



Международный журнал информационных технологий и энергоэффективности

Сайт журнала: <http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.65:004.623

ПРИМЕНЕНИЕ МЕТОДОВ РЕСТРУКТУРИЗАЦИИ И ДУБЛИРОВАНИЯ ДЛЯ ОТОБРАЖЕНИЯ РЕЗУЛЬТАТОВ ЗАПРОСОВ В ООСУБД НИКА

Тищенко В.А.

ФГУ "ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР "ИНФОРМАТИКА И УПРАВЛЕНИЕ" РОССИЙСКОЙ АКАДЕМИИ НАУК", Москва, Россия, (119333, город Москва, ул. Вавилова, д.44 к.2), ОЧУ ВО "ПРАВОСЛАВНЫЙ СВЯТО-ТИХОНОВСКИЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ" (115184, город Москва, Новокузнецкая ул., д. 23б), e-mail: vtischenko@isa.ru

Альтернативным методом представления данных по отношению к методам отображения, инкапсулируемыми объектами БД, является реструктуризация иерархических объектов базы данных. В частности в ООСУБД НИКА данный метод позволяет изменить ключ основного массива для представления результатов запроса в виде дерева, упорядоченного на основе другого ключа. Данный метод представляет вершины дерева результата в нужном пользователю порядке. Ограничением процесса реструктуризации является то, что терминальная вершина, выбираемая в качестве нового ключа, должна уникальным образом идентифицировать элемент массива. Для иллюстрации процесса реструктуризации приводится пример объекта 'Дела' типа массив, в котором изменяется ключевая вершина 'Номер' на ключевую вершину 'ФИО'. Метод дублирования используется для расширения отображения ключей дополнительной информацией, содержащей значения терминальных вершин с подчиненного уровня.

Ключевые слова: Реструктуризация иерархических объектов, запросная система ООСУБД НИКА, методы отображения объектов.

APPLICATION OF RESTRUCTURING AND DUPLICATION METHODS FOR REPRESENTATION QUERY RESULTS IN NIKA OODBMS

Tishchenko V.A.

FEDERAL RESEARCH CENTER "INFORMATICS AND MANAGEMENT" OF THE RUSSIAN ACADEMY OF SCIENCES", Moscow, Russia, (119333, Moscow, Vavilova str., 44, bldg. 2), ST. TIKHON'S ORTHODOX HUMANITARIAN UNIVERSITY (115184, Moscow, Novokuznetskaya str., 23b), e-mail: vtischenko@isa.ru

An alternative method of data representation in relation to the representation methods encapsulated by DB objects is the restructuring of hierarchical database objects. In particular, in the NIKA OODBMS this method allows changing the key of the main array to present the query results as a tree ordered based on another key. This method presents the nodes of the result tree in the order required by the user. A limitation of the restructuring process is that the terminal node selected as the new key must uniquely identify the array element. To illustrate the restructuring process, an example of the 'Case' object of the array type is given, in which the key vertex 'Number' is changed to the key vertex 'FullName'. The duplication method is used to expand the representation of keys with additional information containing the values of terminal nodes from a subordinate level.

Keywords: Restructuring of hierarchical objects, request system of OODBMS NIKA, methods of objects representation.

Введение

Идею реструктуризации иерархических объектов развивает в своей статье [1] Абитебул. Исходной предпосылкой для реструктуризации является понятие “релятивизма данных”, т.е. одни и те же данные могут быть представлены различными способами посредством различных схем описания данных. Абитебул основывает релятивизм данных на структурном преобразовании типов. Рассмотрим преобразование типа, которое приводит к массиву с ключом, соответствующим другой терминальной вершине той же самой структуры. Тогда с преобразованием типов можно связать функцию преобразования объектов этих типов. Такая функция называется *функцией реструктуризации*.

Обоснование применения функции реструктуризации

Методы (спецификации) отображения объектов БД в гипертекстовые документы описаны в статье [2]. В [3,4] показано, что существует только четыре общих типа представления данных на двумерной плоскости (в документе): последовательность — одномерное представление, таблица — двумерное представление, иерархия — n-мерное представление и их комбинации. Среди методов отображения объектов существует метод RN (rename), который присваивается элементу массива типа структура. Он не меняет общего типа представления данных, а позволяет отображать ключ элемента массива в виде одной или нескольких подчиненных терминальных вершин. При этом элементы массива остаются упорядоченными в соответствии с исходным ключом. Проблему, связанную с изменением порядка следования элементов всего массива и идентификации элементов по новому ключу, решает реструктуризация.

Формальное определение функции реструктуризации

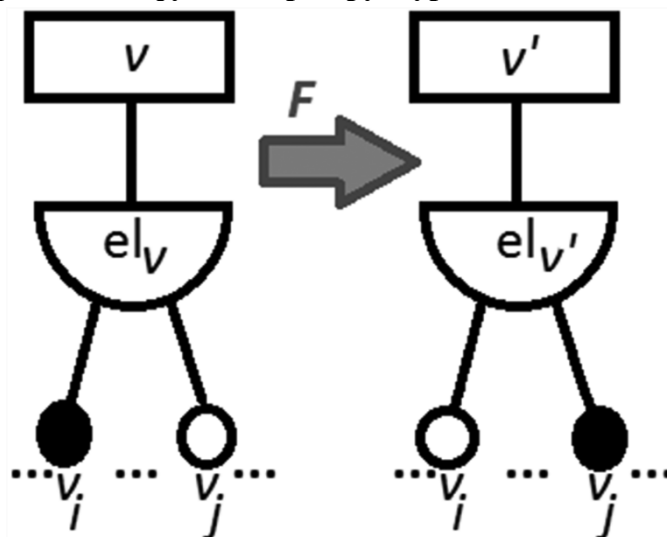


Рисунок 1.- Преобразование типа F , изменяющее ключ массива (закрашенный кружок)

Более формально, в соответствии с определением в статье [2] тип объекта БД — это граф с тремя основными подмножествами вершин: V^* — терминальные вершины, $V^\#$ — массивы, V° — структуры. Вершины указанных поименованных типов составляют иерархию вершин в БД. Типы естественным образом именуются названиями вершин. Также определяют два дополнительных типа вершин: V^p — ссылка на шаблон и V^r — ссылка на значение. Вершины типа ссылка на значение позволяют строить граф произвольной сложности. Таким образом, имеются базовые типы данных V^* , конструкторы новых типов $V^\#$, V° и повторное использование

типов V^p, V^r . Тип нетерминальных вершин определяется рекурсивно через типы подчиненных им вершин. Тогда преобразование типа, изменяющее ключ массива имеет вид $F: T(v) \rightarrow T(v')$, где $v, v' \in V^\#$, т.е. $T(v)=[T(e_{lv})]$ и $T(v')=[T(e_{lv'})]$, причем $T(e_{lv})=\{T(v_1), \dots, T(key_v), \dots, T(v_n)\}$, $T(e_{lv'})=\{T(v_1), \dots, T(key_{v'}), \dots, T(v_n)\}$, $key_v=v_i$, $key_{v'}=v_j$, $v_i, v_j \in V^*$, $i \neq j$, терминальные вершины v_i и v_j должны однозначно идентифицировать элементы массивов v и v' соответственно (см. Рис.1).

Здесь символ el обозначает элемент массива типа структура, а key — ключевую вершину массива. Массивы v и v' отличаются только ключами и *мощность* их структур данных будет одинаковой, но упорядочены они будут по разным ключам v_i и v_j соответственно. Остальные вершины структуры e_{lv} , если это необходимо, преобразование F тождественно переводит в самих себя в структуре $e_{lv'}$. Реструктуризацию, изменяющую ключ массива v , определим как функцию f преобразования объектов типа $T(v)$ в объекты $T(v')$: $f(O(v))=O(v')$, причем типы $T(v)$ и $T(v')$ связаны преобразованием типов F в определенном выше смысле. Функция реструктуризации f преобразует каждый объект типа $O(v)$ в соответствующий объект типа $O(v')$.

Реструктуризация посредством OOML

(a) Home page NIKA_ROOT Дела ▾
Дела
001.107
 Вихрев Илья Николаевич
 Фамилия Вихрев
 Год рождения 1869
 Место рождения Владимирская губ., г.Суздаль
 Из мещан г.Суздаль
 ПЕРИОДЫ ЖИЗНИ [1]
 Реабилитация
 ① Документы

001.1074
 Браиловский (Брайловский?) Александр Николаевич
 Фамилия Браиловский
 Год рождения 1880
 протоиерей

(b) Home page NIKA_ROOT Прославленные новомученики ▾
Прославленные новомученики
Абакумова (Аббакумова) Надежда Петровна
 Номер ос4.1549
 =>
Абиссов Ален
 Номер
 =>
Абрамов Мин
 Номер ос9.039
 =>
Абросимов Федор Семенович
 Номер о12.140
 =>
Августа (Защук Лидия Васильевна)
 Номер окп.32
 =>
Августин (Беляев Александр Александрович)
 Номер о32.962

(c) Home page NIKA_ROOT ОписанияГрупДел ▾
ОписанияГрупДел
d01.1
 "дело епископа Гавриила (Абалымова) и др., 1931г."
 Архив ЦА ФСБ РФ. Д.Р-35500.
 КолЧеловек 33
 Прибл.КолЧеловек 1 епископ
 ОсновныеОбвиняемые
 ПериодСледствия
 Ор
 СписокПострадавших

d01.10
 "Дело "Южно-Русского Синода", под руководством митр
 Архив Архив УКГБ Краснодарского края. Д.П-48721.
 КолЧеловек 31
 Прибл.КолЧеловек 1 митрополит. 1 епископ. 1 архиепископ. нево

(d) Home page NIKA_ROOT ГрупповыеДела ▾
ГрупповыеДела
"Антисоветское церковное подполье", 1943
 Номер_пт d01.8
 =>
"Дело "Антисоветской группы черносотен"
 ГрупповыеДела
 ГрупповоеДело
 Номер_пп
 => -> .ОписанияГрупДел.*(-Номер_пп)

"Дело "Южно-Русского Синода", под руко
 Номер_пт d01.10
 =>
"Дело 1934г. Оболенского М.Ф. и др. Ленин
 Номер_пт d01.13

Рисунок 2. - Пример реструктуризации результатов запроса к массиву 'Дела' в виде массива 'Прославленные новомученики': ключ 'Номер' (а) изменяется на ключ 'ФИО' (b); пример реструктуризации массива 'ОписанияГрупДел' в виде массива 'ГрупповыеДела': ключ 'Номер_пп' (с) изменяется на ключ 'ГрупповоеДело' (d)

Object Oriented Markup Language (OOML) — объектно-ориентированный язык разметки [5] был создан для описания документного интерфейса ООСУБД НИКА. OOML является предшественником SGML [6]. Преимуществом OOML перед SGML является то, что схема документа описывается отдельным файлом (макетом), а сам документ хранится отдельно. Посредством языка OOML можно осуществлять реструктуризацию БД, выкачивая данные из БД по одному макету, а затем загружая те же самые данные в БД по другому макету.

В качестве примеров реструктуризации можно рассмотреть две дополнительные ветки в БД “За Христа пострадавшие” [7]: ‘Прославленные новомученики’ и ‘Групповые Дела’. В первом примере дополнительная корневая вершина создается как результат выгрузки данных в документ по исходному макету результатов запроса по полю ‘ЧинСвятости’ массива ‘Дела’ и последующей загрузкой документа в БД по результирующему макету. При этом исходный и результирующий макеты отличаются порядком полей ‘Номер’ и ‘ФИО’. В исходном макете первое поле (оно является ключевым) — ‘Номер’, а в результирующем макете первое поле — ‘ФИО’. Остальные вершины не загружаются в дополнительную ветку, а получаются посредством ссылочной вершины ‘=>’ по полю ‘Номер’ из массива ‘Дела’. В результирующей ветке элементы массива будут упорядочены по полю ‘ФИО’, а не по полю ‘Номер’ как в массиве ‘Дела’ (Рисунок 2 а, b). Во втором примере создается массив ‘Групповые Дела’ как результат переупорядочивания в результате реструктуризации массива ‘Описания Групп Дел’. Аналогично первому примеру в реструктуризации участвуют две терминальные вершины: ‘Номер_пп’ и ‘Групповое Дело’ (Рисунок 2 с, d). На Рисунке 2 также для каждого пункта приводится соответствующий фрагмент схемы описания данных. Определенные выше преобразование F и функция реструктуризации f отображают соответственно типы и объекты (a) в (b) и (c) в (d).

Построение индексов

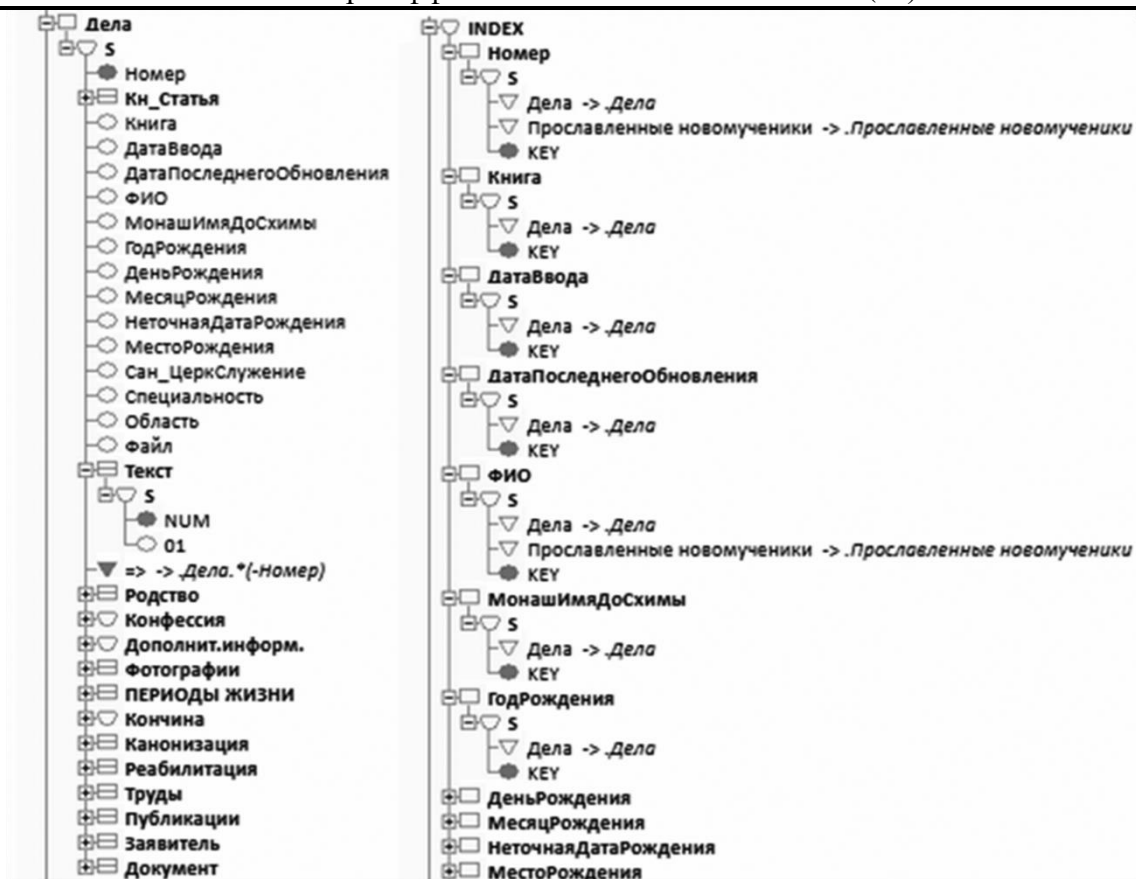


Рисунок 3. - Индексирование в ООСУБД НИКА: слева — массив ‘Дела’, справа — ‘INDEX’

Интересным применением реструктуризации служит построение индексов для атрибутивных вершин основного массива, т.е. тех терминальных вершин, которые могут участвовать в запросах. На основе описанного выше механизма, проиллюстрированного примерами, под корневой структурой ‘INDEX’ строятся одноименные с атрибутивными вершинами массивы (индексы), являющиеся результатом реструктуризации основного массива. Ключами индексов являются соответствующие атрибутивные вершины. Все значения ключей в индексах упорядочены в возрастающем порядке. Как видно из Рисунка 3 атрибутивным вершинам основного массива ‘Дела’: ‘Номер’, ‘Книга’, ‘ДатаВвода’, ‘ДатаПоследнегоОбновления’, ‘ФИО’ и т.д. соответствуют одноименные массивы под вершиной ‘INDEX’. Структура ‘INDEX’ используется как набор точек входа в основной массив, а также для выполнения запросов к БД.

Отображение результатов запроса

Приведенные примеры реструктуризации не затрагивают структуры основного массива, а представляют собой некоторые “настройки” над основным массивом в той же БД. Однако существуют случаи, в которых требуется сделать реструктуризацию самого основного массива и построить на основе реструктуризованного массива отдельную БД, в которой все подмассивы индексов основного массива будут перестроены на основе нового ключа.

Результатом запроса [8] является массив, структура которого повторяет структуру объекта запроса, являющегося корневым массивом или его подмассивом. Интерфейс информационно-поисковой системы представляет собой обозреватель интернет, являющийся гипертекстовой системой. О близости моделей гипертекста и ООСУБД НИКА говорится в статье [9]. Результат запроса отображается в виде иерархии вершин объекта запроса, которые удовлетворяют

условиям запроса. В случае *суррогатных* ключей в основном массиве и его подмассивах просмотр иерархии результатов запроса может быть затруднительным. Результат применения реструктуризации к основному массиву ‘Дела’ показан на Рисунке 4.

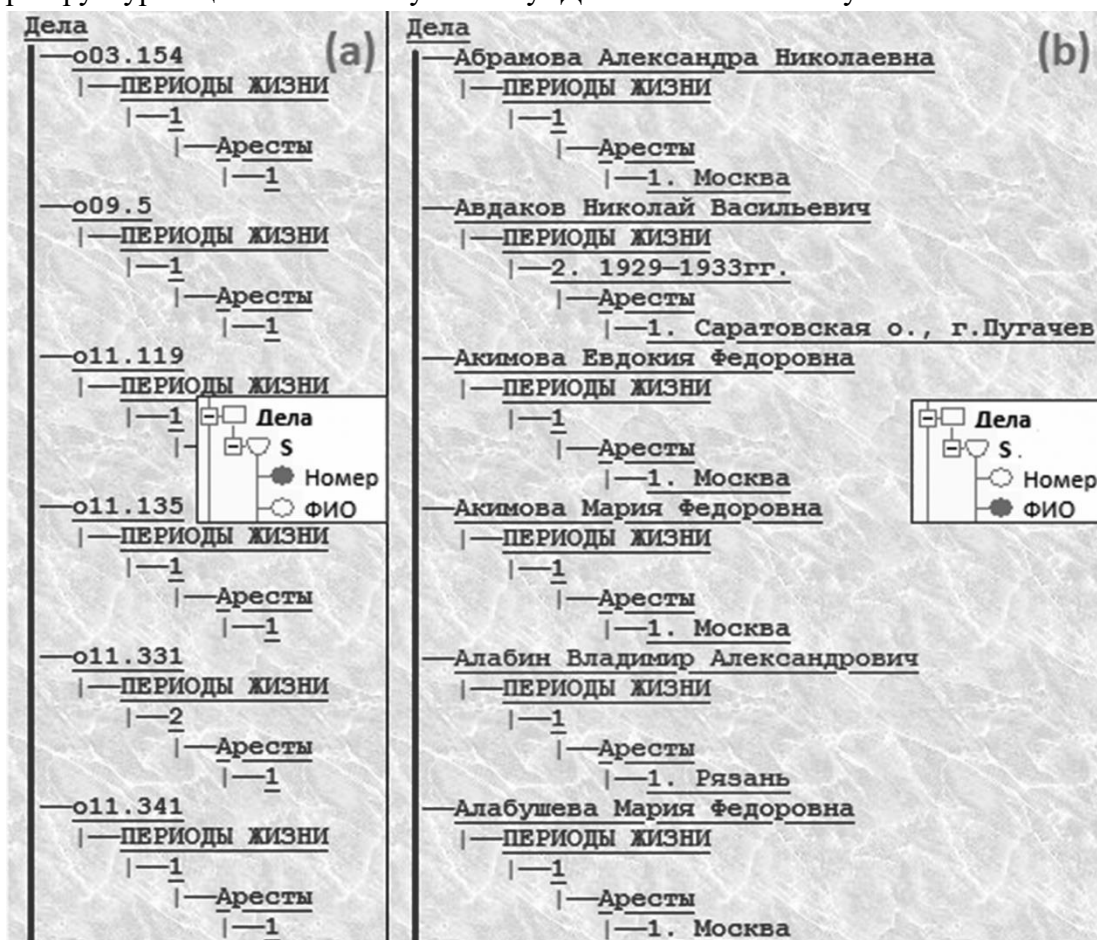


Рисунок 4.- Результат запроса до (а) и после (b) реструктуризации со схемными вставками

На Рисунке 4 также приведены фрагмент исходной схемы (а) и результирующей схемы (b), на которых обозначены ключевые вершины ‘Номер’ (а) и ‘ФИО’ (b).

Метод дублирования данных

Все подмассивы основного массива ‘Дела’ являются нумерованными. Результат запроса на Рисунке 4,а получается малоинформативным для подмассивов ‘ПЕРИОДЫ ЖИЗНИ’ и ‘Аресты’. Для повышения информативности выдаваемого результата метод реструктуризации был дополнен методом дублирования данных отдельных подчиненных терминальных вершин в ключи подмассивов. Вопрос, связанный с нарушением целостности БД при дублировании данных (например, при модификации сдублированных терминальных полей) и т.п. обсуждается в статье [10]. Результат применения метода показан на Рисунке 4,б. Если для данного элемента массива существуют терминальные вершины ‘Период жизни’ и ‘МестоАреста’, то они добавляются к ключу соответствующего подмассива.

Другие виды реструктуризации

Кроме рассмотренного в статье вида реструктуризации существует множество иных способов изменения иерархической структуры БД. Например, можно объединять в структуры

по смыслу предметной области определенные вершины, вводя новые уровни иерархии. Возможен и обратный процесс — удаление структур и расположение содержащихся в них вершин на одном уровне. При этом мощность структуры данных основного объекта, в который вложены эти вершины остается одинаковой, просто выделяются или расформируются подобъекты. В отличие от упомянутых видов реструктуризации рассмотренный вид реструктуризации основного массива “по ключу” требует перестройки всех индексов для атрибутивных вершин. Элементы массивов этих индексов содержат тип объекта “ссылка на шаблон”, повторно использующий тип основного массива. Это означает, что при смене ключа на другую терминальную вершину в основном массиве во всех “ссылках на шаблон” на этот массив произойдет такое же структурное изменение.

Цель рассмотренного преобразования улучшить читаемость отображаемой для пользователя иерархической структуры данных, полученной в результате запроса, а также всех подструктур индексов основного массива. Метод реструктуризации основного массива “по ключу” был дополнен методом дублирования данных для модификации суррогатных ключей подмассивов основного массива содержательными данными предметной области. Результаты данного исследования применяются в информационно-поисковой системе при публикации БД в интернет.

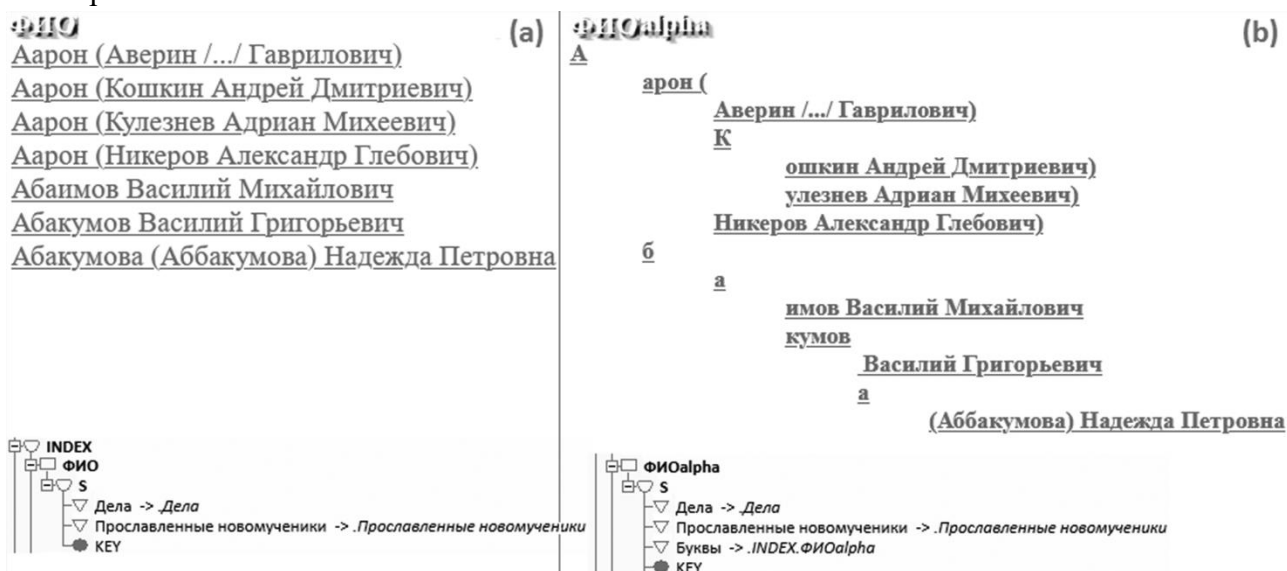


Рисунок 5. - Реструктуризация индекса по полю ‘ФИО’. Внизу схемы описания данных

Наконец, можно привести примеры видов реструктуризации, затрагивающие не только структурные вершины, но и терминальные. Неключевую вершину ‘ФИО’ можно разделить на три терминальных вершины ‘Фамилия’, ‘Имя’, ‘Отчество’, по которым строятся отдельные индексы. Если в этом примере добавить четвертую вершину — ‘Пол’, то увеличится мощность структуры данных при реструктуризации:

Более сложным является вид реструктуризации посредством разделения ключевых вершин на части для построения многоуровневых индексов. Пример такого классификатора показан на Рисунке 5. Этот вид реструктуризации является предметом отдельного исследования. Он используется для построения оптимальных многомерных классификаторов по лексикографическому признаку [11,12].

Список литературы

1. Abiteboul S., Hull R. Restructuring hierarchical database objects. Theoretical Computer Science. v.62, 1988, pp.3-38.
2. Емельянов Н.Е., Тищенко В.А. Методы отображения объектов для построения web-сервера объектно-ориентированной базы данных // Развитие безбумажных технологий в организационных системах / Сборник трудов ИСА РАН / Под ред. д.т.н. проф. Арлазарова В.Л. и д.т.н. проф. Емельянова Н.Е. М.: URSS. 1999. С. 96-109.
3. Емельянов Н.Е. “Теоретический анализ документного интерфейса”, препринт - М: ВНИИСИ, 1987.
4. Богачева А.Н., Емельянов Н.Е. "Семантическая модель документа // Системные исследования. Ежегодник 2001 / "Едиториал УРСС", М.2003, С. 360-375.
5. Bogacheva A.N., Emeljanov N.E., Romanov A.P. Object Oriented Markup Language and Restructuring Hierarchical Database Objects // Proceeding ADBIS '95 Proceedings of the Second International Workshop on Advances in Databases and Information Systems. pp. 137-142, June 27 - 30, 1995.
6. Price L.A. Practical SGML as an introduction to SGML // ACM SIGDOC Asterisk Journal of Computer Documentation, Vol.20, Issue 2, pp.36-38. <https://doi.org/10.1145/381815.381861>
7. База данных «За Христа пострадавшие». <http://martyrs.pstbi.ru/> .
8. Богданов А.С., Емельянов Н.Е., Ерохин В.И., Романов Б.Л. Реализация запросной системы на основе XPath для ООСУБД НИКА // Труды ИСА РАН. М.: Едиториал УРСС. 2003.С. 130-146.
9. Емельянов Н.Е., Тищенко В.А. Представление гипертекста в СУБД НИКА // Технология программирования и хранения данных / Сб. трудов ИСА РАН. Т.45. Под ред. чл.-корр. РАН Арлазарова В.Л. и д.т.н. проф. Емельянова Н.Е. - М. 2009. С.17-36.
10. Арлазаров В.Л., Емельянов Н.Е. Революция 2005 года в реляционных базах данных // Технология программирования и хранения данных / Сб. трудов ИСА РАН. Т.45. Под ред. чл.-корр. РАН Арлазарова В.Л. и д.т.н. проф. Емельянова Н.Е. - М. 2009. С.10-18.
11. Тищенко В.А. OPC-trie: спецификация оптимального классификатора для СУБД НИКА // труды ИСА РАН, 2021. Т. 71. Вып. 1. С.67-71.
12. Тищенко В.А. Структура OPC-trie как новый тип индекса в СУБД НИКА // Труды ИСА РАН, 2021. Т. 71. Вып. 4. С.76-81.

References

1. . Abiteboul S., Hull R. Restructuring hierarchical database objects. Theoretical Computer Science. v.62, 1988, pp.3-38.
2. Emelyanov N.E., Tishchenko V.A. Object mapping methods for building an object-oriented database web server. Development of paperless technologies in organizational systems // Proceedings of the ISA RAS. Ed. by Doctor of Technical Sciences prof. V.L. Arlazarov and Doctor of Technical Sciences prof. Emelyanov – М.: URSS. 1999. pp. 96-109.
3. Emelyanov N.E. Theoretical analysis of the document interface. М., preprint VNIISI, 1987, p.40
4. Bogacheva A.N., Emeljanov N.E. Semantic model of a document // System research. Yearbook 2001 – М.: URSS. 2003. pp. 96-109.
5. Bogacheva A.N., Emeljanov N.E., Romanov A.P. Object Oriented Markup Language and Restructuring Hierarchical Database Objects // Proceeding ADBIS '95 Proceedings of the Second

- International Workshop on Advances in Databases and Information Systems. pp. 137-142, June 27 - 30, 1995.
6. Price L.A. Practical SGML as an introduction to SGML // ACM SIGDOC Asterisk Journal of Computer Documentation, Vol.20, Issue 2, pp.36-38. <https://doi.org/10.1145/381815.381861>
 7. Database 'for Christ suffered' <http://martyrs.pstbi.ru/> .
 8. Bogdanov A.S., Emelianov N.E., Erokhin V.I., Romanov B.L. Implementation of a query system based on XPath for the OODBMS NIKA // Organizational management and artificial intelligence. Proceedings of the ISA RAS. Ed. by Doctor of Technical Sciences prof. Arlazarov V.L. and d.t.s. prof. Emelyanov N.E. M.: URSS. 2003. pp. 130-146.
 9. Emelyanov N.E., Tishchenko V.A. Representation of hypertext in the NIKA DBMS // Technology of programming and data storage / Sat. Proceedings of the ISA RAS. T.45. Ed. Corresponding Member RAS Arlazarov V.L. and Doctor of Technical Sciences prof. Emelyanov N.E. - M. 2009. pp. 17-36.
 10. Arlazarov V.L., Emelyanov N.E. The 2005 revolution in relational databases // Technology of programming and data storage / Sat. Proceedings of the ISA RAS. T.45. Ed. by Corresponding Member RAS Arlazarov V.L. and Doctor of Technical Sciences prof. Emelyanov N.E. - M. 2009. pp. 10-18.
 11. Tishchenko V.A. OPC-trie: specification of an optimal classifier for the NIKA DBMS // Proceedings of ISA RAS, 2021, 71, No. 1. pp.67-71.
 12. Tishchenko V.A. OPC-trie structure as a new type of index in NIKA DBMS // Proceedings of ISA RAS, 2021, 71, No. 1. pp. 76-81.
-