

Рисунок 12- Электронное изображение арсенида галлия при измененных сферической (слева) и хроматической (справа) абберациях

3 Анализ результатов

Проанализировав полученные изображения, заметим, что такие параметры как сферическая и хроматическая абберации, а также напряжение и дефокусировка влияют на качество изображения, но не влияют на его общую структуру. Для повышения качества изображения необходимо подобрать оптимальные значения параметров. Для этого установим зависимость максимальной интенсивности от выбранных параметров (рисунок 4).

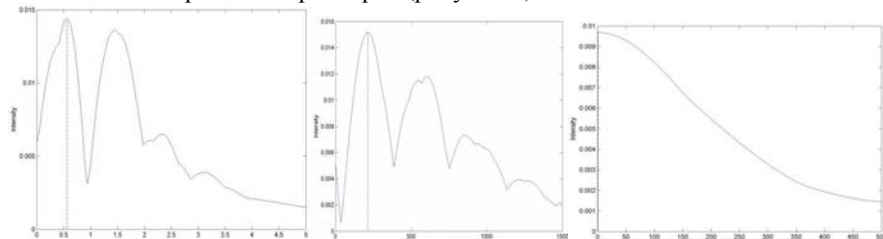


Рисунок 13- Графики зависимости максимальной интенсивности от параметров микроскопа

Оптимальными значениями будем считать сферическую абберацию 0,6 мм, дефокусировку 220 ангстрем и хроматическую абберацию 50 ангстрем. На рисунке 5а представлено натурное изображений полученное в электронном микроскопе высокого разрешения, на рисунке 5б соответствующее модельное изображение при оптимальных значениях параметров. Можно отметить высокую степень похожести.

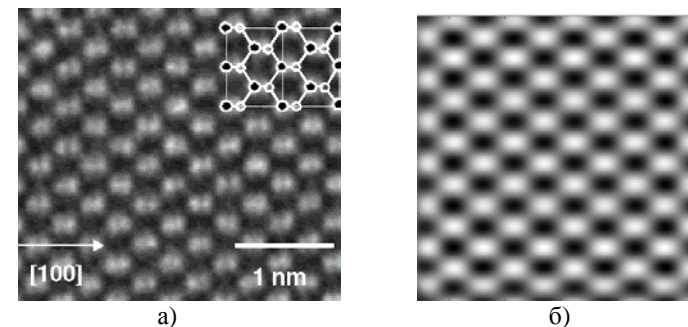


Рисунок 14- Изображение арсенида галлия в электронном микроскопе а) натурное изображение б) результаты моделирования

Заключение

В результате реализации метода многих слоев были получены изображения кристаллической решетки при различных параметрах микроскопа. Полученные изображения вполне точно отображают структуру рассмотренной кристаллической решетки.

Также были определены параметры, при которых изображение имеет наиболее высокое качество. Метод многих слоев для компьютерного моделирования является наиболее удобным методом.

Литература

1. Кларк, Э. Р. Микроскопические методы исследования материалов / Э. Р. Кларк, К. Н. Эберхардт. – М.: Техносфера, 2007. – 367 с.
2. Куприянов, А.В. О наблюдаемости кристаллических решеток по изображениям их проекций [Текст] / А.В. Куприянов, В.А. Сойфер // Компьютерная оптика. – 2012. – Т. 36, № 2. – С. 249-256.
3. P. Stadelmann Image Calculation Techniques, Lausanne, Switzerland, 1993. – P. 1-27.
4. Earl J. Kirckland Advanced computing in electron microscopy, Cornell University Ithaca, N.Y. 2006 P. 77-93, P. 115-198

Д.Г. Шамиданов, А.А. Бармин, М.А. Шилина

ПРАВИЛА ФОРМИРОВАНИЯ УНИВЕРСАЛЬНОЙ ФОРМАЛЬНОЙ МОДЕЛИ КОНТЕНТА ИНФОРМАЦИОННОГО ПРОСТРАНСТВА

(Уфимский государственный авиационный технический университет)

В статье рассмотрены проблемы построения универсальной формальной модели доступа к данным. Рассматриваются правила формирования модели контента со структурированными и неструктурированными данными.



Введение

В процессе функционирования корпоративных автоматизированных информационных систем накапливается большое количество разнородных, несвязанных между собой данных. Для поддержки принятия решений руководству, зачастую, необходимы объединенные данные всех систем. В целях обеспечения единого и связного доступа к данным, предлагается формирование универсальной формальной модели контента, отличительной особенностью которой является представление в атрибутивной форме метаданных и данных гетерогенных хранилищ, соответствующих правилам, определённым пользователями в аспекте их поиска и применения. Подход к построению данной модели рассматривается с точки зрения формирования правил по структурированию контента с применением известных информационных технологий и поисковых алгоритмов.

Целью работы является выработка подхода к построению универсальной формальной модели доступа к данным различных хранилищ.

Для достижения поставленной цели необходимо выполнение следующей задачи: разработка правил структурирования контента и выработка на их основе подхода к построению формальной модели контента информационного пространства, отражающей многомерный характер данных.

Правила формирования модели контента по структурированным источникам

В качестве примера формирования модели контента по структурированному источнику используется база данных веб-портала кафедры АСУ.

Таблица базы данных в данном случае – это отражение сущности предметной области. Каждая таблица состоит из множества записей. Записи представляют собой строку в таблице, а строка состоит из ячеек. Ячейки, в свою очередь, состоят из названия, типа и значения, что соотносится с понятием *data* – содержимого информационного пространства – в предлагаемой модели контента [1].

Данные предметной области описывают в данном примере работу кафедры ВУЗа по выпуску студентов и защите выпускных квалификационных работ (ВКР).

В рамках предлагаемой модели контента конкретная запись в таблице (на примере таблицы *diploms*) будет определять понятие содержимого информационного пространства *data* следующими элементами:

- типом атрибута (*t*): тип поля в таблице (текстовый);
- наименованием атрибута (*n*): название ячейки – заголовок столбца в таблице (*dipl_name*);
- значением атрибута (*v*): хранимое значение (“Название темы ВКР”).

Данные информационного пространства ($D_i = \{d_j^i\}$) будут определены:

- языком (*lang*): русский;
- кодировкой данных (*enc*): utf-8;
- форматом представления (*format*): текстовый;



- содержимым (*data*): значение записи в таблице базы данных.

Тогда каждый элемент данных информационного пространства предметной области может быть представлен в виде кортежа из четырёх элементов:

$$d_j^i = \langle lang, enc, format, data \rangle,$$

где d_j^i – *i*-й экземпляр данных, *lang* – язык, *enc* – кодировка данных, *format* – формат представления, *data* – содержимое.

Описание данных предметной области при помощи данного представления позволяет утверждать, что данные представляются в атрибутивной форме. После формирования атрибутивной модели, с её помощью можно будет выполнять запросы к информационному пространству на языке, на котором сформирована модель. Например, для задачи по отбору документов за определённый период необходимо сначала преобразовать запрос, сформированный на естественном языке, в запрос, основанный на правилах взаимодействия с хранилищами данных. В результате преобразования данные могут быть получены и, впоследствии, проанализированы.

Формирование по неструктурированным источникам

Примером неструктурированных данных, требующих анализа, может являться файловое хранилище с множеством произвольных текстов на естественном языке. Прежде чем строить модель контента по таким данным, их необходимо сначала структурировать. Перевод неструктурированных текстовых массивов в структурированные возможен с помощью решений, основанных на технологии текстомайнинга [3].

Текстомайнинг позволяет находить новые знания в неструктурированных текстовых массивах при помощи статистического анализа совместного появления слов в текстовых файлах.

Например, используя программу WordStat, неструктурированные данные могут быть классифицированы методом иерархической кластеризации. С помощью данного метода можно построить древовидную структуру: те ключевые слова, которые наиболее часто выпадают друг с другом, связаны короткими линиями; те, которые вместе выпадают редко, – длинными. Программа разбивает массив на кластеры, которые логически или семантически связаны. Метод также позволяет выявить и новые, неочевидные связи, например два фактора, которые часто выпадают с третьим.

Проклассифицировав источники данных с неструктурированной информацией с помощью текстомайнинга, данные можно будет сопоставить с предлагаемой формальной моделью контента.

Файл будет определять понятие содержимого информационного пространства *data*, где элементами будут:

- тип атрибута (*t*): тип файла в хранилище (.txt, .doc, .pdf, .odt);
- наименование атрибута (*n*): название файла;
- значение атрибута (*v*): место файла в иерархической кластеризации.

Данные информационного пространства ($D_i = \{d_j^i\}$) будут определены:

- языком (*lang*): язык текстов файлового хранилища;



- кодировкой данных (*enc*): кодировка файлов;
- форматом представления (*format*): текстовый;
- содержимым (*data*): текст файла.

Классификация источников неструктурированных данных также возможна при помощи полнотекстового индекса, формируемого при использовании технологии полнотекстового поиска. Суть работы алгоритма полнотекстового индекса в том, что из каждого документа информационного пространства извлекается текст, разбивается на слова, и для каждого слова создается отдельная таблица индекса. Появляется связь конкретного слова и документов, в которых это слово встречается [4].

При создании полнотекстового индекса для каждого документа запись, в которой указано, содержит ли он конкретный термин предметной области. В результате возникает бинарная инцидентность «термин-документ», которая может быть представлена в форме матрицы. Теперь в зависимости от порядка просмотра этой матрицы (по строкам или по столбцам) можно получить вектор термина, в котором указано, в каких документах он встречается, либо вектор документа, в котором указано, какие термины в нем употребляются [2].

Заключение

Современные организации являются сложными системами, информационное обеспечение управления которых обеспечивается широким кругом программного обеспечения. Для единообразного доступа к данным в работе предложены правила по формированию универсальной формальной модели доступа к данным. Модель позволит единообразно представить контент как структурированных, так и неструктурированных информационных ресурсов.

В работе рассмотрены правила формирования универсальной модели доступа к данным по структурированным и неструктурированным источникам. На их основе предлагается подход к построению формальной модели контента информационного пространства, отражающей многомерный характер данных.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-37-00064 мол_а.

Литература

1. Куликов Г.Г., Бармин А.А. Business Intelligence – методологический и информационно-технологический аспект // Информационные технологии и системы [Электронный ресурс]: тр. Пятой Междунар. науч. конф., Банное, Россия, 24-28 февр. 2016 г. (ИТиС – 2016) : науч. электрон. изд. С. 243-248.
2. Куликов Г.Г., Шилина М.А., Старцев Г.В., Бармин А.А. Структурирование контента информационного пространства технического университета с использованием процессного подхода и семантической идентификации // Вестник УГАТУ, 2014. № 4 (65). С. 115-124.



3. Александр Беленький. Текстомайнинг. Извлечение информации из неструктурированных текстов [Электронный ресурс] – режим доступа: <http://compress.ru/Article.aspx?id=19605>

4. Younet – Блог о веб-разработке. Полнотекстовый индекс: Full text index [Электронный ресурс] – режим доступа: <http://younet.kz/blog/mysql/polnotekstovyyj-indeks-full-text-index>

Е.Д. Шикина, Е.А. Кагорин

ДИСПЕТЧЕРИЗАЦИЯ ПОТОКА ЗАЯВОК СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ПОМОЩЬЮ ТРАНСПОРТНОЙ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

(Самарский государственный технический университет)

Рассмотрим систему массового обслуживания (СМО) с различными каналами (имеющими, к примеру, разную пропускную способность или отдельные очереди). В отличие от СМО с неразличимыми каналами (имеющими одинаковую пропускную способность и общую очередь) оргграф её состояний G имеет нелинейную структуру, так как число возможных состояний определяется не только суммарным количеством заявок, присутствующих в системе (в очередях и на обслуживании), но также их распределением в очередях отдельных каналов. [1-4]

При этом оргграф G представим линейной цепочки

$$G = G_0 \leftrightarrow G_1 \leftrightarrow G_2 \leftrightarrow \dots \leftrightarrow G_i \leftrightarrow G_{i+1} \leftrightarrow \dots \leftrightarrow G_N,$$

составленной из подграфов G_i , $i=0 \div N$, отвечающих наличию в очередях и каналах суммарно i заявок. Здесь N – максимальное число заявок в системе, равное сумме числа каналов и ёмкостей их очередей.

Тогда каждая соседняя пара $G_i \leftrightarrow G_{i+1}$ образует двудольный оргграф с парными антиколлинеарными дугами, отражающими переходы между состояниями «слева-направо» при получении очередной входной заявки и «справа-налево» при выходе из системы обслуженной заявки.

Например, подграф G_0 состоит из единственного состояния простоя СМО (каналы и их очереди пусты), как и подграф G_N , состоящий из единственного состояния отказа СМО (каналы и их очереди полностью заняты). Прочие подграфы G_i содержат большее число отдельных состояний.

Таким образом, каждый двудольный оргграф $G_{i-1} \leftrightarrow G_i$, $i=1 \div (N-1)$, можно интерпретировать как реализацию двух типов (не)замкнутых транспортных задач линейного программирования:

- 1) $G_{i-1} \rightarrow G_i$ с $|G_{i-1}|$ поставщиками и $|G_i|$ потребителями;
- 2) $G_{i-1} \leftarrow G_i$ с $|G_i|$ поставщиками и $|G_{i-1}|$ потребителями.

В качестве аналога удельных транспортных затрат при транспортировании от поставщика к потребителю выступают интенсивности переходов «слева-