

ИНФОРМАЦИОННО-ЛОГИЧЕСКАЯ МОДЕЛЬ И АЛГОРИТМ АВТОМАТИЗИРОВАННОГО ПОДБОРА РЕЦЕНЗЕНТОВ СТАТЕЙ ДЛЯ НАУЧНОГО ЖУРНАЛА

А.В. Антонова, Р.Н. Голых

Бийский технологический институт (филиал) ФГБОУ ВО «АлтГТУ им. И.И. Ползунова», г. Бийск

Рецензирование научных публикаций играет ключевую роль в научной сфере и является неотъемлемым этапом в процессе публикации. В настоящее время наблюдается непрерывный рост числа статей во многих научных областях. Этот рост создает значительное напряжение на процесс рецензирования научной литературы, так как количество доступных публикаций превышает возможности рецензентов. Назначение квалифицированных рецензентов для оценки статей является сложной и ответственной задачей. Организаторы процесса рецензирования должны учитывать специализацию рецензентов, чтобы обеспечить соответствие их экспертизы теме статьи. Для обеспечения соответствия специализации рецензентов теме статьи одним из подходов является использование классификаторов. Недостатком классификаторов является то, что каждый автор должен указывать коды классификаторов для статьи, при этом, как правило, многие научные журналы монодисциплинарные, или, наоборот, один рецензент не является специалистом по всем областям. Поэтому для подбора рецензентов возникает необходимость анализировать содержимое материала публикации, не прибегая к классификаторам. Анализ содержимого материалов по тексту требует разработки систем искусственного интеллекта, который анализирует содержание текста, причем отдельно взятая фраза из текста ничего не говорит о тематике статьи, а ключевые слова могут относиться ко множеству статей и ко множеству тематик. Поэтому наиболее простой способ выявления тематики статьи – по названию. Поэтому возникает необходимость в разработке информационно-логической модели и, в конечном итоге, информационной системы, которая будет подбирать необходимых рецензентов статьи. При этом необходимо учитывать морфологию слов в названии статьи.

Ключевые слова: модель, информационная система, рецензирование, статья.

ВВЕДЕНИЕ

Рецензирование научных публикаций играет ключевую роль в научной сфере и является неотъемлемым этапом в процессе публикации. В настоящее время наблюдается непрерывный рост числа статей во многих научных областях. Этот рост создает значительное напряжение на процесс рецензирования научной литературы, так как количество доступных публикаций превышает возможности рецензентов. Назначение квалифицированных рецензентов для оценки статей является сложной и ответственной задачей. Организаторы процесса рецензирования должны учитывать специализацию рецензентов, чтобы обеспечить соответствие их экспертизы теме статьи. Для обеспечения соответствия специализации рецензентов теме статьи одним из подходов является использование классификаторов. Недостатком классификаторов является то, что каждый автор должен указывать коды классификаторов для статьи, при этом, как правило, многие научные журналы монодисциплинарные, или, наоборот, один рецензент не является специалистом по всем областям. Поэтому для подбора рецензентов возникает необходимость анализировать содержимое материала публикации, не прибегая к классификаторам. Анализ содержимого материалов по тексту требует разработки систем

искусственного интеллекта, который анализирует содержание текста, причем отдельно взятая фраза из текста ничего не говорит о тематике статьи, а ключевые слова могут относиться ко множеству статей и ко множеству тематик. Поэтому наиболее простой способ выявления тематики статьи – по названию. Поэтому возникает необходимость в разработке информационно-логической модели и, в конечном итоге, информационной системы, которая будет подбирать необходимых рецензентов статьи. При этом необходимо учитывать морфологию слов в названии статьи.

ОСНОВНАЯ ЧАСТЬ

Для разработки информационно-логической модели первоначально были рассмотрены основные подходы, алгоритмы, методы, инструменты, используемые для автоматизации подбора рецензентов [1–4].

1. **Mathematics Subject Classification 2010** для определения предметной области, к которой относятся результаты исследования, представленные в публикации. Метод реализован в виде сервиса по технологии плагинов OJS и может быть использован в любом математическом журнале, функционирующем на платформе OJS. Метод апробирован в информационной системе научного журнала Lobachevskii Journal of Mathematics.

Алгоритм автоматизированного подбора рецензентов научных статей реализован следующим образом: изначально создается и пополняется база экспертов-рецензентов, работающих с журналом. Для этого в автоматическом режиме осуществляется поиск в библиографических базах данных статей, написанных каждым экспертом. Из этих статей извлекаются коды классификатора предметных областей MSC2010, которые затем включаются в профиль соответствующего эксперта. Если код MSC2010 уже присутствует в профиле эксперта, то его вес увеличивается. Далее, при поступлении новой статьи в журнал, из ее метаданных также извлекаются коды MSC2010. На основе этих кодов формируется рейтинг экспертов, наиболее соответствующих тематике статьи. Рейтинг эксперта определяется степенью совпадения кодов MSC2010 статьи и профиля эксперта, с учетом весов кодов в профиле. При этом эксперты, являющиеся соавторами данной статьи, исключаются из списка возможных рецензентов. После формирования списка рекомендованных рецензентов, система автоматически отправляет им уведомления о необходимости провести рецензирование. Аналогичные уведомления рассылаются авторам статьи и редакторам журнала. Таким образом, весь процесс подбора и информирования рецензентов происходит в автоматическом режиме, без участия редакционного персонала. Реализация данного алгоритма осуществлена в виде отдельного программного модуля (плагина), интегрированного с платформой управления научными журналами Open Journal Systems (OJS). Это позволяет использовать разработанный метод в любом журнале, функционирующем на базе OJS [1].

2. Метод Peer2ref. Основная проблема, с которой сталкиваются авторы и издатели – это выбор подходящих рецензентов для анонимного рецензирования научных статей. Это становится все сложнее из-за растущей специализации ученых в узких областях исследований. Анализ литературы помогает найти подходящих рецензентов, но это трудоемкий и сложный процесс, осложняемый неоднозначностью имен авторов.

Peer2ref работает следующим образом.

Сначала из текста статьи (обычно из аннотации) выделяются ключевые слова и понятия. Это делается с помощью специального алгоритма, который определяет, какие слова наиболее важны для описания содержания.

Затем эти ключевые слова используются для поиска в базе данных авторов, чьи статьи содержат схожие понятия. Эта база данных была создана заранее и содержит профили ученых, основанные на ключевых словах из их публикаций. Чтобы создать эту базу данных экспертов, авторы Peer2ref взяли все

статьи, опубликованные за последние 10 лет в более чем 3800 журналах, индексируемых в MEDLINE.

Чтобы правильно найти нужных авторов, разработчики Peer2ref решили проблему одинаковых фамилий и инициалов. Они присвоили каждому автору уникальный идентификатор, основываясь на его соавторах.

Для каждого уникального автора в базе данных был создан профиль, отражающий его экспертизу. Этот профиль состоит из ключевых слов, извлеченных из аннотаций статей, которые он публиковал.

Когда пользователь вводит новую статью, Peer2ref сравнивает ее ключевые слова с профилями авторов в базе данных. Те авторы, чьи профили наиболее похожи на ключевые слова статьи, предлагаются в качестве потенциальных рецензентов.

Таким образом, когда пользователь вводит свою статью в Peer2ref, система сравнивает ее ключевые слова с профилями экспертов в базе данных и предлагает наиболее подходящих рецензентов [2].

Предложенная информационно-логическая модель основана на алгоритме Peer2ref за тем исключением, что для ускорения подбора рецензентов анализируется лишь названия статьи абсолютно по всем словам. При этом если рецензент по тематике статьи не имеется в базе данных, то осуществляется подбор рецензентов по смежным тематикам.

Согласно информационно-логической модели пользователь осуществляет ввод названия статьи (или выбор названия из базы статей).

В свою очередь, информационная система выполняет следующие действия:

- идет разделение названия статьи на слова;
- система получает название статьи из формы;
- система разбивает строку названия на отдельные слова, используя пробелы и знаки препинания как разделители;
- система использует phpMorphy для извлечения корней каждого слова;
- система проверяет наличие совпадений корней введенных слов со словами из названия статьи;
- система преобразовывает слова из названия статьи, которые имеют совпадение с корнями, в базовую форму;
- система сопоставляет базовые формы слов с терминами из базы данных рецензентов;
- система проверяет наличие совпадений базовых форм слов с терминами в базе данных рецензентов;
- система запускает SQL-запрос для поиска рецензентов, соответствующих терминам;
- подсчитывает совпадения и определяет рецензентов;
- система подсчитывает количество совпадений для каждого рецензента;
- система определяет подходящих рецензентов на основе количества совпадений;

• система формирует таблицу с подходящими рецензентами, количеством совпадений и соответствующими терминами, выводит таблицу на веб-страницу.

Исходные данные для работы системы представляются следующим образом:

$c : [1; n] \cap N \rightarrow C$; – символы в названии статьи (C – множество символов UTF-8, n – количество символов в названии статьи);

R – множество рецензентов;

W – множество ключевых слов;

$T \subseteq R \times W$ – соответствие рецензентов и ключевых слов.

Математически алгоритм работы информационной системы заключается в следующем:

1. Построение массива слов (включая предлоги): $w : [1; m] \rightarrow N \times N$;

где m – количество слов.

$w(i) = (w_1(i); w_2(i))$ – номер первого символа слова из названия статьи и номер последнего символа соответственно.

Для массива слов выполняются следующие условия:

$$\forall i \in [1; m] \cap N; \forall j \in [w_1(i); w_2(i)] \cap N$$

$$c(j) \neq space;$$

$$\forall i \in [1; n] \cap N | \forall k \in [1; m] \cap N$$

$$i \notin [w_1(k); w_2(k)]:$$

$$c(i) = space.$$

2. Подсчёт количества совпадений ключевых слов для каждого рецензента:

$$K(r) =$$

$$= \sum_{(s,v) \in T \cap (\{r\} \times W)} \sum_{i=1}^m P(v, w_1(i), w_2(i));$$

$$K : R \rightarrow N$$

где θ – функция Хэвисайда; P – функция, равная 1, если совпадают корни ключевого слова v и слова – подстроки символов из c с индексами от $w_1(i)$ до $w_2(i)$ включительно.

3. Выбор рецензентов согласно следующему выражению:

$$M = \max_{r \in R} K(r);$$

$$R_{corr} = \begin{cases} M = 0, \emptyset \\ M \neq 0, K^{-1}(M) \end{cases}$$

Если $M = 0$, то соответствующее уведомление направляется в редакцию журнала о том, что статья,

скорее всего тематике журнала не соответствует или отсутствуют необходимые рецензенты.

Фрагменты кода и результаты работы информационной системы автоматизированного подбора рецензентов для научного журнала представлены на рис. 1.

```
// First we include phophory library
require_once(__DIR__ . "/vendor/autoload.php");

// set some options
$options = array(
    // storage type, follow types supported
    // phophory::STORAGE_FILE - use file operations (read, fseek) for dictionary access, this is very slow...
    // phophory::STORAGE_SHM - load dictionary in shared memory (using shmop.php extension), this is preferred mode
    // phophory::STORAGE_DB - load dict to memory each time when phophory initializes,
    // this is useful when shmop ext is not activated.
    // Speed same as for phophory::STORAGE_SHM type
    'storage' => phophory::STORAGE_FILE,
    // Extend graphql for getAllForecastInfo method call
    'with_grantab' => false,
    // Enable prediction by suffix
    'predict_by_suffix' => true,
    // Enable prediction by prefix
    'predict_by_prefix' => true
);

// Create phophory instance
try {
    $phophory = new phophory(null, 'ru_RU', $options);
    $phophory = new phophory($phophory_filesBundleId($phophory::getDefaultBundleId(), 'ru_RU'), $options);
} catch (PhophoryException $e) {
    die("Error occurred while creating phophory instance: " . $e->getMessage());
}
```

```
// Подсчёт количества совпадений для каждого рецензента
foreach ($unique_reviewers as $id => $data) {
    $unique_reviewers[$id]['count'] = count(array_unique($data['terms']));
}

// Сортировка рецензентов по количеству совпадений (по убыванию)
usort($unique_reviewers, function($a, $b) {
    if ($a['count'] == $b['count']) {
        return 0;
    }
    return $b['count'] - $a['count']; // Сравниваем значения count
});

// Вывод таблицы
echo "<table border=1>";
echo "<tr><th>Рецензент</th><th>Количество совпадений</th><th>Термины</th></tr>";
foreach ($unique_reviewers as $reviewer) {
    // Выводим информацию о рецензенте
    $unique_terms = array_unique($reviewer['terms']);
    $count_unique_terms = count($unique_terms);
    echo "<tr>";
    echo "<td>". $reviewer['reviewer']["id"] . "</td>";
    echo "<td>". $count_unique_terms . "</td>";
    echo "<td>";
    foreach ($unique_terms as $term) {
        echo $term . "<br>";
    }
    echo "</td>";
    echo "</tr>";
}
echo "</table>";
mysql_close($conn);
```

а) фрагменты программного кода

Рецензент	Количество совпадений	Термины
Ляпкин Алексей Иванович	2	Наноалмаз Гальванический

б) результаты работы информационной системы

Рис. 1. Фрагменты кода (а) и результаты работы (б) информационной системы автоматизированного подбора рецензентов для научного журнала

ЗАКЛЮЧЕНИЕ

Разработана информационно-логическая модель для создания системы подбора рецензентов статей научного журнала (предварительно выполнен анализ аналогичных систем [1–4]).

Разработано формальное математическое описание и алгоритм функционирования информационной системы.

Выполнена реализация информационной системы и осуществлено её функциональное тестирование.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Елизаров, А.М. Метод автоматизированного подбора рецензентов научных статей, реализованный в информационной системе научного журнала [Текст] / А.М. Елизаров, Е.К. Липачев, Ш.М. Хайдаров // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23–28 сентября 2019 г., г. Новороссийск). – М.: ИПМ им. М.В.Келдыша, 2019 – С. 318-328.

2. Peer2ref: A new online tool for locating peer reviewers // 0-blogs-biomedcentral-com [Электронный ресурс]. – Электрон. текст. дан. – Режим доступа: <https://0-blogs-biomedcentral-com.brum.beds.ac.uk/bmcblog/2012/09/11/peer2ref-a-new-online-tool-for-locating-peer-reviewers/> – Загл. с экрана.

3. Елизаров, А.М. Рекомендательная система поиска экспертов для проведения научного рецензирования в математическом журнале [Текст] / А.М. Елизаров, Е.К. Липачёв, Ш.М. Хайдаров // Электронные библиотеки. – 2020. – №4. – С. 708–732.

4. Ахметов, Д.Ю. Автоматизация редакционных процессов в информационной системе управления электронными научными журналами [Текст] / Д.Ю. Ахметов, А.М. Елизаров, Е.К. Липачёв // Электронные библиотеки. – 2015. – №1-2. – С. 32–45.

Антонова Анжелика Владиславовна – магистрант, Бийский технологический институт (филиал) ФГБОУ ВО АлтГТУ

Гольх Роман Николаевич – профессор кафедры МСИА, Бийский технологический институт (филиал) ФГБОУ ВПО АлтГТУ, тел. (923)162-93-27, e-mail: gm@bti.secna.ru.

INFORMATION-LOGICAL MODEL AND ALGORITHM FOR AUTOMATED SELECTION OF REVIEWERS OF ARTICLES FOR A SCIENTIFIC JOURNAL

A.V. Antonova, R.N. Golykh
Biysk Technological Institute, Biysk

Peer review of scientific publications plays a key role in the scientific field and is an integral stage in the publication process. Currently, there is a continuous increase in the number of articles in many scientific fields. This growth creates significant stress on the process of peer review of scientific literature, since the number of available publications exceeds the capabilities of reviewers. The appointment of qualified reviewers to evaluate articles is a complex and responsible task. Organizers of the review process must take into account the specialization of reviewers to ensure that their expertise matches the topic of the article. One approach to ensure that the specialization of reviewers matches the topic of the article is to use classifiers. The disadvantage of classifiers is that each author must specify classifier codes for the article, while, as a rule, many scientific journals are monodisciplinary, or, conversely, one reviewer is not an expert in all areas. Therefore, in order to select reviewers, it is necessary to analyze the content of the publication material without resorting to classifiers. Analysis of the content of materials by text requires the development of artificial intelligence systems that analyze the content of the text, and a single phrase from the text does not say anything about the topic of the article, and keywords can relate to many articles and many topics. Therefore, the easiest way to identify the topic of an article is by title. Therefore, there is a need to develop an information-logical model and, ultimately, an information system that will select the necessary reviewers of the article. In this case, it is necessary to take into account the morphology of words in the title of the article.

Index terms: model, information system, review, article.

REFERENCES

1. Elizarov, A.M. Method of automated selection of reviewers for scientific articles, implemented in the information system of a scientific journal [Text] / A.M. Elizarov, E.K. Lipachev, Sh.M. Khaidarov // Scientific service on the Internet: proceedings of the XXI All-Russian scientific conference (September 23-28, 2019, Novorossiysk). - M.: IPM im. M.V. Keldysh, 2019. – P. 318–328.
2. Peer2ref: A new online tool for locating peer reviewers // 0-blogs-biomedcentral-com [Electronic resource]. - Electronic. text. given. – Access mode: <https://0-blogs-biomedcentral-com.brum.beds.ac.uk/bmcblog/2012/09/11/peer2ref-a-new-online-tool-for-locating-peer-reviewers/> – Title from the screen.
3. Elizarov, A.M. A recommendation system for finding experts for conducting scientific peer review in a mathematical journal [Text] / A.M. Elizarov, E.K. Lipachev, Sh.M. Khaidarov // Electronic libraries. – 2020. – No. 4. – P. 708–732.
4. Akhmetov, D.Yu. Automation of editorial processes in the information system for managing electronic scientific journals [Text] / D.Yu. Akhmetov, A.M. Elizarov, E.K. Lipachev // Electronic libraries. – 2015. – No. 1-2. – P. 32–45.

Antonova Anzhelika Vladislavovna – Master's student, Biysk Technological Institute

Golykh Roman Nikolaevich – professor, Biysk Technological Institute, tel. (923)162-93-27, e-mail: gm@bti.secna.ru