

7. Об утверждении типового паспорта безопасности территорий субъектов Российской Федерации и муниципальных образований: Приказ МЧС России от 25.10.2004 г. №484 (ред. от 11.09.2013).
8. Williams G.J., Simoff S.J. Data Mining: Theory, Methodology, Techniques, and Applications: Springer. 2006. 329 p.
9. Ничепорчук В.В., Пенькова Т.Г. Паспорт территорий – динамический инструмент анализа опасностей // Научные и образовательные проблемы гражданской защиты. 2014. № 1. С. 3-8.
10. Gorban A., Pitenko A., Zinovyev A. ViDaExpert: user-friendly tool for nonlinear visualization and analysis of multidimensional vectorial data: Cornell University Library. 2014 URL: <http://arxiv.org/abs/1406.5550>.
11. Using ArcViewGIS: The Geographic Information System of Everyone: ESRI Press. – 1999. 350 p.
12. Abdi H., Williams L. Principal Components Analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010. N. 2(4). PP. 439-459.
13. Peres-Neto P., Jackson D., Somers K. How many principal components? Stopping rules for determining the number of non-trivial axes revisited // Computational Statistics & Data Analysis. – 2005. N. 49(4). PP. 974-997.
14. Jain A., Dubes R. Algorithms for Clustering Data. Michigan State University: Prentice Hall, 1988. 320 p.

Analysis of natural and technogenic safety based on principal component analysis and cluster analysis techniques (in case of Krasnoyarsk region)

Penkova Tatiana Gennadyevna, PhD, Senior Research Associate

This paper presents a comprehensive analysis of natural and technogenic safety indicators of the Krasnoyarsk region in order to explore geographical variations and patterns in occurrence of emergencies by applying the multidimensional analysis techniques – principal component analysis and cluster analysis – to data of the Territory Safety Passports.

Keywords: data mining, principal component analysis, cluster analysis, natural and technogenic safety

УДК 004.42

**ОБРАБОТКА МНОГОМЕРНЫХ ДАННЫХ
СРЕДСТВАМИ ТЕХНОЛОГИИ ТОФИ**

*Сауле Кайроллиевна Сагнаева, к.ф.-м.н., доцент,
Евразийский национальный университет им.Л.Н.Гумилева
Тел.: +7 705 709 9703, e-mail: sagnaeva_tar@mail.ru,*

*Шынар Ерланкызы, магистрант
Евразийский национальный университет им.Л.Н.Гумилева
Тел.: +77476917300, e-mail: shinara.92@mail.ru
fit.enu.kz*

В статье рассматриваются вопросы анализа и визуализации многомерных данных средствами технологии ТОФИ. Описаны технологические особенности построения хранилища данных ТОФИ, многомерных кубов и срезов кубов. Основное назначение технологии ТОФИ – поддержка аналитической деятельности, произвольных запросов пользователей-аналитиков.

Ключевые слова: OLAP, построение хранилищ данных, оперативная аналитическая обработка данных, технология ТОФИ.

Введение

Технология комплексного многомерного анализа данных получила свое название OLAP (On-Line Analytical Processing) и является ключевым компонентом организации хранилищ данных. Концепция OLAP предложена Эдгаром Коддом в 1993 году, и на основе требований, изложенных Коддом в 1995 году был сформирован тест FASMI (Fast Analysis of Shared Multidimensional Information – быстрый анализ разделяемой многомерной информации). Технология ТОФИ разработана казахстанской компанией «КСИ «Фактор» и предназначена для создания информационно-аналитических систем поддержки принятия и оценки управленческих решений субъектов управления и бизнеса [1]. В работе [2] приводится анализ соответствия многомерного анализа средствами ТОФИ требованиям Кодда к OLAP-средствам. К сожалению, в ТОФИ отсутствуют средства data-mining, что не позволяет отнести ТОФИ к BI-системам, хотя в ТОФИ реализованы ETL-приложения и имеется хранилище данных. ТОФИ предоставляет возможность представления данных в виде многомерных кубов (стандартных и нестандартных), и обработка данных производится собственным инструментом ТОФИ. При этом снижаются затраты на создание моделей и систем преобразования данных из реляционной модели в объектную.



С.К. Сагнаева

Процесс создания хранилища данных в ТОФИ для рассматриваемой предметной области включает в себя два этапа – определение показателей, которые необходимы для многомерного анализа, и построение информационной модели с помощью выявленных показателей. Иными словами, необходимо определить *объекты, переменные и субъекты* оценивания и

тем самым ответить на следующие вопросы:

– Какие элементы рассматриваемой предметной области требуют оценки и, возможно, изменений, направленных на повышение их качества? (*Объекты оценивания*).

– Какие именно параметры (характеристики) этих элементов требуют оценки? (*Переменные оценивания*).

– Какие участники процесса и заинтересованные лица имеют наиболее развернутое представление о качестве рассматриваемых элементов по этим параметрам, а также возможностях повышения их качества? (*Субъекты оценивания*).

Технология ТОФИ предоставляет средства просмотра информации в различных срезах, автоматического получения агрегированных данных, выполнения операций свертки, детализации, сравнения во времени, и это позволяет отнести ее к OLAP-системам.

Общее описание технологии ТОФИ

Наименование ТОФИ формируется на основании первых букв основных сущностей технологии: «Тип объекта», «Отношения между типами объектов», «Факторы», «Измерители». ТОФИ-технология представляет собой как методологию изучения предметной области, так и инструментарий, автоматизирующий предлагаемую методологию [3]. Инструментарий для построения ТОФИ-модели поддерживает как различные возможности моделирования данных и их обработку, так и механизмы взаимодействия пользователя с ними. Описание состояния любой сложной системы в ТОФИ представляет собой описание свойств объектов и отношений между объектами. Описание изучаемой системы (модель) в ТОФИ представляется в виде совокупности объектов, которые определенным образом взаимодействуют между собой и изменяют свои состояния. Обь-



Ш. Ерланкызы

екты в модели существуют постольку, поскольку им отнесены некоторые факторы, измерители, объекты, отношения между объектами и атрибуты для описания их состояния. Модель предметной области, построенная по технологии ТОФИ, называется ТОФИ-моделью. В дальнейшем рассмотрим процесс построения ТОФИ-модели НИР вуза.

Инструменты моделирования ТОФИ

Главная особенность технологии ТОФИ заключается в том, что первичными являются свойства, а объекты определяются как совокупность свойств. Поведение объекта определяется его возможностью изменять свои свойства и свойства других объектов. Эта особенность определяет процесс моделирования состояния сложной системы по технологии ТОФИ следующим образом: на первом этапе моделировании определяются все свойства как абстрактные понятия; на втором этапе с помощью этих свойств создаются группы свойств, приводящие в дальнейшем к дифференциации типов объектов, а также, при необходимости, и отношений между типами; и, наконец, производится объявление типов объектов, самих объектов и отношений между ними.

Для составления динамических моделей в ТОФИ-технологии используются понятие периода: объекты предметной области обычно существуют в некотором временном интервале, с течением времени одни объекты или свойства «умирают», другие – «рождаются». Поэтому модели реальной предметной области являются динамическими.

Для описания состояния сложных систем в технологии ТОФИ используются следующие сущности ТОФИ:

Объект – любая субстанция или сущность рассматриваемой предметной области, которую необходимо рассматривать с заданной точки зрения. Все выделенные объекты группируются в «типы объектов». Объект в технологии ТОФИ выступает и в качестве владельца данных, и в качестве свойства, когда один объект является свойством другого объекта.

Тип объекта – совокупность свойств, с помощью которых описываются состояния конкретных объектов этого типа. Тип объекта группирует конкретные объекты и определяет все их качественные и количественные свойства.

Отношение между типами объектов – совокупность типов объектов, определенным образом связанных между собой и которая имеет такие свойства, как тип объектов.

Отношение между объектами – совокупность объектов и отношений между объектами, которая является экземпляром отношения между типами объектов и состоит из участников-объектов, вступающих в это отношение. Отношение между объектами выступает в технологии ТОФИ и в качестве владельца данных, и в качестве свойства, когда отношение между объектами является свойством другого объекта или отношения между объектами.

Свойство – характеристика объекта или отношения между объектами, фиксирование которой определяет его состояние. Свойства бывают количественные, качественные и прочие.

Фактор – качественное свойство объекта или отношения между объектами предметной области, которое является перечислением возможных состояний.

Измеритель – количественное свойство объекта или отношения между объектами предметной области. Каждый измеритель имеет единицу измерения.

Единицы измерения – мера, с помощью которой измеряются количественные данные. Единицы измерения в свою очередь могут являться свойством. Такая ситуация возможна в случае, когда измеритель меняет свою единицу измерения от одного объекта к другому.

Атрибуты – все те свойства объекта или отношения между объектами, которые не могут трактоваться как факторы, измерители. Примерами атрибутов могут быть фотографии объектов, графические схемы объектов, местонахождение (адрес) объектов и другие.

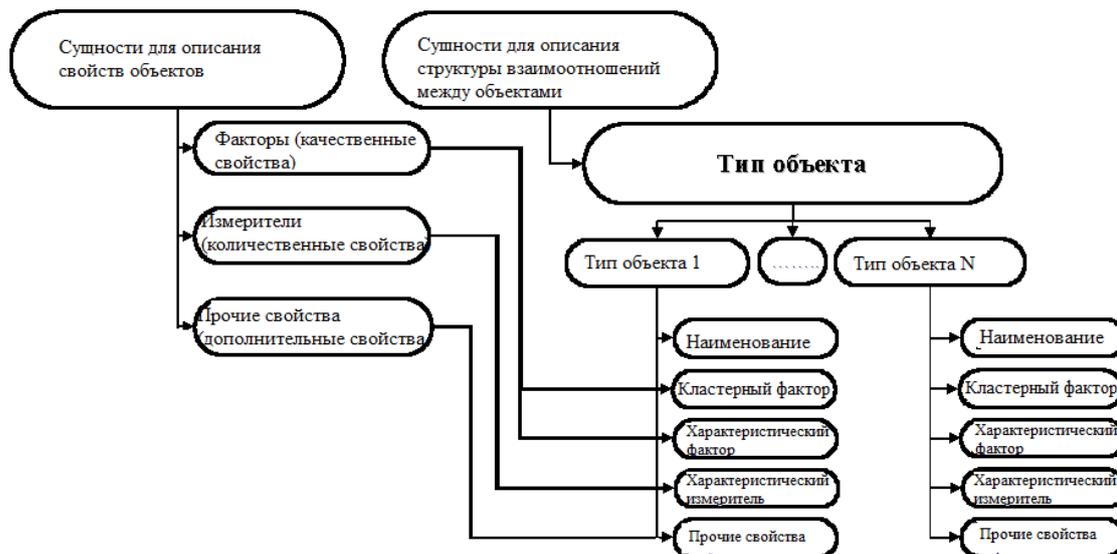


Рис.1. Главные сущности технологии ТОФИ

Стандартные кубы

Основу ТОФИ составляет ее информационное хранилище, которое фактически состоит из двух уровней: реляционного и многомерного. На реляционном описывается физическая структура хранения данных, на многомерном – логическое многомерное представление данных для пользователей системы. Сущность ТОФИ «Стандартный куб» формируется из измерений периодов, измерений свойств и измерений объектов и отношений. Каждой ячейке куба ставятся в соответствие конкретные данные (O, P(s,q,T), где O – объект или отношение между объектами, P – свойство, s – статус данных, q – поставщик данных, T – период или актуальный интервал жизни значения. Все стандартные кубы группируются в группы стандартных кубов. Каждая группа стандартных кубов, так называемая «родительская группа», может содержать другие группы (подгруппы) стандартных кубов и стандартные кубы.

Измерениями стандартного куба являются совокупность измерений периодов, свойств, объектов и отношений. Один стандартный куб может иметь одно или несколько измерений объектов и отношений, одно или несколько измерений свойств, ноль или одно измерение периодов. На элементы измерения стандартного куба могут быть установлены фильтры с помощью дополнительных условий. Для каждого измерения (кроме измерения периодов) может быть указана дата, на которую нужно формировать измерение. Если такая дата не указана, измерение формируется на дату, указанную для куба. Если и для куба не указана, то измерения формируются на последнюю дату последнего периода измерения периодов.

Из измерений стандартного куба формируется сам куб. При этом набор измерений свойств образует прокуб свойств, а набор измерений объектов и отношений формирует прокуб объектов и отношений. Каждой ячейке прокуба свойств (на ребрах, гранях и внутри) ставится в соответствие некоторое свойство с указанием, при необходимости, статуса данных и поставщика данных. Каждой ячейке прокуба объектов и отношений (на ребрах, гранях и внутри) ставится в соответствие некоторое отношение между объектами (на гранях и внутри) или объект или отношение между объектами (на ребрах). Ребрами являются измерения прокуба, а гранями являются гиперкубы размерности больше одного (в том числе и внутренность прокуба).

В ТОФИ имеется возможность вычислять значения расширенных ячеек (атрибут «Расширенные ячейки» с использованием арифметических выражений или алгоритмов.

Над стандартными кубами определены операции объединения, пересечения, сечения и т.д.

Нестандартные кубы

Нестандартные кубы или ТОФИ-запросы представляют собой набор SQL-запросов или вызов некоторого сервиса из источника данных или из собственной базы данных. Нестандартный куб формируется как объединение (в теоретико-множественном смысле) результатов всех SQL-запросов, входящих в нестандартный куб, или как результат, возвращаемый сервисом.

Построение модели ТОФИ

Возможности слоя «Окно модели» технологии ТОФИ позволяют выполнить построение информационной модели. Важным этапом построения информационной модели в технологии ТОФИ является выделение факторов, которые являются качественными свойствами объекта.

Создание нового фактора выполняется во вкладке «Факторы». На вкладке «Факторы» создается новый фактор «Журналы» в соответствии с рис.2.

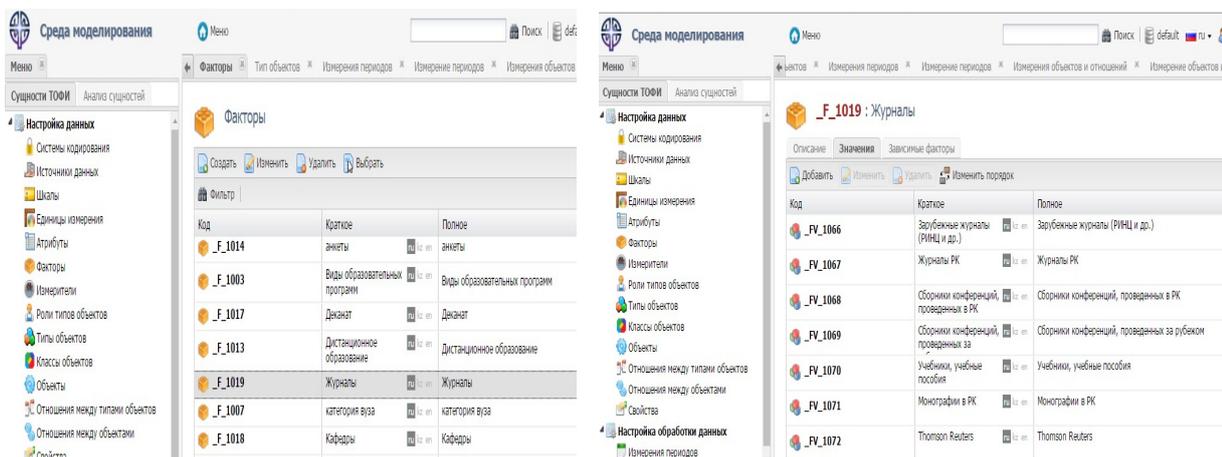


Рис.2. Добавление фактора «Журналы» и его значений в модель

Сущность «Измерители» предназначена для описания количественных свойств объектов модели. Для создания нового измерителя, например, «Количество публикаций», необходимо определить, является измеритель «мягким» или «жестким», т.е. будет зависеть от фактора или нет, в соответствии с рис.3.

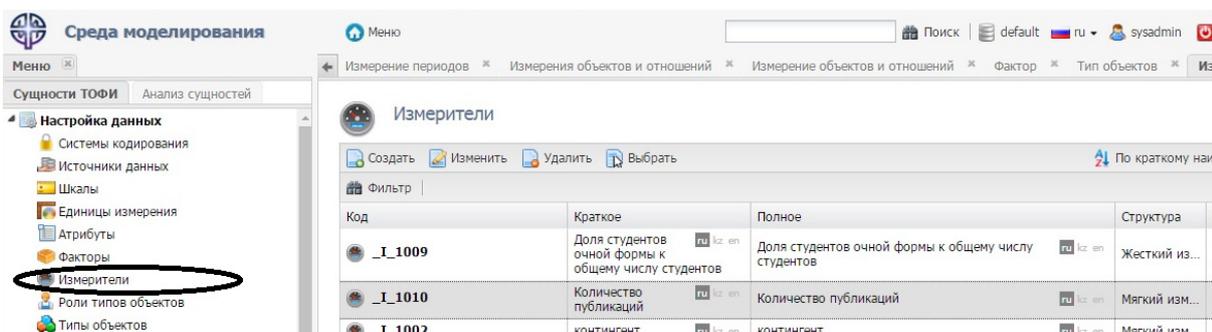


Рис.3. Добавление «мягкого» измерителя в модель

Данный измеритель определен как «мягкий», зависящий от фактора «Журналы», и принимает значения в соответствии с рис.4.

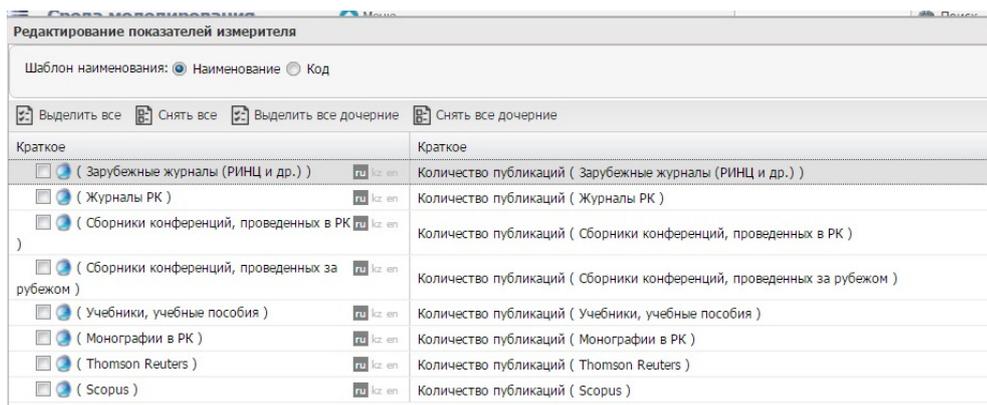


Рис.4. Показатели «мягкого» измерителя

В модели определен тип объектов «Кафедры ФИТ» в соответствии с рис.5.

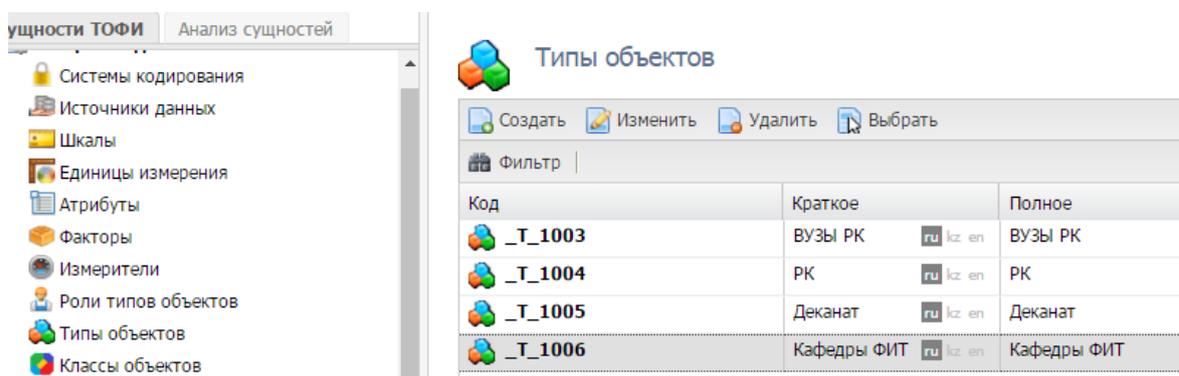


Рис.5. Добавление типа объекта в модель

В качестве кластерного фактора выбран фактор «Кафедры ФИТ» и определены классы в соответствии с рис.6.

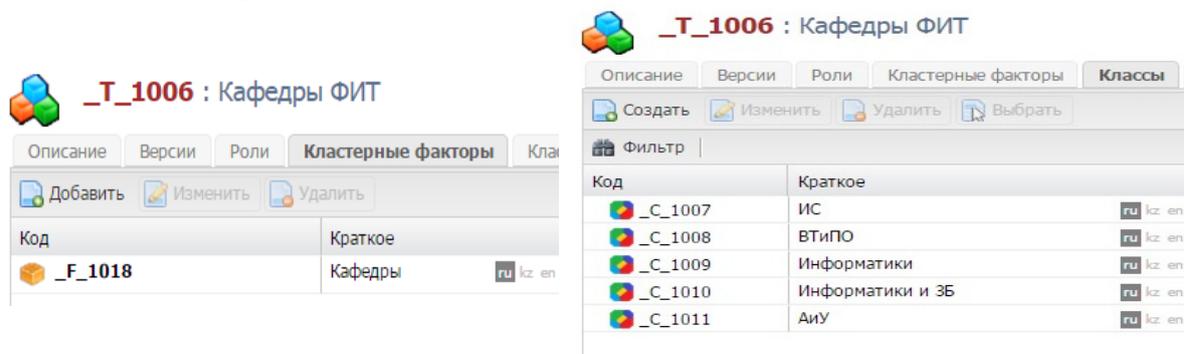


Рис.6. Определение классов с использованием кластерного фактора

Элементы класса задаются непосредственным вводом в окне модели в соответствии с рис.7.

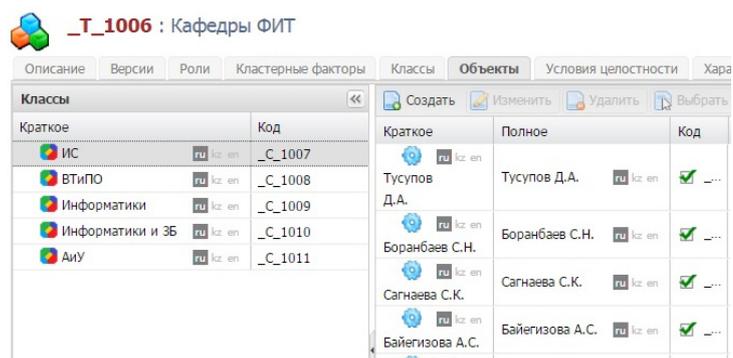


Рис.7. Объекты класса «ИС»

Для каждого типа объектов определяются его характеристические группы типа объекта в соответствии с рис.8.

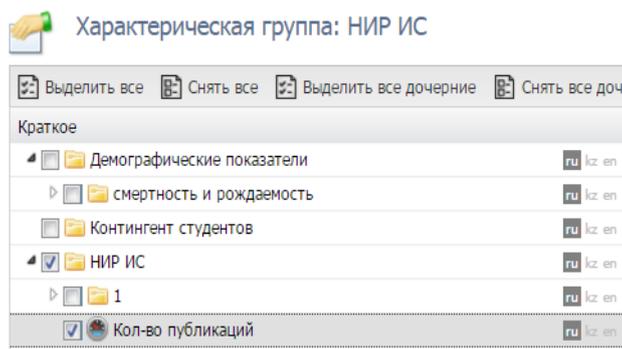


Рис.8. Формирование характеристической группы типа объекта

Построение многомерных кубов ТОФИ

Для работы с многомерными кубами в технологии ТОФИ предназначен раздел «Стандартные кубы», которые позволяют проводить многомерный анализ данных.

Для построения стандартных кубов ТОФИ предварительно формируются три вида измерений: периодов, объектов и отношений между объектами, периодов.

Для построения измерения из объектов и отношений необходимо задать тип объектов, классы объектов или отношение между объектами, условия фильтрации объектов в соответствии с рис.9.

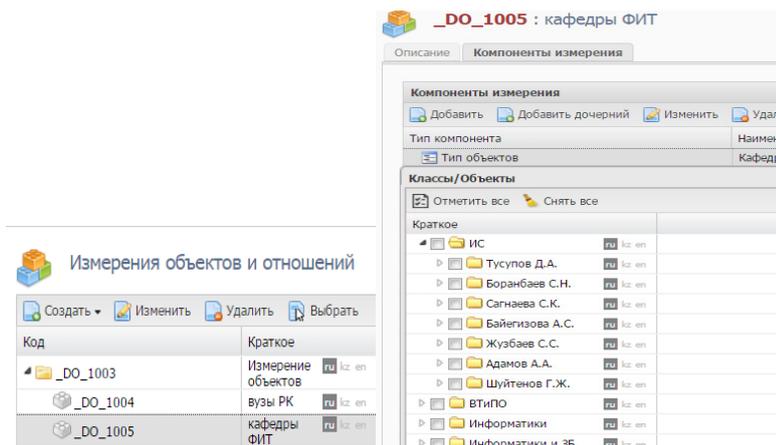


Рис.9. Измерение объектов и компоненты измерения

Для построения измерения из периодов задается тип периода, начальная дата и конечная дата в соответствии с рис.10.

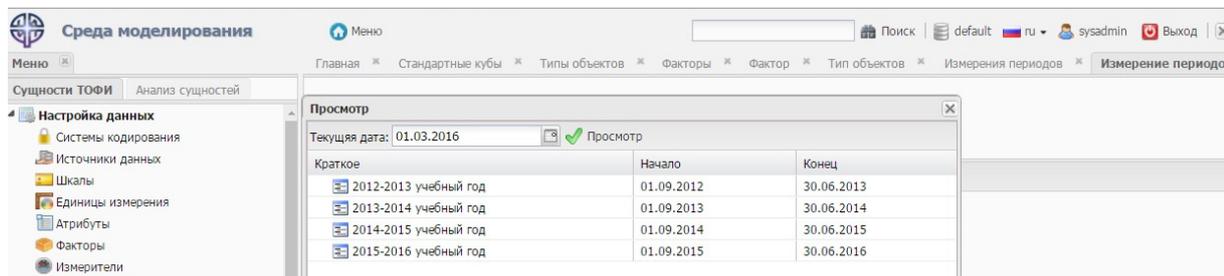


Рис.10. Измерение периодов

С помощью сущности «Стандартные кубы» создаются прокубы (прототипы) измерителей, которые состоят из определенного перечня показателей. Просмотр прокуба позволяет увидеть разрез куба без измерений периодов и объектов. Прокуб не содержит данных в соответствии с рис.11.

Свойства ППС			Значение
Свойства ППС	Количество публикаций (кол-во)	Зарубежные журналы (РИНЦ и др.) (кол-во)	0
		Журналы РК (кол-во)	0
		Сборники конференций, проведенных в РК (кол-во)	0
		Сборники конференций, проведенных за рубежом (кол-во)	0
		Учебники, учебные пособия (кол-во)	0
		Монографии в РК (кол-во)	0
		Thomson Reuters (кол-во)	0
		Scopus (кол-во)	0
		<i>Количество публикаций (кол-во)</i>	<i>0</i>

Рис.11. Просмотр прокуба «НИР»

На вкладке «Стандартные кубы» создаются стандартные кубы. Представление куба можно просмотреть в соответствии с рис.12.

Представление куба (ППС ИС)						
Действия		Тусупов Д.А.				
ППС ИС		2011-12 уч. год	2012-13 уч. год	2013-14 уч. год	2014-15 уч. год	
Свойства ППС	Количество публикаций (кол-во)	Зарубежные журналы (РИНЦ и др.) (кол-во)	3	3	5	2
		Журналы РК (кол-во)	5	4	2	1
		Сборники конференций, проведенных в РК (кол-во)	5	4	2	1
		Сборники конференций, проведенных за рубежом (кол-во)	4	3	4	3
		Учебники, учебные пособия (кол-во)	2	1	1	1
		Монографии в РК (кол-во)	1	1	1	1
		Thomson Reuters (кол-во)	1	1	1	1
		Scopus (кол-во)	1	1	1	1
		<i>Количество публикаций (кол-во)</i>				

Рис. 12. Представление куба

Авторы считают, что с помощью инструмента многомерного анализа и средств визуализации ТОФИ появляется возможность быстрого и удобного анализа состояния предметной области. Большим преимуществом данной технологии является свойство расширяемости, т.е. по мере необходимости пользователь может добавить новые показатели, скрыть ненужные или добавить новые.

Литература:

1. *Габбасов М.Б.* Новая информационная технология моделирования и мониторинга состояния сложных систем // Тезисы докладов международной конференции «Современные проблемы прикладной математики и механики: теория, эксперимент и практика», посвященной 80-летию академика Н.Н.Яненко. – Новосибирск, 2001.
2. *Жуат М.Ж., Бакаев Д.Д., Сагнаева С.К.* OLAP в технологии ТОФИ // Материалы научно-практической конференции «Применение математического моделирования и информационных технологий в исследованиях социально-экономических проблем», посвященной 50-летию к.ф.-м.н., доцента Габбасова М.Б. – Астана, 2011. С. 172-177.
3. *Габбасов М.Б.* Архитектура и возможности технологии ТОФИ //Журнал «Вычислительные технологии», часть 1. Совместный выпуск Вестника восточно-казахстанского государственного технического университета им. Д. Серикбаева и журнала «Вычислительные технологии» института вычислительных технологии СО РАН. – Усть-Каменогорск, 2013. С. 270–277.

Processing of multidimensional data by means of technology TOFI

Saule Kairollievna Sagnayeva, cand., Associate Professor, L.N.Gumilyov Eurasian National University, ENU

Shynar Erlankyzy, undergraduate, L.N.Gumilyov Eurasian National University, ENU

The article deals with the analysis and visualization of multidimensional data Tofi technology means. We describe the technological features of the construction of the Tofi data warehouse, multidimensional cubes and slices of cubes. The main purpose of the Tophi technology is to support analytical activities, ad-hoc query users and analysts.

Keywords: OLAP, data warehousing, online analytical processing, TOFI technology.

УДК 519.95

СВЯЗЬ МЕЖДУ СТРУКТУРОЙ И ТАКСОНОМИЕЙ ГЕНОМОВ ХЛОРОПЛАСТОВ ХВОЙНЫХ

Михаил Георгиевич Садовский, доктор физико-математических наук

Тел.: +79029904597, e-mail: msad@ism.krasn.ru

Институт вычислительного моделирования СО РАН

Анна Игоревна Чернышова, студентка 4 курс

Тел.: +79620709211, e-mail: anna2121695@gmail.com

Сибирский федеральный университет, ИФБиТ

В работе представлены предварительные результаты по изучению связи между структурой и таксономией геномов хвойных хлоропластов, полученные методом динамических ядер. Данная связь была выявлена. Показаны особенности применяемых методов, а также представлены выводы по полученным результатам.

Ключевые слова. ДНК, слово, частота, распределение, корреляция, эволюция, порядок.

Введение

В современном мире особый интерес в биологическом мире представляет изучение структуры геномов. Данная работа посвящена изучению связи между структурой и таксономией геномов хвойных хлоропластов.



А.И. Чернышова

Ранее уже были представлены результаты по выявлению связи между структурой и функцией биологических макромолекул [1]. В работе [2] была выявлена связь между структурой и таксономией геномов хлоропластов. Анонсируя результаты, скажем, что связь у хвойных хлоропластов также выявлена. Под структурой здесь подразумевается частотный словарь толщины три W_3 символьной последовательности, соответствующей ДНК. В свою очередь, под частотным словарём толщины 3 понимается список всех троек $v_1 v_2 v_3$ подряд идущих символов с указанием их частот. Всякий частотный словарь W_3 отображает геном в 64-мерное метрическое про-



М.Г. Садовский