

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математического и компьютерного моделирования

Разработка и проектирование информационной системы для

учета публикационной активности научных сотрудников

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 5 курса 561 группы

направления 09.03.03 Прикладная информатика

механико-математического факультета

Арыкова Уланбека Кубанычбековича

Научный руководитель
доцент, к.ф.-м.н., доцент

Е.Ю. Крылова

Зав. кафедрой
зав. кафедрой, д.ф.-м.н., доцент

Ю.А. Блинков

Саратов 2023

Введение. Публикационная активность сотрудников является одним из основных факторов конкурентоспособности университета, устойчивого роста соответствующих показателей в отечественных и мировых рейтингах вузов. Индикатором исследовательской деятельности педагога являются научные публикации как средство обмена результатами научного поиска и распространения потоков научной информации, что отражается и на показателях продуктивности и результативности научной деятельности вуза. Основными методами исследования являются анализ количественной характеристики публикационной активности педагогов вузов страны, анализ библиометрических данных и их классификация, статистический и сравнительный анализ.

Цель данной бакалаврской работы заключается в проектировании и разработке базы данных для информационной системы учета публиктивной активности научных сотрудников. Для выполнения поставленной задачи:

- будет проведен анализ предметной области;
- построена модель информационной системы средствами UML;
- будет проведено моделирование информационной системы;
- на языке SQL будет реализована база данных информационной системы, отвечающая потребностям предметной области;
- будут построены витрины данных и реализован ряд наиболее часто необходимых запросов.

Основное содержание работы. Публикационная активность – важнейший фактор конкурентоспособности университета, устойчивого роста соответствующих показателей в отечественных и мировых рейтингах вузов. Публикационная активность – результат научно-исследовательской деятельности автора, коллектива, воплощенный в виде научной публикации – оценивается по показателям:

- общее число публикаций;
- индекс цитируемости публикаций;
- импакт-фактор журнала, где опубликована работа.

Общее число публикаций – наиболее обобщенный показатель, получаемый из библиографических баз систем цитирования.

Индекс цитируемости публикаций – основной показатель, который обозначает полное количество ссылок на работы, где данный исследователь фигури-

рует в качестве автора или соавтора, в других источниках, распределенное по годам. Индекс цитирования – широко принятый в научном мире показатель «значимости» трудов конкретного ученого.

Импакт-фактор журнала – численный показатель значимости и популярности периодического издания. Этот механизм был применен впервые в 60-ых годах 20 века в США. Специалисты Института информации предложили оценивать качество периодики по количественной характеристике, а именно, по числу цитирований или ссылок на статьи в данном журнале.

Для анализа публикационной активности мирового научного потока существуют международные и отечественные базы (системы) цитирования. Наиболее известными международными базами данных цитирования являются Web of Science и Scopus.

Web of Science – политематическая реферативно-библиографическая и наукометрическая база данных, которой индексируются около 12 500 журналов, из которых около 170 – российские.

Эти ресурсы включают ссылки на полные тексты в первоисточниках и все библиографические ссылки, встречающиеся в каждой публикации, что позволяет в краткие сроки получить самую полную библиографию по интересующей теме (глубина архива – с 1980 года).

Для наглядности представлены основные положения методики расчета в соответствии с рисунком 1.

С 2015 г. на платформе Web of Science размещена база данных лучших научных журналов России – Russian Science Citation Index (RSCI).

RSCI – отдельная база данных, не входящая в основное ядро базы Web of Science Core Collection, однако она полностью интегрирована с поисковой платформой WoS. Список журналов, включенных RSCI, можно посмотреть в каталоге журналов НЭБ e-Library.

Научная электронная библиотека – ведущая российская организация в области предоставления научной информации, аналитических данных и сервисов. Является крупнейшим российским агрегатором научной информации, при помощи которого университеты, научно - исследовательские организации и инновационные компании получают доступ к онлайн - библиотеке научных статей и книг, индексам цитирования и сервисам поиска информации.



Рисунок 1 — Основные положения

Российский индекс научного цитирования (РИНЦ) – национальная библиографическая база данных научного цитирования созданная на базе НЭБ e-LIBRARY. РИНЦ предназначен для оперативного обеспечения научных исследований актуальной справочно-библиографической информацией, является мощным аналитическим инструментом, позволяет осуществлять оценку результативности и эффективности деятельности научно-исследовательских организаций, ученых, уровень научных журналов и т.д.

Scopus позволяет проводить научный поиск по интересующей тематике, предоставляет информацию по цитируемости статей, включает авторские профили с подробной информацией об авторе и оценкой его научной деятельности, профили организаций с подробной информацией об организации и оценкой ее научной деятельности. Количество публикаций автора и цитирований статей в Scopus можно узнать бесплатно в свободном доступе в «Scopus Preview» по адресу: <http://www.scopus.com/search/form/authorFreeLookup.uri>.

Анализ предметной области – одна из первых задач, с решением которых приходится столкнуться разработчику программной системы. На этапе концептуального проектирования осуществляется сбор, анализ и редактирование требований к данным.

Для осуществления удобного и рационального учета публикационной активности необходимо хранить в первую очередь основную информацию об авторах, такую как имя, фамилия, e-mail, ученую степень и звание и т.д.

Одним из важных условий является то, что каждый автор публикации должен быть привязан к определенному структурному подразделению. В связи с этим, необходимо хранить всю информацию о структурном подразделении и занимаемой должности автора в нем.

Одному автору могут принадлежать несколько публикаций соответственно необходимо производить учет всех публикаций и хранить информацию об их доступности, языке написания, авторе и т.д. Так же публикуемая работа должна принадлежать определенной предметной области.

В свою очередь все публикации условно делятся на книги, статьи в журналах и охранные документы. Если о книгах нужно хранить информацию об ее типе (это может быть монография, учебное пособие и т.п.), дате написания и международный стандартный книжный номер, то, в свою очередь, о статьях в журналах и о самих журналах необходимо хранить такую информацию, как уникальный международный номер, идентифицирующий периодическое печатное или цифровое издание и дату релиза.

Отдельно стоит подчеркнуть связь журналов с базами цитирования, так как необходимо хранить всю информацию о включении рассматриваемой работы в определенную базу цитирования.

Соблюдение всех вышеуказанных требований дает возможность быстро найти необходимую информацию.

На основе анализа, проведенного на этапе создания модели, построена схема базы данных, разрабатываемой информационной системы. Она удовлетворяет всем, выявленным на предыдущем этапе, информационным потребностям пользователей, а так же хранит список с данными, необходимыми для их идентификации.

ER-диаграммы (или ER модели) полагаются на стандартный набор символов для отображения сущностей, их атрибутов и связей, как показано в соответствии с рисунком 2.

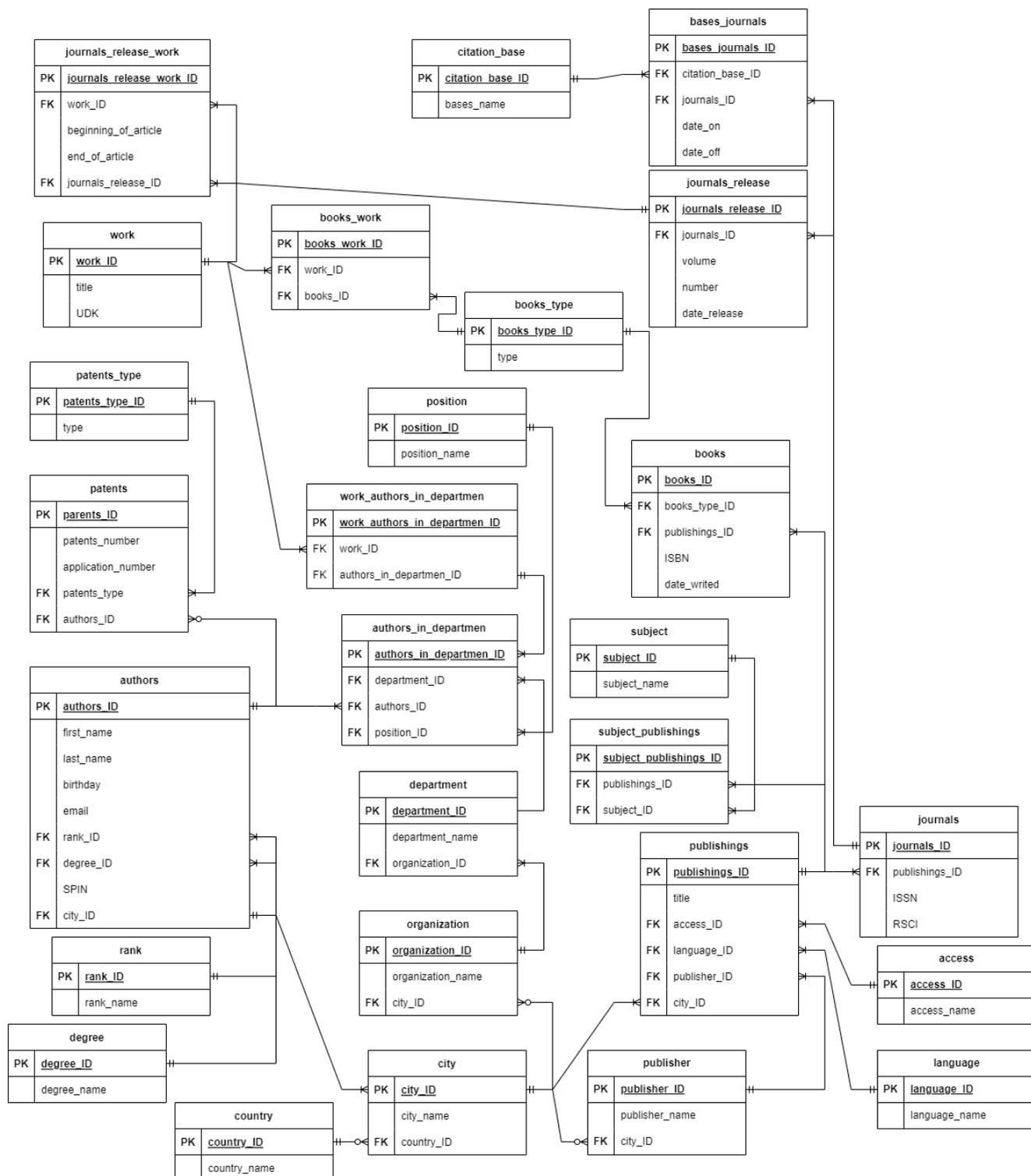


Рисунок 2 — ER-диаграмма информационной системы

В основе моделируемой базы данных лежат несколько основных таблиц:

- authors – содержит, как и все таблицы данной базы, суррогатный первичный ключ и информацию об авторе: ФИО, дата рождения, e-mail, звание, степень, персональный идентификационный код автора в SCIENCE INDEX, город его рождения;

- organization – представляет рассматриваемую организацию; помимо первичного ключа описывается названием организации и соответственно городом, в котором она находится; данная таблица имеет ограничения: информация вводимая в поля email и SPIN должна соответствовать шаблонам, прописанным в коде создания таблиц;

- department – структурное подразделение организации с самостоятельными задачами, функциями и ответственностью за выполнение возложенных задач и функций; содержит два атрибута: наименование и внешний ключ таблицы *organization*, с наложенным на их совокупность ограничением уникальности;

- publishings – представляет публикации, является родительской сущностью и далее разделяется на *книги* и *журналы*, в общем виде и хранит помимо первичного ключа такую обобщенную информацию, как наименование данной публикации, ее доступность, язык, издательство и город, в котором опубликована;

- books – если работа публикуется в книге или является книгой, помимо общей информации, необходимы данные описывающие именно книги – уникальный номер, тип книги и дата ее написания;

- journals – аналогичная ситуация и с журналом; здесь атрибутами таблицы будут: международный стандартный номер, который служит для идентификации газет, журналов и любых категорий периодических изданий на любом носителе, как печатном, так и электронном, с ограничением UNIQUE; библиографическая БД (RSCI), так называемая русская полка журналов на платформе WoS, также с ограничением уникальности.

Обе эти таблицы помимо первичного ключа будут хранить внешний ключ таблицы «publishings», т.к. хранящаяся там информация может относиться как к *книгам*, так и к *журналам*;

- work – хранит информацию о работе, как о произведении.

Атрибут – это информационное отображение свойств объекта. Например, объект «книга» в данной диаграмме характеризуется атрибутами: «ISBN», «дата написания» и др. Но, если атрибут может представлять собой некий перечень возможных значений, как «тип книги» может быть «учебным», «научным», «справочным» и т.д., для каждого такого атрибута создается отдельная таблица, в которой должен храниться перечень всех возможных значений. В данном же случае следующие таблицы будут иметь только два поля: суррогатный первичный ключ и наименование того или иного «перечня атрибутов». В рамках данной работы каталожными являются следующие таблицы:

- degree – учёная степень – степень квалификационной системы в науке, позволяющей ранжировать научных деятелей на отдельных этапах академической карьеры; в настоящее время в Российской Федерации присуждают учёные степени: «кандидата наук» и «доктора наук»;

- rank – учёное звание, в отличие от учёной степени, подтверждается не только определённым уровнем специалиста как исследователя, но и соответствие этого специалиста конкретной научно-преподавательской должности; в настоящее время в Российской Федерации присуждают учёные звания «доцента» и «профессора» (по кафедре, по специальности);

- position – должность – структурная единица в организации или вне её, замещаемая физическим лицом, отвечающим установленным квалификационным требованиям, несущим должностные обязанности и наделённым должностными полномочиями, в соответствии с руководящими документами в той или иной сфере деятельности; перечень должностей профессорско-преподавательского состава ВУЗа может выглядеть следующим образом: «ассистент», «преподаватель», «доцент», «профессор» и т.д.;

- country – таблица «страна» хранит первичный ключ и наименование страны;

- language – то же самое относится и к таблице «язык», необходимой для получения информации о языке публикации;

- publisher – издательство - организация, осуществляющая подготовку и выпуск публикации по заказу автора, посредника или по собственной инициативе;

- citation_base – международная система цитирования (или же – база цитирования) – наукометрическая реферативно-библиографическая база данных; главной функцией систем цитирования является предоставление пользователям информации относительно частоты референций к тем или иным статьям;

- access – доступность – состояние публикаций, при котором субъекты, имеющие права доступа, могут реализовывать их беспрепятственно; может быть «открытым», «ограниченным», «платным» и т.д.;

- subject – данная таблица необходима для того, чтобы показать к какой предметной области относиться та или иная работа;

- books_type – типы книги, как уже упоминалось ранее, может быть «учебным», «научным», «справочным» и т.д.;

- patents_type – патенты же бывают трех видов: «патент на изобретение», «на промышленный образец» и «на полезную модель»;

Следующие таблицы выделяются среди вышеперечисленных «таблиц - атрибутов», но помимо упомянутых двух полей хранят некоторую дополнительную информацию:

- city – таблица *город* хранит внешний ключ таблицы «country», что дает информацию о стране, в которой находится этот город;

- journals_release – хранит всю информацию, касающуюся выпуска журнала;

- patents – хранит такую информацию о патенте, как «номер заявки», кому патент принадлежит, внешний ключ таблицы «вид патента» и т.д.;

Так же данная диаграмма содержит такой вид таблиц, когда между двумя сущностями имеет место связь многие ко многим, которая реализуется при построении архитектуры БД посредством таблицы связи:

- authors_in_department – связывает таблицы *авторы* и *структурное подразделение*, показывая роль автора в структурном подразделении, например занимаемую им должность;

- work_authors_in_department – демонстрирует связь публикуемой *работы* и *структурного подразделения*, от которого работа публикуется;

- journals_release_work – связывает таблицы *выпуск журнала* и *работа*;

- books_work – представляет связь между таблицами *книги* и *работа*;

- bases_journals – демонстрирует взаимосвязь журналов и баз цитирования;

- subject_publishings – определяет к какой предметной области относиться рассматриваемая публикация;

Система управления базами данных (СУБД) - комплекс программ и языковых инструментов для управления базой данных и обеспечения взаимодействий с прикладными программами. Первые СУБД появились ещё в 1970-х, и сегодня их используют в каждой второй компании: от небольших интернет-магазинов до Facebook, Google и Amazon. Схема работы данной системы представлена в соответствии с рисунком 3.

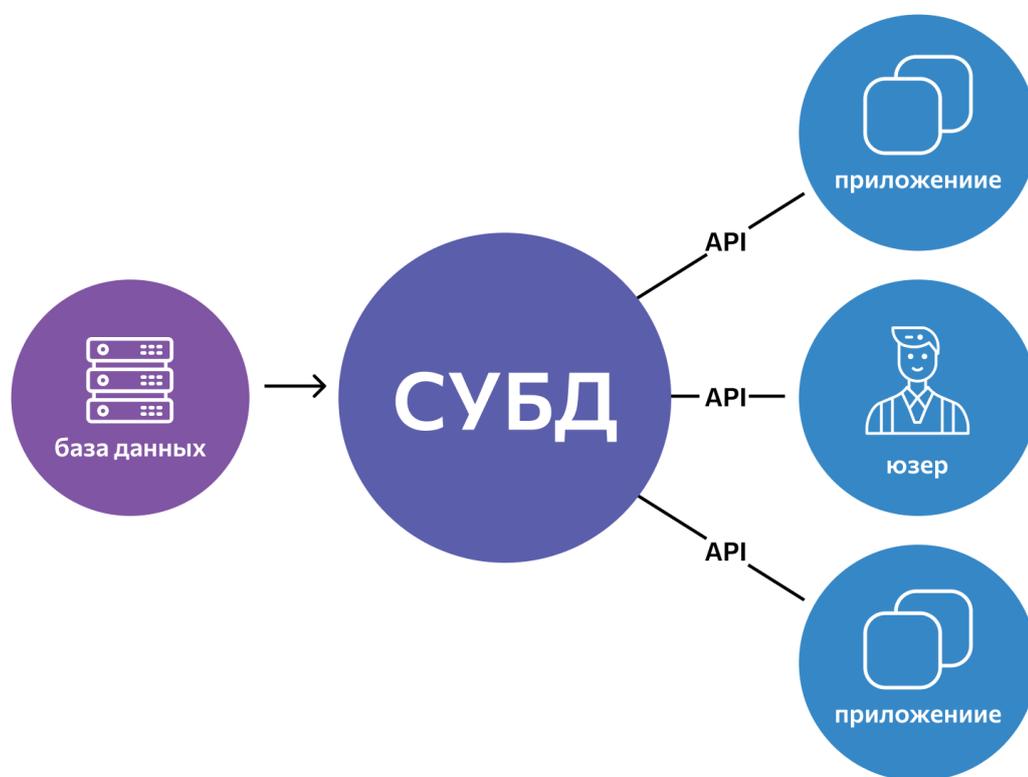


Рисунок 3 — Схема работы СУБД

Проектирование баз данных состоит в построении комплекса взаимосвязанных моделей данных. Основные задачи проектирования баз данных можно сформулировать следующим образом:

- обеспечение хранения в БД всей необходимой информации;
- обеспечение возможности получения данных по всем необходимым запросам;

- сокращение избыточности и дублирования данных;
- обеспечение целостности базы данных.

Чтобы управлять базами данных и находить нужную информацию, запросы к ним пишут на специальных языках. Самый популярный из них – SQL (от англ. Structured Query Language – «язык структурированных запросов»).

SQL – это язык программирования, используемый в большинстве реляционных баз данных для запросов, обработки и определения данных, а также контроля доступа.

Язык SQL позволяет:

- получать доступ к данным в системах управления реляционными базами данных;
- описывать данные и их структур;
- определять данные в базах данных и управлять ими;
- взаимодействовать с другими языками через модули SQL, библиотеки и предварительные компиляторы;
- создавать и удалять базы данных и таблицы;
- создавать представления, хранимые процедуры и функции в БД;
- устанавливать разрешения на доступ к таблицам, процедурам и представлениям.

При создании базы данных для информационной системы учета публичивной активности был использован PostgreSQL. Далее предлагается рассмотреть код создания отдельных сущностей и необходимые для сохранения логики проекта ограничения.

На начальном этапе с помощью команды «CREATE TABLE» создаются таблицы, которые не имеют внешних ссылок. В качестве примера в соответствии с листингом 1 приведен код создания таблицы «contry».

```
1 CREATE TABLE country
2 (
3   country_ID integer PRIMARY KEY,
4   country_name varchar(25) NOT NULL,
5   UNIQUE (country_name)
6 );
```

Листинг 1 – Код создания таблицы «contry»

Для заполнения таблиц используется команда «INSERT». Значения должны быть определены в предложении «VALUES» для каждого столбца. «VALUES» – определяет строку значений, которые будут вставлены в таблицу. Пример заполнения данных в таблицу «authors» представлен в соответствии с листингом 2.

```
1 insert into authors
2 values (1, 'Aleksey', 'Ivanov', '13.10.1987', 'a.
   ivanov@mail.ru', 1, 1, '5132-8521', 1)
3
4 insert into authors
5 values (2, 'Mihail', 'Ignatov', '19.04.1983', 'ignatov.
   mihail@yandex.ru', 1, 2, '3521-2356', 2)
6
7 insert into authors
8 values (3, 'Larisa', 'Dmitrieva', '06.01.1976', '
   dmitrieva76@yandex.ru', 2, 2, '3245-8782', 3)
```

Листинг 2 – Заполнение таблицы «authors» данными

Далее, если все введенные данные удовлетворяют ограничениям всей таблицы или отдельно взятого столбца, с помощью команды SELECT – оператора запроса в языке SQL, возвращающего выборку из базы данных, как показано соответствии с листингом 3.

```
1 SELECT first_name, last_name, email, spin FROM authors;
```

Листинг 3 – Витрина данных таблицы «authors»

Представлен пример заполненных данными столбцов *имя*, *фамилия*, *почта* и *spin-код* таблицы «authors», в соответствии с рисунком 4.

Заключение. Публикационная активность сотрудников в нашей стране является одним из важнейших показателей развития науки в соответствии с целевыми показателями, установленными Министерством образования и науки Российской Федерации. В ряде отечественных и зарубежных исследований отмечалось, что применение методов материального стимулирования

	first_name character varying (30) 🔒	last_name character varying (30) 🔒	email character varying (30) 🔒	spin character varying (9) 🔒
1	Aleksey	Ivanov	a.ivanov@mail.ru	5132-8521
2	Mihail	Ignatov	ignatov.mihail@yandex.ru	3521-2356
3	Larisa	Dmitrieva	dmitrieva76@yandex.ru	3245-8782

Рисунок 4 — Пример заполнения таблицы «авторы» данными

научных сотрудников и преподавателей может дать достаточно значимый эффект влияния на публикационную активность.

В данной бакалаврской работе в процессе проектировании информационной системы для учета публикационной активности гостиницы были разработаны UML-диаграмма прецедентов, отражающая варианты использования информационной системы разными пользователями. В процессе работы в PostgreSQL разработана БД, в которой можно хранить и манипулировать с информацией об авторах и их публикационной активности.

Результатом данной бакалаврской работы является спроектированная и разработанная информационная система для учета публикационной активности сотрудников. Таким образом, все поставленные задачи данной работы выполнены и цель достигнута.