

области, как информационные технологии, невозможно. Не удалось выявить такую методику расчета, которая бы полностью удовлетворяла решению задачи определения экономической эффективности автоматизированной информационной системы при внедрении тезауруса.

Нами была предложена собственная методика оценки экономической эффективности внедрения АИС с тезаурусом. Основой этой методики является проведение экспертного опроса по методике СВ-90, затем расчет полученного эффекта по методике компании JS Penney. Расходы по внедрению тезауруса определялись по методу "оценки портфелей", в завершение рассчитывалось отношение полученного результата к необходимым затратам.

Расчеты для АИС с тезаурусом при общем количестве пользователей 150 человек показали величину возможного годового экономического эффекта около 600 тыс. рублей.

СПИСОК ЛИТЕРАТУРЫ

1. Михайлов А. И., Черный А. И., Гиляревский Р. С. Основы научной информации.— М.: Наука, 1965.— 656 с.
2. Жмайло С. В. К вопросу об определении тезауруса // НТИ. Сер. 1.— 2003.— № 12.— С. 20–25.
3. Михайлов А. И., Черный А. И., Гиляревский Р. С. Основы информатики.— М.: Наука, 1968.
4. Batty D. Controlled vocabulary and thesauri in support of online information access // D.— Lib Magazine, ISSN 1082-9873, November, 1998.

5. Svenonius E. Definitional approaches in the design of classification and thesauri and their implications for retrieval and for automatic classification // Knowledge organization for information retrieval. Proc. Of the 6 International study conference on classification research.— London, 1997.

6. Азарова И. В., Митрофанова О. А., Синопальникова А. А., Ушакова А. А., Яворская М. В. Разработка компьютерного тезауруса русского языка типа WordNet // Доклады научной конференции "Корпусная лингвистика и лингвистические базы данных" / Под ред. А. С. Герда.— СПб., 2002.— С. 6–18.

7. Добров Б. В., Лукашевич Н. В. Тезаурус и автоматическое концептуальное индексирование в университетской информационной системе РОССИЯ // Третья всероссийская конференция по Электронным Библиотекам "Электронные библиотеки: перспективные методы и технологии, электронные коллекции".— Петрозаводск, 2001.— С. 78–82. См. также: rcdl2001.krc.karelia.ru/papers/papers/dobrov_lukashevich/dobrov_paper.rtf

8. Тезаурус Медиалингва [Электрон. ресурс].— Режим доступа: <http://www.medialingua.ru/thesaurus.html>

9. Тезаурус RCO [Электрон. ресурс].— Режим доступа: http://www.rco.ru/product.asp?ob_no=19

10. Горькова В. И. Информетрия (количественные методы в научно-технической информации) // Итоги науки и техники ВИНТИ.— Сер. Информатика.— 1988.— Т. 10.— 328 с.

11. Черный А. И. Общая методика построения тезаурусов // НТИ. Сер. 2.— 1968.— № 5.— С. 9–33.

Материал принят к опубликованию 21.12.06.

УДК 004.891:(088.8)ВИНИТИ

А. А. Батюшко, Г. А. Куц, В. В. Омерда, Т. П. Пономаренко

Система экспертной оценки патентных документов в производственном цикле ВИНТИ РАН

Описаны некоторые программно-технологические решения по созданию на основе Международной патентной классификации (МПК) и Рубрикатора ВИНТИ модели экспертной оценки отбора необходимых документов для их использования в производственном процессе ВИНТИ. Влияние на модель технологических параметров позволяет определить вид выдачи документа для его дальнейшей обработки (полный текст, титульный лист и патентная формула, библиографический формуляр и др.). Включение системы в технологический процесс позволит снизить расход материальных и временных ресурсов и снять с квалифицированных сотрудников ряд рутинных ручных операций.

Создание ядра индексов МПК, используемых в ВИНТИ, послужит основой для проведения дальнейших исследовательских работ по применению различных классификационных схем при обработке входного потока НТЛ и увязке их с Рубрикаторм ВИНТИ.

ВВЕДЕНИЕ

Одна из составных частей входного потока научно-технической литературы (НТЛ) в ВИНТИ РАН — патентные документы, которые, наряду с правовым статусом юридического документа, за-

крепляющим права изобретателя, являются ценнейшим источником мировой НТЛ и занимают существенное место в общем потоке материалов, поступающих в ВИНТИ. Без патентной документации нельзя обеспечить в базах данных (БД) и реферативном журнале (РЖ) ВИНТИ полноту ин-

формации по соответствующим отраслям науки и техники. Информация такого рода особенно активно используется специалистами при подготовке обзоров или на начальных стадиях организации работ, связанных с оценкой тенденций развития уровня науки и техники, дает возможность следить за текущим потоком по всем видам НТЛ с наименьшими затратами времени.

В соответствии с многолетней практикой и существующими договоренностями ВИНТИ формирует входной поток патентных документов промышленно развитых стран на базе фонда Всероссийской патентно-технической библиотеки. ВПТБ предоставляет во временное пользование отечественные и зарубежные описания изобретений, патентные документы текущего поступления.

В настоящее время изменение патентной ситуации в мире привело к значительным изменениям в объемах и видах издаваемых патентных документов. В мировой практике произошел перевод традиционных патентных фондов с бумаги на электронные виды носителей.

Так как в ВИНТИ отсутствует система обработки патентных документов на электронных носителях, институт по-прежнему получает из ВПТБ их бумажные аналоги.

ТЕХНОЛОГИЯ ОБРАБОТКИ ПАТЕНТНЫХ ДОКУМЕНТОВ

В настоящее время в ВИНТИ ежегодно обрабатывается до 100 тыс. патентных документов, получаемых из ВПТБ. Технология обработки содержит множество трудоемких и ресурсоемких ручных операций. Сотрудники отбирают патентные документы в ВПТБ, упаковывают их и отправляют в ВИНТИ. Так как материалы предоставляются во временное пользование на 12 дней, производится полное ксерокопирование патентных документов. За этим следуют разбор и проверка качества ксерокопий, упаковка оригиналов и отправка их обратно в ВПТБ, где их опять разбирают и подкладывают на те же места, откуда документы были изъяты. Помимо огромного ручного труда, требуется большой расход материалов на ксерокопирование. Дальнейшая обработка патентных документов в ВИНТИ происходит с ксерокопий.

При этом выполняются следующие технологические операции.

1. *Отбор в ВПТБ патентных документов* осуществляется по первичности приоритета в соответствии с Рубрикаторм ВИНТИ. ВПТБ предоставляет патентные документы в следующем виде (табл. 1):

Таблица 1

Страна	Вид документа
США	Патенты
Германия	Заявки
Франция	Заявки
Россия	Патенты
Великобритания	Заявки
Австрия	Патенты
Швейцария	Патенты
ЕПВ	Заявки

Отобранные документы пересчитываются, оформляются в папки и регистрируются в журнале ВПТБ.

2. *Ксерокопирование отобранных патентных документов.* После доставки из ВПТБ в ВИНТИ полученные материалы еще раз просматриваются на предмет соответствия Рубрикатору ВИНТИ, так же происходит проверка на наличие страниц, рисунков, патентных формул. Проверенные документы отправляются на полное ксерокопирование.

Патенты ксерокопируются в Производственно-издательском комбинате ВИНТИ (ПИК).

После возврата материалов из ПИК их разбирают: отделяют оригиналы от копий, проверяют качество и совпадение количества оригиналов и копий, формируют новые упаковки для отправки в ВПТБ.

3. *Сдача оригиналов в ВПТБ.* Распаковку пачек с патентными документами, подкладку их на то же место, регистрацию в журнале ВПТБ (проставление даты сдачи) осуществляют сотрудники ВИНТИ.

4. *Научная систематизация и частичная библиографическая обработка* (работа с копиями). Научная систематизация заключается в определении тематики документа с тем, чтобы направить его на дальнейшую обработку: реферирование, индексирование, рубрицирование — в один из отделов научной информации (ОНИ) по отраслям науки и техники. Таких отделов в ВИНТИ шестнадцать: по астрономии, автоматике и радиоэлектронике, биологии, геологии и горному делу, географии и геодезии, информатике и полиграфии, математике, механике машиностроению, металлургии, транспорту, химии и химической технологии, физике, энергетике и электротехнике, экономике и глобальным проблемам. На документе проставляются штампы ОНИ, штамп основного ОНИ подчеркивается.

Частичная библиографическая обработка заключается в оформлении ряда полей на копии патентного документа с тем, чтобы обеспечить их правильный ввод и отражение в БД ВИНТИ:

- подчеркивание кодов ИНИД (международно-согласованные цифровые коды для идентификации библиографических данных), в которых содержится сведения о следующих библиографических полях: номер поданной заявки, дата подачи заявки, дата публикации, изобретатель, патентообладатель;
- оформление таких полей библиографического описания патента, как автор, сокращенное название и пр.;
- простановка штампов страны, вида документа.

5. *Внесение данных в систему учета патентных документов.* Обработка патентных документов в ВИНТИ не автоматизирована. Это единственный вид документов, не регистрируемый в автоматизированной системе ВИНТИ АС "ВХОД". Но в систему поступают некоторые данные: номер технологической порции, количество патентных документов в порции, страна, ФИО обработавшего порцию сотрудника, дата ввода, распределение документов по тематике в ОНИ. По результатам ввода ежемесячно снимаются сводки о выработке сотрудников, распределении документов по странам и по ОНИ.

Как уже указывалось выше, патентные документы единственная часть входного потока

ВИНИТИ, обработка которой не автоматизирована.

В то время как сериальные издания и издания книжного типа регистрируются в АС "ВХОД", получают штрих-коды, уникальные системные идентификаторы и технологические атрибуты, патенты неизвестно ничего, кроме ряда количественных характеристик, необходимых для учета выработки сотрудников.

В данной работе не ставилась цель рассмотреть возможность включения в АС "ВХОД" электронных описаний патентных документов, проходящих обработку в ВИНИТИ. Здесь описаны некоторые программно-технологические решения, позволяющие снизить расход материальных и временных ресурсов и освободить квалифицированных сотрудников от ряда рутинных ручных операций.

СОВЕРШЕНСТВОВАНИЕ ТЕХНОЛОГИИ ОБРАБОТКИ ПАТЕНТНЫХ ДОКУМЕНТОВ

Документы в ВПТБ отбираются по Рубриктору ВИНИТИ. Причем востребованность индексов МПК, тем и предметов для отражения в БД и РЖ изменяются как по объективным, так и по субъективным факторам. Эту информацию при отборе документов сотрудник держит в памяти, что приводит к некоторой избыточности и неточности отобранного потока.

На этапе научной систематизации лишние документы отбраковываются, но из-за сжатости сроков обработки ксерокопии с них уже изготовлены.

Для реферирования патентного документа по ряду отраслей науки и техники его полный текст не нужен, достаточно иметь титульный лист с библиографическими данными, рефератом и общим чертежом. Однако ксерокопии полных текстов и для таких документов уже изготовлены.

В силу субъективных обстоятельств референты и редакторы ряда ОНИ работают с ограниченным перечнем языков и стран, но и по "необрабатываемым" документам ксерокопии уже изготовлены.

Поэтому для облегчения труда предлагается на основе МПК и Рубриктора ВИНИТИ создать модель экспертной оценки отбора необходимых документов для их использования в производственном процессе ВИНИТИ. Влияние на модель технологических параметров позволит определить вид выдачи документа для его дальнейшей обработки (полный текст, титульный лист и патентная формула, библиографический формуляр и др.).

Использование индексов МПК патентного документа позволит дать рекомендации для первичной разметки по Рубриктору ВИНИТИ.

Факторы, влияющие на технологию обработки

Признак технологической обработки (ПТО) определяет на первом этапе разработки полноту копирования патентных документов с целью сокращения сроков и материальных затрат, но с учетом обеспечения сотрудников Института всей необходимой для работы информацией.

Влияние потребностей ОНИ в патентных документах. Референты и редакторы ВИНИТИ работают с документами не полного спектра индексов МПК, а с его определенной частью. Для учета этих особенностей от ОНИ поступают заказы на технологическую обработку или отказ от нее для документов с теми или иными индексами МПК. Перечень языковых возможностей тоже ограничен, что находит отражение в Заявке на отбор патентных документов, форма которой приводится (табл. 2).

Влияние страны публикации патентного документа. Представление описаний изобретений патентных документов разных стран отличаются друг от друга (табл. 3)

Влияние тематики патентного документа. В зависимости от тематики для подготовки реферата и индексирования могут быть востребованы как полные тексты патентного документа (например, химия, машиностроение), так и только титульные листы с рефератом и патентной формулой.

Таблица 2

Заявка на отбор патентных документов для ОНИ

№	Тематика	Индекс МПК	Рубрика ВИНИТИ	Языки 1. англ. 2. нем. 3. фр. 4. рус.	Срок	Полнота копирования
						1. полный текст+чертеж. 2. полный текст без чертеж. 3. тит лист+ 1 стр. текста 4. др. варианты
1						
2						
3						

Таблица 3

Страна	Расположение материалов
ФРГ, Франция, Австрия, Швейцария, Россия, Болгария	1. Титульный лист с библиографическими данными, рефератом и общим чертежом. 2. Текст описания изобретения с патентной формулой в конце текста. 3. Чертежи.
США, Великобритания	1. Титульный лист с библиографическими данными, рефератом и общим чертежом. 2. Чертежи. 3. Текст описания изобретения с патентной формулой в конце текста.
Австралия	1. Титульный лист с библиографическими данными, без реферата и чертежей. 2. Реферат на отдельной странице. 3. Описание изобретения с патентной формулой в конце. 4. Чертежи.

На основании этих факторов вводим признаки технологической обработки:

- документ копируется полностью;
- копируется первая страница;
- документ не обрабатывается;
- на документ готовится библиографический формуляр.

В процессе работы могут возникнуть и другие варианты ПТО.

Установление связи параметров — индекс МПК, ОНИ, страна, язык и признаки технологической обработки, дает возможность разделить обработку патентных документов на технологические потоки, сэкономить на ксерокопировании материалов, сократить сроки обработки.

На основании отражения документов с конкретными индексами МПК в выпусках реферативного журнала и базах данных ВИНТИ может быть принято решение об их востребованности и простановке того или иного технологического признака.

Автоматическая простановка признака не используется, решение принимается в результате анализа данных и служебных записок от ОНИ.

Технология с участием экспертной системы

Пронумерованные папки с оригиналами патентных документов поступают из ВПТБ к техническому оператору сектора обработки патентных документов ВИНТИ.

Оператор, работающий с программой экспертной оценки "Патент", вводит с документа в базу данных АС "ВХОД":

- страну (выбирается из словаря) и язык; язык выбирается из словаря по коду страны; если языков несколько, существует возможность выбора одного из них;
- рабочий номер патента — так называемый рабочий номер документа (РНД), для некоторых

стран его можно сканировать. РНД уникальный, цифровой, максимальная длина 9 Б;

- нормализованный индекс МПК; индекс МПК вводится по шаблону и контролируется по словарю-классификатору; по индексу МПК определяется название индекса и связанные с ним рубрики ВИНТИ и ОНИ.

Ответом программы является признак технологической обработки: документ копируется полностью с чертежом или без, копируется первая страница, документ не обрабатывается, на документ готовится библиографический формуляр.

В соответствии с признаком технологической обработки оператор помещает документ в один из технологических потоков обработки.

В процессе работы программы происходит учет патентного документа — запись введенных данных:

РНД + МПК + ПТО + ФИО оператора +
+ дата обработки.

К этим данным добавляется уникальный идентификатор штрих-код и канал поступления — ВПТБ.

Такую регистрацию нельзя назвать полной, так как библиографическое описание документа не полагается в АС "ВХОД". Но уникальность рабочего номера патента всегда позволит объединить имеющуюся информацию с появившейся библиографической.

Таким образом, технический оператор, не владеющий ни классификационными системами МПК, ни Рубрикаторм ВИНТИ, ни иностранными языками перераспределяет потоки патентных документов по технологическим участкам, исключая излишнее копирование материалов.

Бесспорное достоинство этой технологии — возможность учета патентных документов в АС "ВХОД".

Работоспособность экспертной системы поддерживается словарной базой: классификатора МПК, Рубрикаторм ВИНТИ, таблицы связи между ними и признаками технологической обработки.

Если у ОНИ изменилась потребность в конкретных индексах МПК (например, востребованы новые), то простановка соответствующего технологического признака позволит включить их в обработку. И наоборот, простановка признака "не обрабатывать" у устаревшего индекса исключит из потока связанные с ним патентные документы.

Автоматизация научной систематизации (тематической разметки)

Возможность автоматизации тематической разметки патентных документов реализуется на основании таблицы связи МПК и Рубрикаторм ВИНТИ.

Копии патентных документов, прошедшие учет в АС "ВХОД", поступают к сотруднику отдела научной систематизации. Сотрудник, работающий с программой "Патент", вводит с документа рабочий номер патента, по которому подбираются уже