных системах глобальной вычислительной сети Интернет.

Диалоговые системы, включающие алгоритмы смысловой обработки и выдачи пользователям наиболее релевантной информации, основываются на большой языковой модели и используют методы обучения как с учителем, так и с подкреплением. Например, это ChatGPT v.4, Writesonic, Neeva, Phind, Browse AI, Gato и др. Основная задача поиска и обработки научно-технической информации – это максимально сузить границы поиска для получения пертинентной информации. Чат-боты с такой задачей не справляются, так как в них заложены универсальные механизмы обработки всех видов информации. И как только границы поискового запроса сужаются, чат-боты начинают выдавать все более релевантный массив, что не соответствует прикладным задачам поиска (например, научно-технической информации).

Несмотря на многообразие существующих методов и подходов, они не удовлетворяют требованиям, связанным с возможностью выявления на ранних стадиях поиска терминологической структуры предметной области информационной потребности.

МЕТОДИКА ИССЛЕДОВАНИЯ

Предложим гипотезу о том, что структура текста может использоваться для определения или уточнения таких его характеристик, как популярность, авторство, выделение «токсичных» фрагментов, эмоциональный окрас, тип текста и т.п. Такие характеристики позволяют задавать дополнительные критерии поиска.

Для проверки этой гипотезы необходимо выделить из текста части речи (I): прилагательное (*ADJ*), предлог (*ADP*), наречие (*ADV*), вспомогательный глагол (*AUX*), сочинительный союз (*CCONJ*), определительные артикли (*DET*), сущность (*ENTITY*), существительное (*NOUN*), частица (*PART*), местоимение (*PRON*), знак препинания (*PUNCT*), подчинительный союз (*SCONJ*), глагол (*VERB*), и их позиции в тексте и количество раз, когда они встречались. Для сравнения текстов разного объема надо вычислить:

1) значения среднего положения частей речи $P^{(i)} = \frac{\sum_{n=1}^{N^{(i)}} p_n^{(i)}}{N^{(i)}}, \forall i \in I$, где $N^{(i)}$ – число раз, которое встретилась i-я часть речи в рассматриваемом тексте, $p - n$-я позиция -й части речи;

2) косинусную меру [23] на основании данных о числе используемых частей речи для сравнения близости текстов и введенных шаблонных значений одного и другого классов (при бинарной классификации) – $CM^{(1)}$ и $CM^{(2)}$.

Для введения шаблонных значений просуммируем число частей речи всех текстов одного класса и полученное значение будем использовать как первый шаблон, а второй шаблон получим путем суммирования числа частей речи всех текстов другого класса. Такой подход использует свойства косинусной меры, значение которой пропорционально углу отклонения вектора, проведенного из начала координат в заданную точку пространства от вектора, направленного в точку, соответствующую шаблону (т. е. абсолютные значения не будут оказывать существенного влияния).

В результате получим форму представления данных для рассматриваемой группы текстов (табл. 1).

После подготовки данных задача бинарной классификации может быть решена одним из методов машинного обучения. Классификатор (kNN, SVM, C5.0, Naïve Bayes), согласно теореме [24], выбирается эмпирическим путем основе одной или группы метрик, используемых в задачах классификации (Accuracy, ROC, F_β, Recall, Precision).

Таблица 1

Форма представления текстовых данных для их анализа*

№ текста	ADJ	ADP	ADV	AUX	CCONJ	DET	ENTITY
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	$P_1^{(ADJ)}$	$P_1^{(ADP)}$	$P_1^{(ADV)}$	$P_1^{(AUX)}$	$P_1^{(CONJ)}$	$P_1^{(DET)}$	$P_1^{(ENTITY)}$
…	…	…	…	…	…	…	…
n	$P_n^{(ADJ)}$	$P_n^{(ADP)}$	$P_n^{(ADV)}$	$P_n^{(AUX)}$	$P_n^{(CONJ)}$	$P_n^{(DET)}$	$P_n^{(ENTITY)}$
…	…	…	…	…	…	…	…
N	$P_N^{(ADJ)}$	$P_N^{(ADP)}$	$P_N^{(ADV)}$	$P_N^{(AUX)}$	$P_N^{(CONJ)}$	$P_N^{(DET)}$	$P_N^{(ENTITY)}$

Продолжение таблицы 1

NOUN	PRON	PUNCT	SCONJ	VERB	$CM^{(1)}$	$CM^{(2)}$	Класс текста
9	**10**	**11**	**12**	**13**	**14**	**15**	**16**
$P_1^{(NOUN)}$	$P_1^{(PRON)}$	$P_1^{(PUNCT)}$	$P_1^{(SCONJ)}$	$P_1^{(VERB)}$	$CM_1^{(1)}$	$CM_1^{(2)}$	0/1
…	…	…	…	…	…	…	…
$P_n^{(NOUN)}$	$P_n^{(PRON)}$	$P_n^{(PUNCT)}$	$P_n^{(SCONJ)}$	$P_n^{(VERB)}$	$CM_n^{(1)}$	$CM_n^{(2)}$	0/1
…	…	…	…	…	…	…	…
$P_N^{(NOUN)}$	$P_N^{(PRON)}$	$P_N^{(PUNCT)}$	$P_N^{(SCONJ)}$	$P_N^{(VERB)}$	$CM_N^{(1)}$	$CM_N^{(2)}$	0/1

* *Примечание: ADJ* – прилагательное, *ADP* – предлог, *ADV* – наречие, *AUX* – вспомогательный глагол, *CCONJ* – сочинительный союз, *DET* – определительные артикли, *ENTITY* – сущность, *NOUN* – существительное, *PART* – частица, *PRON* – местоимение, *PUNCT* – знак препинания, *SCONJ* – подчинительный союз, *VERB* – глагол

Определение принадлежности текста к тому или иному классу зависит от решаемой задачи (определение авторства, принадлежности текста к некоторому типу текстов, региону).

Может оказаться, что в результате подготовки выделяются новые особенности и/или классы текста. Для такой проверки предварительно следует применить методы поиска ассоциативных правил (например, Apriori) и кластеризации (например, k-means).

Сложность решения задачи бинарной классификации текстов может быть связана с несбалансированностью выборки. Тогда необходимо использовать методы кросс-валидации для данных с низким расслоением значений параметров текстов, что может быть проверено путем вычисления коэффициентов Джини для отдельных параметров. Исключение параметров и отсутствие расслоения значений снизит шум и должно положительно сказаться на получаемых результатах.

РЕШЕНИЕ ЗАДАЧ БИНАРНОЙ КЛАССИФИКАЦИИ ТЕКСТОВ

Для проверки гипотезы о возможности использования структуры текстов для выделения дополнительных характеристик при поиске информации проведем ряд экспериментов.

Эксперимент 1. Определение популярности текста. Если предположить, что при анализе структуры текста существует такое количество просмотров, после превышения которого однозначно определяется к какой категории (больших или низких просмотров/качества) относится текст, то можно провести эксперимент, разметив данные таким образом, чтобы разделить их на два класса, означающих большое и малое количество просмотров.

Возьмем данные с сайта Kaggle, содержащие тексты докладов и информацию о количестве их просмотров (https://www.kaggle.com/datasets/thegupta/ted-talk). Стенограммы подготовлены на английском языке и содержат 4609 текстов с количеством просмотров от 0 до 65678748 и медианным значением длины транскрипции доклада в символах, равным 8819. Для анализа будем определять части речи и составим таблицу, в которой для каждого вычисления будет приведено среднее место положения частей речи.

На основе этих данных будем решать задачу бинарной классификации текстов. Использование любых методов машинного обучения не меняет качественную картину точности определения популярности текста.

Из полученных данных (рис. 1) видно, что при изменении границы, по которой будет определяться категория, к какой относится текст, существует некоторая область смешения, когда не удается разделить тексты на категории.

Эксперимент 2. Определение категории текста. Возьмем разбитые по тематическим категориям новостные тексты, приведенные на сайте Kaggle (https://www.kaggle.com/datasets/qusaybtoush1990/english-news). Выборка содержит 1490 записей на английском языке от 250 до 350 записей для каждой категории (класса) при медианном значении длины текста в символах равном 1961.

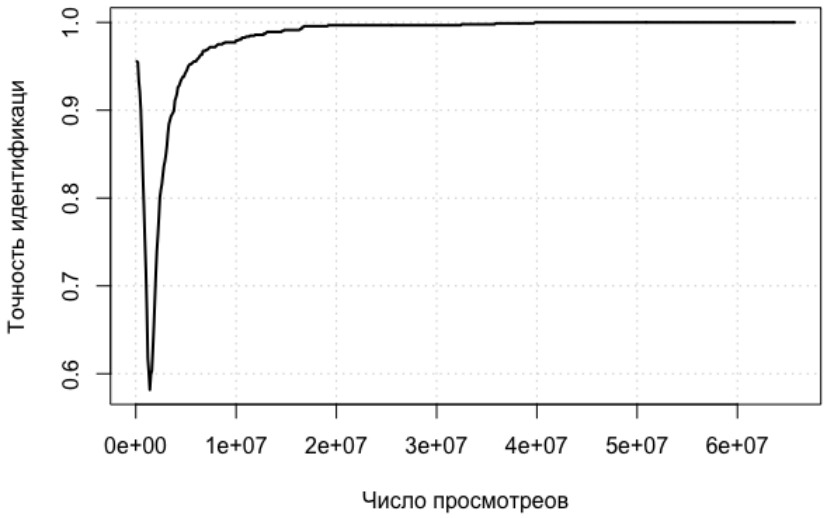

Рис. 1. График точности (accuracy) определения класса, к которому относится текст, при изменении границы разделения по числу просмотров текстов на классы популярных и непопулярных методом *k*NN

Таблица 2

Точность (Accuracy) классификаторов при выделении выбранной категории из всего множества новостных текстов*

kNN	SVM	C5.0	Naïve Bayes	gbm
Категория tech соответствует классу =1				
0,67	0,78	0,82	0,74	0,82
Категория politics соответствует классу =1				
0,59	0,70	0,74	0,71	0,75
Категория entertainment соответствует классу =1				
0,62	0,72	0,76	0,58	0,71
Категория sport соответствует классу =1				
0,64	0,80	0,79	0,77	0,81
Категория business соответствует классу =1				
0,56	0,76	0,80	0,67	0,76

* *Примечание:* **kNN –** метод *k* ближайших соседей; **SVM –** метод опорных векторов; **C5.0 –** алгоритм построения дерева решений; **Naïve Bayes –** наивный Байесовский классификатор; **gbm –** метод градиентного спуска

Для определения категории текста выберем одну из таких категорий, как бизнес, спорт, развлечения, политика, технологии. Таким образом сведем эту задачу к задаче бинарной классификации текстов, присваивая выбранной категории значения класса =1, а остальным категориям значения класса = 0, чтобы данные были сбалансированы, и получим результаты, приведенные в табл. 2.

Несмотря на то, что разные методы показывают разную точность для разных категорий текста видно, что получаемые результаты не случайны. Таким образом, можно говорить о том, что категорию текста получается определить на основе предлагаемого подхода.

Дополнительные эксперименты

Успешный способ определения авторства с использованием описанного подхода описаны (см. например [22]) на примере полных стихов А. Блока и С. Есенина на языке оригинала (русский язык).

Определение эмоционального окраса текста. Рассмотрим данные представленные для анализа на конкурсе RuSentNE-2023 (https://codalab.lisn.upsaclay.fr/competitions/9538#results), проводимом совместно МГУ им. М.В. Ломоносова и Университетом Ньюкасла (Великобритания). Набор данных содержит 6637 записей, размеченных как: -1 – негативный окрас, 0 – без эмоционального окраса и +1 – позитивный окрас. Медианное значение длины текста в символах равно 138. Тексты составлены на русском языке (табл. 3 и 4).

Анализ группы текстов на специально подготовленных сбалансированных выборках показал, что эмоциональный окрас определить не удается. Получаемые результаты близки к 50%, что свидетельствует о случайном характере получаемых ответов при рассмотрении классов по отдельности. При использовании несбалансированных выборок текстов модели склонны ставить всей или почти всей выборке признак одного из классов.

Таблица 3

Точность (Accuracy) классификаторов при выделении выражений с позитивным окрасом в класс 1, а остальных выражений в класс 0

kNN		SVM		C5.0		Naïve Bayes		gbm	
Класс		Класс		Класс		Класс		Класс	
1	0	1	0	1	0	1	0	1	0
0,583	0,533	0,577	0,568	0,565	0,590	0,005	0,989	**0,607**	**0,556**

Таблица 4

Точность (Accuracy) классификаторов при выделении выражений с негативным окрасом в класс 1, а остальных выражений в класс 0

kNN		SVM		C5.0		Naïve Bayes		gbm	
Класс		Класс		Класс		Класс		Класс	
1	0	1	0	1	0	1	0	1	0
0,562	0,536	0,597	0,508	0,706	0,400	0,547	0,571	0,547	0,584

Одной из причин таких результатов может быть то, что подобранные выражения состоят из одного предложения (заголовков новостных лент), и их понимание без контекста затруднено. Это отражает сложность решения такой задачи в общем виде.

Определение географического региона, в котором подготовлен текст. Еще одна задача классификации, которая может помочь выявить область поиска – это определение региона, к которому относится текст новости, опубликованной в газете. Если провести эксперимент, аналогичный предыдущим, на основе данных с сайта – http://news.pfo-perm.ru/, собранных за два месяца (апрель и май 2023 г.), то получим результат, не позволяющий сделать выводы о принадлежности текста к заданному региону (точность ~ 0,5). Точность работы моделей повышается до устойчивых 60% (т.е. становится не случайной), если рассматривать макрорегионы (например, уральский регион, в который входят Пермский край, Удмуртская республика, Тюменская, Свердловская, Кировская и Челябинская области). В этом случае число записей на русском языке составило 5191 при медианном значении длины новости 1627 символов.

Такие результаты оказываются недостаточными для решения задачи фильтрации текстов по регионам в системах поиска.

Определение категории текста. Возьмем с сайта *BBC* новостные взяты, разбитые по тематическим категориям, приведенным на сайте Kaggle (https://www.kaggle.com/datasets/hgultekin/bbcnewsarchive). Выборка содержит 1096 записей на английском языке при медианном значении длины текста в символах равном 2277,5. Особенность данного набора новостных блоков текстов заключается в типовой структуре их построения, когда сообщается, что комментатор/автор узнал, что что-то произошло или получен какой-то результат значения экономического показателя, счет в матче и т.п.

Для определения категории текста будем выбирать одну из таких категорий, как бизнес, спорт, развлечения, политика, технологии, а задача, как и в другом эксперименте с новостями сводилась к задаче бинарной классификации.

Выявление «токсичных» комментариев. Подборка с токсичными комментариями, расположенная на сайте Kaggle (https://www.kaggle.com/code/aleksandrsahkanevich/17-toxic-comments), содержит 14412 записей на русском языке с медианным значением длины текста в символах, равным 101. При этом получаем результаты, которые не позволяют однозначно выделить токсичные комментарии (точность ~0,5).

ДИСКУССИЯ ПО РЕЗУЛЬТАТАМ ИССЛЕДОВАНИЯ

При работе в сети Интернет качество выдачи информации зависит от семантического запроса, который мы сформулировали, и алгоритма, который формирует ранжированное множество текстов для выдачи [25].

Определение таких характеристик текста, как популярность (в том числе и потенциальная), региональная принадлежность и авторство (или похожесть на стиль региона/автора), тип (репортаж, газетная статья, художественное произведение, эссе, научный текст и т.п.) позволяет ввести новые параметры поиска и тем самым повысить соответствие выдачи ожидаемым результатам.

Проведенные эксперименты позволяют предположить, что качество идентификации принадлежности текста к тому или иному классу падает при уменьшении объема этого текста, не зависит от языка и может быть описано характеристикой, приведенной на рис. 2.

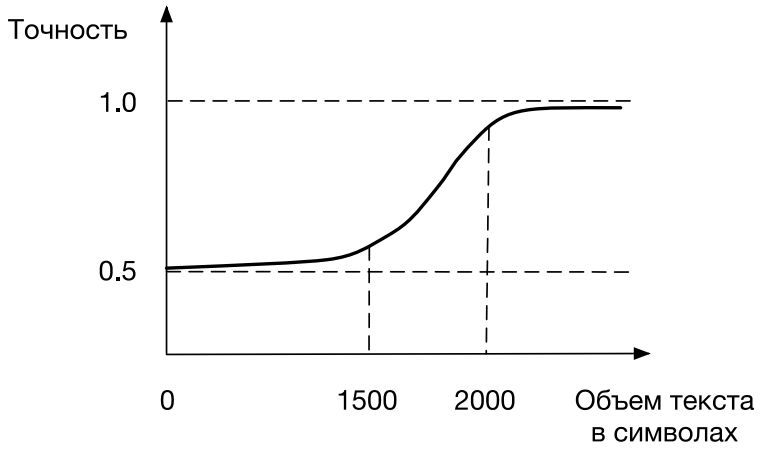

Рис. 2. Изменение точности решения задач классификации в зависимости от объема текста

Эти же эксперименты показывают, что приведенная зависимость может нарушаться, если неправильно выбраны другие характеристики текстовых данных. Такими характеристиками могут быть, например, количество просмотров, после которого текст идентифицируется как популярный, или схожесть текстов. В одном из дополнительных экспериментов задача классификации не решалась, так как новостные заметки, вне зависимости от категории, готовились по единому шаблону.

Таким образом, можно сделать вывод, что предлагаемый нами подход не во всех случаях дает результат, и его следует применять совместно с традиционными подходами, использующими технологии Bag of Words, TF-IDF, Word2Vec, тогда он будет приводить к повышению качества идентификации традиционных подходов так как они привязываются к словам и не учитывают их использование после проведения операций токенизации и/или лемматизации, и/или стемминга.

Дополнительные результаты может дать и обобщение такого подхода до структур предложений [20, 21], использование разных типов которых привнесет новую информацию об изучаемых текстах. В этом случае в качестве признаков, наличие и число которых будет проверяться, можно использовать типы предложений по:

- цели высказывания – повествовательное, вопросительное, побудительное;
- форме наклонения – изъявительное, условное, повелительное;
- виду залога – активный, пассивный;
- виду временной группы – прошедшее простое время, настоящее простое время, настоящее в процессе, прошедшее в процессе, настоящее с результатом, прошедшее с результатом, настоящее с результатом и процессом, прошедшее с результатом и процессом, будущее простое, будущее в прошедшем, будущее в прошедшем с результатом, близкое будущее, близкое будущее в прошлом, незамедлительное будущее, незамедлительное будущее в прошлом;
- характеру связи между предметом речи и тем, что о нем сообщается – утверждение, отрицание;
- форме сказуемого – простое (выраженное смысловым глаголом или связочным глаголом), с модальной оценкой (модальные глаголы), с аспектной оценкой (аспектные глаголы), с модальной и аспектной оценкой (модальный +аспектные глаголы), со смысловой оценкой (глаголы со значением предположения, предложения, желания, рекомендации, совета), с модальной и смысловой оценкой (модальный глагол+ глагол со значением предположения, предложения, желания, рекомендации, совета).

Можно применять дополнительные типы классификаций в зависимости от языка (например, для английского языка может быть введена классификация по словам, меняющим структуру предложений), а также рассматривать каждое предложение как входящее в целый ряд классификаций.

ЗАКЛЮЧЕНИЕ

При поиске товаров в интернет-магазине у нас появляется возможность задать такие жесткие критерии, как тип товара, цвет, размер и т.д. Это позволяет сузить поиск, однако порядок выдачи может не совпадать с нашими ожиданиями. В таких информационных средах как Интернет при поиске нет возможности жестко задать характеристики. На практике поиск осуществляется по ключевым словам с использованием правил их комбинирования на основе логических операций, операций над множествами, изменения словоформ.

Практика показывает, что этого недостаточно, так как время на поиск нужного нам источника может оказаться очень большим – несмотря на то, что мы представляем то, что ищем язык написания запроса (дескриптора) не в полной мере передает наши ожидания, а это накладывается на механизмы, реализованные в поисковых машинах. Таким образом мы сталкиваемся с ситуацией, когда необходимо введение новых критериев идентификации искомой

информации. Особенно важным это становится для текстовой информации, которая формируется из разнородных источников с разными правилами структурирования.

СПИСОК ЛИТЕРАТУРЫ

1. Максимов Н.В., Голицына О.Л., Монанков К.В., Лебедев А.А., Баль Н.А., Кюрчева С.Г. Средства семантического поиска, основанные на онтологических представлениях документальной информации // Научно-техническая информация. Сер. 2. – 2019. – № 7. – Р. 8–19.
2. Васина Е.Н., Попов И.И. Модели и методы автоматизации обработки и анализа документальной информации // Известия Российского экономического университета им. Г.В. Плеханова. – 2012. – № 3. – С. 44–50.
3. Хорошилов А.А., Дмитришин А.Н. Принципы создания машинных грамматик для использования в промышленных системах обработки текстовой информации // Информатизация и связь. – 2020. – № 1. – С. 41–47.
4. Белоногов Г.Г., Гиляревский Р.С., Селетков С.Н., Хорошилов А.А. О путях повышения качества поиска текстовой информации в системе Интернет // Научно-техническая информация. Сер. 2. – 2013. – № 8. – С. 1–11.
5. Хорошилов Ал-др А., Хорошилов Ал-ей А., Котляр А.В., Гареев М.Ш., Кокутин С.Н. Технологии автоматизированной обработки неструктурированной текстовой информации // Первая международная научно-техническая конференция «The 2017 Symposium on Cybersecurity of the Digital Economy (CDE'17)». – Университет Иннополис, Республика Татарстан, Россия: Издательский Дом «Афина», 2017. – С. 419-435.
6. Белоногов Г.Г., Гиляревский Р.С., Хорошилов Ал-др А., Хорошилов Ал-ей А. Развитие систем автоматической смысловой обработки текстовой информации // Нейрокомпьютеры: разработка, применение. – 2010. – № 8. – С. 4–13.
7. Трусов В.А. Концептуальный подход к поиску и семантической обработке научно-технической информации в распределенных системах Интернета // Научно-техническая информация. Сер. 2. – 2021. – № 4. – Р. 1-11.
8. Serban I.V., Sordoni A., Bengio Y., Courville A., Pineau J. Building end-to-end dialogue systems using generative hierarchical neural network models // AAAI'16: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. – 2016. – Vol. 30, № 1. – P. 3776–3783.
9. Sutskever I., Martens J., Hinton G. Generating text with recurrent neural networks // Proceedings of the 28th International Conference on Machine Learning (ICML-11). – Madison, WI, United States: Omnipress, 2011. – P. 1017–1024.
10. Micarelli A., Gasparetti F., Biancalana C. Intelligent search on the internet // Reasoning, Action and Interaction in AI Theories and Systems: Essays Dedicated to Luigia Carlucci Aiello. – 2006. – P. 247–264.
11. Kim W., Choi D.W., Park S. Agent based intelligent search framework for product information using ontology mapping // Journal of Intelligent Information Systems. – 2008. – Vol. 30. – P. 227–247.
12. Bernard L., Einspanier U., Haubrock S., Hübner S., Kuhn W., Lessing R., Lutz M., Visser U. Ontologies for intelligent search and semantic translation in spatial data infrastructures // Photogrammetrie, Fernerkundung, Geoinformation. – 2003. – Vol. 6. – P. 451–462.
13. Zhang X., Zhao J., LeCun Y. Character-level convolutional networks for text classification // Advances in neural information processing systems (NIPS 2015). – 2015. – Vol. 28.
14. Мельчук И.А. Опыт теории лингвистических моделей «смысл-текст»: семантика, синтаксис. – Москва: Школа «Языки русской культуры», 1999. – 345 с.
15. Иорданская Л.Н., Мельчук И.А. Смысл и сочетаемость в словаре. Языки славян. – Москва, 2007. – 672 с.
16. Апресян Ю.Д. Языковая картина мира и системная лексикография. Языки славян. – Москва, 2006.
17. Austin J.L., Urmson J.O. How to Do Things with Words. – Cambridge, Massachusetts: Harvard University Press, 2009.
18. Витгенштейн Л. Избранные работы. Территория. – Москва, 2005.
19. Osgood C.E., Suci G., Tannenbaum P. The Measurement of Meaning. – Urbana: University of Illinois Press, 1957. – 342 p.
20. Новикова А. В., Мыльников Л.А. Вопросы реализации машинного перевода текстов деловой коммуникации для языковой пары «Русский – Английский язык» // Научно-техническая информация. Сер. 2. – 2017. – № 6. – С. 26-36; Novikova A.V., Mylnikov L.A. Problems of machine translation of business texts from Russian into English // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 3. – P. 159–169.
21. Novikova A. Direct Machine Translation and Formalization Issues of Language Structures and Their Matches by Automated Machine Translation for the Russian-English Language Pair // Proceedings of International Conference on Applied Innovation in IT. – 2018. – Vol. 6, № 1. – P. 85–92.
22. Мыльникова А.В., Мыльников Л.А. Вопросы дистрибутивно-смыслового анализа скелетных структур текстов в задачах автоматизированной обработки языковых данных // Научно-техническая информация. Сер. 2. – 2023. – № 5. – С. 21–30.
23. Мыльников Л.А. Статистические методы интеллектуального анализа данных. – Санкт-Петербург: БХВ-Петербург, 2021. – 240 с.
24. Wolpert D.H., Macready W.G. No free lunch theorems for optimization // IEEE Transactions on Evolutionary Computation. – 1997. – Vol. 1, № 1. – P. 67–82.

25. Page L., Brin S., Motwani R., Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report. – Stanford, United States: Stanford InfoLab. – 1998. – 17 p.

Материал поступил в редакцию 01.08.23.

МЫЛЬНИКОВА Анна Вячеславовна – кандидат филологических наук, доцент кафедры Иностранные языки и связи с общественностью Пермского национального исследовательского политехнического университета
e-mail: Novikova@yandex.ru

ТРУСОВ Владимир Александрович – доктор технических наук, доцент, старший научный сотрудник лаборатории междисциплинарных эмпирических исследований Национального исследовательского университета – «Высшая школа экономики»; доцент кафедры Микропроцессорные средства автоматизации Пермского национального исследовательского политехнического университета
e-mail: VATrusov@hse.ru

МЫЛЬНИКОВ Леонид Александрович – кандидат технических наук, доцент, заведующий научно-учебной лабораторией междисциплинарных эмпирических исследований Национального исследовательского университета – «Высшая школа экономики»; доцент кафедры Микропроцессорные средства автоматизации Пермского национального исследовательского политехнического университета
e-mail: LAMylnikov@hse.ru

ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 025.43:[311.311:002]

Ю.В. Мохначева

Взаимодействие термина «библиометрический анализ» с другими высокочастотными ключевыми словами в темах *SciVal*

Представлены результаты исследования высокочастотных ключевых терминов (ВКТ) в темах SciVal (Scopus), в числе которых присутствовал термин «Библиометрический анализ». Собранный за май–ноябрь 2022 г. массив из ВКТ по 181 теме SciVal, в которой термин «Библиометрический анализ» фигурировал в качестве ВКТ, анализировался по трем примерно равным группам со-слов, сформированным по числу суммарного пересечения в темах. Такое разделение ВКТ хорошо вписалось в закономерность С. Брэдфорда, благодаря чему было сформировано «ядро» ВКТ по исследуемым темам. В итоге, были определены наиболее близкие по содержанию темы.

Результаты нашего исследования способствуют пониманию взаимосвязей между различными исследовательскими темами посредством анализа динамики со-слов, подтверждая гипотезу о том, что с помощью сетей со-слов из разных дисциплин можно выявить общность между ними и чем больше совпадений между двумя ключевыми словами тем теснее взаимосвязь различных тем.

Ключевые слова: *высокочастотные ключевые термины, библиометрический анализ, со-слова, анализ ключевых терминов, терминологический анализ, темы SciVal, взаимодействие ключевых терминов*

DOI: 10.36535/0548-0027-2023-10-4

ВВЕДЕНИЕ

Изучение исследовательских фронтов – это чрезвычайно актуальное направление в библиотечно-информационной науке [1]. Впервые концепцию исследовательских фронтов предложил Д. Прайс [2], а в 1994 г. Ю. Гарфилд в своей работе [3] определил исследовательские направления как кластеры совместного цитирования. В этом контексте особое значение приобрел ресурс SciVal[1] (Elsevier), где каждую тему представляет набор публикаций с общим сфокусированным интеллектуальным интересом, и каждая отдельная публикация может относиться только к одной из них. Темы состоят из трех отличительных ключевых фраз: как правило, первые две представляют собой высокочастотные ключевые фразы, которые выбираются для обеспечения макроуровневого описания темы исследования, а третья – более конкретное описание темы. В каждом из тематических кластеров представлено от 5000 и более проиндексированных в БД Scopus публикаций с 1996 г. по настоящее время. Кластеры создаются на основе частоты прямого цитирования между темами соответствующего кластера [1]. Актуальность темы определяется процентилем: чем он выше, тем больше внимания привлекает тема. Процентиль актуальности рассчитывается по количеству цитирований и просмотров публикаций за последние два года, что отражает характеристики повышенного внимания и новизны, поэтому он может отражать направления исследований. Темы с процентилем актуальности превышающем 90, рассматриваются как «исследовательские фронты», а выше 99 – как «горячие исследовательские фронты» [1].

Взаимодействия между дисциплинами со временем эволюционируют и концентрируются в основном на соседних дисциплинах [4]. В работе [5] показана эволюция библиотечного дела и информатики путем анализа тем исследований в журнальных статьях по этому направлению. Анализ проводился для пяти периодов, охватывающих 1996–2019 гг. Результаты исследования показали, что: библиотечное дело со временем стало менее распространенным, поскольку в период 2000–2005 гг. отсутствовали тематические кластеры, имеющие отношение к библиотечным во-

[1] SciVal – аналитический инструмент на основе данных Scopus. – URL: https://www.scival.com/

просам; библиометрия и особенно анализ цитирования в разные периоды отличалась высокой стабильностью, о чем свидетельствуют стабильные подкластеры и согласованные ключевые слова. До последнего времени «Поиск информации» был доминирующей областью. Тем не менее, как утверждают авторы [5], области изучения библиотечно-информационной науки существенно изменились, а фокус исследований сместился на «Информационные вопросы».

В работе [6] рассматривалась эволюция библиотечно-информационной сферы на основе отслеживания перечня авторских ключевых слов в исследовательских статьях, опубликованных в период 1971-2015 гг. Автор обнаружил, что с 1970-х гг. к 2015 г. этот перечень эволюционировал от проектирования и управления информационными системами до анализа научных коммуникаций; хранения и поиска информации; доступов к информации; управления информацией и знаниями; обучения пользователей. Было отмечено, что библиотечно-информационная область эволюционировала в такие новые направления, как: управление знаниями, информационная грамотность и библиометрия. Между тем, контекст исследований стал больше склоняться к академическим библиотекам, лишь в некоторой степени затрагивая публичные библиотеки. Автор публикации предполагает, что библиометрия и связанные с ней методы будут доминировать в изучении библиотечно-информационной науки. По мнению автора, интерес к библиометрии вызван необходимостью ранжирования и оценки областей науки в университетах на уровне правительств, библиотек и научных организаций.

Результаты работы [6] подтверждаются авторами статьи [7], которые проанализировали ключевые слова, библиографические связи и совместные цитирования для отслеживания изменений в тематике исследований по библиотечно-информационной науке за четыре периода между 1995 г. и 2014 г. Изучив 580 высокоцитируемых публикаций по библиотечно-информационной науке, авторы [7] обнаружили, что две темы: «Поиск информации и извлечение информации» и «Библиометрия» присутствовали во всех четырех периодах, однако доля статей уменьшалась для «Поиска информации» и увеличивалась для «Библиометрии». Несмотря на то, что количество тематических блоков, связанных с библиометрией, колебалось в разные периоды, это направление зарекомендовало себя в библиотечно-информационной области как стабильное. Авторы отмечали, что ряд одних и тех же тем повторялись в разные периоды, а новые появлялись редко, однако библиометрия (особенно – анализ цитирования) отличалась стабильностью по периодам и выросла в одну из наиболее доминирующих областей по библиотечно-информационному направлению.

В работе [8] определены основные темы исследований в библиотечно-информационной области: «Библиометрия, наукометрия и информетрика»; «Анализ цитирования»; «Поиск информации»; «Информационное поведение»; «Библиотеки»; «Исследования пользователей»; «Анализ социальных сетей»; «Вебометрия». Авторы обнаружили, что самой масштабной темой в период 2011–2015 гг. был «Библиометрический анализ».

По мнению уже упомянутых [6–8] и других авторов [9], «Библиотечное дело» постепенно теряет свое доминирование в библиотечно-информационной сфере, что подтверждается сокращением числа кластеров, связанных с библиотечными вопросами, а исследователи в этой области больше склонны цитировать публикации по дисциплинам, не относящимся к библиотечно-информационному направлению [9].

Ключевые слова являются базовыми элементами в публикациях и часто используются в библиометрических исследованиях как для выявления структуры знаний, так и для представления их концепций [10, 11]. Некоторые авторы предполагают [12, 13], что два ключевых слова, встречающихся в одной и той же статье, указывают на связь между темами, которые они представляют, и, чем выше частота совпадения ключевых слов, тем ближе темы исследований. Ключевое слово может представлять основное содержание статьи, а частота его встречаемости и совместное использование могут в некоторой степени отражать тематику, сфокусированную в специальной области [13]. Чем больше совпадений между двумя ключевыми словами, тем теснее их взаимосвязь [14], а если ключевые слова сгруппированы в один и тот же кластер, то они, скорее всего, будут отражать идентичные темы [12].

В 2004 г. была опубликована статья [15], в которой предложена концепция обнаружения новых тенденций в научных областях – Emerging Trend Detection (ETD), подразумевающая как выявление новых тем, так и их взаимосвязи в предметной категории. Для определения тенденций исследований авторы разработали метод, заключающийся в анализе социтирования и совместного употребления ключевых слов – со-слов[2].

Изучению возникновения новых и эволюционирования уже существующих научных направлений на основе анализа употребляемости ключевых слов с применением экспертных, лингвистических, библиометрических и математических методов посвящен ряд публикаций [16–25].

В последние годы термин «Библиометрический анализ» широко проник в темы, непосредственно и опосредованно связанные с библиометрией, превратившись в один из наиболее употребляемых ключевых терминов. Этому поспособствовала потребность: в определении новых тенденций в науке, в построении моделей сотрудничества, в получении возможности изучения интеллектуальной структуры конкретной предметной области в литературе [26].

Несмотря на широкое проникновение термина «Библиометрический анализ» в самые разные научные направления, он все-таки более релевантен для библиотечно-информационной области. В публикации [25] представлены результаты анализа 3308 высокочастотных ключевых терминов (ВКТ)[3] по 239 темам в

[2] со-слова – перевод с англ. «co-words», что означает общее употребление одних и тех же ключевых слов.

[3] Здесь и далее под высокочастотными ключевыми терминами (ВКТ) подразумеваются ключевые слова, отображенные в топ-списках по каждой из тем SciVal (до 100 слов на одну тему). Перечни этих терминов доступны на карточках публикаций в Scopus в разделе: «Темы SciVal».

библиотечно-информационной сфере. Было обнаружено, что в некоторых темах наблюдалась положительная динамика употребляемости определенных ключевых терминов, в то время как в других происходил обратный процесс. Кроме того, в ряде случаев было замечено «перетекание» ряда ВКТ (снижение встречаемости в одной теме и рост в другой) из одной темы в другую, что может свидетельствовать о смещении акцентов в научных направлениях в соответствии с меняющейся конъюнктурой исследований. В работе [25] показано, что перечень тем, по которым термин «Библиометрический анализ» фиксировался в числе ВКТ, разделился на две категории: темы из библиотечно-информационной сферы и темы по самым разным областям знания вне библиотечно-информационных наук. Основываясь на результатах предыдущих исследований [25, 27, 28], и принимая во внимание утверждения о том, что два ключевых слова, встречающиеся в одной и той же статье, указывают на связь между темами, которые они представляют, и чем выше частота совпадений ключевых слов, тем ближе темы [13, 14], мы провели исследование ВКТ «Библиометрический анализ» в темах SciVal.

ВЫСОКОЧАСТОТНЫЙ КЛЮЧЕВОЙ ТЕРМИН «БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ» В ТЕМАХ SCIVAL

Объект исследования – массив ВКТ (со-слов), пересекающихся в темах SciVal, в которых термин «Библиометрический анализ» входит в число ВКТ.

Цели исследования.

1. Выявление наиболее близких исследовательских тем SciVal с ВКТ «Библиометрический анализ» по числу пересечений с другими ВКТ.
2. Анализ динамики употребляемости ВКТ «Библиометрический анализ» в темах SciVal.
3. Выявление наиболее активно развивающихся тем SciVal с ВКТ «Библиометрический анализ».

Методология и методы исследования

Информационная платформа исследования – база данных (БД) Scopus[4] (Elsevier). Сбор данных: с мая по ноябрь 2022 г. На первом этапе по БД Scopus был проведен поиск публикаций по ключевой фразе: ‘библиометрический анализ’. Отобраны первые 1000 публикаций, отсортированные по нисходящей релевантности. Из карточек документов отбирались темы SciVal, в которых термин «Библиометрический анализ» присутствовал в числе ВКТ. Выявленные таким образом темы со всеми своими характеристиками заносились в таблицу. Аналогичный отбор тем проводился по связанным публикациям: просматривались первые 50 документов, отсортированных по нисходящей релевантности. Собранный массив из ВКТ по всем выявленным темам анализировался с целью определения пересечений одних и тех же ВКТ в разных темах. Таким образом, в один массив был отобран 2271 ВКТ, пересекающийся в анализируемых темах два и более раза. Эти ВКТ были разделены на три приблизительно равные группы, благодаря чему появилась возможность определения наиболее близких по содержанию тем.

Сбор данных проводился в течение полугода, поэтому динамика употребляемости ВКТ в течение этого периода претерпела некоторые изменения. Так, за время сбора информации термин «Библиометрический анализ» выбыл из числа ВКТ по двадцати темам. Тем не менее, при определении пересечений ВКТ анализировались все выявленные темы, включая в том числе выбывшие.

Для выявления круга наиболее активно развивающихся тем, связанных с ВКТ «Библиометрический анализ», для каждой из исследуемых тем был определен коэффициент динамики развития темы – T_{lv}: чем меньше значение этого коэффициента, тем активнее развивается тема. Коэффициент T_{lv} определяется на основе трех показателей: числа ВКТ с положительной динамикой роста в теме – $BKT_{pos\ din}$; числа ВКТ в пограничной зоне развития (т. е. с 0% ростом) – BKT_{0din}; числа ВКТ с отрицательной динамикой роста – $BKT_{neg\ din}$:

$$T_{lv} = N\ BKT_{neg\ din} + N\ BKT_{0\ din} : N\ BKT_{pos\ din},$$

где N – число ВКТ в теме SciVal.

Результаты исследования

Состав ВКТ по темам SciVal меняется со временем: какие-то термины выбывают из их числа, а какие-то, наоборот, занимают их место. По этой причине репертуар тем с ВКТ «Библиометрический анализ» за период с мая по ноябрь 2022 г. значительно скорректировался. Так, в ноябре 2022 г. по сравнению с маем – июнем 2022 г. из отобранных тем выбыли двадцать: 1) 3D Printers, Spare Parts, Supply Chain; 2) Academic Staff, Workload, Education; 3) Biodiversity, Intergovernmental Panel on Climate Change, Environmental Assessment Process; 4) Climate Change, Sociology, Crisis Management; 5) Configuration Management, Matlab/Simulink, Solar Cells; 6) Economic Systems, Professional Training, Public Administration; 7) *Editor, Self-Citation, Introduction*; 8) *Editorial Boards, Innovation Studies, Citation Index*; 9) Excess Deaths, Excess All-Cause Mortality, COVID-19; 10) *Information Science, Citations, South Africa*; 11) Knowledge Management; Organizational Culture; Social Media; 12) Labrador Retriever, Computer Use, Animals; 13) Mutation Fibrin, Fragment D, Metacarpal Bones; 14) Neuroethics, Philosophical Methodology, Morality; 15) Nursing Informatics, Education, Yearbooks; 16) Operations Research, Soft Systems Methodology, Critical Systems Thinking; 17) Plasmatrons, Ion Selective Electrodes, Arc of A Curve; 18) Synthetic Biology, Technoscience, Biotechnology; 19) Ulifloxacin, Ammonium Acetate, Quinolone Derivative; 20) Water Footprint, Water-Energy Nexus, Nexus.

В числе выбывших – три темы (выделены курсивом), непосредственно связаны с библиометрическими исследованиями, а бо́льшая часть – 17 тем никак не относились к библиотечно-информационной сфере и, вероятнее всего, термин «Библиометрический анализ» фигурировал там в качестве метода для изу-

[4] Доступ для российских пользователей закрыт с 01 января 2023 г. в связи с санкционной политикой Европейского Союза.

чения качественных и количественных динамических характеристик научных направлений.

Всего была отобрана 181 тема SciVal, где за период май–ноябрь 2022 г. термин «Библиометрический анализ» присутствовал в числе ВКТ. По этим темам все ВКТ были собраны в единый массив (6296 терминов). Дальнейший анализ проводился на массиве ВКТ, которые, как минимум, дважды пересекались в темах – 2271 ВКТ с общим суммарным числом пересечений = 11680. Не имели пересечений с другими темами 4025 ВКТ, и они в исследовании не учитывались.

Массив ВКТ с двойными и более пересечениями в 181 теме был разделен на три приблизительно равные группы в зависимости от числа пересечений (примерно по 3700-4000 суммарных пересечений на одну группу):

группа 1 – ВКТ с наибольшим числом пересечений в темах: 157 ВКТ с пересечением в 13-93 темах; суммарное число пересечений в темах = 3757 (табл. 1);

группа 2 – ВКТ со умеренным числом пересечений в темах: 518 ВКТ с пересечением в 5-12 темах; суммарное число пересечений в темах = 3834;

группа 3 – ВКТ с минимальным числом пересечений в темах: 1596 ВКТ с пересечением в 2-4 темах; суммарное число пересечений в темах = 4089.

Таким образом, разделение ВКТ на три группы довольно хорошо вписалось в закономерность С. Брэдфорда[5]: 1596:518≈518:157≈3. В табл. 1 представлен перечень ВКТ первой группы.

Данные табл. 1 показывают, какие ВКТ в 181 теме SciVal пересекались наиболее часто, что свидетельствует о сильной смысловой связи между этими терминами. Данная статистика отображает пересечения ВКТ на уровне со-слов.

Таблица 1

Первая группа высокочастотных ключевых терминов с наибольшими пересечениями в темах SciVal с ВКТ «Библиометрический анализ»

ВКТ	Число пересечений с другими темами SciVal с ВКТ «Библиометрический анализ»
Periodicals	93
Publications	93
Citations	77
Research	76
Publishing	64
Editorial	61
Article	57
Science	52
Journal Impact Factor	51
Biomedical Research	49
Education	44
Universities	44
Perspective	43
Innovation	42
Literature Review	41
Editor	40
Writing	39
Human	35
Systematic Review	35
Research Personnel	34
Scientific Literature	34
Citation Analysis	32
Faculty	32
India	32
Teaching	32
Web of Science	32
Peer Review	31
Review	31
Medicine	30
Digital	29
First International	29
Electronic Publishing	28
Scientific Publications	28
Students	28
Bibliometric Study	27
Global	27
Reflection	27
Research Output	27
Scholars	27
Scopus	27
Book	26
Conference	26
COVID-19	26
Editorial Policies	26
Industry	26
Reviewers	26
Awards and Prizes	25
Commentary	25
Research Productivity	25
Societies	25
Artificial Intelligence	24
Big Data	24
Higher Education	24
Introduction	24
Leadership	24
Open Access	24
Scientists	24
Ethics	23
Libraries	23
Organizations	23
Scholarly Communication	23
Special Issue	23
Trends	23
Careers	22
Commerce	22
Congresses	22
Curricula	22

[5] С. Брэдфорд установил, что количество журналов в третьей зоне будет примерно во столько раз больше, чем во второй зоне, во сколько раз число наименований во второй зоне больше, чем в первой. Обозначим Т1 – число журналов в 1-й зоне, Т2 – во 2-й, Т3 – в 3-й зоне. Если n – отношение количества журналов 2-й зоны к числу журналов 1-й зоны, то закономерность, вскрытая С. Брэдфордом, может быть записана так: Т1 : Т2 : Т3 = 1 : n : n^2, или: Т3: Т2 = Т2: Т1 = n. [29, 30].

ВКТ	Число пересечений с другими темами SciVal с ВКТ «Библиометрический анализ»
Challenges	21
Mapping	21
Ranking	21
Sustainability	21
Technology	21
China	20
Data Mining	20
Impact	20
Medical Education	20
Medical Literature	20
Social Sciences	20
Academies and Institutes	19
Brazil	19
Evidence Based Practice	19
Information Systems	19
Library and Information Science	19
Medical Societies	19
Production	19
Publishing Research	19
Scientific Research	19
Digital Libraries	18
Fellowships and Scholarships	18
Future	18
Health Science	18
Information Dissemination	18
Information Science	18
Latin America	18
Manuscripts	18
Nation	18
Pediatrics	18
Social Network Analysis	18
Competency	17
Evolution	17
Graduate	17
Library Science	17
Medical	17
Metric	17
Sustainable Development	17
Case Study	16
Collaborative	16
Development	16
Engineering	16
Governance	16
Letters to the Editor	16
Public Health	16
Research Support	16
Small and Medium-sized Enterprises (SMEs)	16
Theory	16
Academic	15
Assessment	15
Contribution	15
Data Science	15
Document	15
Hirsch Index	15
History	15
Nursing	15
Research Peer Review	15
Sustainable	15
Anthropogenic Effect	14
Bibliometric Indicators	14
Clinical Research	14
Educational Research	14
Engineering Education	14
Pandemic	14
Policy	14
Social Media	14
Academic Journals	13
Altmetrics	13
Cardiology	13
Content Analysis	13
Context	13
Diversity	13
Editorial Board	13
Entrepreneurship	13
Funding	13
Health Services Research	13
Humanities	13
Information Technology	13
Interdisciplinary	13
Key Words	13
Knowledge Management	13
Law	13
Library Information	13
Physicians	13
Program	13
Research Performance	13
Research Trends	13
SciSearch	13
Sociology	13
Tourism	13

Рассмотрим пересечения ВКТ на уровне тем. В табл. 2 приведены темы SciVal, связь между которыми оказалась максимальной за счет наибольшего числа со-слов, т.е. тех ВКТ из 181 темы, которые имели наибольшее число пересечений в этих темах. Видно, что чем больше число пересечений ВКТ (со-слов), тем выше вероятность того, что темы находятся внутри или около границ одного исследовательского поля, между ними существует сильная смысловая связь и что помимо тем библиотечно-информационного направления, к которым относится «Наукометрия», – а именно в этой области «Библиометрический анализ» является одним из основных методов исследования, – здесь фигурируют несколько тем по другим направлениям: «Медицина», «География» и др. Это свидетельствует о проникновении наукометрических методов в различные научные области как на постоянной, так и на эпизодической основе.

Таблица 2

**Топ-50 наиболее тесно связанных между собой тем SciVal по числу пересекающихся в них высокочастотных ключевых терминов (со-слов) из первой группы*.
Сортировка по нисходящему совместному пересечению ВКТ первой группы**

Тема	Число пересекающихся ВКТ в темах с «Библиометрическим анализом» в качестве ВКТ			
	Группа КТ 1	Группа КТ 2	Группа КТ 3	Всего: КТ1 + КТ2 + КТ3
Co-Authorship; Scientific Collaboration; Bibliometric Analysis	53	32	10	95
Hirsch Index; Self-Citation; Bibliometric Analysis	51	33	9	93
Hospitality; Tourism and Hospitality; Hospitality Management	51	21	18	90
Library Science; Tenure; Land Information System	50	27	14	91
Readership; Altmetrics; Journal Impact Factor	50	26	12	88
Reviewers; Peer Review; Manuscripts	49	28	16	93
Scholarly Publishing; Scientometrics; COVID-19	48	26	21	95
Beauties; Citations; Bibliometric Analysis	48	25	15	88
Intellectual Structure; Bibliometric Analysis; Scientometrics	46	34	17	97
Citations; Bibliometric Analysis; Journal Impact Factor	46	32	16	94
Open Access Publishing; Publishers; Scholarly Communication	46	26	16	88
Ranking; Journal Ranking; AACSB	45	26	21	92
Bibliometric Analysis; Research Evaluation; Book Publishers	45	31	15	91
Nursing Informatics; Education; Yearbooks	44	28	23	95
European Regional Development Fund; Pediatrics; ERDF	44	32	15	91
Open Access Publishing; Institutional Repository; Preprints	44	31	14	89
Citations; Research Productivity; Bibliometric Analysis	44	20	21	85
Communication; Doctoral Thesis; Spanish Universities	43	33	13	89
Market Research; Marketing; Academic Journals	42	25	23	90
Editor; Self-Citation; Introduction	42	20	23	85
Medical Writing; Biomedical Research; Co-Authorship	41	34	14	89
Publication Ethics; Editor-In-Chief; Scientific Misconduct	41	28	17	86
Information Systems Research; Congresses; Knowledge Management	39	34	18	91
Knowledge Organization; Paul Otlet; Brazil	38	38	17	93
Acknowledgement; Funding; Thanking	38	31	21	90
Amparo; Physical Education; Brazil	38	36	13	87
Scientific Publications; Biomedical Research; Bibliometric Analysis	37	33	20	90
Clinical Trials; Medical Journals; Ethical Issues	36	33	20	89
Citations; Co-Authorship; Economic Journals	35	31	18	84
Excellence; Bibliometric Analysis; Bibliographic Information	35	15	17	67
Medical Students; Earning A Doctorate; Scientific Publications	34	24	28	86
Engineering Education; Educational Research; Congresses	33	32	25	90
Hyperlink; Ranking; Search Engines	33	19	30	82
Scientometrics; Biomedical Research; Research Output	33	22	26	81
Research Personnel; Research Productivity; Women in Science	32	29	26	87
Abstracts; Clinical Trials; Reporting Bias	32	35	17	84
Brazil; Nursing; Theses	32	28	24	84
Journal Impact Factor; Case Report; Larynx	31	34	21	86
Biomedical Research; Bibliometric Analysis; Arab Countries	31	23	25	79
Bibliometric Analysis; Female Scientist; Map Projections	30	33	26	89
Scientific Workforce; Career Outcomes; Career Choice	30	31	26	87
Research Performance; Institute of Technology; Scientometrics	30	18	29	77
Research Personnel; Data Reuse; Librarians	30	25	22	77
Congresses; Citation Counts; Engineering Research	30	26	18	74
Interamerican Society; Argentina; Latin America	30	21	22	73
Middle Income Country; Health Research; Income	29	30	28	87
Geography; Neoliberalization; First International	29	20	21	70
Preprints; Space-Time Adaptive Processing; Biomedical Research	28	39	19	86
Scientometrics; Hirsch Index; Medical Journals	27	27	33	87
Bibliometric Analysis; Citation Index; Research Trends	27	22	35	84

* *Примечание:* см. табл. 1

Таблица 3

Наиболее активно развивающиеся темы (T_{lv}<1) с высокочастотным ключевым термином «Библиометрический анализ» и их процентиль актуальности*

№ п/п	Тема	T_{lv}	Процентиль известности темы
1	Smart Cities; Big Data; Internet of Things	0,01	99,863
2	Water Footprint; Water-Energy Nexus; Nexus	0,01	99,835
3	Platforms; Collaborative Consumption; Peer to Peer	0,03	99,651
4	Supply Chain; Environmentally Preferable Purchasing; Green Practices	0,03	99,933
5	Big Data; Business Intelligence; Supply Chain Management	0,04	99,667
6	Business Model Innovation; Innovation; Digital Transformation	0,05	99,825
7	Intellectual Structure; Bibliometric Analysis; Scientometrics	0,06	99,432
8	Research Personnel; Research Productivity; Women in Science	0,07	98,878
9	Social Entrepreneurship; Innovation; Impact Investing	0,07	99,47
10	Information Modeling; Facilities Management; Construction Industry	0,08	99,836
11	Hirsch Index; Self-Citation; Bibliometric Analysis	0,099	98,818
12	Academic Medicine; Medical School; Female Physician	0,12	99,022
13	Industrial Symbiosis; Sustainable Development; Circular Economy	0,12	99,88
14	*Editorial Comments; Spelling; Pagination*	*0,14*	*72,791*
15	Co-Authorship; Scientific Collaboration; Bibliometric Analysis	0,19	98,296
16	*Knowledge Organization; Paul Otlet; Brazil*	*0,19*	*87,5*
17	Industry 4.0; Cyber Physical System; Manufacturing	0,25	98,833
18	*Ranking; Journal Ranking; AACSB*	*0,27*	*88,634*
19	Reviewers; Peer Review; Manuscripts	0,28	94,847
20	International New Ventures; Born Global; Small And Medium-Sized Enterprises (SMEs)	0,3	99,277
21	Open Access Publishing; Institutional Repository; Preprints	0,32	95,526
22	Citations; Research Productivity; Bibliometric Analysis	0,33	91,057
23	Hospitality; Tourism and Hospitality; Hospitality Management	0,35	96,3
24	Prefabricated Buildings; China; Precast Concrete	0,35	97,209
25	*Medical Students; Earning A Doctorate; Scientific Publications*	*0,37*	*74,116*
26	Consort; Meta-Analysis; Reporting	0,41	99,452
27	Information Literacy; Library Instruction; Librarians	0,41	96,706
28	Nursing Informatics; Education; Yearbooks	0,45	93,202
29	Capacity Building; Capacity Development; Income	0,47	95,682
30	Technology Roadmapping; Data Mining; Technological Competitiveness	0,49	98,242
31	Impotence; Phosphodiesterase V Inhibitor; Tadalafil	0,5	94,635
32	Topic Model; Data Mining; Text Classification	0,56	98,693
33	Biorefining; Bioenergy; Bioeconomy	0,59	96,991
34	Citations; Bibliometric Analysis; Journal Impact Factor	0,67	93,733
35	Readership; Altmetrics; Journal Impact Factor	0,72	97,816
36	Continuous Auditing; Financial Statement Audit; Big Data	0,79	92,408
37	Research Personnel; Data Reuse; Librarians	0,85	96,21
38	*Bibliometric Analysis; Research Evaluation; Book Publishers*	*0,92*	*86,797*
39	*Library Science; Tenure; Land Information System*	*0,92*	*84,683*

* *Примечание:* по состоянию на ноябрь 2022 г.

Увеличение числа публикаций по различным областям знания (вне библиотечно-информационной сферы) с применением термина «Библиометрический анализ» вызвано, скорее всего, потребностью оценки их состояния и развития, и вряд ли в таких случаях можно говорить о полноценной междисциплинарности исследований.

Интенсивность развития тем можно выявить с помощью коэффициента T_{lv}[6]: чем меньше значение T_{lv}, тем интенсивнее развивается тема. В табл. 3 значения T_{lv} тем SciVal с ВКТ «Библиометрический анализ» приводятся с указанием их процентиля актуальности[7].

Данные табл. 3 показывают, что 33 из 39 тем (85%) имели низкий коэффициент T_{lv} и высокие показатели актуальности. Из них: 15 (38%) с T_{lv}<1 оказались «горячими исследовательскими фронтами», а 18 (46%) – «исследовательскими фронтами» [1]. Лишь у 9 (15%) тем процентиль актуальности оказался ниже 90 и соответствовал интервалу значений от 72,8 до 87,5. Мож-

[6] Формулу расчета см. в разделе: «Методология и методы исследований» настоящей статьи.

[7] Известность тем SciVal выражается в виде индикатора «процентиль» (Topic prominence percentile, eng.), который определяется системой Scopus.

но предположить, что в это число могли попасть совсем еще новые «нераскрученные» темы или же наоборот, те, которые начинают «угасать».

В табл. 3 представлены темы SciVal с коэффициентом $T_{lv}<1$, показывающим, что в этих темах доля ВКТ с положительной динамикой употребляемости преобладала над негативной и нулевой.

Однако помимо развивающихся немало и «затухающих» тем, в которых все ВКТ не только не имели положительного роста динамики, но и часто наблюдалось активное ее снижение. Таких тем оказалось 46 – четверть от всех анализируемых. В это число попали такие, казалось бы, актуальные темы в библиотечно-информационной области как: «Academic Libraries, Bibliometric Analysis, Library and Information Science» (процентиль актуальности – 37,308); «Scientometrics, Public Policy Issue, Citations» (процентиль актуальности – 48,328); «Scholarly Communication, Hubs, Open Access» (процентиль актуальности – 49,268).

181 анализируемая тема разделилась по числу ВКТ с релевантностью «единица»[8] на три группы тем, в которых:

1) лишь один ВКТ имел релевантность «единица» – 155 тем (85%);

2) несколько (два и более, но не все) ВКТ имели релевантность «единица» – 16 тем (9%);

3) все ВКТ имели релевантность «единица» – 10 тем (5,5%). Эта группа обладала двумя характерными особенностями: (1) все эти темы не имели ни одного ВКТ с положительной динамикой употребляемости, а, напротив, у многих (в некоторых случаях у всех) наблюдалось 100%-е снижение; (2) в данной группе тем общее число ВКТ (согласно SciVal) заметно меньше, чем во всех остальных: от 5 до 41 ВКТ против 100 ВКТ, характерных для большинства тем.

Из 181 темы только в 29-и (16%) ВКТ «Библиометрический анализ» имел коэффициент релевантности «единица», а положительная динамика его употребляемости в этом случае наблюдалась лишь в 8-и темах: 1. «Citations, Research Productivity, Bibliometric Analysis» (+591,40%); 2. «Intellectual Structure, Bibliometric Analysis, Scientometrics» (+305,40%); 3. «Biomedical Research, Bibliometric Analysis, Arab Countries (+137,50%); 4. «Hirsch Index, Self-Citation, Bibliometric Analysis» (+100,90%); 5. «Scientific Publications, Biomedical Research, Bibliometric Analysis» (+76,50%); 6. «Co-Authorship, Scientific Collaboration, Bibliometric Analysis» (+73,70%); 7. «Citations, Bibliometric Analysis, Journal Impact Factor (+60,70%); 8. «Beauties, Citations, Bibliometric Analysis» (+4%).

Как уже отмечалось, термин «Библиометрический анализ» широко проник в самые разные научные направления в качестве метода оценки состояния научных исследований в той, или иной научной области. Тем не менее, этот термин более характерен для библиотечно-информационной сферы. Подтверждением этого вывода может служить то, что наибольшее снижение употребляемости ВКТ «Библиометрический анализ» с релевантностью «единица» в темах наблюдалось вне библиотечно-информационного поля, где, по-видимому, он использовался в качестве метода оценки развития науки и научных направлений: «China, Understanding, Combination» (-100%); «Data Fusion, Absorbing, Multimode» (-100%); «Electronic Medical Record, Phenotype, Information Retrieval» (-100%); «Evidence, Control, Research» (-100%); «Food Technology, Knowledge Map, Privacy Policies» (-100%); «Iodine-131, Extract, Antiinflammatory Activity» (-100%); «Physical Disability, Disabled Persons, Health» (-100%); «Waste Water, Pyrolysis, Response Surface Methodology» (-100%). И только в одной теме библиотечно-информационной тематики наблюдалось значительное снижение употребляемости ВКТ «Библиометрический анализ»: «Scientometrics, Public Policy Issue, Citations» (-100%), что, вероятно, связано с «перетеканием» этого ключевого термина в близкую по содержанию, но уже с другим названием тему.

ЗАКЛЮЧЕНИЕ

В настоящей статье представлены результаты исследования взаимодействий высокочастотных ключевых терминов в темах SciVal (Scopus). В качестве основы для изучения служил набор тем SciVal, в которых ключевой термин «Библиометрический анализ» являлся высокочастотным. Это позволило определить «ядро» высокочастотных со-слов для тем SciVal, связанных с «Библиометрическим анализом».

Предложенная нами схема анализа пересечений ВКТ в темах, затрагивающих общую проблему (или метод исследования), позволяет выявить широту их междисциплинарности. Кроме того, появляется возможность установления круга наиболее актуальных и активно развивающихся научных тем в контексте с определенным ключевым термином, в данном случае – термином «Библиометрический анализ».

Результаты выполненной работы способствуют пониманию взаимосвязей между различными исследовательскими темами посредством анализа динамики со-слов и подтверждают гипотезу о том, что с помощью сетей со-слов из разных дисциплин можно выявить общность между ними и чем больше совпадений между двумя ключевыми словами, тем теснее взаимосвязь различных тем.

СПИСОК ЛИТЕРАТУРЫ

1. Xia W., Jiang Y., Zhu W., Zhang S., Li T. Research Fronts of Computer Science: A Scientometric Analysis // Journal of Scientometric Research. – 2021. – Vol. 10, № 1. – P. 18-26. DOI: https://doi.org/10.5530/jscires.10.1.3.
2. Price DJDS. Networks of Scientific Papers: The pattern of bibliographic references indicates the nature of the scientific research front // Science. – 1965. – Vol. 149, № 3683. – P. 510-515. DOI: https://doi.org/10.1126/science.149.3683.510.
3. Garfield E. Research fronts // Current Contents. – 1994. – Vol. 41, № 10. – P. 3-7.

[8] Ключевой фразе присваивается значение «релевантности» в диапазоне от 0 до 1.

4. Deng S., Xia S. Mapping the interdisciplinarity in information behavior research: a quantitative study using diversity measure and co-occurrence analysis // Scientometrics. – 2020. – Vol. 124. – P. 489-513. DOI: https://doi.org/10.1007/s11192-020-03465-x.
5. Han X. Evolution of research topics in LIS between 1996 and 2019: an analysis based on latent Dirichlet allocation topic model // Scientometrics. – 2021. – Vol. 125. – P. 2561-2595. DOI: https://doi.org/10.1007/s11192-020-03721-0.
6. Onyancha O.B. Forty-Five Years of LIS Research Evolution, 1971–2015: An Informetrics Study of the Author-Supplied Keywords // Publishing Research Quarterly. – 2018. – Vol. 34. – P. 456-470. DOI: https://doi.org/10.1007/s12109-018-9590-3.
7. Chang Y.-W., Huang M.-H., Lin C.-W. Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses // Scientometrics. – 2015. – Vol. 105. – P. 2071-2087. DOI: https://doi.org/10.1007/s11192-015-1762-8.
8. Liu P., Wu Q., Mu X., Yu K., Guo Y. Detecting the intellectual structure of library and information science based on formal concept analysis // Scientometrics. – 2015. – Vol. 104. – P. 737-762. DOI: https://doi.org/10.1007/s11192-015-1629-z.
9. Huang M.-H., Chang Y.-W. A comparative study of interdisciplinary changes between information science and library science // Scientometrics. – 2012. – Vol. 91. – P. 789-803. DOI: https://doi.org/10.1007/s11192-012-0619-7.
10. Chen G., Xiao L. Selecting publication keywords for domain analysis inbibliometrics: A comparison of three methods // Journal of Informetrics. – 2016. – Vol. 10. – P. 212-223. DOI: https://doi.org/10.1016/j.joi.2016.01.006.
11. Su H.N., Lee P.C. Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in technology foresight // Scientometrics. – 2010. – Vol. 85, №1. – P. 65-79. DOI: https://doi.org/10.1007/s11192-010-0259-8.
12. Chen X., Chen J., Wua D., Xie Y., Li J. Mapping the research trends by co-word analysis based on keywords from funded project // Procedia Computer Science. – 2016. – Vol. 91. – P. 547-555. DOI: https://doi.org/10.1016/j.procs.2016.07.140.
13. Zong Q.J., Shen H.Z., Yuan Q.J., Hu X.W., Hou Z.P., Deng S.G. Doctoral dissertations of Library and_Information Science in China: A co-word analysis // Scientometrics. – 2013. – Vol. 94. – P. 781-799. DOI: https://doi.org/10.1007/s11192-012-0799-1.
14. An X.Y., Wu Q.Q. Co-word analysis of the trends in stem cells field based on subject heading weighting // Scientometrics. – 2011. – Vol. 88. – P. 133-144. DOI: https://doi.org/10.1007/s11192-011-0374-1.
15. Kontostathis A., Galitsky L.M., Pottenger W.M., Roy S., Phelps D.J. A Survey of Emerging Trend Detection in Textual Data Mining // In: Berry M.W. (eds.) Survey of Text Mining. – New York, NY: Springer, 2004. – P. 185-224. DOI: https://doi.org/10.1007/978-1-4757-4305-0_9.
16. Губанов Д.А., Макаренко А.В., Новиков Д.А. Методы анализа терминологической структуры предметной области (на примере методологии) // Управление большими системами. – 2013. – Вып. 43. – С. 5-33. DOI: https://doi.org/10.1134/S0005117914120133.
17. Small H., Boyack K. W., Klavans R. Identifying emerging topics in science and technology // Research Policy. – 2014. – Vol. 43, № 8. – P. 1450-1467. DOI: https://doi.org/10.1016/j.respol.2014.02.005.
18. Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature // Journal of the American Society for Information Science and Technology. – 2005. – Vol. 57, № 3. – P. 359-377. DOI: https://doi.org/10.1002/ asi.20317.
19. Wang X., Cheng Q., Lu W. Analyzing evolution of research topics with NEViewer: a new method based on dynamic co-word networks // Scientometrics. – 2014. – Vol. 101. – P. 1253-1271. DOI: https://doi.org/10.1007/s11192-014-1347-y.
20. Zhao D. The Knowledge Base and Research Front of Information Science 2006–2010: An Author Cocitation and Bibliographic Coupling Analysis // Journal of the Association for Information Science and Technology. – 2014. – Vol. 65, № 5. – P. 995-1006. DOI: https://doi.org/10.1002/asi.23027.
21. Cheng Q., Wang J., Lu W., Huang Y., Bu Y. Keyword-citation-keyword network: a new perspective of discipline knowledge structure analysis // Scientometrics. – 2020. – Vol. 124. – P. 1923-1943. DOI: https://doi.org/10.1007/s11192-020-03576-5.
22. Сысоев А.Н., Цветкова В.А., Тютюнова В.С. Лингвистические методы анализа данных в задачах наукометрии // Научно-техническая информация. Сер. 1. – 2018. – № 9. – С. 22-27.
23. Антопольский А.Б., Савчук С.О., Тамеев А.А. О разработке онтологии поисковых терминов по лингвистике // Информационные ресурсы России. – 2020. – № 4(176). – С. 2-7.
24. Власова С.А., Каленов Н.Е. Автоматизированная система анализа ключевых терминов // Программные продукты и системы. – 2022. – Т. 35, № 3. – С. 420-427. DOI: https://doi.org/10.15827/ 0236-235X.139.420-427.
25. Мохначева Ю.В., Цветкова В.А. Российские публикации по библиотечно-информационным наукам в Scopus // Научные и технические библиотеки. – 2022. – № 3. – С. 14-38. DOI: https://doi.org/10.33186/1027-3689-2022-3-14-38.
26. Donthu N., Kumar S., Mukherjee D., Pandey N., Lim W.M. How to conduct a bibliometric analysis: An overview and guidelines // Journal of Business Research. – 2021. –Vol. 133. – P. 285-296. DOI: https://doi.org/10.1016/j.jbusres.2021.04.070.
27. Mokhnacheva Yu.V., Tsvetkova V.A. Terminological Analysis of Publications as a Method of Research Trends in Science // Third International Conference on Science & Technology Metrics (STMet

2021), Virtual, December 06-08, 2021, 2022. – С. 1-5. DOI: https://doi.org/10.6025/stm/2021/3/1-5.
28. Мохначева Ю.В., Цветкова В.А. Развитие тематики научных исследований на основе терминологического подхода (на примере темы «Иммунология и микробиология» по данным Scopus - SciVal) // Научно-техническая информация. Сер. 1. – 2021. – № 6. – С. 22-28. DOI: https://doi.org/10.36535/0548-0019-2021-06-3.
29. Гиляревский Р.С. Библиометрия как понимание социальных закономерностей научных коммуникаций // Труды СПбГИК. – 2022. – Т. 226. – С. 15-23.
30. Handbook of Special Brinkmanship and Information Work: 2nd ed. / gen. ed W. Ashworth. – London: Aslib, 1962. – P. 364.
31. SciVal Quick Reference Guide. – Elsevier, 2019. – 20 p.

Материал поступил в редакцию 21.07.23.

МОХНАЧЕВА Юлия Валерьевна – кандидат педагогических наук, ведущий научный сотрудник Библиотеки по естественным наукам Российской академии наук (БЕН РАН), Москва
e-mail: j-v-m@yandex.ru

УВАЖАЕМЫЕ КОЛЛЕГИ!

ВИНИТИ РАН предлагает Вашему вниманию
База данных (БД) ВИНИТИ РАН в режиме online

База данных (БД) ВИНИТИ РАН — Федеральная база отечественных и зарубежных публикаций по естественным, точным и техническим наукам. Генерируется с 1981 г., обновляется ежемесячно, пополнение составляет более 600 000 документов в год.

БД ВИНИТИ РАН включает 26 тематических фрагментов, состоящих из более чем 190 разделов.

Документы БД содержат библиографию, ключевые слова, рубрики и реферат первоисточника.

На основе БД ВИНИТИ пользователям доступны следующие продукты:

- online доступ к базе данных круглосуточно, без выходных;
- выполнение тематического поиска специалистом ВИНИТИ по запросу заказчика;
- по заявкам предоставляются любые наборы тематических фрагментов БД ВИНИТИ или их разделов на любых видах электронных носителей, или через FTP-сервер;
- для ознакомления с возможностями поиска имеется демо-версия базы данных bd.viniti.ru.

База данных ВИНИТИ зарегистрирована Российским агентством по правовой охране программ для ЭВМ, баз данных и топологий интегральных микросхем (РосАПО) (Свидетельство № 960034 от 23.09.1996г.)

Подробную информацию Вы можете получить:

Адрес: 125190, Россия, Москва, ул. Усиевича, 20, ВИНИТИ РАН

Телефон: 8 499-152-54-81

E-mail: feo@viniti.ru

УВАЖАЕМЫЕ КОЛЛЕГИ!

ВИНИТИ РАН предлагает Вашему вниманию Реферативный Журнал в электронной форме

РЖ в электронной форме (ЭлРЖ) выпускается по всем разделам естественных, технических и точных наук.

Каждый номер ЭлРЖ является полным аналогом печатного номера РЖ по составу описаний документов, их оформлению и расположению. Он сопровождается оглавлением, указателями.

ЭлРЖ представляет собой информационную систему, снабженную поисковым аппаратом и позволяющую пользователю на персональном компьютере:

- читать номер РЖ, последовательно листая рефераты;
- просматривать рефераты отдельных разделов по оглавлению;
- обращаться к рефератам по указателям авторов, источников, ключевых слов;
- проводить поиск документов по словам и словосочетаниям;
- выводить текст описаний документов во внешний файл.

ЭлРЖ могут быть:

- записаны на DVD-ROM;
- передаваться через FTP-сервер (клиенту предоставляется логин и пароль с доступом к FTP-серверу ВИНИТИ, с которого он скачивает заказанные журналы).

Электронные реферативные журналы можно заказать за текущий год с любого номера, а также за предыдущие годы.

Подробную информацию Вы можете получить:

Адрес: 125190, Россия, Москва, ул. Усиевича, 20, ВИНИТИ РАН

Телефон: 8 499-152-62-11

E-mail: feo@viniti.ru

ВИНИТИ РАН

Центр научно-информационного обслуживания

Информационные услуги, предоставляемые ЦНИО ВИНИТИ РАН:

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

ВИНИТИ РАН располагает следующими информационными ресурсами:

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

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

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

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

Проводится индивидуальное обслуживание пользователей в читальном зале ЦНИО ВИНИТИ РАН.

Подробную информацию Вы можете получить:
Адрес: 125190, Россия, г. Москва, ул. Усиевича, 20, ВИНИТИ РАН;
Телефоны: 499-155-42-17, 499-155-42-43;
E-mail: cnio@viniti.ru