

А. А. Рагойша
Онлайновый патентный поиск в учебном процессе
на химическом факультете Белорусского государственного университета

Работа с литературными первоисточниками является для студента новым видом деятельности и поэтому сопряжена со значительными объективными трудностями. Научные публикации кардинально отличаются от учебных пособий не только уровнем сложности материала, но и своей структурой, фрагментарностью, недогматичностью, разнообразием форматов и рассеянностью в информационном пространстве. Студенты сравнительно быстро осваивают базовые принципы работы с научными статьями, поскольку в своем предыдущем опыте они находят этому типу публикации, хотя и отдаленный, но аналог — статью в популярном журнале. Патентная литература же специфична, и начинающие пользователи непроизвольно пренебрегают ею, причем не из-за ее низкой информационной ценности, а по причине отсутствия соответствующих навыков.

Необходимость ознакомления студентов с патентными документами обусловлена следующими факторами:

- Уникальность информации. По разным оценкам, 70–95 % научной и технологической информации, содержащейся в патентах, не публикуется в иной научной литературе [1].
- Достоверность. Патент относится к классу рецензируемой научной литературы; воспроизводимость опубликованного в патенте материала — обязательное условие при выдаче патента.
- Детализация. Патент, как правило, является многостраничным документом, содержащим подробное описание методики изготовления и применения объекта. Это выгодно отличает патент от статьи в научном журнале, где существуют ограничения в объеме публикации.
- Новизна. Если изобретатель предполагает возможность практического внедрения результатов своих исследований, то он документально фиксирует их в форме патента; повторная публикация в форме научной статьи нередко происходит через длительный период времени [1].
- Доступность. Патентные документы ведущих стран переведены из печатного в электронный формат. В Интернете предоставляется бесплатный доступ к десяткам миллионов патентных документов, причем количество их возрастает постоянно [2].
- Приемлемая система обнаружения информации. Патентный поиск может быть осуществлен достаточно быстро, с использованием стандартизированных поисковых программ.

Патентные документы выполняют две функции: юридическую и информационную. Юридические аспекты (срок действия патента, охраняемая территория и предметная область и т. д.) особенно важны на стадиях оформления патентной заявки или внедрения изобретения в производство. Такая деятельность предполагает высокий квалификационный уровень исполнителя, и у большинства студентов, особенно младших курсов, она может вызывать лишь умозрительный интерес.

Информационное наполнение патента представляет ценность для широкого круга исследователей, поэтому патентный документ должен быть для студента настолько же понятным объектом, каковым является научная статья, тезис доклада на

конференции или фрагмент реферативного журнала. В такой же степени понятной должна быть методика обнаружения и извлечения документа из информационного массива.

Включение патентного поиска в учебный процесс предполагает решение двух самостоятельных, но взаимосвязанных проблем:

- ознакомление студента с совершенно новым для него типом источников научной информации;
- обучение навыкам информационного поиска в научной электронной базе данных.

Для решения указанных выше проблем при последовательной реализации принципа "от простого к сложному" потребовалось бы создание учебной базы данных, содержащей несколько уровней сложности, тренинг на моделях-имитациях и на заключительном этапе переход к реальным офлайновым и онлайн-объектам. Такой подход позволил бы подготовить высококвалифицированного специалиста по информационному поиску, однако, кроме потребности в больших затратах времени, таил бы в себе опасность подмены конечной цели работы. Поисковый процесс произвольно воспринимался бы студентом как самодостаточный, но не как вспомогательный элемент в получении информации.

Иной путь — обучение с использованием реальной базы данных — привлекает внимание студентов к содержанию извлекаемых документов. В этом случае, однако, затруднения обнаруживаются на стадии поиска, а также при смене информационного ресурса. Каждая из известных онлайн-баз данных характеризуется своими особенностями функционирования, которые необходимо знать пользователю. Успешность взаимодействия с одним ресурсом не гарантирует полное освоение иных.

Представляется целесообразным проводить обучение методике работы с патентной литературой в три стадии.

На первом, базовом, уровне студенты знакомятся с патентным документом как источником научной информации и с основными приемами текстового патентного поиска в пределах одной базы данных. Предполагается, что по завершении первого уровня обучения пользователь приобретает способность решать отдельные тривиальные задачи, связанные с патентной информацией.

На втором уровне проводится наработка навыков работы с информационным центром, содержащим достаточно материала для полноценной научной деятельности.

Цель третьего уровня обучения — ознакомление со специфическими ресурсами и вспомогательными поисковыми инструментами.

Место в учебном процессе, продолжительность обучения на каждом из уровней могут варьироваться в зависимости от уровня предварительной подготовленности учащихся и степени их мотивации к данной работе. Так, например, на химическом факультете Белорусского государственного университета в конце 1990-х годов ознакомление с патентными базами данных проводилось только на выпускном курсе и только на базовом уровне. В результате роста компьютерной грамотности студентов базовый уровень был перемещен первоначально на третий, а в последние годы — на второй курс. Подобный дрейф претерпевают и более сложные учебные элементы, первоначально ориентированные исключительно на дипломников.

На всех стадиях обучения мы работаем только с бесплатными онлайн-ресурсами.

Мы проанализировали достоинства и недостатки доступных в Интернете

патентных баз данных и в качестве объекта изучения, предназначенного для базового уровня, выбрали информационный массив **USPTO** — патентного бюро США [3].

Выбор обосновывается следующими соображениями:

- Большой объем. Массив содержит полные тексты всех патентов, выданных патентным бюро США за полтора столетия, а также патентных заявок 2000-х годов.
- Репрезентативность. Среди авторов широко представлены не только американские, но и зарубежные изобретатели, и базы данных в значительной степени отражают общемировое состояние патентной информации.
- Простота функционирования и интуитивно понятный интерфейс.
- Соответствие стандарту, де-факто принятому в Интернете.
- Наличие небольшого количества нестандартных элементов, которые можно использовать в качестве иллюстраций при совершенствовании навыков поиска.

На химическом факультете Белорусского государственного университета ознакомление студентов с началами патентного поиска проводится в рамках учебного курса «Поиск химической информации в электронных базах данных и в Интернете». В ходе практикума студенты-второкурсники получают краткие сведения о содержании баз данных *USPTO*, правилах формулирования запроса и алгоритме анализа извлеченных результатов. Соответствующие инструкции изложены в онлайн-овых методических пособиях [4].

По имеющимся в тексте гиперссылкам студент вызывает веб-страницы, открывающиеся в новых окнах браузера, и работает параллельно с методическим пособием и с удаленным сайтом.

Информационный массив *USPTO* выполнен в форме двух идентичных построению, но независимых баз данных, содержащих патенты и патентные заявки. С позиций учебного процесса, этот ресурс — удобный объект для сравнительного анализа документов обоих типов.

Бланк *Quick Search* — пример удачного интерфейса поисковой программы. При внешней простоте бланк позволяет с минимальными затратами труда формулировать запросы средней степени сложности.

На стадии заполнения бланка решаются следующие учебные задачи:

- закрепление знаний о структурных элементах страниц подобного рода (редактируемое поле, поле записи, выпадающий список, булевы операторы, навигационное меню);
- закрепление знаний об использовании операторов для назначения логической связи между поисковыми терминами;
- трансформация мысленно сконструированного поискового задания в формальную комбинацию терминов и операторов;
- принятие решения об оптимальном выборе полей поиска.

Список записей, обнаруживаемых в результате поиска, имеет стандартное строение и не требует детального обсуждения на занятии.

Поскольку патент является, как правило, новым для студентов типом документов, мы подробно рассматриваем и его строение, и соответствующую терминологию (русскоязычную и англоязычную). Учащийся должен научиться различать в патенте страницу библиографического описания (*Front Page, First Page*) с полями «автор» (*Inventor*), «номер публикации» (*Patent Number*), «название» (*Title*), «даты подачи и опубликования» (*Filing, Issue dates*), а также структурные части: «реферат» (*Abstract*), «формулу изобретения» (*Claims*), «описание» (*Description*).

Студенту предлагается самостоятельно разобраться в назначении блока ссылок *References Cited* (американские патенты — прототипы) и *Referenced By* (более поздние патенты США, ссылающиеся на данный).

Занятие завершается выполнением контрольного задания.

Пример задания: «Для чего используется карбамид в патентованной жевательной резинке?».

Учащиеся должны обнаружить соответствующие патенты и их проанализировать. Правильный ход решения предполагает:

- использование в запросе фразы “*chewing gum*”, а не набора слов *chewing, gum*;
- учет существования синонимов *carbamide* (карбамид) и *urea* (мочевина);
- оптимальный выбор полей поиска (*Title* — 0 результатов, *Full Text* — слишком большой список, *Abstract* — оптимальный вариант).

При правильном алгоритме работы студент выводит на экран два документа, рефераты которых невелики по объему и вполне понятны даже пользователю, имеющему минимальные познания в английском языке.

После извлечения и просмотра текстов патентов в аудитории проводится устное обсуждение химического поведения запатентованных композиций:

- клатрат карбамида и пероксида водорода предлагается использовать для отбеливания зубов (студенты ответ находят самостоятельно);
- мочевина ферментативно превращается в аммиак, который нейтрализует органические кислоты в полости рта (как правило, без помощи преподавателя студенты правильное решение не предлагают).

Поисковый бланк *Advanced Search* в данной части учебного курса используется только для иллюстрации иного подхода к формулированию поискового задания.

Бланк предполагает запись задания в виде логического выражения, в котором поисковые термины связаны между собой операторами, а порядок выполнения действий назначается скобками. Такой подход позволяет тонко регулировать направление поиска, но из-за большого количества элементов, специфических только для этой базы данных (коды полей, формат указания полей, формат дат и пр.), затраты усилий на освоение нестандартного для Интернета синтаксиса нецелесообразны.

В настоящее время наибольшее количество патентной информации, доступной бесплатно, сконцентрировано в информационном комплексе **esp@cenet** [5] Европейского патентного ведомства (*EPO — European Patent Office*), и именно этот ресурс особенно полезен исследователю. Базы данных сервиса *esp@cenet* содержат патентные документы более восьмидесяти национальных и международных патентных бюро. Объем имеющегося здесь материала и временной охват по разным странам варьируется в широких пределах — от библиографических описаний до факсимильных копий и от считанных годов до десятков лет.

Не умаляя достоинства *esp@cenet*, отметим, что некоторые элементы структуры, интерфейса и принципов действия поисковой программы не являются интуитивно понятными для начинающего пользователя. Учитывая информационную ценность данного источника, мы знакомим младшекурсников с ним, однако детальное его изучение проводится в спецкурсе на стадии подготовки к дипломной работе. *Esp@cenet* — основной объект *второго* уровня патентного поискового практикума.

На центральном сервере *esp@cenet* размещены три независимых базы данных

(EP, WIPO, Worldwide); предметом нашего рассмотрения является универсальная *Worldwide*. Из четырех поисковых бланков, имеющих на сайте в настоящее время, мы рекомендуем к использованию *Advanced Search* как самый удобный: на нем в явной форме указаны поля, доступные для поиска, и образцы формулирования соответствующих частей запроса. Каждую стадию поисковой работы сервер сопровождает контекстными подсказками (столбец *Quick Help* у левого края окна); к сожалению, этот полезный материал лишь частично включен в автономное *Руководство пользователя (Help)*.

Объем знаний студента-старшекурсника позволяет сместить акценты обучения с технических аспектов на научные. Направление поиска привязывается к теме конкретной дипломной работы; формулирование текста запроса предваряется анализом вариантов и нюансов терминологии, используемой авторами разных научных школ в журнальных публикациях. В извлеченных патентах мы детально рассматриваем содержимое раздела *Description*, причем внимание учащихся концентрируем на тех фрагментах, где особенно явно проявляются достоинства патентной литературы (детализация методик, варьирование параметров процессов и т. п.).

База данных *Worldwide* и поисковая программа имеют специфические черты, незнание которых может привести к неудаче в обнаружении информации:

- Поиск ведется не по всему тексту патента, а только по основным полям библиографического описания и по реферату.
- База данных многоязычна, и использование только английских терминов в запросе не гарантирует полноту извлечения информации. Текстовый поиск возможен только с использованием латинского алфавита; в частности, кириллические документы обнаруживаются косвенным путем или по классификационным кодам.
- Поиск по полю *Title* может оказаться безуспешным из-за того, что только у половины документов это поле заполнено, а текст названия на английском языке имеется у трети документов.
- В большинстве полей бланка оператором по умолчанию является *AND*, но в полях *Publication*, *Application*, *Priority numbers* оператор по умолчанию — *OR*; в поле *Publication Date* операторы вовсе не допускаются.
- База данных может содержать несколько идентичных (или практически идентичных) документов, например, если изобретатель патентует одну и ту же работу в разных странах. Поисковая программа выводит на экран один из аналогов, остальные вызываются по гиперссылкам. Список патентов-аналогов, для которых имеются факсимильные копии, размещается на закладке библиографического описания в группе *Also published as*.

Иные особенности поведения программы не столь критичны, поэтому обязательному запоминанию не подлежат; сведения о них (например, справку о синтаксисе запроса) студент может почерпнуть из онлайн-методического пособия.

Поиск по ключевым словам, по нескольким причинам, не всегда приводит к оптимальному результату; альтернативная методика предполагает использование классификационных кодов в запросе. Студенты знакомятся с основами патентной классификации, а также с сайтами, на которых находятся инструменты соотнесения тематики документа и его места в международной классификационной системе (*IPC*).

Для закрепления полученных знаний выполняется узконаправленный

тематический поиск по комбинированному запросу, состоящему из ключевых слов и классификационных кодов.

Esp@cenet — это информационный комплекс, в который входят не только сервисы, размещенные на центральном сервере, но и патентные службы европейских стран, членов *EPO*. Центральная база данных *Worldwide* не является механической суммой национальных, и для решения некоторых поисковых задач национальная база данных оказывается предпочтительной. В частности, сайт *Российского патентного ведомства на esp@cenet* (<http://ru.espacenet.com/>) имеет русскоязычный интерфейс и позволяет использовать в запросе термины на русском языке для поиска патентных документов Российской Федерации (только в базе данных *RU—esp@cenet*).

В Интернете присутствует заметное число бесплатных патентных ресурсов, потенциально полезных пользователю. Примеры, иллюстрирующие разные типы источников, мы обсуждаем на *третьем* уровне обучения патентному поиску.

Google Patents (<http://www.google.com/patents>) — образец использования технологии *Google* в области патентной информации. В настоящее время эта база данных содержит только патенты и патентные заявки США и в значительной степени дублирует сервисы *USPTO*. Преимущества *Google Patents* заключаются в удобном интерфейсе, возможности ведения текстового поиска по старым патентам и получении факсимильной копии патента не постранично, а единым файлом. Основные недостатки — ограниченные функции поискового бланка и наличие ошибок, возникших при оптическом распознавании отсканированных документов.

SurfIP (<http://www.surfip.gov.sg/>) — пример метапоисковой системы. Ресурс интересен как источник патентной информации стран Юго-Восточной Азии.

PAJ (<http://www19.ipdl.inpit.go.jp/PA1/cgi-bin/PA1INIT>) — патентная база данных Японии; основная ее особенность заключается в функции машинного перевода текста патента с японского на английский язык.

Роспатент (<http://www.fips.ru/russite>) — пакет российских патентных баз данных. В практикуме используется для сравнительного анализа ресурсов нескольких близких по содержанию источников (совместно с *esp@cenet* и *RU—esp@cenet*).

SureChem (<http://www.surechem.org/>) — инструмент для обнаружения патентов, содержащих сведения о химическом веществе, по структурной формуле.

В связи с динамичностью Интернета, список характерных ресурсов мы обновляем ежегодно.

В заключение подчеркнем, что занятия по поиску патентной информации носят вспомогательный характер; их цель состоит в ознакомлении с рабочими инструментами, необходимыми для активного использования в курсах химических дисциплин.

Литература

1. M. Bregonje. Patents: A unique source for scientific technical information in chemistry related industry? - World Patent Information, 2005, 27, No.4, pp. 309–315.
2. M. Blackman. WIPO News. - World Patent Information, 2009, 31, No.1, pp. 75–77.
3. USPTO. Patent Full-Text and Full-Page Image Databases - <http://patft.uspto.gov/>.
4. А. А. Рагойша. Алфавит веб-поиска для химиков. - <http://www.abc.chemistry.bsu.by/>.
5. European Patent Office. esp@cenet. - <http://ep.espacenet.com/>.