

Информационная модель учета временного фактора в информационно-справочных системах*

Леонова Ю.В., Федотов А.М.

Институт вычислительных технологий СО РАН, Новосибирск
juli@ict.nsc.ru, fedotov@sbras.ru

Аннотация

Основным предметом рассмотрения данной статьи является информационная модель учета временного фактора в информационно-справочных системах (ИСС). ИСС должна обеспечивать выполнение запросов на какой-нибудь момент времени в прошлом, то есть создание среза истинности фактов на произвольную дату. Учет временного фактора предлагается реализовать на основе двух зависимостей: 1) версии документов, связанные с изменениями атрибутов документов на выбранных интервалах времени; 2) отношения родитель-потомок между новыми и старыми объектами.

Введение

Информационные ресурсы, переведенные в электронную форму, приобретают новое качество, которое обеспечивает им более широкое распространение и эффективное использование. Важными аспектами успешной научной деятельности являются информационная поддержка подготовки проектов, обеспечение научных мероприятий справочной информацией.

Создание информационно-справочных систем поддержки научных исследований, формирования и реализации управленческих решений для крупных научных сообществ как СО РАН, является одним из наиболее действенных способов интеграции научных коллективов.

При создании таких систем наиболее остро стоит проблема получения и использования оперативной информации.

Проанализировав сформулированные в работах [10-14] требования, предъявляемые к информационно-справочным системам, можно сделать вывод, что одним из краеугольных элементов при построении ИСС является использование метаданных для формирования структуры и свойств элементов информационной системы. Использование метаданных – важный фактор обеспечения интероперабель-

ности при разработке информационных ресурсов. Соответственно главный элемент любой ИСС – унифицированный каталог, содержащий метаданные документов.

Основным предметом рассмотрения данной статьи является информационная модель учета временного фактора в информационно-справочных системах. ИСС должна обеспечивать выполнение запросов на какой-нибудь момент времени в прошлом, то есть создание среза истинности фактов на произвольную дату.

Постановка задачи

Информационную систему можно определить как множество связанных между собой документов [8]. Документы содержат информацию о некоторых объектах и явлениях реального мира. Такими объектами могут быть персоны, организации, учреждения, подразделения, статьи, проекты, отчеты по проектам.

Определение: *Документ* – это отображение в электронном виде объекта реального мира: статьи, проекта, организации, подразделения, персоны и т. д.

Согласно определению в “Философской энциклопедии” [1] *объект* – предмет, явление, процесс или их отдельные стороны, существующие в реальной действительности, на которые направлена деятельность, мысль, чувство и т.д. Объект противопоставляется субъекту в его предметно-практической и познавательной деятельности.

Субъект – индивид или социальная группа, выступающие носителем предметно-практической деятельности и познания. В качестве объекта выступают, прежде всего, предметы и явления реального мира. В качестве объекта может выступать и сам субъект.

Субъектом (субъектами) осуществляется деятельность, его действия направлены на объект (объекты), действия осуществляются с помощью разных средств, способов, технологии. Здесь проявляются те особенности деятельности, о которых писал А.Н. Леонтьев: “в деятельности... происходит переход объекта в его субъективную форму, образ; вместе с тем в деятельности совершается также переход деятельности в её объективные результаты, в её продукты. Взятая с этой стороны, деятельность выступает как процесс, в котором осуществляются взаимопереходы между

полосами "субъект-объект" [2]. В научной деятельности это конкретизируется в том, что объект (научное исследование) переходит в субъективную форму - объект, проект, в конце деятельности получается объективный результат (отчет по проекту).

Таким образом, документы, входящие в ИСС, могут описывать предметы, субъекты, деятельность, результаты деятельности.

Объекты реального мира являются динамическими. Это означает, что информация об объектах, предоставляемая ИСС, изменяется во времени. Однако в зависимости от отражаемой стороны объект может быть статичным или динамичным.

Информация о некоем объекте содержится в системе либо непосредственно в виде документа, который ее представляет, описывает или моделирует, либо в виде упоминаний об этом объекте, которые имеются в других документах, т.е. содержат опосредованную информацию об этом объекте [3]. Поясним смысл основных видов документа.

Документ-Описание – это документ, который содержит описание реального объекта, например, описание некоторой организации или конкретной персоны, информация о которых используется в информационной системе. Этот документ-описание содержит информацию о некотором объекте, но при этом сам не является объектом. Отметим, что описательными документами также являются описания информационных ресурсов в каталоге или описание библиографических источников. Таким образом, метаданные полностью характеризуют документ-описание, т.е. данный вид документа генерируется из метаданных.

Документ-Представление – это документ, который является непосредственно конкретным объектом (например, статья, книга или фотография). Кроме того, к этому классу относятся документы, которые содержат информацию о некотором объекте и при этом сами являются объектом, например, библиографические карточки. Таким образом, данный документ имеет метаданные и ссылку на соответствующее представление, т.е. метаданные являются частью документа.

Документ-Отображение – это документ, который является отображением другого документа, т.е. его точной копией или электронным образом. Например, PDF (или PS)-файл статьи, сканированный документ и т.д. Таким образом, метаданные и документ полностью отделены друг от друга.

Документ-Моделирование – это документ, который моделирует некоторый реальный объект, например, с помощью компьютерной программы.

Следовательно, документами в нашем понимании, в частности, являются компьютерные программы, алгоритмы, растровые и векторные изображения и их описания. Научная статья и ее библиографическое описание (библиографическая карточка) в нашем понимании являются различными, но связанными документами.

Отметим, что в ИСС могут присутствовать документы двух типов: документы-представления и документы-описания.

Поскольку ИСС оперирует непосредственно не с документами, а с метаданными документов, то для построения модели временного изменения объектов достаточно ограничиться документами-описаниями.

Характеристики документов

Согласно определению ИСС описание документа D представляется как набор атрибутов A_k , в соответствии с выбранной схемой данных¹:

$$D = \{A_x\}_{x=1..k} \quad (1)$$

где k – количество атрибутов в описании документа. Каждый атрибут имеет имя, тип и значение (содержание):

$$A_x = \{N_x, T_x, V_x\}. \quad (2)$$

Значения атрибута принадлежат некоторому множеству значений:

$$V_x \in \{v_x^m\}_{m=1..M}, \quad (3)$$

где M – количество значений соответствующего атрибута.

Для учета временного фактора в рассматриваемом случае документы ИСС должны содержать данные (атрибуты), связанные с промежутками времени. Важной характеристикой документа является *текущее время*, позволяющее сравнивать исторические данные в прошлом и настоящим моментом времени. Изменения значения атрибута в момент времени t будем считать временным изменением.

Рассмотрим примеры документов, у которых возможны изменяемые атрибуты.

- 1) **Персона.** Рассмотрим динамику изменения свойств объекта, описывающего персону. Во время жизни персона может изменить фамилию, адрес, ученую степень, продвигаясь по служебной лестнице, при назначении на новую должность персону меняет свой статус, персону может переходить из одного института в другой и т.д.
- 2) **Подразделение.** К основным динамическим свойствам объекта "подразделение" относятся: название подразделения, контактные данные, состав сотрудников, наличие головных и подчиненных подразделений. Особо отметим случай слияния нескольких подразделений в одно и, наоборот, разделения подразделения на несколько других подразделений. В этом случае происходит создание новых подразделений.
- 3) **Приказ.** Динамика изменения объекта "приказ" отражает движение документов в организации с момента их получения или создания до завершения исполнения. После создания приказ проходит этап утверждения (подписания). Кроме того, приказ может быть отменен другим приказом. Таким образом, приказ имеет три состояния:

¹ см. раздел Схема данных

создан, утвержден, отменен. При этом содержание приказа не меняется.

С другой стороны приказ можно рассматривать как бумажный документ (документ-представление). На нем могут быть надписи, визы, исправления и т.п. Это тот же объект, но с другими метаданными, описывающими изображение (приказа). В этом случае отображается другая сторона объекта “приказ” – его изображение, которое может меняться. На приказе могут появляться новые подписи, пятна и, как следствие, появляются новые версии изображения приказа. При этом содержательная сторона приказа не меняется.

Заметим, что хранение изображений документов специфично для архивных ИС. В ИСС приказ обычно рассматривается как документ-описание.

- 4) **Проект.** Динамика изменения проекта аналогична динамике изменения приказа и подразделения. После создания проект проходит этап утверждения. У проекта может меняться руководитель, список исполнителей. Возможны случаи слияния нескольких проектов в один и, наоборот, разделения проекта на несколько других проектов. У проекта также имеется дополнительный атрибут – результаты (отчеты), которые могут накапливаться во времени.
- 5) **Документация.** Изменения объекта “документация” в большей мере соответствует этапам документооборота, когда документы проходят этап интенсивного обсуждения и редактирования в процессе согласования. У документации может быть несколько различных состояний. Например, документ может сначала быть идентифицирован как «Release» (рабочий). После утверждения, состояние может быть установлено в «Current» (принятый) или позже в «Stable» (изданный).

Схема данных

Основой ИСС является схема данных, которая определяет основные документы ИСС.

Рассмотрим информационную модель каталога X.500 [4]. Информация размещается в каталоге X.500 в виде *объектов*. По смыслу, каждый содержащийся в каталоге объект отражает факт существования некоторой сущности реального мира: организации, человека, сервера и т.д. Объект характеризуется своим набором *атрибутов*. Классы объектов определяются в *схеме* каталога. **Схема** – формальное описание содержимого и структуры каталогов – представляет собой набор связанных между собой классов, которые в свою очередь состоят из набора атрибутов. В объявлении класса задается его имя и перечисляются атрибуты, которые используются в объектах данного класса, а также фиксируются те из них, задание значений в которых обязательно. Кроме того, могут быть указаны один или несколько родительских классов, определения которых будут наследоваться данным классом. Все атрибуты, упо-

минаемые как обязательные в родительских классах, становятся обязательными и в порождаемом классе. То же самое относится и к необязательным атрибутам. В качестве базиса схемы данных для построения ИСС нами были выбраны рекомендации X.500.

Проиллюстрируем сказанное следующим примером (рис. 1), в котором объявляются три класса, показанные на следующей диаграмме: person (человек) и organizationalPerson (должностное лицо) и internetPerson (пользователь Интернет).

Класс person определяет наиболее общие атрибуты, которые присущи любому человеку: полное имя (commonName, например, фамилия и имя); фамилия (surname); произвольное текстовое описание (description); DN-ссылка на другие объекты, представляющие этого человека в каталоге (seeAlso); телефонный номер (telephoneNumber); пароль для входа в систему (userPassword).

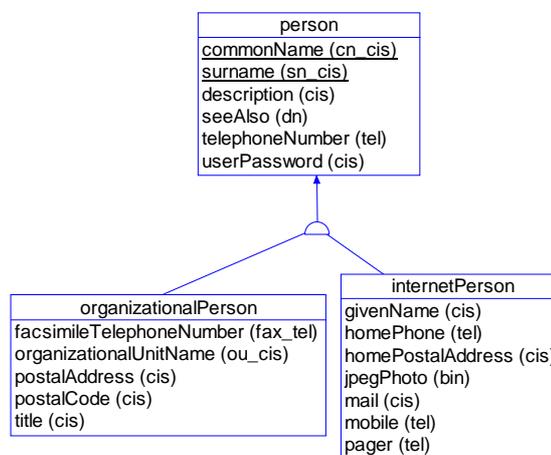


Рис.1. Описание объекта персона (X.500)

Класс person является родительским для двух производных от него классов. Класс organizationalPerson добавляет дополнительные атрибуты, характеризующие человека как должностное лицо: номер факса, наименование подразделения организации, где он работает, почтовый адрес и код, должность. С другой стороны класс internetPerson характеризует человека как пользователя Интернет, вводя атрибуты: имя (givenName), домашний адрес и телефон, фотография, адрес электронной почты, номера мобильного телефона и пейджера.

Резюмируя сказанное, можно отметить, что схему данных документа можно представить в виде иерархии. На вершине иерархии находится наиболее общее описание объектов: UID – уникальный идентификатор объекта и его основные свойства – состояние, содержание. Расширяя схему данных документа, можно конкретизировать различные стороны описания объекта. При этом документы-потомки, наследуют атрибуты у документов-предков.

На некотором этапе добавление локальных расширений схемы данных в узлы дерева приводит к типизации конкретных объектов ИСС: персон,

организаций, публикаций и т.д. Для более детального отражения необходимых сторон объектов выполняется дальнейшее расширение схемы данных документов.

Например, рассмотрим описание персоны (рис. 2). Базовый элемент схемы данных UID, добавляя локальное расширение VCard, типизируем описание объекта – персоны. Далее добавляем локальные расширения, связанные с решаемыми задачами, можно получить документ, связанный с производственной деятельностью персоны или медицинскую карту персоны.

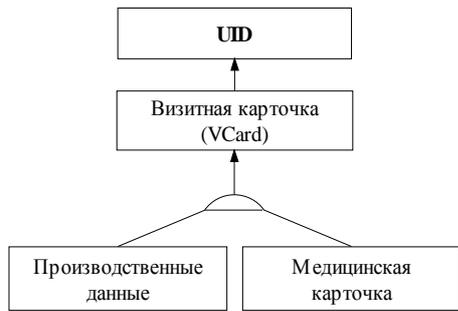


Рис.2 Типизация описания персоны

Состояние

Как говорилось ранее, объекты не являются статичными. Объекты в информационном пространстве появляются, исчезают и изменяются.

При построении ИСС возникает необходимость в хранении истории изменения метаданных объектов и предоставления информации в задаваемых временных срезах. Для учета временных характеристик мы отталкивались от понятия “состояния”. Однако понятие “состояния” имеет многосмысловое и неоднозначное толкование в зависимости от контекста.

Состояние – абстрактный многозначный термин, в общем, обозначающий множество стабильных значений переменных атрибутов объекта.

Свойства “состояния”:

- описывает переменные атрибуты объекта;
- стабильно до тех пор, пока над объектом не будет произведено действие;
- если над объектом будет произведено некоторое действие, его состояние может измениться.

Под действием в ИСС понимается воздействие на объект, вызывающее изменение его атрибутов. Как правило, действие над объектом в информационной системе выполняется администратором.

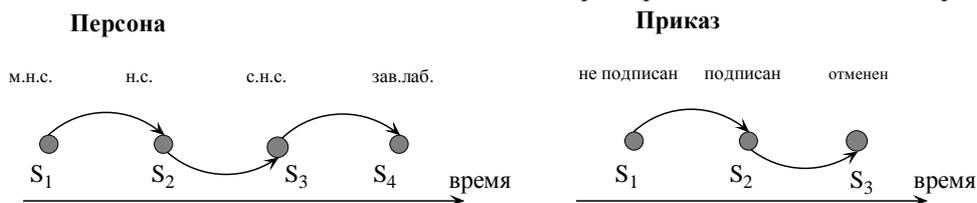


Рис.3 Пример графа переходов для объектов персона и приказа

Кроме того, состояние объекта может измениться в результате изменения окружающей среды, т.е. было произведено действие над другим объектом, например, если закрылся институт, в котором работал человек.

Напомним, что документы в информационной системе представляются как набор атрибутов. Значения атрибута принадлежат некоторому множеству значений (3). Элементы этого множества значений v_x будем называть *состояниями* S_x

$$S_x^m \equiv v_x^m, \quad (4)$$

Примеры состояний:

- приказа: подписан, не подписан;
- статьи: опубликована, не опубликована;
- персоны: множество должностей – инженер, м.н.с, н.с., с.н.с., зав.лаб., ученый секретарь, ...

Для некоторых атрибутов возможен переход из одного состояния в другое (рис. 3).

Последовательная смена состояний объекта – процесс. Процесс смены состояний объекта явно или неявно связан с определенными датами или промежутками времени. Например, для объекта персоне атрибут “адрес” может изменяться с течением времени [5, 6]. Но при этом предыдущие значения адреса должны оставаться доступными для приложений (или пользователей), которые могли бы извлекать полностью или частично “адресную” историю персоны. Так, адрес содержит как текущий и предыдущий адрес (и, возможно, будущий адрес), но и временную информацию, связанную с каждым значением адреса.

Следовательно, посредством временной модели данных мы можем получить интегрированные знания о том, где персоне жила и когда это было. Отношения между фактами, т.е. адресом персоны и временем показано на рис. 4.

Если рассматривать данные, представленные в ИСС, как некоторое отражение текущего состояния действительности для моделируемого мира, то каждый документ может восприниматься как некоторый факт, который является истинным в определенный момент или интервал времени [7]. Для каждого факта можно указать тот промежуток времени, когда этот факт являлся истинным в моделируемом мире, представленном в базе данных.

Таким образом, с объектами ассоциируются так называемые *периоды существования*, представляющие собой отрезки времени, когда имеют силу конкретные состояния некоторого объекта.

Рассмотрим подробнее представление временных характеристик в ИСС. Как говорилось ранее,

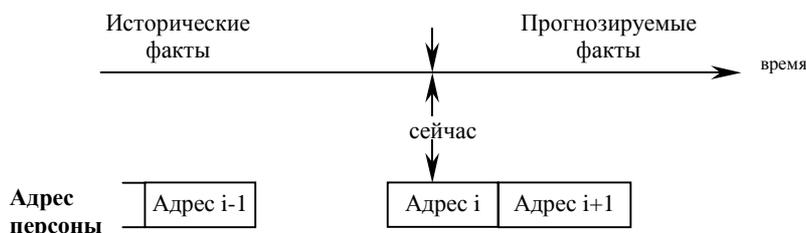


Рис.4. Факты и время

состояние объекта меняется при изменении его атрибутов. В результате изменения атрибутов возможны две ситуации: уточнение описания и порождение нового описания. Смена состояния объекта соответствует порождению нового описания, т.е. документа. Отметим, что принятие решения о появлении нового документа является внешним действием по отношению к системе, его принимает администратор системы. Отметим, что принятие решения о появлении нового документа является внешним действием по отношению к системе, его принимает администратор системы. Информация об эволюции объектов при каждом изменении состояния объекта должна сохраняться в системе. Для любого объекта, который был создан в момент времени t_{start} и закончил свое существование в момент времени t_{end} , в системе будут сохранены все его состояния на временном интервале $[t_{start}, t_{end}]$.

Версии

В ИСС возможны два типа изменений объектов:

- 1) изменение объекта, выражающееся в создании нового объекта;
- 2) изменение объекта, выражающееся в модификации атрибутов документа.

Как правило, порождение нового объекта происходит при разделении объекта на несколько других объектов или, наоборот, при слиянии нескольких объектов в один. Связи между родительскими и дочерними объектами отсутствуют, но у объекта присутствует ссылка на его предшественников (родителей), позволяющая при необходимости восстановить связь “родитель-потомок”.

Изменения второго типа разбиваются на два вида модификаций документов: сильная и слабая модификация.

Слабой модификацией будем называть уточнение описания объекта, выполняемое, например, при исправлении ошибок.

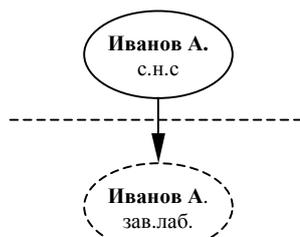


Рис.5. Порождение новой версии документа “персона”

Сильной модификацией будем называть порождение нового документа, выполняемое при “важных” изменениях объекта таких, как изменение общественного статуса персоны (смена должности, адреса) или списочного состава, названия лаборатории.

Отметим, что слабые изменения в отличие от сильных изменений не накладываются на историческую шкалу.

Администратор определяет наборы атрибутов для сильной и слабой модификации. Очевидно, что слабая и сильная модификации – это интуитивные понятия и решение об отнесении исправления документа к сильной или слабой модификации принимается лицом, ответственным за содержание (администратор института).

Так, изменение в адресе организации название улицы с “проспекта Науки” на “проспект акад. Лаврентьева” для персоны, работающей в данной организации, будет считаться слабой модификацией, поскольку данная модификация не меняет общественный статус персоны.

Сильная модификация вызывает создание версии документа, соответствующей (описывающий объект) данному отрезку времени.

Тем самым, документ можно представить как отображение $D(t)$ (функцию) множества версий документа на временную ось.

При выполнении запроса на представление информации во временном срезе, необходимо выдавать версию (версии) документа, значения атрибутов которого соответствуют заданным временным характеристикам. Последняя версия документа является текущей версией объекта.

Поясним данные принципы на трех примерах.

1) Допустим, у персоны Иванов А. (рис. 5) изменилась должность. Эта модификация является сильной модификацией, поскольку меняется статус персоны. Поскольку персона всегда соответствует одному объекту, документ, описывающий персону, укладывается в модель версий. Соответственно при

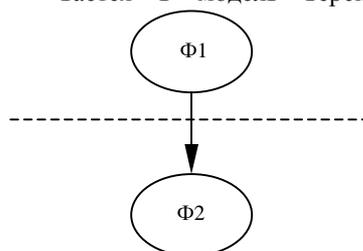


Рис.6. Порождение нового объекта “фотография”

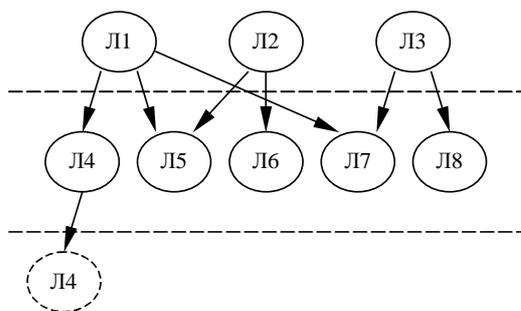
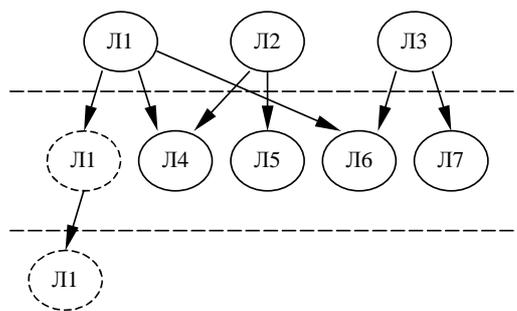


Рис. 7 Изменение объекта “лаборатория”.

а) порождение новых объектов



б) порождение версии объекта

сильной модификации всегда происходит порождение новой версии документа.

2) Рассмотрим объект фотография Ф1 (рис.6). Здесь объектом является само изображение, а его метаописание – это классификационные признаки фотографии. При изменении классификационных признаков фотографии порождается версия, поскольку само изображение не меняется. Так, принадлежность фотографии к альбому является классификационным признаком. Фотографию можно перемещать из альбома в альбом, приписывать к разным альбомам, но при этом само изображение остается тем же. При изменении изображения порождается новый объект фотография Ф2.

3) Допустим, существуют три лаборатории Л1, Л2 и Л3 (рис. 7а). Каждая лаборатория соответствует объекту информационной системы. Рассмотрим динамику изменения объектов.

В результате разделения и слияния лабораторий Л1, Л2 и Л3 образовалось 5 новых лабораторий Л4, Л5, Л6, Л7 и Л8. Поскольку происходит деление объектов, отображение лаборатории не укладывается в модель версий. В этом случае лаборатории Л1, Л2 и Л3 ликвидируются и появляется 5 лабораторий, соответствующих новым объектам. При этом у новых лабораторий есть предшественники.

Затем лаборатория Л4 была переименована и появилась новая версия документа. Поскольку решение о появлении нового объекта или появлении версии принимает лицо, ответственное за содержание, то возможна и другая структура объектов (рис. 7б). Например, от лаборатории Л1 отпочковалась лаборатория Л4, т.е. породился новый объект, но при этом лаборатория Л1 не прекратила свое существование.

Реализация информационной модели

Схема данных ИСС разделяется на две части: часть, содержащая базовое описание объектов, и дополнительная часть, связанная с локальными расширениями. Расширяя схему данных объекта, на каком-то этапе добавляется временное расширение, которое может иметь свои локальные расширения (рис. 8).

Таким образом, в документах ИСС выделяются два типа атрибутов (рис.9):

- 1) множество атрибутов, которые допускают сильные модификации;
- 2) множество атрибутов, которые не допускают сильные модификации.

Установление связей в документах между локальными расширениями осуществляется посредством поименованных связей, снабженных метаданными. Для представления поименованных связей используется модель направленных связей [8].

Для отображения отношений на реляционную структуру выбрана модель “таблица на свойство” (представление свойств с использованием специальной ключевой таблицы) [9], поскольку такое представление является унифицированным и не зависит от конкретного свойства.

Заключение

Подводя итог, можно сделать вывод, что предложенная авторами модель реализует линейную структуру хранения истории, которая обеспечивает доступ к документам вне зависимости от временных характеристик.

Литература

- [1] Философский энциклопедический словарь. М: Советская Энциклопедия, 1980.
- [2] Леонтьев А.Н. Деятельность. Сознание. Личность.- Изд. 2-е - М.: Политиздат, 1977.
- [3] Шокин Ю.И., Федотов А.М., Леонова Ю.В. Принцип динамического формирования документов в информационных системах, на примере интегрированной распределенной информационной системы (ИРИС) СО РАН // Труды Четвертой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Дубна, 15-17 октября 2002 г. Т.2, С.159-169.
- [4] М.К.Валиев, Е.Л.Китаев, М.И.Слепенков Служба директорий LDAP как инструментальное средство для создания распределенных информационных систем // Препринт ИПМ РАН, №23, Москва, 2000.
- [5] B. Skjellaug. Temporal Data: Time and Object Databases. Research Report 245, Department of Informatics, University of Oslo, April 1997.

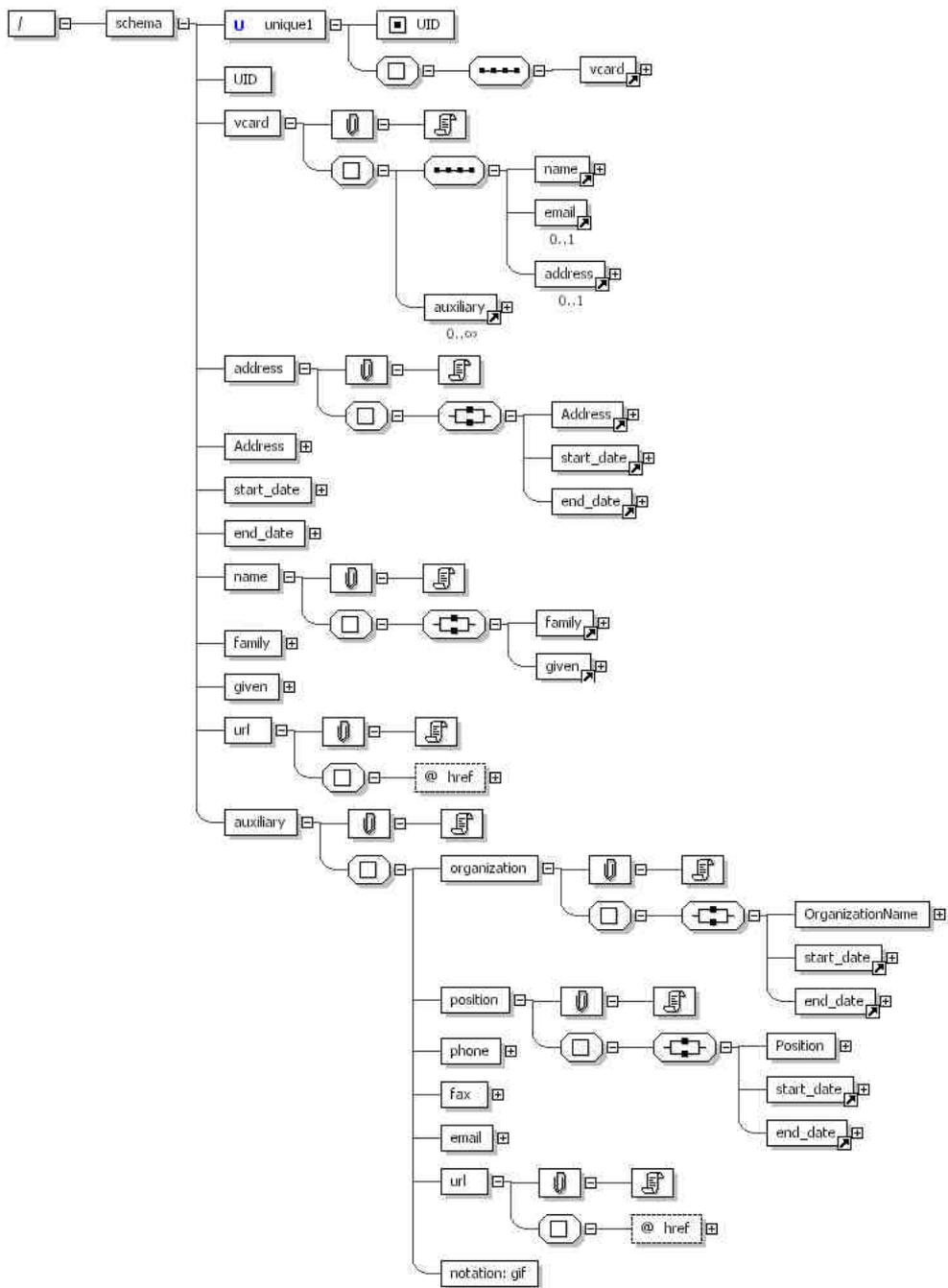


Рис. 8. Схема данных документа с временными локальными расширениями

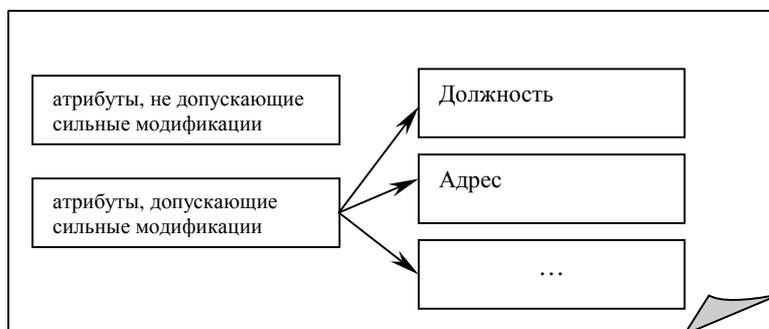


Рис.9. Типы атрибутов в документе

- <http://citeseer.ist.psu.edu/skjellaug97temporal.html>
- [6] В.Б. Барахнин, Ю.В. Леонова Информационная модель отношений между документами в информационной системе // Вычислительные технологии, Спец. выпуск: Труды IX рабочего совещания по электронным публикациям (EI-Pub) – 2005. – Т.10. – С. 129-137.
- [7] Z. Pan and J. Heflin. DLDB: Extending Relational Databases to Support Semantic Web Queries // Technical Report LU-CSE-04-006, Dept. of Computer Science and Engineering, Lehigh University, 2004.
- [8] Challenge '97 - толчок в сторону X.500 // Сети, №9, 1996. <http://www.osp.ru/nets/1996/09/104.htm>
- [9] X.500 На пути к информационному раю // Мир Internet, №5, 2002.
http://www.iworld.ru/magazine/index.phtml?do=show_number&m=94701987
- [10] Игнатъева М. Итоги развития X.500 в 1997 году // Сборник АДЭ, 1998.
http://www.ctel.msk.ru/x500/Publ/my_p.pdf
- [11] Агошков С.В., Бездушный А.Н., Галочкин М.П., Кулагин М.В., Меденников А.М., Серебряков В.А. Интегрированная Система Информационных Ресурсов (ИСИР) РАН – подход к созданию интегрированных цифровых библиотек // Труды Первой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Санкт-Петербург, 19-22 октября 1999 г. С.254-270.
- [12] Ю.В. Леонова, А.М. Федотов Требования, предъявляемые к информационным системам для научного сообщества // Труды Всероссийской научной конференции “Научный сервис в сети Интернет: технологии распределенных вычислений”. – М.: Изд-во МГУ, 2005. – С.240-242.

Information model of the account of the temporary factor in information-reference systems

Leonova Yu.V., Fedotov A.M.

The main subject of consideration given article is an information model of the account of the temporary factor in information-reference systems (IRS). IRS must provide execution an inquiry for some moment of time in past that is to say making the cut to true fact to free date. The account of the temporary factor is offered realize on base two dependencies: 1) to versions document, connected with change attribute document on chosen time lag; 2) relations parent-descendant between new and old objects.

*Работа выполнена при частичной поддержке РФФИ: проекты 06-07-89060, 06-07-89038, 07-07-00271, президентской программы «Ведущие научные школы РФ» (грант № НШ-9886.2006.9) и интеграционных проектов СО РАН.