



Общероссийский математический портал

В. А. Минин, И. М. Зацман, В. А. Хавансков, С. К. Шубников, Архитектурные решения для систем вычисления индикаторов тематических взаимосвязей науки и технологий, *Системы и средства информ.*, 2013, том 23, выпуск 2, 260–283

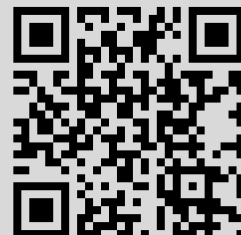
DOI: 10.14357/08696527130217

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.223.170.103

19 ноября 2024 г., 21:24:11



АРХИТЕКТУРНЫЕ РЕШЕНИЯ ДЛЯ СИСТЕМ ВЫЧИСЛЕНИЯ ИНДИКАТОРОВ ТЕМАТИЧЕСКИХ ВЗАИМОСВЯЗЕЙ НАУКИ И ТЕХНОЛОГИЙ*

В. А. Минин¹, И. М. Зацман², В. А. Хавансков³, С. К. Шубников⁴

Аннотация: Дано описание архитектурных решений для вычисления индикаторов тематических взаимосвязей науки и технологий. Предлагаемые решения являются основой создания информационных систем, которые не имеют аналогов в российской научно-технической сфере. Их создание необходимо для мониторинга и оценивания программ научных исследований и принятия решений на всех этапах программной научно-технической деятельности. Описана архитектура системы в целом и концептуальные решения по отдельным функциональным подсистемам. Эти решения основаны на методологии определения индикаторов тематических взаимосвязей, которая разработана в ИПИ РАН. Предлагаемые архитектурные решения обладают рядом принципиальных отличий от имеющихся зарубежных аналогов. Во-первых, появляется возможность выбора системы классификации областей знаний. Во-вторых, обеспечивается формирование и применение терминологических портретов рубрик для выбранной классификационной системы.

Ключевые слова: мониторинг и оценивание программ; взаимосвязи науки и технологий; методология определения индикаторов взаимосвязей; информационные системы; архитектурные решения

DOI: 10.14357/08696527130217

1 Введение

Исследования взаимосвязей науки и технологий с использованием количественных индикаторов ведутся не один десяток лет. Их первоначальной целью была разработка и применение методик выявления случаев переноса и использования в сфере технологического развития и инноваций тех научных знаний, которые оказывают существенное влияние на разработку новых технических решений. Например, в конце прошлого века Е. Мансфилд проводил исследования в этой области, взяв в качестве исходных данных сведения о работе американских фирм [1–3]. Его работы содержат эмпирические данные о взаимосвязях

* Работа выполнена при частичной поддержке РГНФ (грант № 12-02-12019в).

¹ Российский фонд фундаментальных исследований, minin@rfbr.ru

² Институт проблем информатики Российской академии наук, iz_ipi@a170.ipi.ac.ru

³ Институт проблем информатики Российской академии наук, havanskov@a170.ipi.ac.ru

⁴ Институт проблем информатики Российской академии наук, sergeysh50@yandex.ru

результатов научных исследований и технологических инноваций. На основе анализа этих данных он пришел к выводу, что 10% технологических новшеств не были бы изобретены или же были бы разработаны с большой задержкой, если бы были выполнены без использования результатов соответствующих научных исследований.

Причина пристального внимания к индикаторному оцениванию процессов переноса знаний заключается в том, что, с одной стороны, финансирование научных исследований, ориентированных на развитие научно-технического задела в области перспективных технологий, связано с большим риском, с другой стороны, есть опасность упустить новые прорывные решения и потерять конкурентоспособность. Это касается конкурентоспособности как отдельных предприятий в конкретной технологической сфере, так и государства в целом.

Таким образом, ключевым фактором необходимости создания аналитико-информационных систем (АИС) индикаторного оценивания взаимосвязей науки и технологий выступает потребность в обоснованных приоритетах финансирования технологически ориентированных исследований.

Для определения этих приоритетов требуются объективные критерии, основанные на индикаторах частотности случаев переноса знаний из разных областей научных исследований в сферу технологического развития и инноваций. Один из наиболее сложных аспектов в изучении взаимосвязей науки и технологий состоит в количественном индикаторном оценивании процессов переноса знаний [4–8].

Проведенный в ИПИ РАН анализ зарубежного опыта показал, что вычисление количественных индикаторов взаимосвязей науки и технологий требует автоматизированной обработки больших объемов слабоструктурированных полнотекстовых описаний изобретений и сопоставления результатов обработки патентной информации как с названиями источников научных публикаций (журналов или материалов конференций), так и с названиями самих статей, хранящихся в научных электронных библиотеках. Это сопоставление дает возможность определить тематику тех научных направлений, к которым относятся научные публикации, цитируемые в описаниях изобретений [9–12].

Разработанный в ИПИ РАН вариант методологии определения значений индикаторов взаимосвязей науки и технологий позволяет вычислять их значения как для целых областей знаний, так и для отдельных направлений исследований [12]. На его основе в настоящее время в ИПИ РАН создается АИС, предназначенная для индикаторного оценивания взаимосвязей науки и развития информационных технологий (ИТ), описанию архитектуры которой посвящена данная статья.

2 Задачи аналитико-информационной системы

Целью разработки АИС является обеспечение мониторинга и индикаторного оценивания инновационно-технологического потенциала направлений ориенти-

рованных фундаментальных научных исследований. Понятие «инновационно-технологический потенциал направлений ориентированных фундаментальных научных исследований» интерпретируется в контексте следующих основных задач, решение которых должна обеспечивать разрабатываемая АИС.

В рамках первой задачи для любой выбранной технологической области (группы технологий) с использованием АИС планируется вычислять индикаторы как коэффициенты корреляции группы технологий с теми направлениями научных исследований, которые оказывают влияние на развитие технологий именно этой группы, т. е. рассматривается одна группа технологий и N направлений научных исследований (далее кратко — задача 1Т_НН).

В рамках второй задачи для любого выбранного направления научных исследований планируется вычислять коэффициенты корреляции этого направления с теми группами технологий, на которые оказывает влияние это направление, т. е. M групп технологий и одно направление исследований (далее кратко — задача МТ_1Н).

В случае решения второй задачи можно получить решение и более общей задачи: для любых N направлений научных исследований вычислить коэффициенты корреляции этих направлений с теми группами технологий, на которые они оказывают влияние, т. е. M групп технологий и N направлений научных исследований (далее кратко — задача МТ_НН). Таким образом, под оценением инновационно-технологического потенциала направлений фундаментальных научных исследований понимается вычисление указанных коэффициентов корреляции в процессе решения трех перечисленных задач.

К сфере использования результатов АИС относится программно-целевое планирование, финансирование и организация научных исследований, ориентированных на развитие приоритетных и критических технологий. Круг потенциальных пользователей АИС включает руководителей и специалистов органов исполнительной власти федерального и регионального уровня, государственных академий наук и государственных научных фондов, отвечающих за планирование научной деятельности, распределение финансовых ресурсов, мониторинг, оценивание результатов, эффективности и результативности их расходования, а также экспертов и ученых, участвующих в оценивании результатов, эффективности и результативности программно-целевой деятельности в научно-технической сфере.

К настоящему времени существует несколько вариантов методологии оценивания взаимосвязей науки и технологий с использованием количественных индикаторов как коэффициентов корреляции.

Перечислим основные варианты методологии оценивания взаимосвязей науки и технологий, отмечая примеры реализованных (реализуемых) на практике вариантов, для которых решены теоретические и методологические вопросы.

1. Методология оценивания в статике взаимосвязей отдельных групп технологий при фиксированной системе классификации областей знаний и составляющих их направлений научных исследований на заданном интервале времени. Этот вариант является одним из наиболее проработанных с теоретической и методологической точек зрения. Он был реализован, например, в Российском фонде фундаментальных исследований (РФФИ). С его помощью вычислены индикаторы взаимосвязей областей знаний и критических технологий для того массива проектов РФФИ, руководители которых выявили и описали эти взаимосвязи в отчетных формах по своим проектам.
2. Методология оценивания в статике взаимосвязей по широкому спектру групп технологий при фиксированной системе классификации областей знаний и составляющих их направлений научных исследований на заданном интервале времени (с использованием *основного индекса* Международной патентной классификации — МПК). Этот подход хорошо проработан теоретически и опробован на практике. Он был реализован в рамках информационной системы, разработанной по заказу Еврокомиссии с целью анализа многолетнего массива европейских и американских запатентованных изобретений и вычисления коэффициентов корреляции. Этот вариант используется в настоящее время в директоратах Еврокомиссии.
3. Методология оценивания в статике взаимосвязей по всем группам технологий при фиксированной системе классификации областей знаний и составляющих их направлений научных исследований на заданном интервале времени (с использованием *основного и дополнительных индексов* МПК). Этот вариант хорошо проработан теоретически. Продемонстрирована его реализуемость в лабораторных условиях. Он был разработан в ИПИ РАН в рамках проекта по гранту РГНФ № 06-02-04043а «Методы мониторинга и оценки инновационного потенциала и результативности направлений научных исследований». В текущем проекте по гранту РГНФ № 12-02-12019 планируется реализовать разработанную в ИПИ РАН методологию в рамках проектирования макета АИС с экспериментальным массивом полнотекстовых описаний изобретений отечественных и зарубежных заявителей, относящихся к рубрикам МПК для ИТ.

Наибольший интерес при оценивании взаимосвязей науки и технологий представляют не отдельные технологии, а группы технологий по приоритетным направлениям научно-технологического развития, среди которых выделяется группа ИТ [13]. Тематика каждого научного направления, результаты которого используются при разработке технологий, задается в виде одной или нескольких рубрик выбранной системы классификации областей знаний. Тематика каждой анализируемой группы технологий задается в виде списка рубрик МПК. Наиболее часто используются списки рубрик МПК из номенклатуры, разработанной Фраунгоферским институтом системотехники и инновационных исследований

(Fraunhofer Gesellschaft-Institute für Systemtechnik und Innovationsforschung — FhG-ISI) [8].

В номенклатуре FhG-ISI группа ИТ, для которой в настоящее время разрабатывается макет АИС, предназначенный для индикаторного оценивания тематических взаимосвязей, описывается следующими тремя рубриками МПК:

- (1) класс¹ G06 — «Вычисление; счет» (эта рубрика МПК включает оптические вычислительные устройства, обработку цифровых данных с помощью компьютеров, аналоговые и гибридные компьютеры);
- (2) подкласс G11C — «Запоминающие устройства статического типа»;
- (3) подкласс G10L — «Анализирование или синтезирование речи; распознавание речи».

Таким образом, макет АИС будет содержать патентные информационные ресурсы² по трем рубрикам МПК: G06, G11C и G10L. Используя эти ресурсы, система должна будет по запросам экспертов определять индикаторы тематических взаимосвязей для изобретений по ИТ, имеющих коды рубрик G06, G11C и G10L. Для вычисления значений индикаторов в макете будут использоваться следующие информационные ресурсы:

- классификатор научных направлений, исследования по которым взаимосвязаны с развитием ИТ;
- массив библиографических описаний изобретений по ИТ, запатентованным и опубликованным в России в течение заданного периода времени, с проставленными рубриками МПК;
- файл библиографических описаний цитируемых научных публикаций, извлеченных из полнотекстовых описаний изобретений по ИТ, запатентованным и опубликованным в России в течение заданного периода времени;
- база данных (БД) структурированных библиографических описаний цитируемых научных публикаций, для каждой из которых будут проставлены рубрики научных направлений.

Следует отметить, что наличие временных полей в структурах данных библиографических описаний изобретений (дата подачи заявки, дата выдачи патента) и библиографических данных статей (год ее публикации) позволяет кроме значений индикаторов взаимосвязей вычислять и временные характеристики, которые показывают время отклика технологического развития на результаты научных исследований в соответствующих областях знаний.

¹Все рубрики МПК образуют иерархическую структуру и делятся на 5 категорий: разделы (верхний уровень иерархии), классы, подклассы, группы и подгруппы (нижний уровень иерархии).

²Полные описания изобретений к патентам доступны в открытом доступе на веб-сайте Федерального института промышленной собственности (ФИПС) по адресу <http://www1.fips.ru/wps/portal/Registers>.

3 Информационные ресурсы аналитико-информационных систем и технология их обработки

Методы и подходы, предлагаемые для решения трех задач АИС (обозначенных ранее как задачи *IT_NH*, *MT_1H* и *MT_NH*), практически полностью совпадают. Основное отличие состоит в принципах формирования информационных ресурсов, которые необходимы для решения каждой из этих задач. Для решения первой задачи необходимо ввести в АИС массив полнотекстовых описаний запатентованных в России изобретений по одной выбранной группе технологий в течение заданного периода времени (т. е. того периода, для которого необходимо получить значения индикаторов), например по группе ИТ. Для решения второй и третьей задач необходимо загрузить массив описаний изобретений по *M* группам технологий, запатентованных в России в течение заданного периода времени.

После загрузки в АИС необходимых описаний изобретений сначала должен быть сформирован массив библиографических описаний научных публикаций, цитируемых в описаниях изобретений. Именно здесь возникает необходимость формирования структурированных данных на основе слабоструктурированных описаний изобретений. Затем необходимо провести рубрицирование этих научных публикаций в рамках заданной классификации научных направлений. Далее средствами АИС выполняется автоматическое сопоставление частот появления рубрик МПК, поставленных в загруженных описаниях изобретений, и частот появления рубрик цитируемых в этих изобретениях научных публикаций (рубрик заданной классификации научных направлений).

Значения частот рубрик научных документов одного направления исследований, цитируемых в изобретениях с рубриками МПК заданной группы технологий, представляют собой значения индикаторов взаимосвязей этого направления с рубриками МПК заданной группы технологий. Отметим, что значения индикаторов существенно зависят от выбранной классификации научных направлений.

Предлагаемая процедура оценивания инновационно-технологического потенциала направлений ориентированных фундаментальных научных исследований предполагает использование в качестве входных импортируемые информационные ресурсы. В упрощенном виде входные и выходные информационные ресурсы и их взаимодействие на разных этапах представлены на рис. 1.

Загружаемые в АИС патентные ресурсы публикуются Роспатентом на CD/DVD носителях и доступны на его сайте. Для привязки библиографических ссылок к направлениям научных исследований в макете АИС используется Классификатор РФФИ.

Выходным информационным ресурсом служит БД структурированных библиографических ссылок на публикации в описаниях изобретений, каждая из

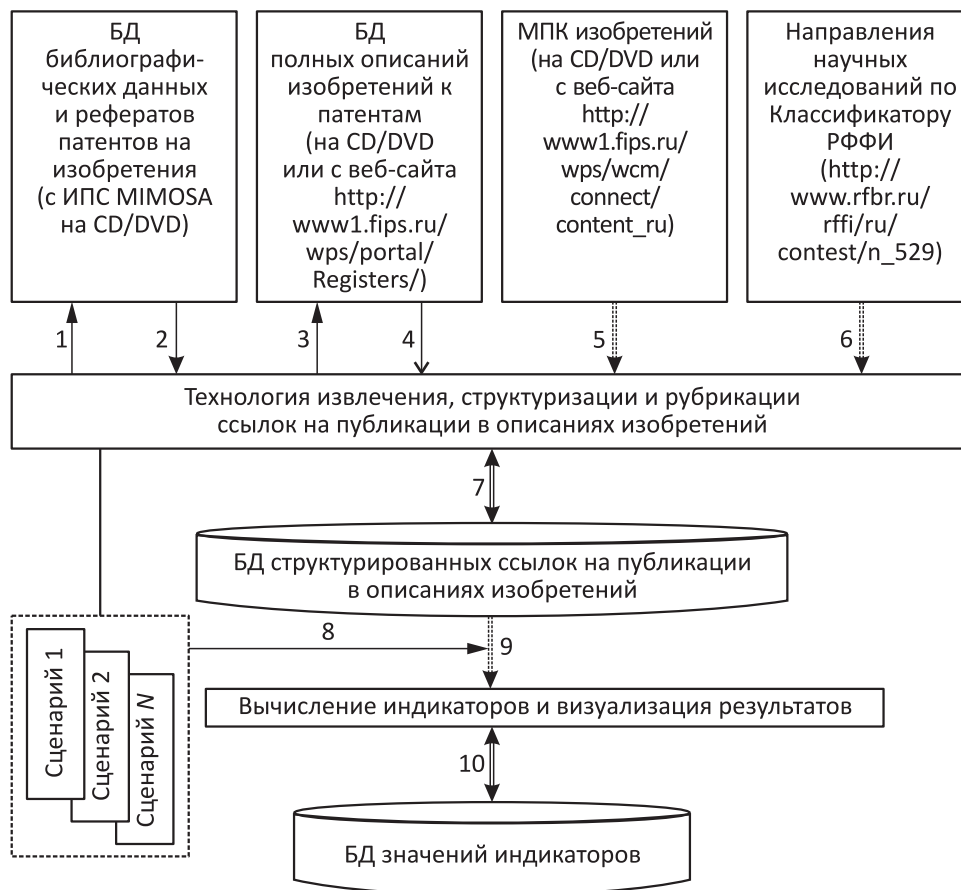


Рис. 1 Входные и выходные информационные ресурсы макета АИС

которых привязана, с одной стороны, к индексу МПК, с другой стороны, к индексу направления научных исследований.

На рис. 1 показаны следующие информационные массивы:

- БД библиографических данных и рефератов патентов на изобретения (поставляется вместе с информационно-поисковой системой (ИПС) MIMOSA);
- БД полных описаний изобретений к патентам;
- Международная патентная классификация (МПК) изобретений;
- Классификатор РФФИ;
- БД структурированных ссылок на публикации в описаниях изобретений;
- БД значений индикаторов.

Уточним содержание информационных потоков и процедур, обозначенных стрелками на рис. 1 (номер элемента нижеприведенного списка соответствует номеру стрелки).

1. Составление процедуры (сценария) вычисления значений индикаторов как коэффициентов корреляции. При выполнении каждой процедуры определяется и параметрически описывается исследуемая группа технологий и тот ретроспективный временной период, для которого необходимо вычислить значения индикаторов. С этой целью используются рубрики МПК и даты (года) публикации патентов.
2. Список библиографических данных патентов, отобранных по запросу и включающих рубрики МПК и даты (года) публикации патентов.
3. Формирование запроса в БД полных описаний изобретений (достаточно знать номер патента на изобретение из полученного списка библиографических данных).
4. Получение по запросу полного описания запатентованного изобретения.
5. Использование кодов и названий рубрик МПК.
6. Использование кодов и названий рубрик классификатора направлений научных исследований, в данном случае — Классификатора РФФИ.
7. Формирование и пополнение БД структурированных ссылок на публикации, выделенных из полных описаний изобретений.
8. Группировка структурированных ссылок на публикации в соответствии с параметрами сценария, в частности по направлениям исследований, критическим и приоритетным технологиям.
9. Структурированные ссылки на публикации с присвоенными им рубриками МПК изобретений, в которых упоминается данная публикация, и рубриками классификатора научных направлений, к которым относится данная публикация.
10. Вычисленные значения индикаторов.

В разработанной технологии вычисления значений индикаторов можно выделить два основных этапа. Первый (подготовительный) этап включает формирование БД структурированных ссылок на публикации, включенных в описания изобретений. Он предполагает обработку больших объемов слабоструктурированных описаний изобретений, трудно поддается полной автоматизации и поэтому требует привлечения операторов. Этот этап может повторяться итерационно в целях пополнения БД. Второй этап включает вычисление значений количественных индикаторов и получение на их основе экспертных оценок инновационно-технологического потенциала направлений ориентированных фундаментальных научных исследований, т. е. на этом этапе предполагается привлечение экспертов.



Рис. 2 Архитектура АИС

Для реализации разработанной технологии вычисления значений индикаторов взаимосвязей в настоящее время создается макет АИС, которая состоит из 7 функциональных подсистем (рис. 2), взаимодействующих между собой через структурированные наборы данных, объединенные в БД АИС:

- (1) подсистема создания и ведения классификаторов АИС, включая импорт внешних классификаторов (подсистема «**Классификаторы**»);
- (2) подсистема импорта и формирования БД описаний изобретений, используемых для выделения библиографических ссылок на публикации (подсистема «**Импорт**»);
- (3) подсистема формирования и ведения информационных массивов выделенных неструктурированных библиографических ссылок на публикации, цитируемые в описаниях изобретений (подсистема «**Цитирование**»);
- (4) подсистема формирования и ведения БД структурированных библиографических ссылок на публикации, в том числе присвоение рубрик научных направлений (подсистема «**Рубрицирование**»);

- (5) подсистема формирования и ведения БД значений индикаторов взаимосвязей науки и технологий (подсистема «**Индикаторы**»);
- (6) подсистема построения сценариев оценивания индикаторов взаимосвязей науки и технологий и экспертной верификации вычисленных значений индикаторов (подсистема «**Экспертиза**»);
- (7) подсистема автоматизированного построения аналитических отчетов с количественными индикаторами и качественными экспертными оценками взаимосвязей науки и технологий (подсистема «**Аналитика**»).

4 Проектные решения для функциональных подсистем АИС

Подсистема «Классификаторы» (рис. 3) предназначена для ведения БД классификаторов и решает следующие задачи:

- импорт внешних классификаторов и ведение БД классификаторов;
- составление терминологических портретов рубрик классификаторов.

В рамках АИС для формального описания направлений научных исследований предполагается использовать Классификатор РФФИ, а для кодирования видов и групп технологий — МПК.

Подсистема должна обеспечивать автоматизированный ввод классификаторов с электронных носителей или с сайтов, доступных через Интернет. Учитывается,

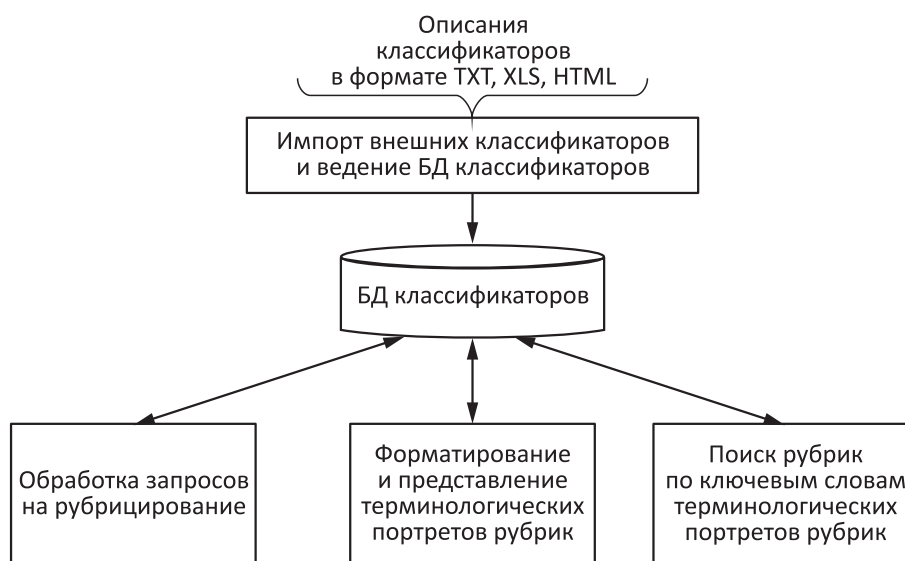


Рис. 3 Архитектура подсистемы «Классификаторы»

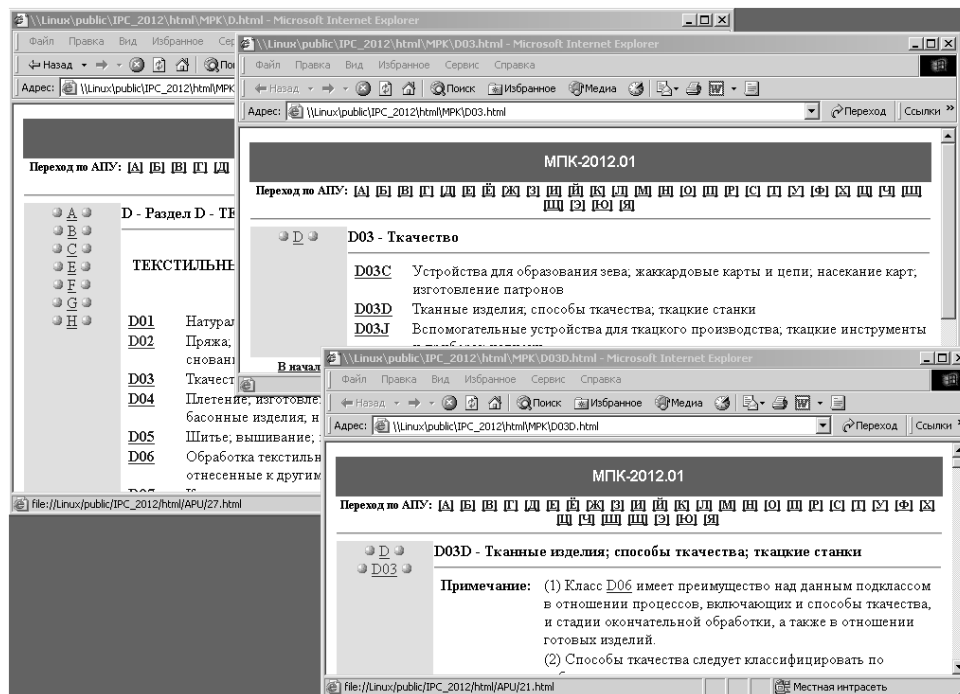


Рис. 4 Фрагмент описания МПК с примерами классов и подклассов из раздела D

что классификаторы могут изменяться во времени. Обеспечивается визуальное отображение классификаторов для пользователей системы, наличие возможности поиска отдельных рубрик, в том числе возможность создания терминологических портретов рубрик.

Импортируемые в АИС электронные описания классификаторов имеют различные структуры и форматы. Например, классификатор РФФИ представляет собой список рубрик в одном PDF-файле. Международная патентная классификация представляет собой коллекцию иерархически связанных файлов в формате HTML (рис. 4). Рисунок 4 включает коды и названия отдельных рубрик следующих двух уровней МПК, которые называются классами и подклассами, входящих в раздел D «Текстиль; бумага». Верхний уровень иерархии включает всего восемь разделов:

- (1) А — удовлетворение жизненных потребностей человека;
- (2) В — различные технологические процессы; транспортирование;
- (3) С — химия; металлургия;
- (4) D — текстиль; бумага;

- (5) Е — строительство; горное дело;
- (6) F — машиностроение; освещение; отопление; двигатели и насосы; оружие и боеприпасы; взрывные работы;
- (7) G — физика;
- (8) H — электричество.

Подсистема «Классификаторы» включает набор структурированных запросов, которые обеспечивают ее взаимодействие с другими подсистемами АИС в целях решения следующих задач:

- выбор рубрики при рубрицировании публикаций и научных журналов;
- формирование ссылок на рубрики при отображении их кодов в текстах, таблицах и диаграммах;
- поиск рубрик по ключевым словам из их терминологических портретов.

Подсистема «Импорт» (рис. 5) служит для формирования БД описаний изобретений, содержащих цитируемые научные публикации. Она предназначена для решения следующих задач:

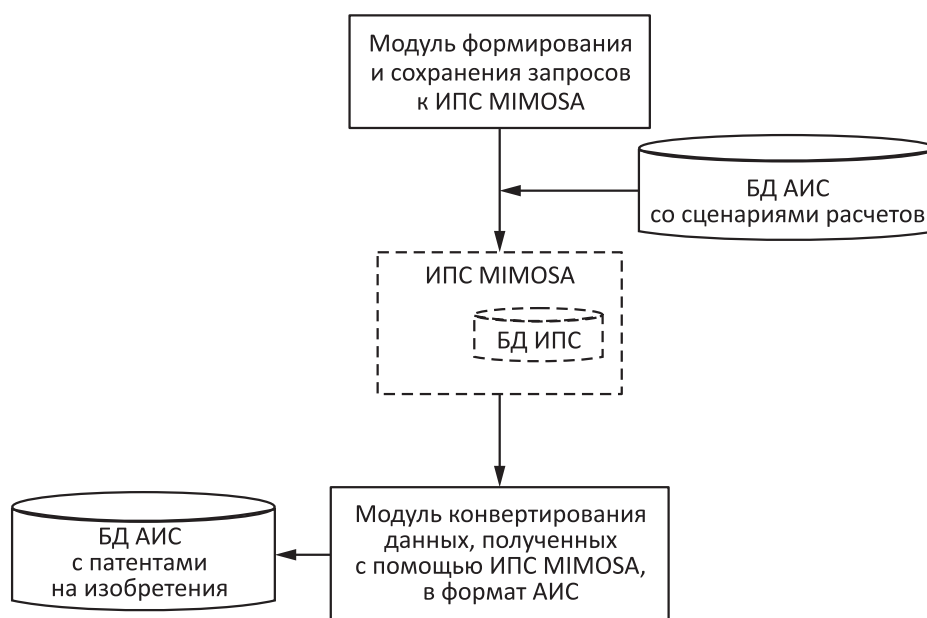


Рис. 5 Архитектура подсистемы «Импорт»

- составление и ведение библиотеки типовых запросов к БД описаний изобретений,
- формирование БД описаний изобретений путем импорта внешних информационных массивов.

Исходя из заданного сценария оценки взаимосвязей направлений научных исследований и технологических областей создается поисковый запрос. В ответ на запрос формируется список патентов на изобретения. В качестве базового программного инструмента на этом этапе используется ИПС MIMOSA, которая позволяет работать с реферативной информацией, содержащейся на оптических носителях, выпускаемых Роспатентом.

Информационно-поисковая система MIMOSA обеспечивает не только поиск информации по запросам, но и позволяет выгружать массивы найденных документов в виде структурированных текстовых файлов или в формате своей БД.

Индексы МПК в ИПС MIMOSA могут содержаться в разных полях, что учитывается при формировании запроса на поиск патентов на изобретения. Эти поля содержат коды рубрик МПК, к которым относится изобретение, но заполняются они разными участниками процесса подачи заявок на выдачу патентов на изобретения (это поле с меткой ICAI, которое заполняется заявителем) и процесса проведения экспертизы заявки по существу (это поля с метками ICA и ICAA, которые заполняются экспертом). Ниже приводится пример построенного запроса для ИПС MIMOSA:

$(ICA = G06^* OR ICAA = G06^* OR ICAI = G06^*) ANDDP = 2012^*$,

где знак «*» служит знакозамениателем¹.

В случае если запрос выполнен (т. е. синтаксис проверен и поиск проведен) и хотя бы один документ найден, то ИПС МИМОЗА предлагает просмотреть список библиографических данных всех найденных документов. В результате выполнения запроса формируется структурированный текстовый файл библиографических описаний патентов, отвечающих построенному запросу. Файл с библиографическими данными содержит рефераты описаний изобретений и URL (uniform resource locator) ссылки для доступа к полному описанию в открытом реестре изобретений Роспатента.

Подсистема «Цитирование» (рис. 6) формирует список структурированных ссылок на научные публикации, цитируемые в описаниях изобретений. Она решает следующие задачи:

- создание и ведение библиотеки шаблонов для поиска ссылок на цитируемые публикации;

¹ Данный запрос имеет следующее содержание: выбрать все документы, которые содержат код МПК G06 в полях «Классификация — расширенный уровень», «Классификация — дополнительная информация», «Классификация — изобретательская информация» и символы «2012» в поле «Дата публикации».

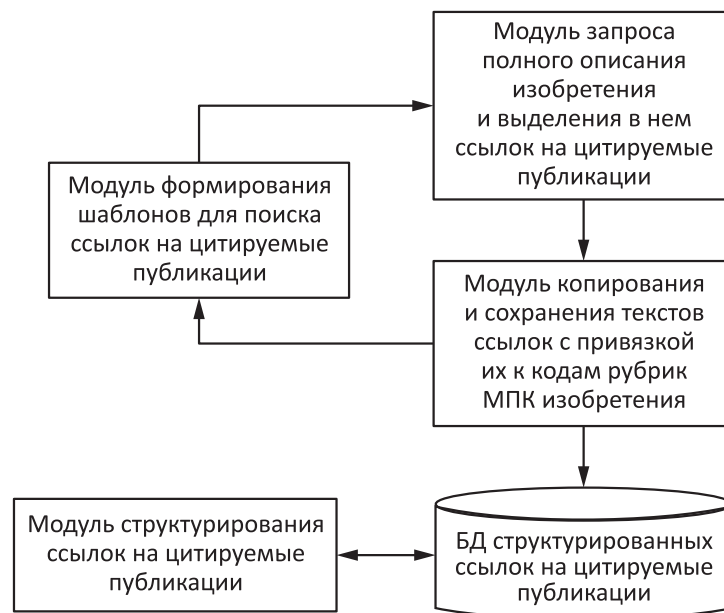


Рис. 6 Архитектура подсистемы «Цитирование»

- автоматизированный поиск ссылок на цитируемые публикации;
- формирование и ведение библиотеки правил для структурирования ссылок на публикации;
- автоматизированное структурирование ссылок;
- формирование БД структурированных библиографических ссылок на цитируемые публикации.

Задача поиска ссылок на публикации решается с помощью автоматического выделения фрагментов неструктурированного текста по заданным признакам. Для этого используются шаблоны разных видов ссылок на публикации, которые строятся на основе признаков, характерных для каждого вида ссылок на публикации. Шаблоны формируются на основе данных анализа текстового массива описаний изобретений. Они могут корректироваться в процессе работы АИС. Все ссылки на публикации в тексте описания изобретения, найденные с помощью шаблонов, выделяются цветом (рис. 7). Подобное выделение дает возможность оператору контролировать точность поиска ссылок на публикации.

В ходе контроля оператором точности и полноты поиска ссылок на публикации накапливается статистика по определению успешно выделенных фрагментов

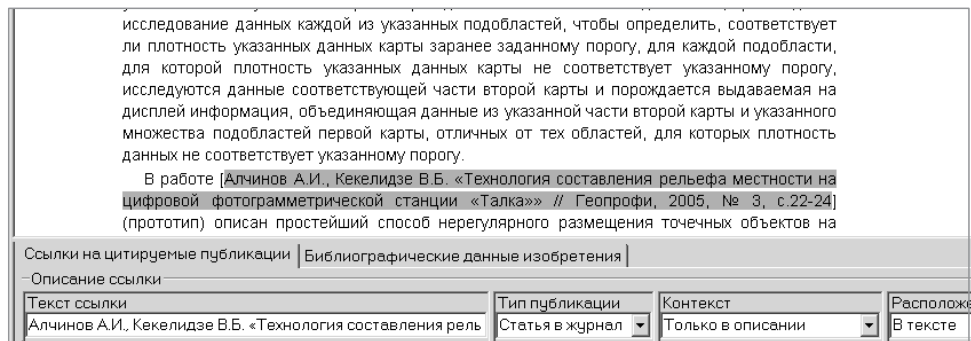


Рис. 7 Выделенная ссылка в неструктурированном тексте описания изобретения

текста, которые действительно оказались ссылками цитирования. Такая статистика позволяет увеличить результативность автоматического выделения ссылок цитирования, что, в свою очередь, служит предпосылкой для повышения результативности процесса автоматизированного поиска библиографических ссылок в описаниях изобретений.

Текст ссылки сохраняется в БД АИС с привязкой к библиографическим данным патента на изобретение (в том числе к рубрикам МПК) и с указанием вида публикации (статья, книга, доклад и пр.). Затем неструктурированный текст ссылки на публикацию (только для статей) структурируется по отдельным полям (авторы, название публикации, название журнала, год публикации, том / выпуск, страницы) с использованием библиотеки правил структурирования.

Для идентификации научных журналов в БД АИС предполагается использовать нормализованный список их названий с перечнем рубрик научных направлений исследований, к которым относятся статьи журналов. Таким образом, привязка ссылки на публикацию к журналу из указанного списка (с использованием процедуры нормализации названия журнала) автоматически показывает набор рубрик научных направлений, наследуемых данной публикацией из атрибутов этого журнала. В случае отсутствия возможности автоматически нормализовать название журнала и привязать ссылку на публикацию к научному журналу, предусмотрена возможность передачи этой ссылки библиографу, осуществляющему нормализацию названия журнала и привязку к нему этой публикации.

Выходом данной подсистемы является массив структурированных ссылок цитирования с привязкой: (а) к рубрикам научных направлений исследования; (б) к библиографическим данным изобретения в целом и к рубрикам МПК в частности. Привязка к библиографическим данным изобретения дает возможность вычислять значения индикаторов взаимосвязей науки и технологий с фильтрацией и /или группировкой по разным атрибутам изобретения (например: по стране патентовладельца, дате подачи заявки, выдачи патента и т. д.).

Подсистема «Рубрицирование» предназначена для классификации структурированных ссылок на цитируемые публикации по научным направлениям заданного классификатора и решает следующие две задачи:

- (1) рубрицирование цитируемых научных публикаций на основе списка нормализованных названий журналов, которые имеют рубрики научных направлений,
- (2) уточнение рубрик цитируемых научных публикаций на основе сопоставления названий публикаций с терминологическими портретами, которые приписаны рубрикам научных направлений.

Операция рубрицирования цитируемых научных публикаций на основе списка нормализованных названий журналов обеспечивает присвоение рубрик только для части цитируемых научных публикаций из-за неполного распознавания видов публикаций. В первой версии АИС предполагается рубрицирование только статей из научных журналов. В процессе рубрицирования доступны: название журнала, название статьи, а также терминологические портреты, которые приписаны рубрикам научных направлений.

В процессе разработки методологии определения значений индикаторов взаимосвязей науки и технологий было выделено четыре типа возможных тематических отношений между изобретениями и публикациями:

- (1) одной рубрике МПК изобретения соответствует одна рубрика научного направления;
- (2) одной рубрике МПК соответствуют несколько рубрик научных направлений;
- (3) нескольким рубрикам МПК соответствует одна рубрика научного направления;
- (4) нескольким рубрикам МПК соответствуют несколько рубрик научных направлений.

Подсистема обеспечивает вычисление индикаторов с учетом всех типов тематических отношений. Сущность изобретения выражается основной рубрикой МПК, а дополнительные носят уточняющий характер. В процессе разработки АИС предполагалось, что проектирование индикаторов с учетом перечисленных типов отношений, различий между основными и дополнительными рубриками является самостоятельной задачей, которая решается вне АИС [14–18]. Методы проектирования индикаторов были рассмотрены в работах [19, 20]. Эти методы были разработаны на основе семиотических моделей [21–26].

В процессе проектирования АИС был проведен анализ двух вариантов рубрицирования публикаций. В первом варианте, в котором используются все рубрики МПК, указанные в библиографических данных изобретения, АИС генерирует матрицу из $L = C_m \times C_s$ взаимосвязей технологий и научных

направлений (C_m — количество рубрик МПК; C_s — число рубрик научных направлений исследований). Особенности формирования этой матрицы наглядно иллюстрирует пример из работы [27], где описан процесс построения матрицы, у которой m изменяется от первой до 441-й рубрики МПК (рассматривались рубрики только одного вида — подклассы МПК), а s — от первого до 187-го направления научных исследований. Значения элементов этой матрицы были вычислены в процессе обработки 656 695 патентов на изобретения, выданных Патентным ведомством США, заявки на которые были поданы в период 1992–1996 гг. Сначала из описаний изобретений были выделены 1 147 160 непатентных ссылок (т. е. ссылки на патенты были исключены). При этом рассматривались только так называемые «front-page публикации», которые цитируются на первых страницах описаний отобранных изобретений. Затем из этих непатентных ссылок были выбраны только те ссылки на журнальные статьи (было отобрано 106 636 ссылок), для которых удалось идентифицировать название журнала и соотнести его с нормативным списком, в котором каждому названию присвоена одна или несколько рубрик научных направлений исследований. В результате обработки этих ссылок были вычислены значения матрицы взаимосвязей технологий и научных направлений. При этом описания изобретений, относящихся к 7% рубрик МПК (31 подкласс), включали 80% ссылок на журнальные статьи, что свидетельствует о весьма неравномерном распределении научных публикаций по областям технологий.

Первый вариант рубрицирования публикаций не дает возможности определить направление исследований в том случае, когда статья опубликована в журнале, которому присвоено несколько рубрик научных направлений исследований. Поэтому в процессе проектирования АИС был разработан второй вариант рубрицирования публикаций, в котором для более точного и полного рубрицирования было предложено использовать процедуру сопоставления терминологических портретов рубрик и названий статей. Очевидно, что результаты расчетов значений матрицы взаимосвязей технологий и научных направлений будут зависеть от выбранного способа рубрицирования.

Подсистема «Индикаторы» предназначена для вычисления значений индикаторов. Эта подсистема решает следующие задачи:

- вычисление частот встречаемости пар рубрика МПК (технологии)– рубрика Классификатора РФФИ (направления научных исследований) в виде матрицы взаимосвязей технологий и научных направлений;
- создание и ведение БД вычисленных значений индикаторов.

Таким образом, в качестве индикаторов взаимосвязей технологий и направлений научных исследований выбраны частоты встречаемости пар индекс МПК и рубрика Классификатора РФФИ, которые определяются на основе ссылок на научные публикации, извлеченные из полнотекстовых описаний изобретений.

Дополнительные атрибуты взаимосвязей

Библиографические данные изобретений	Дата подачи заявки Дата выдачи патента Страна автора изобретения Страна заявителя Страна патентообладателя
Структурные поля ссылок цитирования	Год публикации Страна публикации

Наличие и использование дополнительных атрибутов для каждой полученной взаимосвязи (см. таблицу) позволяет создавать подмножества индикаторов взаимосвязей науки и технологий по странам, периодам подачи заявок, выдачи патентов и т. д.

Подсистема «Экспертиза» предназначена для построения сценариев и экспертной верификации вычисленных значений индикаторов. Эта подсистема решает следующие задачи:

- построение сценариев вычисления значений индикаторов взаимосвязей науки и технологий на основе заданных параметров;
- обеспечение экспертной верификации вычисленных значений индикаторов.

Подсистема «Экспертиза» доступна только для экспертов. Она поддерживает возможность создания, хранения и сопоставления различных вариантов сценарных расчетов. Она также дает возможность сопоставлять вычисленные значения индикаторов с теми информационными ресурсами (описаниями изобретений и ссылками на публикации), в процессе обработки которых было получено каждое значение индикатора.

Для представления результатов работы экспертов в виде аналитических отчетов предлагается использовать подсистему «Аналитика». Эта подсистема выполняет следующие задачи:

- проектирование аналитических отчетов;
- автоматизированное формирование аналитических отчетов, содержащих результаты вычислений значений индикаторов в виде таблиц, диаграмм, графиков и сопровождающего текста (структура аналитического отчета определяется в процессе его проектирования).

5 Заключение

Рассмотренные архитектурные решения для вычисления индикаторов тематических взаимосвязей науки и технологий являются основой создания информационной системы, которая не имеет аналогов в российской научно-технической

сфере. Наиболее близким зарубежным аналогом является европейская система, описание которой дано в работе [27]. Предлагаемые архитектурные решения обладают рядом принципиальных отличий от этого аналога:

- выбор именно тех систем классификации областей знаний, которые используются в процессе принятия решений;
- использование всех научных публикаций, цитируемых в описаниях отобранных изобретений, для определения значений индикаторов взаимосвязей науки и технологий (одновременно планируется вычислять и варианты индикаторов с использованием только «front-page публикаций», что даст возможность сопоставить полученные отечественные результаты с зарубежными);
- формирование и применение терминологических портретов рубрик для используемых систем классификации областей знаний;
- использование ключевых слов из названий публикаций для уточнения рубрики публикации в тех случаях, когда одному источнику публикации приписано несколько рубрик системы классификации областей знаний.

Актуальность создания в России системы для вычисления индикаторов тематических взаимосвязей науки и технологий существенно возросла после принятия Государственной программы РФ «Развитие науки и технологий на 2013–2020 годы», включающей шесть подпрограмм. Для подпрограммы номер 2 «Прикладные проблемно-ориентированные исследования и развитие научно-технического задела в области перспективных технологий» установлены два следующих индикатора:

- (1) коэффициент изобретательской активности (число отечественных патентных заявок на изобретения, поданных в России в расчете на 10 тыс. чел. населения);
- (2) число патентных заявок на изобретения, поданных отечественными заявителями в России из организаций — участников Государственной программы.

Цитируемость в описаниях изобретений публикаций, являющихся непосредственными результатами научных исследований и разработок, не нашла своего отражения в списке индикаторов, которые должны отражать развитие научно-технического задела в области перспективных технологий. Таким образом, в списке утвержденных индикаторов отсутствует целая категория показателей, характеризующих процессы переноса знаний из разных областей научных исследований в отечественную сферу технологического развития и инноваций. Одна из причин отсутствия этой категории заключается в том, что в нашей стране в настоящее время полностью отсутствуют те информационные системы, которые могли бы вычислять индикаторы взаимосвязей научных направлений и технологий, сопоставляя накопленные научные и патентные информационные ресурсы.

Литература

1. *Mansfield E.* Academic research and innovation // *Res. Policy*, 1991. Vol. 20. Iss. 1. P. 1–12.
2. *Mansfield E.* Academic research underlying industrial innovations: Sources, characteristics and financing // *Rev. Econ. Statistics*, 1995. Vol. 77. No. 1. P. 55–62.
3. *Mansfield E.* Academic research and industrial innovation: An update of empirical findings // *Res. Policy*, 1998. Vol. 26. Iss. 7-8. P. 773–776.
4. *Narin F., Noma E.* Is technology becoming science? // *Scientometrics*, 1985. Vol. 7. No. 3-6. P. 369–381.
5. *Schmoch U.* Tracing the knowledge transfer from science to technology as reflected in patent indicators // *Scientometrics*, 1993. Vol. 26. P. 193–211.
6. *Narin F., Olivastro D.* Linkage between patents and papers: An interim EPO/US comparison // *Scientometrics*, 1998. Vol. 41. No. 1-2. P. 51–59.
7. *Tijssen R. J. W., Buter R. K., Van Leeuwen Th. N.* Technological relevance of science: An assessment of citation linkages between patents and research papers // *Scientometrics*, 2000. Vol. 47. No. 2. P. 389–412.
8. *Van Looy B., Zimmermann E., Veugelers R., Verbeek A., Mello J., Debackere K.* Do science–technology interactions pay on when developing technology? An exploratory investigation of 10 science-intensive technology domains // *Scientometrics*, 2003. Vol. 57. No. 3. P. 355–367.
9. *Зацман И. М., Шубников С. К.* Принципы обработки информационных ресурсов для оценки инновационного потенциала направлений научных исследований // *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды 9-й Всеросс. научной конф. RCDL'2007.* — Переславль: Ун-т г. Переславля, 2007. С. 35–44.
10. *Зацман И. М., Курчавова О. А., Галина И. В.* Информационные ресурсы и индикаторы для оценки инновационного потенциала направлений научных исследований // *Системы и средства информатики*, 2008. Вып. 18 (доп.). С. 159–175.
11. *Кожунова О. С.* Цитирование документов в патентах как индикатор взаимосвязи областей науки и технологий // *Системы и средства информатики*, 2012. Т. 22. № 2. С. 106–128.
12. *Минин В. А., Зацман И. М., Кружков М. Г., Норежян Т. П.* Методологические основы создания информационных систем для вычисления индикаторов тематических взаимосвязей науки и технологий // *Информатика и её применения*, 2013. Т. 7. Вып. 1. С. 70–81.
13. *Архипова М. Ю., Зацман И. М., Шульга С. Ю.* Индикаторы патентной активности в сфере информационно-коммуникационных технологий и методика их вычисления // *Экономика, статистика и информатика. Вестник УМО*, 2010. № 4. С. 93–104.
14. *Зацман И. М., Кожунова О. С.* Семантический словарь системы информационного мониторинга в сфере науки: задачи и функции // *Системы и средства информатики*, 2007. Вып. 17. С. 124–141.
15. *Zatsman I., Kozhunova O.* Evaluating for institutional academic activities: Classification scheme for R&D indicators // *10th Conference (International) on Science and*

- Technology Indicators (STI'2008): Book of abstracts. — Vienna: ARC GmbH, 2008. P. 428–431.
16. *Zatsman I., Kozhunova O.* Evaluation system for the Russian Academy of Sciences: Objectives-resources-results approach and R&D indicators // 2009 Atlanta Conference on Science and Innovation Policy Proceedings / Eds. S. E. Cozzens and P. Catalan. <http://smartech.gatech.edu/bitstream/1853/32300/1/104-674-1-PB.pdf>.
 17. *Зацман И. М.* Категоризация результатов и индикаторов программ научных исследований в информационных системах мониторинга // Системы и средства информатики, 2009. Вып. 19 (доп.). С. 200–219.
 18. *Zatsman I., Durnovo A.* Incompleteness problem of indicators system of research programme // 11th Conference (International) on Science and Technology Indicators (STI'2010): Book of abstracts. — Leiden: CWTS, 2010. P. 309–311.
 19. *Zatsman I., Durnovo A.* Program-oriented indicators: Production and application in science // Системы и средства информатики, 2012. Т. 22. № 1. С. 110–120.
 20. *Zatsman I., Durnovo A.* Proactive dictionary of evaluation system as a tool for science and technology indicator development // 17th Conference (International) on Science and Technology Indicators Proceedings. — Montréal: Science-Metrix and OST, 2012. Vol. 2. P. 905–906.
 21. *Зацман И. М., Косарик В. В., Курчавова О. А.* Задачи представления личностных и коллективных концептов в цифровой среде // Информатика и её применения, 2008. Т. 2. Вып. 3. С. 54–69.
 22. *Зацман И. М.* Семиотическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов // Информатика и её применения, 2009. Т. 3. Вып. 2. С. 65–81.
 23. *Зацман И. М.* Нестационарная семиотическая модель компьютерного кодирования концептов, информационных объектов и денотатов // Информатика и её применения, 2009. Т. 3. Вып. 4. С. 87–101.
 24. *Зацман И. М., Дурново А. А.* Моделирование процессов формирования экспертных знаний для мониторинга программно-целевой деятельности // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 84–98.
 25. *Zatsman I.* Tracing emerging meanings by computer: Semiotic framework // 13th Conference (European) on Knowledge Management Proceedings. — Reading: Acad. Publ. Intern. Ltd., 2012. Vol. 2. P. 1298–1307.
 26. *Zatsman I.* Denotatum-based models of knowledge creation for monitoring and evaluating R&D program implementation // 11th IEEE Conference (International) on Cognitive Informatics and Cognitive Computing Proceedings. — Los Alamitos, CA: IEEE Computer Society Press, 2012. P. 27–34.
 27. *Verbeek A., Debackere K., Luwel M., Andries P., Zimmermann E., Deleus D.* Linking science to technology: Using bibliographic references in patents to build linkage schemes // Scientometrics, 2002. Vol. 54. No. 3. P. 399–420.

INFORMATION SYSTEM CONCEPTUAL DECISIONS FOR ASSESSMENT OF LINKAGES BETWEEN SCIENCE AND TECHNOLOGIES

V. A. Minin¹, I. M. Zatsman², V. A. Havanskov³, and S. K. Shubnikov⁴

¹Russian Foundation for Basic Research, Moscow, Russia, minin@rfbr.ru

²IPI RAN, Moscow, Russia, iz_ipi@a170.ipi.ac.ru

³IPI RAN, Moscow, Russia, havanskov@a170.ipi.ac.ru

⁴IPI RAN, Moscow, Russia, sergeysh50@yandex.ru

Abstract: The description of conceptual decisions for assessment of linkages between science and technologies is considered. The decisions are a framework for developing information systems, which will have some new features. These features are not present in other similar information systems of the Russian scientific and technical sphere. Their development is necessary for monitoring and evaluating R&D programs and decision-making at all stages of scientific and technical program activities. The architecture as a whole system and conceptual decisions on separate functional subsystems are described. These decisions are based on methodology for assessment of linkages between science and technologies, which is developed in IPI RAN. Conceptual decisions have a number of fundamental differences from other systems for assessment of linkages between science and technologies. First, there is a possibility of fetching a system for classification of knowledge domains from some classification systems. Second, formation and application of terminological portraits for headings of the chosen classification system is provided.

Keywords: monitoring and evaluating of programs; linkages between science and technologies; methodology for assessment of linkages; information systems; conceptual decisions

DOI: 10.14357/08696527130217

Acknowledgments

The research was partially supported by the Russian Foundation for Humanities (Grant No. 12-02-12019v).

References

1. Mansfield, E. 1991. Academic research and innovation. *Res. Policy* 20(1):1–12.
2. Mansfield, E. 1995. Academic research underlying industrial innovations: Sources, characteristics and financing. *Rev. Econ. Statistics* 77(1):55–62.
3. Mansfield, E. 1998. Academic research and industrial innovation: An update of empirical findings. *Res. Policy* 26(7-8):773–76.

4. Narin, F., and E. Noma. 1985. Is technology becoming science? *Scientometrics* 7(3-6):369–81.
5. Schmoch, U. 1993. Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics* 26:193–211.
6. Narin, F., and D. Olivastro. 1998. Linkage between patents and papers: An interim EPO/US comparison. *Scientometrics* 41(1-2):51–59.
7. Tijssen, R. J. W., R. K. Buter, and Th. N. Van Leeuwen. 2000. Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics* 47(2):389–412.
8. Van Looy, B., E. Zimmermann, R. Veugelers, A. Verbeek, J. Mello, and K. Debackere. 2003. Do science–technology interactions pay off when developing technology? An exploratory investigation of 10 science-intensive technology domains. *Scientometrics* 57(3):355–67.
9. Zatsman, I., and S. Shubnikov. 2007. Principy obrabotki informacionnyh resursov dlja ocenki innovacionnogo potenciala napravlenij nauchnyh issledovanij [Processing principles of information resources for an assessment of innovation potential of the scientific domains]. *Trudy 9 Vserossijskoj Nauchnoj Konferencii “Jelektronnye biblioteki” [9th All-Russian Scientific Conference on Digital Libraries Proceedings]*. Pereslavl': Publishing House of Pereslavl's University. 35–44.
10. Zatsman, I., O. Kurchavova, and I. Galina. 2008. Informacionnye resursy i indikatory dlja ocenki innovacionnogo potenciala napravlenij nauchnyh issledovanij [Information resources and indicators for an assessment of innovation potential of the scientific domains]. *Systems and Means of Informatics* 18(add.):159–75.
11. Kozhunova, O. 2012. Citirovanie dokumentov v patentah kak indikator vzaimosvjazi oblastej nauki i tehnologij [Citing documents in patents as an indicator of science and technology fields interrelations]. *Systems and Means of Informatics* 22(2):106–28.
12. Minin, V., I. Zatsman, M. Krushkov, and T. Norekjan. 2013. Metodologicheskie osnovy sozdaniya informacionnyh sistem dlja vychislenija indikatorov tematicheskikh vzaimosvjazej nauki i tehnologij [Methodological basis for the creation of information systems for the calculation of indicators of thematic linkages between science and technology]. *Informatics and Applications* 7(1):70–81.
13. Arhipova, M., I. Zatsman, and S. Shul'ga. 2010. Indikatory patentnoj aktivnosti v sfere informacionno-kommunikacionnyh tehnologij i metodika ih vychislenija [Indicators of patent activity in the field of information and communication technologies and the method of their calculation]. *Jekonomika, statistika i informatika [Economics Statistics and Informatics]* 4:93–104.
14. Zatsman, I., and O. Kozhunova. 2007. Semanticheskij slovar' sistemy informacionnogo monitoringa v sfere nauki: Zadachi i funkcii [Semantic vocabulary of the system of information monitoring in scientific sphere: Tasks and functions]. *Systems and Means of Informatics* 17:124–41.
15. Zatsman, I., and O. Kozhunova. 2008. Evaluating for institutional academic activities: Classification scheme for R&D indicators. *10th Conference (International) on Science and Technology Indicators (STI'2008) Proceedings. Book of abstracts*. Vienna: ARC GmbH. 428–31.

16. Zatsman, I., and O. Kozhunova. 2009. Evaluation system for the Russian Academy of Sciences: Objectives-resources-results approach and R&D indicators. *2009 Atlanta Conference on Science and Innovation Policy Proceedings*. Eds. S. E. Cozzens and P. Catalan. <http://smartech.gatech.edu/bitstream/1853/32300/1/104-674-1-PB.pdf>.
17. Zatsman, I. 2009. Kategorizacija rezul'tatov i indikatorov programm nauchnyh issledovanij v informacionnyh sistemah monitoringa [Categorization of results and indicators of R&D programs in evaluation systems]. *Systems and Means of Informatics* 19 (add.):200–19.
18. Zatsman, I., and A. Durnovo. 2010. Incompleteness problem of indicators system of research programme. *11th Conference (International) on Science and Technology Indicators (STI'2010): Book of Abstracts*. Leiden: CWTS. 309–11.
19. Zatsman, I., and A. Durnovo 2012. Program-oriented indicators: production and application in science. *Systems and Means of Informatics* 22(1):110–20.
20. Zatsman, I., and A. Durnovo. 2012. Proactive dictionary of evaluation system as a tool for science and technology indicator development. *17th Conference (International) on Science and Technology Indicators Proceedings*. Montréal: Science-Metrix and OST. 2:905–6.
21. Zatsman, I., V. Kosarik, and O. Kurchavova. 2008. Zadachi predstavlenija lichnostnyh i kollektivnyh konceptov v cifrovoj srede [Personal and collective concepts representation in the digital sphere]. *Informatics and Applications* 2(3):54–69.
22. Zatsman, I. Semioticheskaja model' vzaimosvjazej konceptov, informacionnyh ob'ektov i komp'juternyh kodov [Semiotic model for computer coding concepts and information objects]. *Informatics and Applications* 3(2):65–81.
23. Zatsman, I. 2009. Nestacionarnaja semioticheskaja model' komp'juternogo kodirovanija konceptov, informacionnyh ob'ektov i denotatov [Time-dependent semiotic model for computer coding of concepts, information objects, and denotata]. *Informatics and Applications* 3(4):87–101.
24. Zatsman, I., and A. Durnovo. 2011. Modelirovanie processov formirovanija jekspertnyh znaniy dlja monitoringa programmno-celevoj dejatel'nosti [Modeling of processes for creation of expert knowledge for monitoring of goal-oriented programme activities]. *Informatics and Applications* 5(4):84–98.
25. Zatsman, I. 2012. Tracing emerging meanings by computer: Semiotic framework. *13th Conference (European) on Knowledge Management Proceedings*. Reading: Acad. Publ. Intern. Ltd. 2:1298–307.
26. Zatsman, I. 2012. Denotatum-based models of knowledge creation for monitoring and evaluating R&D program implementation. *11th IEEE Conference (International) on Cognitive Informatics and Cognitive Computing Proceedings*. Los Alamitos, CA: IEEE Computer Society Press. 27–34.
27. Verbeek, A., K. Debackere, M. Luwel, P. Andries, E. Zimmermann, and D. Deleus. 2002. Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics* 54(3):399–420.