

УДК 004

Создание и использование базы данных патентов по микро и нанотехнологиям

09, сентябрь 2012

Тихомиров Д. В.

*Научный руководитель : к.т.н, доцент, Резчикова Е.В.
МГТУ им. Н. Э. Баумана, Москва, Россия*

МГТУ им. Н.Э. Баумана
bauman@bmstu.ru

Введение

Актуальная задача заключается в учете технологических тенденций не только для избегания ненужных инвестиций, но и для получения источника для технологического развития.

Так, многие методы были разработаны для учета достижений технологического прогресса, и одним из них является анализ патентной информации. Тем не менее, не для специалистов трудно анализировать патентную информацию, потому что источники патентной информации содержат большое количество специальной технической терминологии. Следовательно, патентная информация должна быть преобразована во что-то более простое и доступное для понимания.

Методы визуализации считаются пригодными для представления патентной информации и результатов ее анализа. Как правило, различные методы визуализации - это один из лучших способов анализа и усвоения данных, потому что методы графического представления часто дают лучшие результаты по сравнению с другими обычными методами.

Это особенно важно для руководителей, которые принимают решения для инвестирования технологии. Методы визуализации более обозримы и интуитивно понятны по сравнению с обычными способами, такими как текстовый, табличный, а также в виде списка.

Эти методы визуализации для патентного анализа в целом называются патентными картами. Патентные карты - это визуализированное выражение общего числа результатов патентного анализа, созданные для восприятия общей патентной информации легко и эффективно.

Как правило, патентный документ содержит множество элементов для анализа, которые могут классифицироваться в виде структурированных и неструктурированных групп элементов. Структурированные элементы - одинаковые по смыслу и формату во всем патентном документе, такие как номер патента, дата подачи, страна, приоритет и т.д..

С другой стороны, неструктурированные элементы – это произвольные и различные по длине и содержанию наборы символов в патентном документе, такие как названия, аннотации, формулы или описания изобретений.

Таким образом, современные методы визуализации для патентного анализа основаны на составлении карт патентной информации с результатами их анализа.

Однако в настоящее время патентные карты имеют определенные недостатки. Во-первых, большинство из них построены на основе времени, на основе рейтинга или матриц, которые содержат лишь один аспект между структурированными элементами или неструктурированными элементами каждого патентного документа.

Очевидно, что системный и сбалансированный подход должен предоставлять собой общую структуру патентной информации более эффективно. Во-вторых, карты являются сложными комплексами патентных документов, хотя используют разные методы.

Поэтому при использовании они делают результаты патентного анализа непонятными и неясными для анализатора.

Следовательно, эти виды патентных карт не дают достаточного интуитивного понимания в выбранной области технологии. И это третий недостаток, который можно отметить. Следовательно, для развивающихся технологий, таких как микро и нанотехнологии, основным становится выявление их преимуществ и оценка развития в будущем.

Таким образом, предлагается альтернативный метод визуализации для патентного анализа, для того, чтобы преодолеть недостатки текущих патентных карт. Предлагаемый подход способствует формированию патентных карт, которые содержат как структурированные, так и неструктурированные элементы патентного документа. Это достигается при помощи алгоритма интеллектуального поиска по базе данных патентной информации, позволяющего осуществлять поиск минимум по шести параметрам, которые вносятся в патентные документы, с дальнейшим отображением при помощи веб-интерфейса. Кроме того, в отличие от ранее описанных методов визуализации патентного анализа, предлагаемый метод основан на формировании списка схожих патентов при помощи ключевых слов в алгоритме поиска из патентных документов. И тем самым он дает четкое представление о патентной информации в более доступной форме.

И в результате этого анализа, он позволяет получить представление о современных технологиях при прогнозировании тенденции высоких технологий в будущем.

Обзор существующих методов анализа

Непосредственно переходя к исследованиям, с использованием патентной информации, можно выделить два главных направления. Одним из них является исследование способов визуализации патентной информации и ее аналитических результатов.

В данной работе уделено внимание рассмотрению и формированию области для исследования, использования патентной информации из существующих патентов.

На основе знаний о патентах, Японское патентное бюро занимается производством и обеспечивая более 200 видов патентных карт для нескольких технологий, начиная с 1997 года (III, 2000). Кроме того, во многих других странах, таких как Корея (Ryoo & Kim, 2005), Италии (Camus & Brancalion, 2003; Фаттори, Pedrazzi и Turra, 2003) и США (Моррис, DeYong, Ву, Салман, и Yemenu, 2002) также предоставляют различные виды патентных карты.

В настоящее время, большинство видов патентного анализа используют патентные цитаты для представления наиболее значимых отношений в патентной информации. Но известно, что патентное цитирование имеет несколько серьезных недостатков. По словам Yoon и Park (2004) , есть четыре основных недостатка патентного цитирования.

Во-первых, трудно охватить общие отношения между патентными документами. Во-вторых, связанные с первым пунктом проблемы, возможности анализа и информативность данных ограничены. В-третьих, цитата не создает возможности учитывать внутренние связи между патентными документами. Наконец, патентное цитирование - это трудоемкая задача, которая использует выборочный поиск по патентному документу.

В связи с этим, Yoon & Park (2004) предложили альтернативный метод патентного анализа, называемый сетевым. Было проведено огромное количество исследований для поиска оптимального метода патентного анализа. Были разработаны нейро-сетевые методы для анализа научно-технической информации (статьи, патенты), а также оказания помощи пользователю в выполнении сложного процесса анализа большого количества подобной информации фирмами Lamirel, Shehabi, Hoffmann, and Francois (2002). На основе текстовых методов анализа Tseng, Wang, Juang, and Lin (2005) создали реальную патентную карту для важной области технологии: углеродных нано-трубок.

Другое основное направление - это патентная классификация. В Black and Ciccolo (2004), была разработана технология машинной классификации текстов, позволяющая автоматически различать патенты, касающиеся биотехнологий и те, которые к ним не относятся. Fall, ToËcsvar ri, Fiervet, and Karetka (2004) представили результаты применения различных алгоритмов обучения компьютеров для экспертных систем при выполнении задачи патентной классификации на Немецком языке.

Методология

Предлагаемый метод визуализации патентной информации сводится к следующему. При помощи параметров запроса (№ патента, ключевое слово, индекс МПК, авторы, фирмы, страна, дата регистрации) для выбранной технологической области, производится кластеризация патентных документов благодаря алгоритму интеллектуального поиска. Далее следует процесс обработки параметров запроса при помощи алгоритма поиска по БД патентной информации. Затем происходит формирование дерева патентных документов в базе данных, содержание которых удовлетворяет параметрам запроса. Последовательность выполняемых операций представлена на Рис. 1.

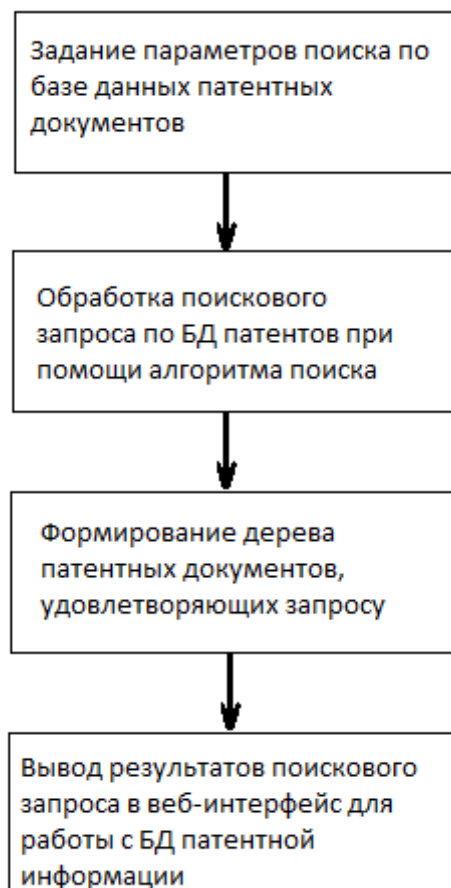


Рис. 1 - Последовательность операций поиска

Построенное дерево, содержащее результаты запроса, выводится в веб интерфейс, разработанный для удобства работы с БД патентной информации.

Параметры поиска

Рассматриваемый метод визуализации патентной информации начинается с введения пользователем параметров поиска по БД патентной информации. Разрабатываемый алгоритм позволяет производить поиск по БД патентной информации минимум по 7 параметрам:

- Нумерационный – номер зарегистрированного патента,
- Предметный – ключевое слово(а), задаваемое(ые) для поиска в патентном документе,
- МПК – индекс международной патентной классификации,
- Авторский – автор(ы) и/или заявители зарегистрированного патента,
- Фирменный – фирма-обладатель патента,
- По странам – страна, где зарегистрирован патент,
- Временной – дата регистрации патента (приоритет).

В результате введения параметров поиска, они передаются для обработки в алгоритм интеллектуального поиска по БД патентной информации.

Обработка запроса алгоритмом поиска

Для наиболее точного поиска по БД патентной документации следует ввести как можно больше параметров поиска. Работа алгоритма поиска по БД патентной информации представлена на рис.2.

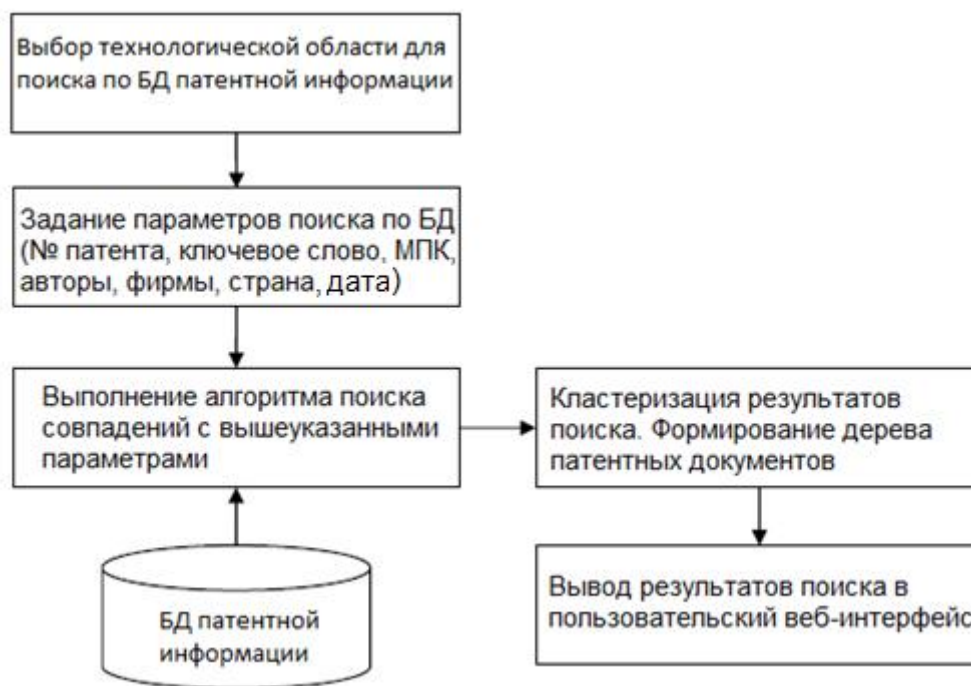


Рис. 2 – Работа алгоритма поиска

Вначале, выбирается технологическая область, в которой необходимо произвести поиск патентной документации. Затем задаются параметры поиска запроса (№ патента, ключевое слово, индекс МПК, авторы, фирмы, страна), чем больше, тем точнее будет поиск. Далее производится поиск по документам, содержащимся в БД патентной информации, осуществляется выборка патентной документации, удовлетворяющей параметрам поиска. После производится кластеризация патентных документов, т.е. автоматическое выявление групп семантически похожих документов среди заданного фиксированного множества патентных документов. Последним шагом является вывод дерева патентных документов, удовлетворяющих параметрам поиска в пользовательский веб-интерфейс.

Формирование дерева патентных документов

На следующем шаге для кластеризованных патентных документов формируется дерево, которое позволяет видеть непосредственную группировку патентных документов по двум параметрам: ключевые слова, дата регистрации патента (см.Рис.3).

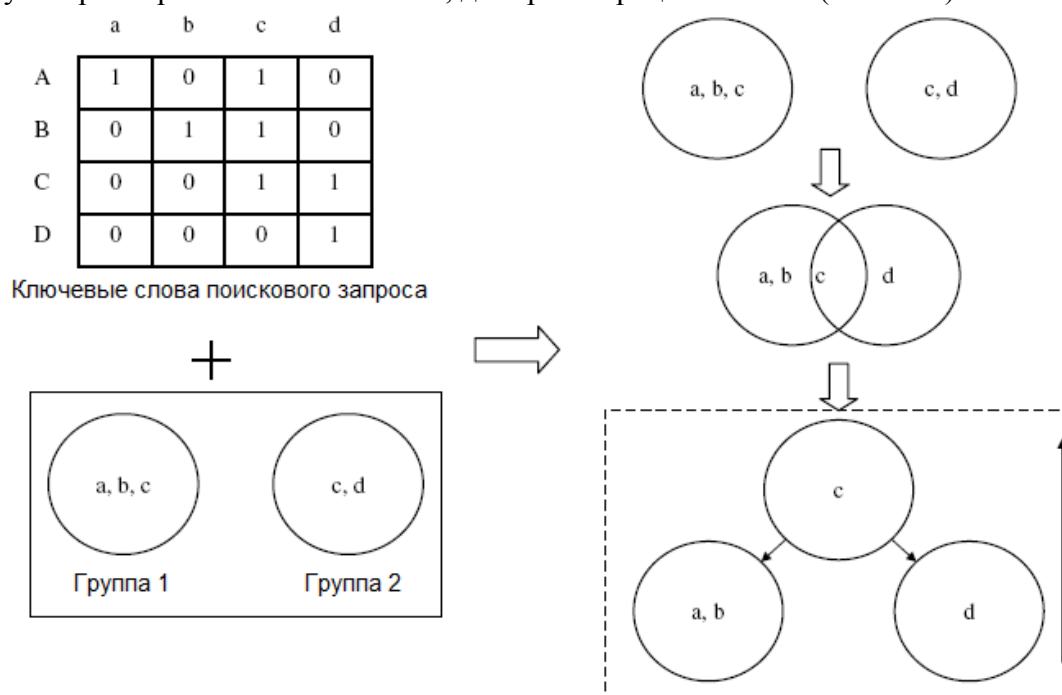


Рис. 3 – Формирование дерева патентных документов по ключевым словам

Например, предположим, что патентные документы "А" и "В" относятся к группе 1. В соответствии с матрицей в рис.3 в патентном документе "А" есть ключевые слова «а» и «с». А патентный документ "В" содержит ключевые слова «b» и «с». Затем для 1-й группы имеем набор из трех ключевых слов 'a', 'b' и 'c'.

Таким образом, исследуются ключевые слова для каждой группы и, используя список ключевых слов для каждой группы, объединяются патентные документы по признаку наличия ключевых слов. Соответственно, для 1-й группы получают ключевые слова "a", "b" и "c". С другой стороны, для 2-й группы получают ключевые слова "c" и "d". Далее видно, что только одна группа имеет ключевое слово "b" и, следовательно, отношения между двумя группами можно представить в виде трех узлов: (a, b), (c) и (d). Здесь общий узел выше, чем другие, поэтому стрелки связи идут от (c) к (a, b) и (d) – рис.4.

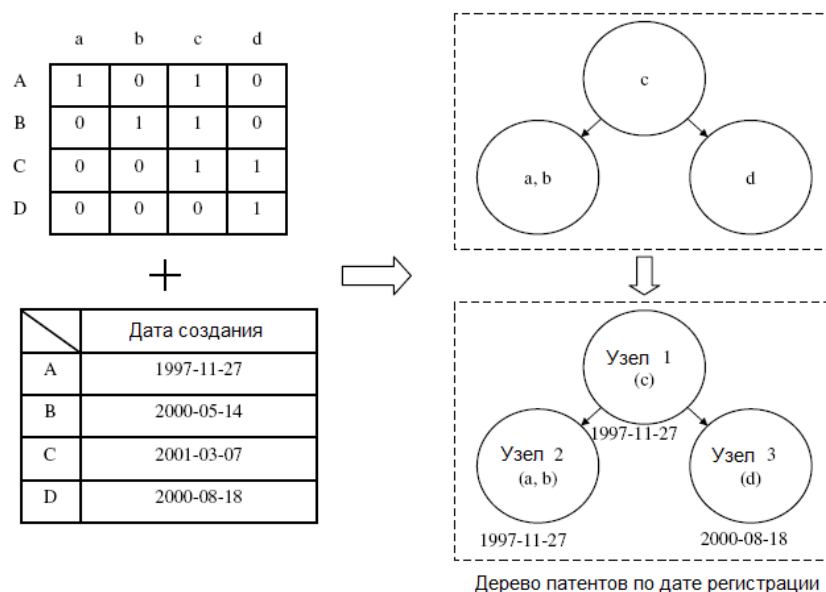


Рис. 4 – Формирование дерева патентных документов по дате регистрации

Изменяя количество и значение ключевых слов, можно легко изменять вид дерева патентной документации. В результате получают дерево с n-узлами.

Далее для каждого узла проверяются даты регистрации патентных документов. Дата регистрации для каждого узла – это наиболее ранняя дата регистрации патентного документа, которые имеют ключевые слова данного узла. На рис.4 представлено формирование дерева патентных документов по дате регистрации.

Предположим, что сформировано дерево патентных документов, как показано на рис.3. Необходимо рассмотреть дату регистрации патентных документов для каждого узла в дереве. Дата регистрации для каждого узла – это наиболее ранняя дата регистрации патентного документа, которые имеют ключевые слова данного узла. Например, на рис 4. узел 2 состоит из ключевых слов "a" и "b". И если "a" относится к патентному документу «А», а "b", в свою очередь, принадлежит патентному документу "В", то датой регистрации для данного узла будет наиболее ранняя дата регистрации патентных документов среди документов "А" и "В". Таким образом, дата регистрации для узла 2 является '1997-11-27'. Кроме того, датой регистрации для узла 3 является '2000-08-18'.

Таким образом, добавляют даты регистрации патентных документов для каждого узла в дерево, сформированное при помощи ключевых слов. При помощи этих двух шагов, сформированное дерево имеет как структурированные элементы – даты регистрации, так и неструктурированные элементы - ключевые слова.

Отображение результатов поиска

Вывод сформированного дерева патентных документов, удовлетворяющих заданным параметрам поиска, производится в виде отображения данных в пользовательском интерфейсе, который позволит легко получать необходимые сведения о выбранном патентном документе, производить просмотр выбранных документы.

Заключение

В данной работе предложен альтернативный метод визуализации для патентного анализа, лишенный недостатков существующих методов формирования патентных карт. По сравнению с другими методами, описанными в разделе 2, разработанный подход позволяет работать как со структурированными элементами, так и с неструктурированными элементами патентных документов. Таким образом, можно

получить четкое представление о патентной информации в наиболее доступной форме. С помощью пользовательского интерфейса, можно определить, какие виды патентов имелись ранее, какие появились недавно, и как эти патенты были объединены и дифференцированы с течением времени.

Кроме того, предлагаемый способ визуализации патентной информации позволяет получить полное представление о новых технологиях. Также он позволяет получить достаточное представление о существующих патентах в выбранной технологической области, для избегания ненужных инвестиций и достоверного поиска информации для патентования новых технических решений.

Литература

- 1 Y. G. Kim, J. H. Suh, S.C. Park. Visualization of patent analysis for emerging technology. Elsevier, 2008, 9