

УДК 004.891:004.9

## ОЦЕНКА АКТУАЛЬНОСТИ ЗАЯВКИ НА ПРОЕКТНОЕ ФИНАНСИРОВАНИЕ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ РАСПРЕДЕЛЕНИЯ ГРАНТОВ

Сироткин А.В., Копченко В.К.

ФГБОУ ВО «Северо-Восточный государственный университет», Магадан,  
e-mail: andrew\_sirotkin@mail.ru, vkopchenko49@gmail.com

В работе рассматривается реализация этапа автоматизированной оценки актуальности заявки, представленной для соискания проектного финансирования. Данный этап является частью процесса рассмотрения и принятия решения по заявкам на выделение грантов, производимого с использованием разрабатываемой авторами автоматизированной системы распределения грантов (АСРГ). Актуальность заявки рассматривается как оценка соответствия множества частных критериев формализованного предлагаемого соискателем решения множеству накопленных аналогов, выступающих в качестве эталонных решений в данной предметной области. Предварительным этапом, необходимым для оценки актуальности, выступает строгая формализация материала заявки по достигаемым показателям для последующей адекватной оценки соответствия их значений актуальным значениям эталонных критериев. Последующим этапом выступает установление состава этих критериев в материалах описания заявки на основе изменённой метрики TF-IDF, с использованием анализа взвешенных термов в тексте заявки. Непосредственно оценка производится путём сравнения значений эталонных частных критериев и сформулированных показателей для аналогичного технического решения. В качестве окончательного этапа анализа выступает принятие решения на основе конкурентного поиска среди заявок той, состав показателей которой наилучшим образом соотносится с максимальным количеством эталонных критериев.

**Ключевые слова:** проектное финансирование, грант, модель, автоматизация, модель заявки

## ASSESSMENT OF THE RELEVANCE OF AN APPLICATION FOR PROJECT FINANCING IN AN AUTOMATED GRANT DISTRIBUTION SYSTEM

Sirotkin A.V., Kopchenko V.K.

North-Eastern State university, Magadan, e-mail: andrew\_sirotkin@mail.ru, vkopchenko49@gmail.com

The paper considers the implementation of the stage of automated assessment of the relevance of the application submitted for project financing. This stage is part of the process of reviewing and making decisions on grant applications made using the automated grant distribution system (ASGD) developed by the authors. The relevance of the application is considered as an assessment of the compliance of a set of particular criteria of the formalized solution proposed by the applicant with a set of accumulated analogues acting as reference solutions in this subject area. The preliminary stage necessary for the assessment of relevance is the strict formalization of the application material according to the achieved indicators for the subsequent adequate assessment of their compliance with the actual values of the reference criteria. The next step is to establish the composition of these criteria in the materials of the application description based on the modified TF-IDF metric, using the analysis of weighted terms in the application text. The evaluation is carried out directly by comparing the values of the reference private criteria and the formulated indicators for a similar technical solution. The final stage of the analysis is the decision-making based on a competitive search among the applications of the one whose composition of indicators best correlates with the maximum number of reference criteria.

**Keywords:** project financing, grant, model, automation, application model

В рамках реализации Указов Президента Российской Федерации от 7 мая 2018 года сформулирована концепция развития цифровых технологий в экономике и государственном управлении. Особое внимание в данной концепции уделяется улучшению условий ведения предпринимательской деятельности, что в немалой степени зависит от государственной, региональной или отраслевой поддержки малого и среднего предпринимательства. Этот факт отмечают различные исследователи, в том числе, например, и для Магаданской области [1].

Одной из форм поддержки инновационных решений и предложений в этой

сфере экономики является проектное финансирование и субсидирование субъектов малого и среднего предпринимательства, осуществляемого на основе распределения грантов. Грантовая поддержка осуществляется на экспертной основе, существует ряд проблем, связанный с принятием решения о выделении гранта, например [2], ведутся соответствующие исследования и поиски решений, направленные на разрешение этих проблем или снижение их влияния на конечный результат, например [3].

Одним из решений, направленных на повышение эффективности работы грантовой системы, может быть применение автоматизированных технических средств,

решающих задачи анализа заявки и принятия решения о выделении гранта. В настоящее время на рынке программного обеспечения отсутствуют специализированные технические решения, ориентированные на заданные задачи, в связи с чем разработка подобной системы представляется весьма актуальной. Применение подобного средства позволит исключить субъективность, повысит скорость и прозрачность процесса принятия решения.

### Материалы и методы исследования

Работа автоматизированной системы распределения грантов (АСРГ) предполагает, по мнению авторов, решение последовательного ряда задач:

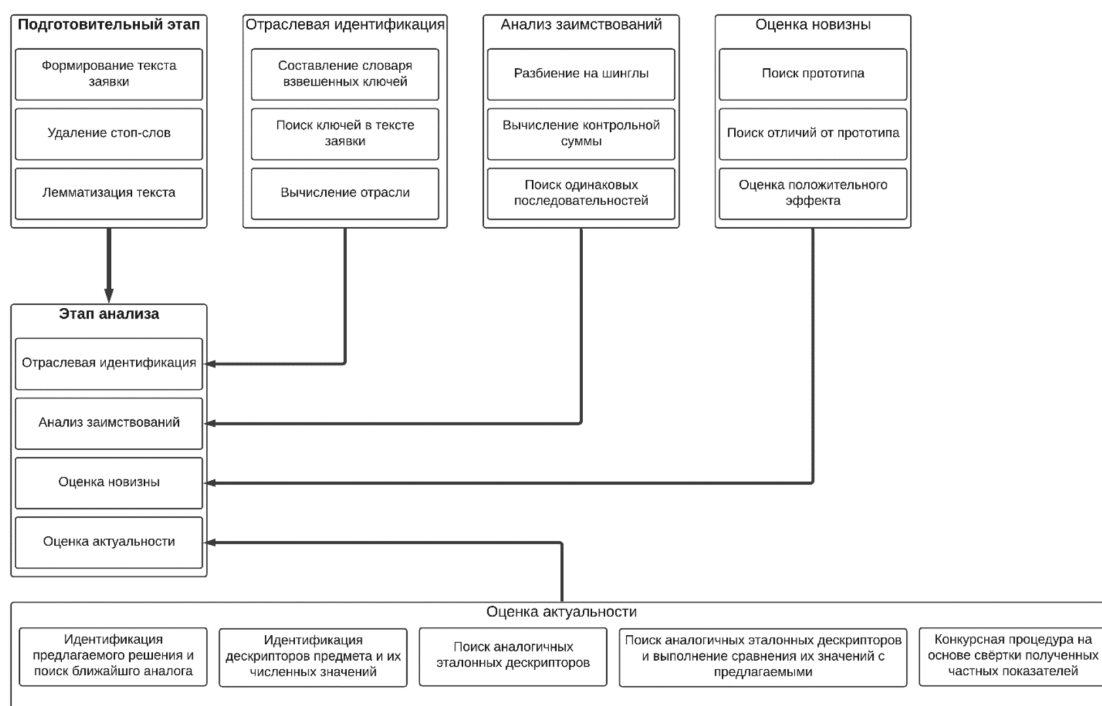
- определение отраслевой принадлежности заявки;
- установление её уникальности и новизны;
- определение актуальности предлагаемого в заявке решения.

Для реализации данных прецедентов необходима разработка соответствующих компонентов АСРГ; первые два из них были разработаны и представлены авторами в работах [4-6]. На рисунке приведена процессная схема работы компонентов АСРГ.

Анализ заявки в АСРГ состоит из двух этапов: подготовительного и аналитического.

На подготовительном этапе текст заявки очищается от стоп-слов, то есть слов, не несущих смысла в отрыве от контекста. К таким словам относятся союзы, предлоги, междометия, местоимения, частицы, HTML-теги. В исключительных случаях из текста могут быть удалены прилагательные. Затем текст очищается от стоп-символов (знаков препинания, служебных символов, числовых знаков, знаков табулирования), выполняется приведение каждого слова в тексте к нижнему регистру, затем – к нормальной форме (для существительных – именительный падеж, единственное число; для прилагательных – именительный падеж, единственное число, мужской род; для глаголов, причастий, деепричастий – глагол в инфинитиве (неопределённой форме) несовершенного вида. Этап анализа состоит из отраслевой идентификации, анализа уникальности, оценки новизны и оценки актуальности.

Задача определения отраслевой принадлежности решается аналогично задаче рубрикации текста с помощью сопоставления каждому терму документа величины, называемой весом, и использования метрики *TF-IDF*, согласно которой вес слова пропорционален частоте использования этого слова в документе и обратно пропорционален частоте использования слова во всех документах коллекции.



Процессная схема работы компонентов

Задача определения уникальности решается с помощью алгоритма поиска нечетких дубликатов, иначе называемого алгоритмом шинглов, в основе которого идет канонизация текста заявки с последующим разделением текста на последовательности слов произвольного размера, но не более 10 слов в каждой последовательности. В настоящее время данная технология достаточно хорошо проработана исследователями (например [7]) и широко используется для численной оценки оригинальности текстовых документов. Основу алгоритма составляет вычисление контрольных сумм (иначе «сигнатур») – уникальных чисел, поставленных в соответствие некоторому тексту. Для каждой последовательности слов рассчитываются 84 значения контрольной суммы с использованием различных функций (*SHA1*, *MD5*, *CRC32*, *FNV* и т.д.). После вычислений контрольные суммы предметного и эталонного текстов сравниваются между собой. Обычно для повышения производительности вычислений сравниваются выборки значений результирующих сумм, кратные 25. Несмотря на то что алгоритм шинглов считается не самым надежным в силу того, что изменение даже одной буквы слова меняет хеш-сумму всей строки, качественная канонизация текста позволяет минимизировать риск пропусков дубликата.

Задача оценки новизны решения может быть решена с помощью поиска прототипа проекта, поиска отличий проекта от прототипа и оценки положительного эффекта отличий. Прототипом можно считать заявку, у которой величина совпавших с текущей заявкой признаков новизны максимальна. В случае отсутствия результатов можно применить оценку важности признаков через сопоставление весов термам описательной части заявки. В случае если величина совпадений новых признаков в рассматриваемой заявке с признаками, предложенными в архивных документах, не превышает некоторого значения (по мнению авторов 50%), проект можно считать соответствующим критерию новизны.

Объектом настоящей работы является разработка модуля АСРГ, предназначенного для определения актуальности решения, заявляемого грантосоискателем.

Актуальность как важность или значимость предмета в системе измерений окружающих систем различными исследователями определяется по-разному в зависимости от предметной области, точки зрения исследователя, целей оценки и пр. Исследователи определяют актуальность

как меру важности, интересности, значимости, например [8].

Наиболее часто при определении актуальности численными методами используются такие подходы, как экспертный и наукометрический.

Известна оценка актуальности научных публикаций, основанная на оценке количества посещений информационного сервиса, учёте активности посетителей, маршрутов перемещений, вероятностных оценок отдельного наблюдения на основе наивного байесовского классификатора (например, [9]) и пр. Данная методика основана на анализе статистических данных посещений информационного интернет-ресурса, на котором размещены оцениваемые объекты, ориентирована на точку зрения посетителя, т.е. фактически реализует метод экспертных оценок, который, на взгляд авторов, не алгоритмизируем в контексте решения поставленной задачи в силу отсутствия широкой публикации заявки грантосоискателя и наличия временных ограничений на принятие решения.

Наукометрическая методика оценки актуальности научных исследований, предложенная, например, в [10], основывается на терминологических методах анализа частоты встречаемости ключевых слов в исследуемом предмете в сравнении с аналогичными работами, признанными актуальными в данной области. В предлагаемой методике учитывается новизна используемых терминов, но не анализируется влияние на целевую функцию, описывающую назначение данного предмета. Тем не менее предлагаемый в работе терминологический подход выглядит привлекательным для целей настоящего исследования и может быть использован после адаптации к выбранным авторами технологиям.

В наукометрической численной оценке актуальности важно учитывать параметр времени, который влияет на «убывание новизны», а значит и актуальности предмета, как показано, например, в [11]. В нашем случае для обеспечения этого критерия предполагается формирование и поддержание на экспертной основе эталонной базы объектов, обладающих высокой актуальностью на текущий момент времени.

Использование терминологических методов наукометрической оценки актуальности заявки предполагает контекстную формализацию предмета, которая, во-первых, позволит отнести предмет к ближайшим аналогам, во-вторых, формализует параметры оценки и их веса в отношении целевой функции предмета.

В этой связи, отталкиваясь от предложенных в работе [11] «терминологических дескрипторов», выступающих в роли качественных критериев терминологического анализа, для оценки актуальности заявки на получение гранта, предполагается использование набора дескрипторов (целевых показателей предмета), описывающих различные качества предмета, и сравнение их численных значений с эталонными аналогами для аналогичного объекта.

В аспекте решения, предлагаемого грантосоискателем, для оценки актуальности необходимо выполнение следующих процедур (рисунок):

- идентификация предлагаемого решения и поиск ближайшего аналога;
- идентификация дескрипторов предмета и их численных значений;
- поиск аналогичных эталонных дескрипторов и выполнение сравнения их значений с предлагаемыми;
- конкурсная процедура на основе свёртки полученных частных показателей.

Идентификация предмета выполняется с использованием накопленной базы аналогов, в качестве которых могут выступать как существующие решения, так и перспективные, отвечающие требованиям времени. Создание подобной базы находится в компетенции экспертов или систем на основе искусственного интеллекта.

Целью идентификации предмета является определение ближайшего аналога для сравнения его параметров с параметрами предмета. Идентификация производится методами терминологического анализа, в частности, в силу краткости описания предмета можно использовать для поиска ближайшего аналога модифицированную метрику  $TF-IDF$ , предложенную авторами в предыдущих публикациях, например в [6]. Пусть  $R$  – описание предмета заявки, содержащее некоторое множество лемматизированных терминов  $Q_R$ ,  $D$  – множество накопленных аналогов  $d$  из соответствующей отрасли, каждому из которых соответствует некоторое множество терминологических ключей  $C_d$ . Тогда поиск аналога, ближайшего к предмету, можно описать как

$$d = \arg \max S(Q_R, C_{d_j}), j \in \overline{[1, |D|]},$$

$$d_j \in D, S = |Q_R \cap C_{d_j}|, \quad (1)$$

где  $S$  – количество совпадений ключевых терминов.

Для повышения точности идентификации можно применить весовую оценку важности признаков, допустим, что на каждом

множестве  $C_{d_j}$  определено множество весов  $W_{d_j}$ . Тогда функция  $S$  примет следующий вид:

$$S = \sum_{k=1}^N w_k; w_k \in W_{d_j}; N = |W_{d_j}|;$$

$$h: W_{d_j} \rightarrow (Q_R \cap C_{d_j}). \quad (2)$$

После нахождения релевантного аналога может быть запущена процедура оценки актуальности заявленного объекта во множестве аналогов, которому принадлежит найденный образец.

Определение актуальности объекта заявки может производиться путём сравнения обобщенного показателя (ОП) дескрипторов проекта с аналогичными показателями для множества эталонных образцов в условиях принудительного ограничения числа дескрипторов для каждой итерации. Актуальность будет доказана, если обобщенный показатель заявляемого объекта превысит аналогичный показатель хотя бы одного эталонного объекта. Исходя из факта существования данного эталонного объекта во множестве аналогов, которые по обобщённому показателю могут его превосходить, логически следует, что данный эталонный объект может превосходить аналоги по неучитываемым в обобщённом показателе дескрипторам, в силу чего заявляемый объект, превосходящий эталон по обобщённому показателю, имеющему доминирующий характер, может конкурировать как с ним, так и с другими аналогами по дескрипторам, не используемым для свёртки ОП. Формализованно этот алгоритм можно представить следующим образом.

Используем метод экспертных оценок (аналогичный, например [12]) для определения матрицы эталонных аналогов  $P \times Z$ , атрибутами которой –  $Z$  – выступают дескрипторы, экземплярами  $P$  – объекты  $p_i, i \in \overline{[1, |P|]}$ , на пересечении строк и столбцов определены численные значения дескрипторов.

Пусть,  $A$  – множество требований (дескрипторов заявленного объекта), такое, что  $A \in Z$ , где  $a_i \in A$  – требование для текущего момента, актуальное в данной предметной области.

Установим неравновесное влияние каждого дескриптора на ОП объекта путём введения  $W$  – множества весов,  $h: W \rightarrow Z, \sum W = 1$ .

Для введения принудительного ограничения числа требований и, соответственно, дескрипторов, введём множество индексов

дескрипторов заявленного объекта и эталонных аналогов, такое, что

$$K = |A|, k \in [1, |K|]; I = |P|,$$

$$i \in [1, |I|]; J = |Z|, j \in [1, |J|].$$

Введём правое и левое ограничения на количество используемых дескрипторов:  $R_{max}$  – максимальное количество используемых дескрипторов,  $R_{min}$  – нижний предел того же множества, при том что  $R$  – множество дескрипторов заявленного объекта.

Для оценки актуальности заявленного объекта необходимо выбрать множество таких дескрипторов, значимость которых для объектов является наибольшей в условиях введённых ограничений:

$$r \in R, R_{min} \leq |R| \leq R_{max}; r_k = \operatorname{argmax}(Z_i W_i). \quad (3)$$

Следующим этапом является оценка актуальности на основе расчёта обобщенного показателя, использующего ту или иную функцию свёртки, мы в данном случае выбрали мультипликативную, которая, на наш взгляд, не имеет очевидной линейной зависимости от количества показателей.

Введём множество ОП эталонных аналогов –  $Q$ ,  $q_i \in Q$ , где  $q_i$  – ОП  $i$ -го проекта:

$$q_i = \prod_{l=1}^{|Z|} Z_l W_l. \quad (4)$$

Соответственно, ОП предмета определяется как

$$q_{пр} = \prod_{l=1}^{|K|} Z_{пр,l} W_{пр,l}. \quad (5)$$

Тогда при выполнении условия

$$\forall i \in [1, |P|] \exists (q_{пр} > q_i) \quad (6)$$

мы считаем предмет заявки актуальным и достойным дальнейшей экспертной оценки для поиска положительного решения на выделение гранта.

### Заключение

Разработанные в ходе исследования модели оценки актуальности могут быть использованы не только в рамках автоматизированной экспертизы заявок на про-

ектное финансирование, но и в качестве уточняющих инструментов при оценке научных публикаций и иных текстов, к которым предъявляются повышенные требования к актуальности. Применение данных моделей в сочетании с разработанными механизмами терминологического анализа позволит не только снизить влияние личных предпочтений экспертов в пользу объективности и качества экспертизы, но и планировать оцениваемые работы в соответствии с учетом актуальных требований отраслевой конъюнктуры.

### Список литературы

1. Акулич О.В. Современное состояние и достаточность государственной поддержки малого и среднего бизнеса в Магаданской области // Экономика: вчера, сегодня, завтра. 2017. № 8. С. 190–202.
2. Бабина К.И. Грант как один из способов финансирования научных учреждений: особенности, проблемы, перспективы // Вестник Саратовского государственного социально-экономического университета. 2014. № 5(54). С. 170–173.
3. Петрулев, Д.С., Гусарова М.С. Государственное регулирование и проблемы грантовой политики в Российской Федерации // Московский экономический журнал. 2019. № 6. С. 2. DOI: 10.24411/2413-046X-2019-16002.
4. Сироткин А.В. Отраслевая идентификация заявок в автоматизированной экспертной системе распределения грантов // Современные наукоемкие технологии. 2019. № 7. С. 99–103.
5. Сироткин А.В. Постановка задачи определения новизны заявки на проектное финансирование в автоматизированной системе распределения грантов // Инновационные технологии управления. 2019. № 6. С. 259–262.
6. Сироткин А.В. Модель определения новизны заявки на проектное финансирование методами терминологического анализа // Современные наукоемкие технологии. 2020. № 4. С. 239–244.
7. Цимбалов А.В., Золотарев О.В. Метод шинглов // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. 2016. № 4. С. 72–79.
8. Баранецкий А.Н. Мера актуальности ценности и мера ценности актуальности // Архонт. 2018. № 1(4). С. 10–17.
9. Силич В.А., Савельев А.О., Марчуков А.В., Алексеев А.А. Методы оценки актуальности научных публикаций, на основе анализа интернет – обращений к научным порталам // Доклады Томского государственного университета систем управления и радиоэлектроники. 2010. № 1–1. С. 283–287.
10. Артамонова Г.В., Галеев А.Р., Баженова Т.С. Альтернативный метод оценки актуальности медико-биологических научных исследований российских НИУ // Менеджер здравоохранения. 2012. № 12. С. 35–41.
11. Бугаев К.В. Возможности определения новизны и актуальности научных тем методами наукометрии (на примере криминалистики) // Научный вестник Омской академии МВД России. 2014. № 3(54). С. 49–54.
12. Орлов А.И. Эконометрика. М.: Экзамен, 2002. 413 с.