

УДК 004.65:621.921.34

**В. Н. Кулаковский, В. Н. Колодницкий\*, В. В. Цегельнюк,  
И. В. Скворцов**

Институт сверхтвердых материалов им. В. Н. Бакуля  
НАН Украины, г. Киев, Украина

\*stmj@ism.kiev.ua

### **Об одном подходе к построению информационной системы управления научными знаниями в предметной области “Сверхтвердые материалы”**

*Рассмотрен подход на основе эвристического метода распознавания образов к построению информационной системы управления научными знаниями, представленными в виде электронных документов различных типов. Описан алгоритм автоматической классификации документов по научным направлениям в предметной области “Сверхтвердые материалы”.*

**Ключевые слова:** сверхтвердые материалы, информационная система, автоматическая классификация, электронные документы.

За более чем полувековой период своей деятельности Институт сверхтвердых материалов им. В. Н. Бакуля НАН Украины достиг важных результатов в области создания сверхтвердых материалов (СТМ), инструментов и изделий функционального назначения на основе СТМ, в разработке эффективных технологий их производства и внедрения в различных областях применения [1].

Сравнительно новым и развивающимся направлением деятельности ИСМ НАН Украины является информатизация научных исследований и применение информационных технологий в создании перспективных и совершенствовании существующих направлений в предметной области “Сверхтвердые материалы” (ПРО “СТМ”) [2, 3]. Для решения подобных задач в сфере информационных технологий осуществляется переход от технологий, ориентированных на исследование количественной стороны информационных процессов, к технологиям, основанным на знаниях как качественных составляющих этих процессов [4]. Для этого необходимо базы данных как совокупность самостоятельных слабо систематизированных научных данных, представленных в виде электронных копий отчетов по темам, текстов диссертаций, патентов и статей, преобразовать в базы знаний, которые содержат структурированную информацию, покрывающую всю область знаний о сверхтвердых материалах, создать иерархическое представление набора понятий и отношений – построить онтологию, и, таким образом, образовать единую знаниеориентированную информационную систему (ЗОИС) [5].

Построение данной системы осуществляется условно в три этапа:

I. Создание классификатора научных направлений, к которым могут относиться те или иные электронные документы.

II. Разработка программного обеспечения для хранения и автоматической

классификации документов (допускающей изменение класса документа в ручном режиме).

III. Заполнение базы данных электронными документами (в том числе и оцифровка бумажных оригиналов).

В рамках первых двух этапов разработана двухуровневая структура базы данных онтологии научной деятельности ИСМ НАН Украины [6], которая состоит из научных направлений, разделов и подразделов. Первый уровень структуры включает главные направления научно-исследовательской деятельности института, утвержденные Президиумом НАН Украины, второй – научные направления непосредственно отделов института.

На базе двухуровневой онтологии научной деятельности ИСМ НАН Украины разработана трехуровневая структура базы знаний и онтологий предметной области “Сверхтвердые материалы”. С целью классификации документов подобраны ключевые слова, которые наиболее корректно характеризуют направления научной деятельности института в целом и научных подразделений в частности. Создан перечень основных научных направлений, их подразделов и научных задач для каждого отдела и лаборатории института.

В результате все научные работы, проводимые в ИСМ НАН Украины, классифицированы по трем направлениям:

1. Изучение влияния высоких давлений на материалы. Технологическое использование высоких давлений в производственных процессах.

2. Исследование физико-химических процессов получения СТМ. Создание новых технологий получения функционально ориентированных материалов и изделий из них.

3. Развитие научных основ создания новейших высоких технологий обработки металлов и неметаллов инструментом из СТМ, разработка методов и технологий применения функционально ориентированных материалов.

Научные направления делятся на подразделы (от 5 до 15, в зависимости от направления, суммарно 30), каждый из них, в свою очередь, состоит из научных задач. Подраздел может состоять из одной или нескольких задач (до 10, суммарно перечислено 111 научных задач, каждая из которых описана характеризующими ее ключевыми словами и словосочетаниями).

Следующим этапом в построении ЗОИС является разработка программно-обеспечения для хранения и автоматической классификации документов.

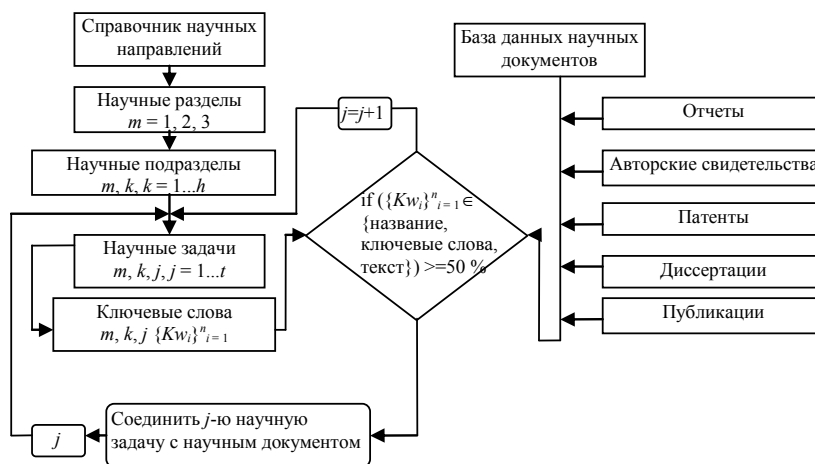
Основой программного обеспечения служит система управления базами данных, которая управляет всеми процессами системы и поддерживает гибкий программный интерфейс между базами. В структуру базы ЗОИС входят основная база и вспомогательные реляционные базы. Основная база знаний ориентирована на размещение информации научно-технического архива института о научных публикациях сотрудников ИСМ НАН Украины, научно-технических отчетов, патентов, авторских свидетельств, диссертаций. Программа управления базами знаний предоставляет программный сервис не только для быстрого внесения необходимой информации в основную базу знаний, но и для создания связи между ней и внешними файлами, содержащими непосредственно сами тексты научных документов, фотографии, рисунки, схемы, таблицы.

Главную роль в данной системе играет справочник научных направлений деятельности института, который состоит из трех уровней: научных разделов, научных подразделов и научных задач (направлений). Каждое научное направление определяется набором ключевых слов, которые характеризуют это научное направление, и, соответственно, относящийся к нему раздел и под-

раздел. Справочник научных направлений представлен в программной системе в виде трех взаимосвязанных между собой баз данных. В настоящей ЗОИС реализована программная модель классификации электронных документов на основе теории распознавания образов [7].

Образами в системе ЗОИС выступают перечисленные выше виды научных документов, которые описываются названием, авторами, ключевыми словами, рефератами, аннотациями и другими характеристиками. Любое научное направление справочника научных направлений, которое характеризуется общими свойствами для всех его элементов, определено как класс образов.

В ЗОИС программно реализован эвристический метод распознавания образов, который основан на сравнении неизвестных образов с эталонами. Точнее, если какой-либо входной образ наиболее соответствует эталону  $i$ -го класса образов, чем какому-либо другому эталону иного класса, то входной образ классифицируется как принадлежащий к  $i$ -му классу образов. В системе ЗОИС осуществлена программная реализация сопоставления подобия. Критерием выступает вхождение 50 % и более набора ключевых слов (эталона) каждого научного направления (класса) в описание характеристик научных документов (образов). Если условия критерия выполняются, то программная система организует ссылочную связь или привязку данного научного документа к выбранному научному направлению. При этом научный документ может относиться к нескольким научным направлениям. Блок-схема программы, реализующей этот алгоритм, представлена на рисунке.



Блок-схема реализации эвристического метода распознавания образов.

Практически же программная система для каждой записи в основной базе данных, содержащей все сведения о научном документе, открывает в соответствующей дочерней реляционной базе записи с информацией о том или ином научном направлении, к которому относится данный научный документ. Так организуется прямая ссылочная связь между документом и научными направлениями из справочника научных направлений. При этом в дочернюю реляционную базу записывается информация о данном научном документе. Такая программная взаимосвязь между справочником научных направлений и основной базой научных документов посредством промежуточной базы данных реализована в программной системе ЗОИС для удобства осуществления поисковых операций в данной системе. Если в дальнейшем осуществлять поиск по справочнику научных направлений, то программный

интерфейс будет давать возможность пользователю сразу находить научные документы, относящиеся к тому или иному научному направлению, и иметь доступ к ним. В случае невыполнения условия критерия, отнесение документа к какому-либо классу выполняется вручную. При этом набор ключевых слов задачи пополняется ключевыми словами документа.

Таким образом, осуществлена структуризация научных знаний в предметной области “Сверхтвердые материалы”, предложен и реализован на основе эвристического метода распознавания образов механизм классификации научной информации по научным направлениям деятельности ИСМ НАН Украины.

*Rozглянуто підхід на основі евристичного методу розпізнання образів до побудови інформаційної системи управління науковими знаннями, представленими у вигляді електронних документів різних типів. Описано алгоритм автоматичної класифікації документів за науковими напрямками в предметній області “Надтверді матеріали”.*

**Ключові слова:** надтверді матеріали, інформаційна система, автоматична класифікація, електронні документи.

*It is reported on the approach based on heuristic of pattern recognition method to building information management system of scientific knowledge presented in the form of electronic documents of various types. An algorithm for automatic classification of documents on scientific fields in the subject area “Superhard materials”.*

**Keywords:** superhard materials, information system, automatic classification, electronic documents.

1. Новиков Н. В. О достижениях Института сверхтвердых материалов им. В. Н. Бакуля НАН Украины в области синтеза и спекания сверхтвердых материалов за 50 лет деятельности // Сверхтв. материалы. – 2011. – № 3. – С. 3–8.
2. Лебедева А. А., Скворцов И. В., Фидаров Т. З. Формы представления знаний и алгоритмы принятия решений в информационных экспертных системах в производстве и применении сверхтвердых материалов и инструментов из них // Эффективность реализации научного ресурсного и промышленного потенциала в современных условиях: Материалы Седьмой ежегодной междунар. пром. конф., 12–16 февр. 2007 г., п. Славское, Карпаты, Украина. – К.: УИЦ «Наука. Техника. Технология», 2007. – С. 125–128.
3. Лебедева А. А., Фидаров Т. З., Скворцов И. В. Подход к построению предметной онтологии для экспертной системы “Сверхтвердые материалы” // Породоразрушающий и металлообрабатывающий инструмент – техника и технология его изготовления и применения: Сб. науч. тр. – К.: ИСМ им. В. Н. Бакуля НАН Украины, 2007. – Вып. 10. – С. 370–374.
4. Ландэ Д. В. Поиск знаний в Internet. Профессиональная работа: Пер. с англ. – М.: Издательский дом “Вильямс”, 2005. – 272 с.
5. Кулаковский В. Н., Колодницкий В. Н., Чистяков Е. М., Цегельнюк В. В. Тезаурус как основа построения онтолого-тезаурусной системы управления знаниями в предметной области “Сверхтвердые материалы” // Искусственный интеллект. Интеллектуальные системы (ИИ-2013): Тез. докл. XIII Междунар. науч.-техн. конф., 23–27 сент. 2013 г., пос. Кацивели (Большая Ялта), АР Крым, Украина. – 2013. – С. 223–226.
6. Лебедева А. А., Гордашник К. З., Колодницкий В. Н. и др. Представление и аналитическая обработка данных в предметной области “СТМ” на основе онтолого-тезаурусного анализа // Породоразрушающий и металлообрабатывающий инструмент – техника и технология его изготовления и применения: Сб. науч. тр. – Вып. 17. – К.: ИСМ им. В. Н. Бакуля НАН Украины, 2014. – С. 370–379.
7. Ванник В. Н., Червоненкис А. Я. Теория распознавания образов. – М.: Наука, 1974. – 416 с.

Поступило 12.09.16