

*Intellectual Technologies
on Transport
No 4*



*Интеллектуальные технологии
на транспорте
№ 4*

*Санкт-Петербург
St. Petersburg
2019*

Интеллектуальные технологии на транспорте № 4, 2019

Сетевой электронный научный журнал, свободно распространяемый через Интернет.
Публикуются статьи на русском и английском языках с результатами исследований и практических достижений в области интеллектуальных технологий и сопутствующих им научных исследований.

Журнал основан в 2015 году.

Учредитель и издатель

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Петербургский государственный университет путей сообщения Императора Александра I» (ФГБОУ ВО ПГУПС)

Сопредседатели редакционного совета

Панычев А. Ю., ректор ПГУПС, С.-Петербург, РФ
Чаркин Е. И., директор по ИТ ОАО «РЖД», Москва, РФ

Главный редактор

Хомоненко А. Д., проф., С.-Петербург, РФ

Редакционный совет

Глухов А. П., вед. НС ГВЦ ОАО «РЖД», Москва, РФ
Дудин А. Н., д.т.н., проф., БГУ, Минск, Беларусь
Илларионов А. В., советн. «РФЯЦ-ВНИИЭФ»,
Саров, РФ
Корниенко А. А., проф., ПГУПС, С.-Петербург, РФ
Ковалец П., проф., Тех. ун-т, Варшава, Польша
Меркурьев Ю. А., проф., РТУ, Рига, Латвия

Нестеров В. М., проф., С.-Петербург, РФ
Пустарнаков В. Ф., ген. дир. «Газинформсервис»,
С.-Петербург, РФ
Титова Т. С., проф., прорект. ПГУПС,
С.-Петербург, РФ
Федоров А. Р., ген. дир. «ДигДез», С.-Петербург, РФ
Юсупов Р. М., проф., чл.-корр. РАН, С.-Петербург, РФ

Редакционная коллегия

Бубнов В. П., проф., С.-Петербург, РФ – зам. гл. ред.
Ададуров С. Е., проф., Москва, РФ
Александрова Е. Б., проф., С.-Петербург, РФ
Атилла Элчи, проф., Аксарай, Турция
Безродный Б. Ф., проф., Москва, РФ
Благовещенская Е. А., проф., С.-Петербург, РФ
Булавский П. Е., д.т.н., доц., С.-Петербург, РФ
Василенко М. Н., проф., С.-Петербург, РФ
Гуда А. Н., проф., Ростов-на-Дону, РФ
Железняк В. К., проф., Полоцк, Беларусь
Заборовский В. С., проф., С.-Петербург, РФ
Зегжда П. Д., проф., С.-Петербург, РФ
Канаев А. К., д.т.н., проф., С.-Петербург, РФ
Котенко А. Г., д.т.н., доц., С.-Петербург, РФ
Куренков П. В., проф., Москва, РФ
Лецкий Э. К., проф., Москва, РФ

Мирзоев Т. А., асс. проф., Джорджия, США
Наседкин О. А., доц., С.-Петербург, РФ
Никитин А. Б., проф., С.-Петербург, РФ
Охтилев М. Ю., проф., С.-Петербург, РФ
Соколов Б. В., проф., С.-Петербург, РФ
Таранцев А. А., проф., С.-Петербург, РФ
Утепбергенов И. Т., проф., Алматы,
Казахстан
Филипченко С. А., доц., Москва, РФ
Фозилов Ш. Х., проф., Ташкент, Узбекистан
Фу-Ниан Ху, проф., Джиангсу, Китай
Хабаров В. И., проф., Новосибирск, РФ
Ходаковский В. А., проф., С.-Петербург, РФ
Чехонин К. А., проф., Хабаровск, РФ
Яковлев В. В., проф., С.-Петербург, РФ
Ялышев Ю. И., проф., Екатеринбург, РФ

Адрес редакции:

190031, Санкт-Петербург, Московский пр., 9, ауд. 1–210
e-mail: itt-pgups@yandex.ru, сайт: <http://itt-pgups.ru>

ISSN 2413-2527

Журнал зарегистрирован Федеральной службой по надзору в сфере связи и массовых коммуникаций,
свидетельство Эл № ФС77-61707 от 07 мая 2015 г.

Журнал зарегистрирован в Российском индексе научного цитирования (РИНЦ).

© Федеральное государственное бюджетное образовательное учреждение
высшего образования «Петербургский государственный университет путей сообщения Императора Александра I», 2019

Разрешается воспроизведение в прессе, а также сообщение в эфире или передача по кабелю опубликованных в составе периодического издания – журнала «Интеллектуальные технологии на транспорте» – статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием автора статьи и сетевого электронного научного периодического издания журнала «Интеллектуальные технологии на транспорте».

Intellectual Technologies on Transport

Issue 4, 2019

Network electronic scientific journal, open access. It publishes articles in Russian and English with the results of research and practical achievements in the field of intelligent technologies and associated research

Founded in 2015

Founder and Publisher

Federal State Educational Institution of Higher Education
«Emperor Alexander I Petersburg State Transport University»

Co-chairs of the Editorial Council

Panychev A. Yu., rector of PSTU, St. Petersburg, Russia
Charkin E. I., director on IT of JSC "RZD", Moscow, Russia

Editor-in-Chief

Khomonenko A. D., Prof., St. Petersburg, Russia

Editorial Council Members

Glukhov A. P., Lead. Res., CCC of JSC «RZD»,
Moscow, Russia

Dudin A. N., Prof., BSU, Minsk, Belarus

Illarionov A. V., advisor, «RFNC-VNIIEF», Sarov,
Russia

Kornienko A. A., Prof., PSTU, St. Petersburg, Russia

Kovalets P., Prof., Tech. University, Warsaw, Poland

Merkuryev Yu. A., Prof., Academician of the
Latvian Academy of Sciences, Riga, Latvia

Nesterov V. M., Prof., St. Petersburg, Russia

Pustarnakov V. F., CEO at «Gazinformservice» LTD.,
St. Petersburg, Russia

Titova T. S., Prof., PSTU, St. Petersburg, Russia

Fedorov A. R., CEO at «Digital Design» LTD.,
St. Petersburg, Russia

Yusupov R. M., Prof., Corr. Member of RAS, St. Petersburg,
Russia

Editorial Board Members

Bubnov V. P., Prof., St. Petersburg, Russia –
Deputy Editor-in-Chief

Adadurov S. E., Prof., Moscow, Russia

Aleksandrova E. B., Prof., St. Petersburg, Russia

Attila Elci, Prof., Aksaray, Turkey

Bezrodny B. F., Prof., Moscow, Russia

Blagoveshenskaya E. A., Prof., St. Petersburg, Russia

Bulavsky P. E., Dr. Sc., As. Prof., St. Petersburg, Russia

Vasilenko M. N., Prof., St. Petersburg, Russia

Guda A. N., Prof., Rostov-on-Don, Russia

Geleznyak V. K., Prof., Polotsk, Belarus

Zaborovsky V. S., Prof., St. Petersburg, Russia

Zegzda P. D., Prof., St. Petersburg, Russia

Kanayev A. K., Prof., St. Petersburg, Russia

Kotenko A. G., Dr. Sc., As. Prof., St. Petersburg, Russia

Kurenkov P. V., Prof., Moscow, Russia

Letsky Ad. K., Prof., Moscow, Russia

Mirzoev T., As. Prof., Georgia, USA

Nasedkin O. A., As. Prof., St. Petersburg, Russia

Nikitin A. B., St. Petersburg, Russia

Okhtilev M. Yu., Prof., St. Petersburg, Russia

Sokolov B. V., Prof., St. Petersburg, Russia

Tarantsev A. A., Prof., St. Petersburg, Russia

Utepbergenov I. T., Prof., Almaty, Kazakhstan

Filipchenko S. A., As. Prof., Moscow, Russia

Fozilov Sh. Kh., Prof., Tashkent, Uzbekistan

Fu-Nian Hu, Prof., Jiangsu, China

Khabarov V. I., Prof., Novosibirsk, Russia

Khodakovskiy V. A., Prof., St. Petersburg, Russia

Chekhnin K. A., Prof., Khabarovsk, Russia

Jakovlev V. V., Prof., St. Petersburg, Russia

Jalyshev Yu. I., Prof., Ekaterinburg, Russia

Editorial adress:

190031, St. Petersburg, Moskovskiy pr., 9, aud. 1–210

e-mail: itt-pgups@yandex.ru, <http://itt-pgups.ru>

ISSN 2413-2527

The journal is registered by the Federal Service for Supervision of Communications and Mass Media,
EL no. FS77-61707 testimony from May 7, 2015.

The journal is registered in the Russian Science Citation Index (RSCI).

© Federal State Educational Institution of Higher Education "Emperor Alexander I Petersburg State Transport University", 2019

The reproduction in the press, as well as a message broadcast or cable published as part of the periodical – journal "Intellectual Technologies on Transport" – articles on current economic, political, social and religious issues with the obligatory indication of the author, and the network of electronic scientific periodical journal "Intellectual Technologies on Transport".

Содержание

Малунова Д. А., Забродин А. В.

Система взаимодействия и поддержки баз данных пользователей системы ГИД "Урал-ВНИИЖТ" 5

Поляков О. М.

Категоризация в естественных языках для глаголов повышенной арности 14

Панков А. В., Дашкевич Я. В.

Интеллектуальный анализ данных для задач идентификации состояний объектов контроля в условиях неопределенности 22

Воробьев Е. Г., Оладе Д. А.

Метод распознавания радиосигналов на основе теории образов 28

Молодкин И. А., Шурыгин Е. М.

Характеристики алгоритма сжатия данных без потерь на основе псевдорегулярных чисел 34

Фоменко Ю. С., Абу Хасан Р., Хомоненко А. Д.

Логическая интеграция данных на примере организации запросов к базе данных авиаперевозок 39

Дергачев А. И., Дергачев С. А., Куранова О. Н., Перепеченов А. М.

Направления информатизации обучения в вузе (на англ.) 48

Кустов В. Н., Станкевич Т. Л.

Трансграничное пространство доверия как составная часть софт-инфраструктуры международной электронной коммерции (на англ.) 53

Памяти коллеги 58

Contents

<i>Malunova D. A., Zabrodin A. V.</i> System of The Interaction and Support of Users Database for GID "Ural-VNIIZHT"	5
<i>Polyakov O. M.</i> Categorization in Natural Languages for Verbs of Increased Arity	14
<i>Pankov A. V., Dashkevich J. V.</i> Data Mining for the Tasks of Identifying the States of Objects Control in Conditions of Uncertainty	22
<i>Vorobiev E.G., Olaode J. A.</i> Method Radio Signals Detection on the Image Recognize Theory.....	28
<i>Molodkin I. A., Shurygin E. M.</i> Characteristics of a Lossless Data Compression Algorithm Based on Pseudo-Regular Numbers.....	34
<i>Fomenko Yu. S., Abou Hasan R., Khomonenko A. D.</i> Logical Data Integration on the Example of Organizing Queries to the Database Air Transportation	39
<i>Dergachev A. I., Dergachev S. A., Kuranova O. N., Perepechenov A. M.</i> Directions of Informatization of Education at the University (in English)	48
<i>Kustov V. N., Stankevich T. L.</i> Trunboundary Trust Space as a Component of an International E-commerce Soft-Infrastructure (in English)	53
In Memory of Colleague	58

Система взаимодействия и поддержки баз данных пользователей системы ГИД "Урал-ВНИИЖТ"

магистрант Д. А. Малунова, к.и.н. А. В. Забродин
Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
darya.malunova@gmail.com

Аннотация. Рассматривается вопрос автоматизации систем диспетчерского управления перевозками на предприятиях ОАО «РЖД», раскрываются этапы процесса ведения списка пользователей системы, отмечены положительные составляющие результата внедрения автоматизации данного производственного процесса.

Основное внимание акцентируется на моделировании предметной области с использованием унифицированного языка моделирования UML и практической реализации проекта на языке JavaScript. Приводятся примеры программной реализации отдельных компонентов системы, обобщается практический опыт внедрения и тестирования разработанной системы в Ярославском информационно-вычислительном центре — структурном подразделении Главного вычислительного центра — филиала ОАО «РЖД» в отделе «Автоматизированных систем диспетчерского управления перевозками».

Статья может быть полезна специалистам в области разработки автоматизированных систем управления.

Ключевые слова: автоматизация систем диспетчерского управления, ведение списка пользователей, АС ГИД, моделирование с использованием UML, MS SQL Server, скрипт синхронизации, JavaScript, web-приложение.

ВВЕДЕНИЕ

Современные информационные технологии помогают не только усовершенствовать технологии на производстве, но и значительно их облегчить. Для получения наиболее высоких результатов своей деятельности предприятия переходят на автоматизацию технологических процессов.

Автоматизация – одно из важнейших направлений развития любой отрасли. Введение автоматизации на предприятии позволяет значительно повысить производительность труда, обеспечить стабильное качество производимых продуктов и услуг, а также исключить ошибки, которые могут быть вызваны человеческим фактором.

При этом существует достаточное количество организаций (предприятий), на которых в силу субъективных

(или объективных) обстоятельств была проведена частичная автоматизация, т. е. автоматизация какого-либо отдельного оборудования и производственных операций, при этом значительная доля операций соотнесена с ручным трудом [1].

В качестве экспериментальной площадки выбран отдел «Автоматизированных систем диспетчерского управления перевозками» Ярославского информационно-вычислительного центра, в котором сотрудники сталкиваются с проблемой, заключающейся в отсутствии единой автоматизированной системы, позволяющей осуществить ведение списка пользователей системы ГИД «Урал-ВНИИЖТ» (АС ГИД).

До некоторого времени процесс администрирования, а именно ведение базы пользователей, велся вручную с помощью таблиц в MS Excel. В интересах администраторов потребовалась разработка универсального модуля, позволяющего экстерриториально иметь доступ к информации о подключенных пользователях, осуществлять их поиск и оперативно отслеживать сроки действия заявок на подключение [2–4].

ФУНКЦИОНАЛЬНАЯ СТРУКТУРА СИСТЕМЫ

На основе проведенного анализа работы сотрудников и систем предприятия выявлены требования для создаваемого модуля, представленные в таблице 1.

Таблица 1

Требования к разрабатываемой системе

Требования к базе данных	Требования к web-интерфейсу
Расположение на едином информационном ресурсе, доступном всем администраторам;	Возможность записи информации, не интегрированной в автоматическом режиме из других систем;
Содержание необходимой информации о каждом пользователе системы;	Разграничение прав доступа (обычный пользователь имеет только возможность просмотра базы, а администратор может не только просматривать данные в базе, но и редактировать их).
Возможность автоматического наполнения при включении сотрудников в соответствующие группы для подключения к АС ГИД.	



Рис. 1. Диаграмма вариантов использования модуля администрирования

На рисунке 1 представлена диаграмма вариантов использования разрабатываемого модуля администрирования в нотации языка UML, учитывающая функциональные требования системы.

Функциональная структура модуля администрирования пользователей АС ГИД представлена в виде диаграммы компонентов на рисунке 2 также в нотации языка UML [5–7].

Web-приложение основано на технологии ASP с использованием языка сценариев JavaScript и включает в себя три составляющие: интерфейс для отображения списка пользователей (gid_all.asp), страницу для отображения журнала изменений (gid_log_all.asp) и файл, вы-

полняющий интеграцию данных из базы пользователей в веб-форму (gid_all_js.asp) [8].

За работу со стороны сервера отвечает система управления базами данных Microsoft SQL Server. База пользователей GID_28 состоит из 2 основных таблиц, которые учитывают выявленные особенности для администратора: одна таблица предназначена для хранения всех данных о подключенных пользователях системы (gid_all_28), а другая — для хранения данных о любых изменениях, связанных с информацией о пользователях (gidlog_all_28) [9, 10].

Скрипт автоматизирует процесс синхронизации данных из базы инвентаризации и Active Directory в базу пользователей (gid_all.exe).

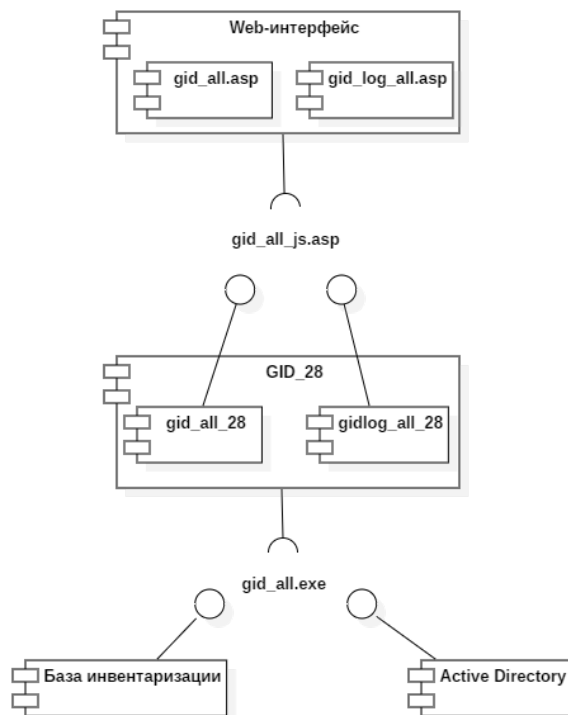


Рис. 2. Функциональная структура подсистемы обработки заявок на подключение к АС ГИД

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Ключевым звеном в данной разработке является скрипт синхронизации gid_all.exe, который наполняет базу GID_28 информацией, полученной из базы инвентаризации (IP-адрес, имя ПК, дата инвентаризации, наименование организации, отдела, должность, телефон) и системы Active Directory (ФИО пользователя, логин, терминальные и сетевые группы). Данные действия автоматизируют процесс синхронизации данных и позволяют исключить ручной ввод.

В выполнении скрипта gid_all.exe заложены следующие основные функции:

- чтение данных из базы инвентаризации;
- анализ групп пользователей из AD;
- сопоставление полученных данных из двух источников на их соответствие, достоверность и актуальность;

- определение принадлежности объекта (пользователя) к группе;
- удаление учетных записей пользователей из базы.

Чтение данных из базы инвентаризации. Функция ReadTable() предназначена для чтения данных из базы инвентаризации и заключается в подсчете количества пользователей, имеющих изначально в базе.

При ее вызове выполняются следующие действия:

- выполнение выборки числа пользователей, в частности их логинов samaccountname, из таблицы gid_all с помощью SQL-запроса;
- считывание всех строк таблицы gid_all в массив с дальнейшей записью данных об SID (уникальных идентификаторов пользователей) в справочник User_SID с логинами пользователей и в User_account с соответствующими им номерами строк в исходной таблице.

Код программной реализации функции представлен на рисунке 3.

```

Sub ReadTable()
    Console.WriteLine("Reading SQL table... ")
    ...
    SQLGID.Open()
    If SQLGID.State = ConnectionState.Open Then
        InsertCommandGID.CommandText = "SELECT COUNT(samaccountname) FROM gid_all_28"
        count = InsertCommandGID.ExecuteScalar
        ReDim data_rows(count, 21)
        InsertCommandGID.CommandText = "SELECT ... FROM gid_all_28"
        SQLReader = InsertCommandGID.ExecuteReader
        Dim i As Integer = -1
        While SQLReader.Read
            i += 1
            data_rows(i, 0) = SQLReader(0) ' rvc
            ...
            data_rows(i, 21) = SQLReader(21) ' dnc
            User_SID.Add(data_rows(i, 7), i)
            User_account.Add(data_rows(i, 1), i)
        End While
    ...
End Sub
    
```

Рис. 3. Код функции чтения данных из базы инвентаризации

Чтение данных из Active Directory. Функция ReadGroups() служит для чтения и выборки групп пользователей из Active Directory. Для полученных данных проводится проверка на соответствие, достоверность и актуальность с данными из базы инвентаризации, учитывая следующие возможные ситуации:

- добавление новой записи (если в базе нет учетной записи с таким логином и идентификатором);
- изменение идентификатора (если учетные записи имеют одинаковые логины, но разные идентификаторы);
- обновление данных о пользователе (пользователь с данным логином и идентификатором уже есть в таблице);

- обновление данных о пользователе при изменении членства в группах;
- обновление старых данных;
- обновление данных, если были изменены ФИО и/или логин пользователя;
- обновление данных, если был изменен только логин пользователя;
- обновление данных, если были изменены только ФИО пользователя.

Код программной реализации функции представлен на рисунке 4.

```

Sub ReadGroups()
    Dim mySearcher As DirectorySearcher = New DirectorySearcher(entry)
    mySearcher.Filter = "(&(objectCategory=user)(...)"
    Console.WriteLine("Searching in Active Directory... ")
    Dim Find_Users As Integer = mySearcher.FindAll.Count
    Console.WriteLine(Find_Users)

    ...

    If SQLConn.State = ConnectionState.Open Then

        ...

        If Not obj.GetDirectoryEntry.Properties("samaccountname").Value Is Nothing Then
            samaccountname = obj.GetDirectoryEntry.Properties("samaccountname").Value
        End If

        ...

        InsertCommandGID.CommandText = "INSERT INTO gid_all_28(...) VALUES (" & obj1 & "," & obj2 & "," & obj3 & "," & obj4 & "," & obj5 & "," &
obj6 & "," & obj7 & "," & term & "," & dsp & "," & dnc & ")" , 1)
        Console.WriteLine(counter & "/" & Find_Users & " INSERT: ...")

        ...

        ElseIf User_account.Contains(samaccountname) And Not User_SID.Contains(SID) Then
            Get_IP_Hostname_RVC(obj, SQLConn, InsertCommandIPDB, samaccountname, "", "")
            If TypeOf obj.GetDirectoryEntry.Properties("MemberOf").Value Is String Then
                GetObjects(obj.GetDirectoryEntry.Properties("MemberOf").Value.ToString)
            Else
                For Each group_dn As String In obj.GetDirectoryEntry.Properties("MemberOf").Value
                    GetObjects(group_dn)
                Next
            End If
            InsertCommandGID.CommandText = "UPDATE gid_all_28 SET " & obj1 & " WHERE samaccountname=" & samaccountname & ""
            InsertCommandGID.ExecuteNonQuery()
            InsLog(samaccountname, displayname, SID, DateToLong(Now), "RECREATED", 5)
            Console.WriteLine(counter & "/" & Find_Users & " RECREATED: ...")

            ...

            ElseIf displayname <> data_rows(User_SID.Item(SID), 2) And samaccountname = data_rows(User_SID.Item(SID), 1) Then
                InsertCommandGID.CommandText = "UPDATE gid_all_28 SET " & obj1 & " WHERE SID = " & SID & ""
                InsertCommandGID.ExecuteNonQuery()
                InsLog(samaccountname, displayname, SID, DateToLong(Now), "CHANGE NAME: " & data_rows(User_SID.Item(SID), 2) & " -> " & displayname,
4)
                Console.WriteLine(counter & "/" & Find_Users & " CHANGE NAME: " & data_rows(User_SID.Item(SID), 2) & " -> " & displayname)
                SQLGID.Close()
            End If
        End If
    End If

    ...

End Sub

```

Рис. 4. Код функции чтения данных из Active Directory

Получение информации о типе доступа пользователя. Функция GetObject() представляет собой процесс получения принадлежности объекта (пользователя) к группе.

Для идентификации пользователей используются следующие маркеры:

- при идентификации по принадлежности подключения используются такие значения, как «1», если пользователь включен в группы терминального или сетевого доступа, и «0», если не включен в данные группы.

- при идентификации пользователей по типу подключения используются такие значения, как «W» (запись), если подключение технологическое, и «R» (чтение), если информационное.

Данные маркеры проставляются в соответствующих столбцах таблицы gid_all в зависимости от типа доступа.

Код программной реализации функции представлен на рисунке 5.

```
Sub GetObjects(ByVal Value As String)
  If InStr(Value, "CN=NRR-GID-DSP-282") = 1 Or InStr(Value, "CN=NRR-GID-DCUP-282-W") = 1 Or InStr(Value, "CN=NRR-TS-GID-282-W") = 1 Then
    obj1 = "W"
  End If
  If InStr(Value, "CN=NRR-GID-DCUP-282-R") = 1 Or InStr(Value, "CN=NRR-TS-GID-282-R") = 1 Then
    obj1 = "R"
  End If
  ...
  If InStr(Value, "CN=NRR-TS-GID-282-R") = 1 Or InStr(Value, "CN=NRR-TS-GID-282-W") = 1 Or InStr(Value, "CN=NRR-TS-GID-284-R") = 1 Or InStr(Value,
"CN=NRR-TS-GID-284-W") = 1 Or InStr(Value, "CN=NRR-TS-GID-285-R") = 1 Or InStr(Value, "CN=NRR-TS-GID-285-W") = 1 Or InStr(Value, "CN=NRR-TS-
GID-SEV") = 1 Or InStr(Value, "CN=NRR-TS-GID-SEV-R") = 1 Or InStr(Value, "CN=NRR-TS-GID-SEV-W") = 1 Or InStr(Value, "CN=NRR-TS-GID-SEV-NOTKS-W")
= 1 Then
    term = 1
  End If
  ...
End Sub
```

Рис. 5. Код функции получения информации о типе доступа пользователя

```
Sub CheckUsers()
  Console.WriteLine(" Searching non-existing users... ")
  ...
  InsertCommand.CommandText = "SELECT ..._ "FROM gid_all_28"
  Dim SQLReader As SqlDataReader = InsertCommand.ExecuteReader
  Dim FindAll_Count As Integer = 0
  While SQLReader.Read
    mySearcher.Filter = "(objectCategory=user)(samaccountname=" & SQLReader(0) & ")"
    FindAll_Count = mySearcher.FindAll.Count
    If FindAll_Count = 0 Then
      InsLog(SQLReader(0), SQLReader(1), SQLReader(2), DateToLong(Now), "DELETE: " ...)
      DeleteUser(SQLReader(0))
    ElseIf FindAll_Count = 1 Then
      ...
      If Not obj.GetDirectoryEntry.Properties("MemberOf").Value Is Nothing Then
        If TypeOf obj.GetDirectoryEntry.Properties("MemberOf").Value Is String Then
          GetObjects(obj.GetDirectoryEntry.Properties("MemberOf").Value)
        Else
          For Each group_dn As String In obj.GetDirectoryEntry.Properties("MemberOf").Value
            GetObjects(group_dn)
          Next
        End If
      End If
    ElseIf FindAll_Count > 1 Then
      Write_log("Найдено более одного совпадения имени " & samaccountname & " в AD")
    End If
    If obj1 = "" And obj2 = "" And obj3 = "" And obj4 = "" And obj5 = "" And obj6 = "" And obj7 = "" And term = 0 And dsp = 0 And dnc = 0 Then
      InsLog(SQLReader(0), SQLReader(1), SQLReader(2), DateToLong(Now), "DELETE: " & ...)
      DeleteUser(SQLReader(0))
      Console.WriteLine("DELETE: " & SQLReader(0))
    End If
  End While
End Sub

Sub DeleteUser(ByVal samaccountname As String)
  ...
  InsertCommand.CommandText = "DELETE FROM gid_all_28 WHERE samaccountname = '" & samaccountname & "'"
  ...
End Sub
```

Рис. 6. Код функции удаления данных о пользователях

Удаление данных о пользователях. Функция CheckUsers() осуществляет выбор учетных записей пользователей для удаления из таблицы gid_all.

При ее вызове выполняются следующие действия:

- создается фильтр для проверки логинов пользователя и специальная переменная FindAll_count для подсчета количества совпадений;

- при сопоставлении учетных записей из базы GID_28 и AD могут возникнуть следующие ситуации:

а) если совпадений не найдено, значит, в базе пользователей есть записи, отсутствующие в AD, и их необходимо удалить;

б) если совпадение одно, то происходит сверка данных;

в) если совпадений найдено больше одного, то лишняя запись удаляется.

Функция DeleteUser() уже удаляет учетные записи пользователя из таблицы gid_all. При ее вызове выполня-

ется соответствующий SQL запрос на удаление учетной записи пользователя с конкретным логином из gid_all.

Код программной реализации функции представлен на рисунке 6.

После успешного завершения выполнения скрипта, таблицы gid_all_28, gidlog_all_28 и gidcnt_all_28 базы данных GID_28 заполняются соответствующими данными о подключенных пользователях [11, 12].

ТЕХНОЛОГИЯ ВЕДЕНИЯ СПИСКА ПОЛЬЗОВАТЕЛЕЙ

Процесс ведения администратором списка пользователей можно условно разделить на два основных этапа.

Первый этап — получение данных о пользователях системы. Для этого необходимо запустить скрипт синхронизации, с помощью которого произойдет наполнение базы данных информацией, полученной из базы инвентаризации и системы Active Directory. Ход выполнения данного процесса представлен на рисунке 7.

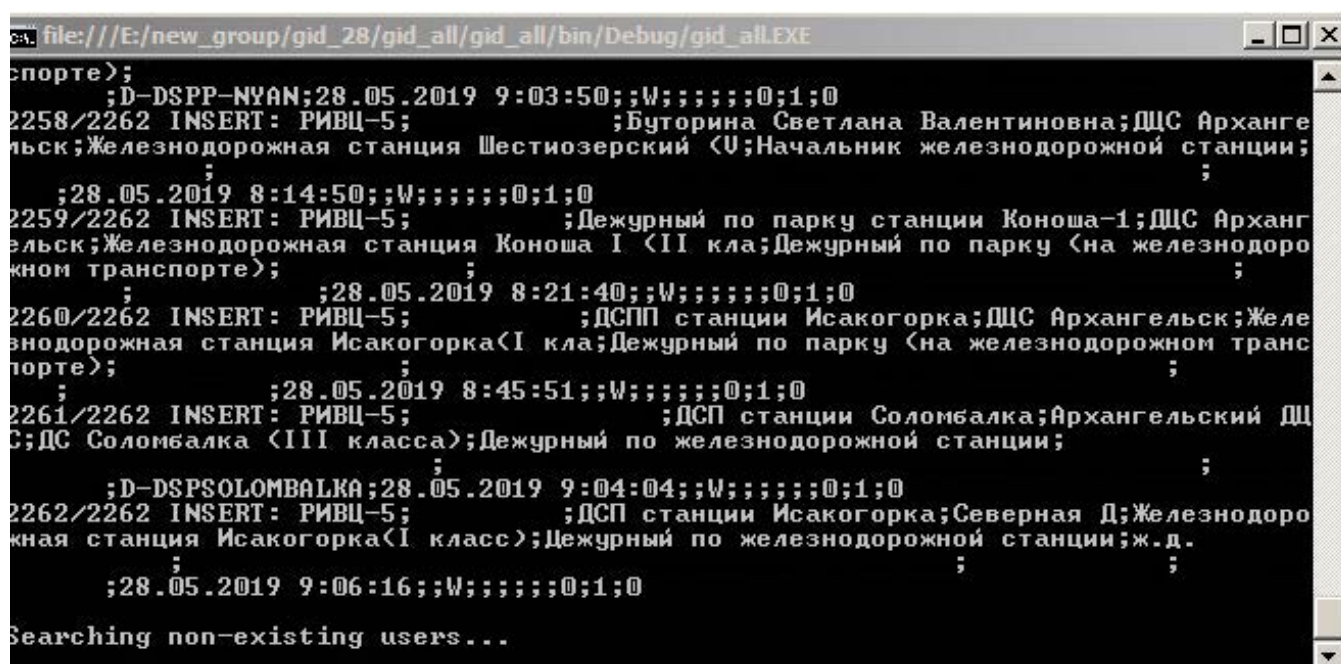


Рис. 7. Процесс заполнения базы пользователей

Второй этап — открытие web-приложения и редактирование полученного списка пользователей.

Для получения необходимой информации о пользователях сначала необходимо выбрать количество отображаемых строк в таблице. С целью обеспечения быстроты действия веб-страницы первоначально загружаются данные для 50 пользователей. После того как произойдет выбор количества отображаемых пользователей, стартовая страница автоматически обновится, и будет представлена таблица с данными о подключенных пользователях. Для администраторов есть возможность редактирования любого поля базы прямо на странице. Внесенные или отредактированные данные автоматически обновляются и сохраняются в базе данных.

На рисунке 8 представлена стартовая страница полученного сайта, на которой отображается заполненный список пользователей.

В дополнение к списку пользователей организована еще такая страница, как «Журнал изменений», представленная на рисунке 9.

Журнал отражает информацию об изменениях, связанных с учетными записями подключенных пользователей системы. Здесь автоматически выгружены такие данные, как: логин, ФИО, уникальный идентификатор пользователя и произошедшие с ним изменения — события и их дата. В зависимости от типа произошедшего изменения информации о пользователе (удаление, изменение, добавление и т. п.) соответствующая строка в таблице окрашивается в различной цветовой гамме. Это позволяет визуально группировать события и улучшает восприятие информации со стороны пользователя.

РЖД Российские железные дороги

Пользователи ГИД Урал: 50
Показать: 50 строк на странице

Обновить Журнал изменений

№	РИВЦ	ФИО пользователя	login	Почтовая папка	Автоответ	Номер заявки	Дата АССОЗ	Ссылка на заявку в АССОЗ	Примечание	IP-адр
1		3 парк ДС Ярослав Главной	and1_bark-doyarg	ur_park3_yurTAR_PARK3	(9 31000 7)64	544734/2018 532344/2018	08.05.2020 04.05.2020	7DD2F40C-5EB9-4234-A7E9-7209B3CB855C_F6520 F8CFD57E-2414-4B14-B45A-656B21FAB476 C930F492-94C3-46DC-8E48-99D0BDBE289A	ХОЛМОВА ЕЛЕНА СЕРГ НАРЫШКИНА ИРИНА В. ЗАНЕГИНА ВИКТОРИЯ И	
2		Абакумов Александр Валерьевич	t-AbakumovAV			658907/2018 1393367/2018	10.04.2020 30.10.2020	D3B8959F-D6E6-4660-A373-E553486B08B2 9E3E CDC8-8808-44D7-8981-4F15055A5029 http://msk-asoz-02.msk.oao.rzd/1240/RequestInfo.asp		
3		Абросимова Наталия Николаевна	d-AbrosimovaNN			780/2018	03.01.2020	1C0A071F-516B-4E5D-BE82-69A03CD92F90	Письмо по ЭП ответствен	
4		Альмеева Юлия Александровна	nik-AltmeevaYA			924287/2017	18.09.2019	20B6C0FC-77CA-4092-A39C-C19625DEB061		
5		Апушкина Светлана Александровна	d-ApushkinaSA	D-GAVRILYUKIN	(9 30560 1)91	1365737/2017	19.12.2019	5BAB82B7-2FBF-4F2C-A096-FFCE2E0395B5		
6		Богданова Татьяна Витальевна	t-BogdanovaTV			440237/2018	11.04.2020	E6B8CDA9-2722-4CF3-83CB-06CCA0511EEB		
7		Богодолова Елена Николаевна	d-BogodolovaEN		(9 31340 6)91	1108263/2018	11.09.2020	C19BB225-SDFE-47C5-B575-23306F98B4F	(0914-32) 7-39-64	
8		Борков Сергей Алексеевич	ds-BorkovSA			863185/2017	31.08.2019	8F1381BD-3D8E-4B1F-A80D-6739522E21E3		
9		Васильев Евгений Сергеевич	ds-VasilevES			1379615/2017 1420021/2017	21.12.2019 09.01.2020	B3B357F1-1558-415E-BB8F-9A2435D3C6A5 9CBB91F1-8341-479B-9F2B-AC3889434289		

Рис. 8. Список пользователей системы АС ГИД

РЖД Российские железные дороги

Журнал изменений

№	login	ФИО пользователя	Дата	Событие
0	t-KomarovAK	Комаров Андрей Николаевич	27.05.2019 15:19:54	CHANGE MEMBERSHIP: (1,0,0,0,0,0,1) -> (1,0,0,1,0,0,1)
1	t-PavlovAG	Павлов Александр Георгиевич	27.05.2019 15:19:40	INSERT: (1,0,0,0,0,1,0,1)
2	d-StavdaAV	Ставида Андрей Владимирович	27.05.2019 15:18:41	INSERT: (0,0,1,0,0,0,0,1)
3	t-ZlatoustovAY	Златоустов Александр Юрьевич	27.05.2019 15:17:37	INSERT: (1,1,1,1,1,1,1,1)
4	d-DoyVandib	ДСП ст.Вандыш	25.05.2019 11:20:05	CHANGE MEMBERSHIP: (0,1,0,0,0,0,0,0) -> (0,1,0,0,0,0,0,1)
5	ds-KuznetsovKP	Кузнецова Ксения Павловна	24.05.2019 11:20:40	INSERT: (0,0,0,1,0,0,0,1)
6	d-MuravevaOI	Муравьева Ольга Сергеевна	24.05.2019 11:19:30	INSERT: (1,0,0,0,0,0,0,1)
7	d-OskochnikovKP	Оскочиников Константин Павлович	24.05.2019 11:19:27	INSERT: (1,0,0,1,0,0,0,1)
8	d-GagarinovaEG	Гагарина Елена Геннадьевна	24.05.2019 11:18:11	INSERT: (0,0,1,0,0,0,0,1)
9	d-KokoshinAA	Кокочин Руслан Анатольевич	24.05.2019 11:17:48	CHANGE MEMBERSHIP: (1,1,1,1,0,0,1,1) -> (0,1,1,1,0,0,1,1)
10	d-PoroshinAS	Порошин Антон Сергеевич	24.05.2019 7:20:03	INSERT: (1,0,0,1,0,0,0,1)
11	d-KokoshinAA	Кокочин Руслан Анатольевич	24.05.2019 7:17:49	CHANGE MEMBERSHIP: (0,0,1,1,0,0,1,1) -> (1,1,1,1,0,0,1,1)
12	tsv-raming	Учебный класс Ярославского ИВЦ	23.05.2019 15:19:24	INSERT: (1,0,0,1,0,0,0,1)
13	tsv-ZhukovaVV	Жукова Вера Валентиновна	23.05.2019 11:17:31	CHANGE MEMBERSHIP: (1,1,1,1,1,0,0,1) -> (1,1,1,1,1,1,1,1)
14	d-DmitryevV	Ильинская Любовь Владимировна	23.05.2019 7:22:17	DELETE: (1,0,0,1,0,0,0,1) РИВЦ-4, 3233303334332F32303138
15	tsv-LyubimovNN	Любицкий Николай Николаевич	22.05.2019 15:20:33	CHANGE MEMBERSHIP: (1,0,0,1,0,0,0,1) -> (0,0,0,1,0,0,0,1)
16	t-ShalunovAI	Шальников Алексей Иванович	22.05.2019 15:20:06	INSERT: (1,0,0,1,1,0,0,1)
17	t-MukhaIV	Муха Сергей Владимирович	22.05.2019 11:19:51	CHANGE MEMBERSHIP: (0,1,0,1,0,1,0,1) -> (0,1,0,0,0,1,0,1)
18	t-KomarovAK	Комаров Андрей Николаевич	22.05.2019 11:19:47	INSERT: (1,0,0,0,0,0,0,1)
19	t-VasilevAV	Васильев Сергей Валерьевич	22.05.2019 7:20:58	INSERT: (1,0,0,1,0,1,0,1)
20	ds-PoroshinAS	Порошин Сергей Владимирович	21.05.2019 15:22:39	DELETE: (0,1,0,0,0,0,0,1) РИВЦ-5, 3138323533332F32303139
21	tsv-VasilevAI	Васильева Ирина Сергеевна	21.05.2019 15:21:53	DELETE: (0,0,1,1,0,0,0,1) РИВЦ-7, 3337333933332F32303137
22	d-RogulovAV	Рогулин Артем Владимирович	21.05.2019 15:21:39	DELETE: (1,0,0,1,1,1,1,1) РИВЦ-1, 345634393732F32303138
23	d-OstankovAA	Останков Алексей Анатольевич	21.05.2019 15:20:04	CHANGE MEMBERSHIP: (0,0,1,0,0,0,0,1) -> (0,0,1,1,0,0,0,1)
24	tsv-VinogradovFN	Виноградов Федор Николаевич	21.05.2019 15:19:16	INSERT: (1,1,1,1,1,1,1,1)
25	d-BelavaIV	Белая Максим Викторович	21.05.2019 15:19:13	INSERT: (1,0,0,0,0,0,0,1)
26	d-ShalunovAI	Шальников Александр Алексеевич	21.05.2019 15:17:44	CHANGE MEMBERSHIP: (0,0,1,1,0,0,0,1) -> (1,0,1,1,0,0,0,1)
27	d-CentRashch-Pc12	Центр расширенной работы место 12	21.05.2019 7:23:27	DELETE: (1,1,1,1,1,1,1,1), 3639303930362F32303138
28	d-YarovAA	Яров Артем Андреевич	21.05.2019 7:23:20	DELETE: (1,0,0,0,0,0,0,1), 313238343431342F32303136553834383430372F32303138
29	t-MuryanII	Мурьян Иван Иванович	20.05.2019 7:22:36	DELETE: (1,1,1,0,0,0,0,1) РИВЦ-5, 3337343938322F32303135553335313330342F32303135313938393732F32303139
30	tsv-KabanovaGV	Кабанова Галина Вениаминовна	20.05.2019 7:22:36	DELETE: (0,1,0,0,0,0,0,1) РИВЦ-5, 3335313034572F3230313755
31	d-SuldanovaEV	Султанова Екатерина Владимировна	18.05.2019 7:23:01	DELETE: (0,0,0,1,0,0,0,1) РИВЦ-1, 313530353236362F3230313855
32	d-GomarovAP	Гомаров Андрей Павлович	18.05.2019 7:21:33	INSERT: (0,0,0,1,0,0,1,1)
33	d-AlmazovaAM	Алмаздак Александра Михайловна	18.05.2019 7:20:01	INSERT: (1,0,0,0,0,0,0,1)
34	d-PopovVS	Попов Василий Степанович	17.05.2019 15:22:25	DELETE: (0,0,1,0,0,0,0,1) РИВЦ-6, 38333834332F32303139
35	d-NizhdorovaAV	Низдорова Виктория Александровна	17.05.2019 15:22:19	DELETE: (1,1,1,1,1,1,1,1) РИВЦ-1, 3231333830392F32303139
36	tsv-cv-01	Сменный инженер сектора ЦТР ЦУТР СП	17.05.2019 15:21:42	DELETE: (0,0,0,1,1,1,1,1) ИВЦ (собств.), 3932303730322F32303138553838323530382F32303138
37	d-VinogradovFN	Виноградов Федор Николаевич	17.05.2019 11:22:17	DELETE: (1,0,0,0,1,1,1,1) РИВЦ-1,
38	tsv-dl-SerZap	Диспетчер локомотивный Северо-западного района ЦУТР СП	17.05.2019 11:21:37	DELETE: (1,1,1,1,1,1,1,1) ИВЦ (собств.), 3938363338322F323031385531353230343342F32303138555555
39	d-SharovIV	Шутов Илья Юрьевич	17.05.2019 11:20:28	INSERT: (0,0,0,1,0,0,0,1)
40	d-EgorovEV	Егоров Евгений Валерьевич	16.05.2019 11:20:21	CHANGE MEMBERSHIP: (0,1,0,0,0,0,0,1) -> (0,1,0,1,0,0,0,1)
41	d-AlexeevNP	Алексеев Николай Павлович	15.05.2019 15:19:12	INSERT: (1,0,0,0,0,1,1,1)
42	d-EgorovEV	Егоров Евгений Валерьевич	15.05.2019 7:20:47	INSERT: (0,1,0,0,0,0,0,1)
43	d-RogulovAV	Рогулов Дмитрий Владимирович	15.05.2019 7:19:39	INSERT: (0,0,1,0,0,0,0,1)
44	t-KabanovaGV	Кабанова Екатерина Васильевна	15.05.2019 7:19:20	INSERT: (1,0,0,0,1,0,1,1)
45	t-NikolaevAA	Некрасов Александр Алексеевич	15.05.2019 7:18:07	INSERT: (1,0,0,1,0,1,0,1)

Рис. 9. Журнал изменений списка пользователей системы АС ГИД

ЗАКЛЮЧЕНИЕ

Рассмотренная подсистема обработки заявок автоматизирует процесс ведения списка пользователей АС ГИД и позволяет в режиме реального времени видеть подключенных пользователей с любого рабочего места администраторов системы. Это дает следующие преимущества:

- возможность контроля больших данных (о более чем 5 000 пользователях);
- возможность экстерриториального использования (т. е. возможность соответствующего доступа ко всем подразделениям ИВЦ, находящимся не только в пределах одной железной дороги);
- увеличение эффективности использования рабочего времени администраторов на 10 %, что приведет к повышению производительности всего отдела.

ЛИТЕРАТУРА

1. Хлебных Л. В., Зубкова М. А., Саукова Т. Ю. Автоматизация производства в современном мире // Молодой ученый. 2017. № 16 (150). С. 308–310. URL: <http://moluch.ru/archive/150/42390> (дата обращения 01.11.2019).
2. Положение об отделе автоматизированных систем диспетчерского управления перевозками Ярославского информационно-вычислительного центра от 04.06.2018.
3. Руководство по организации сопровождения «Автоматизированная система ведения и анализа графика исполненного движения ГИД "Урал-ВНИИЖТ"», 2018.
4. Кузнецов Г. А., Крашенинников С. В., Крайсвитный В. П., Матвеев Д. А. Модернизация системы ГИД "Урал-ВНИИЖТ" // Автоматика, связь, информатика. 2016. № 11. С. 15–19.
5. Леоненков А. В. Самоучитель UML. — 2-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2015. — 418 с.

6. Основы проектирования информационных систем с помощью языка UML: Учебное пособие / А. В. Забродин, В. П. Бубнов. — СПб.: ПГУПС, 2018. — 46 с.

7. Balakina E. P., Baranov L. A., Sidorenko V. G. Automation of Business Processes of the Trains Movement Control on Moscow Underground. In: *Complementary proceedings of the 8th Workshop on Transformation & Engineering of Enterprises (TEE 2014), and the 1st International Workshop on Capability-oriented Business Informatics (CoBI 2014) co-located with the 16th IEEE International Conference on Business Informatics (CBI 2014), Geneva, Switzerland, July 14-15, 2014*, Vol. 1182, 2014. URL: <http://ceur-ws.org/Vol-1182/paper5.pdf> (дата обращения 01.11.2019).

8. Мэрдок К. Л. JavaScript™. Наглядный курс создания динамических Web-страниц = JavaScript™. Your visual blue print for building dynamic Web pages / Пер. с англ. и ред. В.М. Неумоина. — М.: Компьютерное издательство «Диалектика». Издательский дом «Вильямс», 2001. — 288 с.

9. Delaney K. et al. Microsoft SQL Server 2012 Internals, USA, Microsoft Press, 2013, 982 p.

10. Грофф Д. Р., Вайнберг П. Н., Оппель Э. Дж. SQL. Полное руководство. Третье издание. — М.: Издательский дом «Вильямс», 2018. — 960 с.

11. Объектно-ориентированный анализ и проектирование с примерами приложений. Третье издание = Object-oriented analysis and design with application. Third edition / Г. Буч [и др.]; пер. с англ. Д. А. Ключина. — М.: ООО «И.Д. Вильямс», 2008. — 720 с.

12. Haverbeke M. Eloquent JavaScript: A Modern Introduction to Programming, San Francisco, No Starch Press Inc., 2011, 224 p.

System of the Interaction and Support of Users Database for GID "Ural-VNIIZhT"

Master's Degree student D. A. Malunova, PhD A. V. Zabrodin
Emperor Alexander I St. Petersburg State Transport University
St. Petersburg, Russia
darya.malunova@gmail.com

Abstract. Discusses the current issue of automation of dispatching control systems at the enterprises of JSC "Russian Railways", reveals the stages of the process of maintaining the list of users of the system and notes the positive components from the result of the implementation of automation of this production process.

The main focus of the work is on modeling the subject area using unified UML modeling language and practical implementation of the project in JavaScript language. Examples of the program implementation of individual components of the system are also given, practical experience of implementation and testing of the developed system in the Yaroslavl Information and Computing Center, structural division of the Main Computing Center - branch of JSC "Russian Railways" in the department "Automated systems of dispatching control of transportation".

This article will be of interest to specialists in the field of development of automated control systems.

Keywords: automation of dispatch control systems, maintaining a list of users, AS GUIDs, modeling using UML, MS SQL Server, synchronization script, JavaScript, web-application.

REFERENCES

1. Khlebenskikh L. V., Zubkova M. A., Saukova T. Yu. Automation of Production in the Modern World [Avtomatizatsiya proizvodstva v sovremennom mire], *Young Scientist [Molodoy ychenyy]*, 2017, No. 16, Pp. 308–310. Available at: <http://moluch.ru/archive/150/42390> (accessed 01 Nov. 2019).
2. Regulation on the Department of Automated Systems of Dispatch Management of Transportation of the Yaroslavl Information and Communication Center [Polozhenie ob otdel' avtomatizirovannykh sistem dispetcherskogo upravleniya perevozkami Yaroslavskogo informatsionno-vychislitel'nogo tsentra] dated on June 4, 2018.
3. Guidelines for the organization of escort «Automated system for maintaining and analyzing the schedule of executed traffic GID "Ural-VNIIZhT"» [Rukovodstvo po organizatsii soprovozhdeniya «Avtomatizirovannaya sistema vedeniya i analiza grafika ispolnennogo dvizheniya GID "Ural-VNIIZhT"»], 2018.

4. Kuznetsov G. A., Krashennnikov S. V., Kraivitskiy V. P., Matveev D. A. Modernization of the GUIDAL system "Ural-VNIIZhT" [Modernizatsiya sistemy GID "Ural-VNIIZhT"], *Automation, Communication and Informatics [Avtomatika, svyaz', informatika]*, 2016, No. 11, Pp. 15–19.

5. Leonenkov A. V. UML. The self-teacher [Samouchitel' UML], St. Petersburg, BHV-Petersburg, 2015, 418 p.

6. Zabrodin A. V., Bubnov V. P. Basics of designing information systems using the UML language: Study guide [Osnovy proektirovaniya informatsionnykh sistem s pomoshch'yu yazyka UML: Uchebnoe posobie], Saint Petersburg, PSTU, 2018, 46 p.

7. Balakina E. P., Baranov L. A., Sidorenko V. G. Automation of Business Processes of the Trains Movement Control on Moscow Underground. In: *Complementary proceedings of the 8th Workshop on Transformation & Engineering of Enterprises (TEE 2014), and the 1st International Workshop on Capability-oriented Business Informatics (CoBI 2014) co-located with the 16th IEEE International Conference on Business Informatics (CBI 2014), Geneva, Switzerland, July 14-15, 2014*, Vol. 1182, 2014.

Available at: <http://ceur-ws.org/Vol-1182/paper5.pdf> (accessed 01 Nov. 2019).

8. Murdock K. L. JavaScript™. Your visual blue print for building dynamic Web pages [JavaScript™. Naglyadnyy kurs sozdaniya dinamicheskikh Web-stranits], Moscow, Dialektika Computer Publishing, Williams Publishing House, 2001, 288 p.

9. Delaney K., et al. Microsoft SQL Server 2012 Internals, USA, Microsoft Press, 2013, 982 p.

10. Groff J. R., Weinberg P. N., Opper A. J. SQL: The Complete Reference, Third Edition [SQL. Polnoe rukovodstvo. Tret'ye izdanie], Moscow, Williams Publishing House, 2018, 960 p.

11. Booch G. et al. Object-oriented analysis and design with application. Third edition [Ob'ektno-orientirovanny analiz i proektirovanie s primerami prilozheniy. Tret'ye izdanie], Moscow, Williams Publishing House, 2008, 720 p.

12. Haverbeke M. Eloquent JavaScript: A Modern Introduction to Programming, San Francisco, No Starch Press Inc., 2011, 224 p.

Категоризация в естественных языках для глаголов повышенной арности

к.т.н., доцент О. М. Поляков

Санкт-Петербургский государственный университет аэрокосмического приборостроения

Санкт-Петербург, Россия

Road.Dust.SPb@gmail.com

Аннотация. R-лингвистика — новое аксиоматическое направление в когнитивных науках, которое занимается изучением моделирования мира сознанием. Основу аксиоматики составляют аксиомы категоризации. Аксиомы категоризации позволяют построить системы взаимосвязанных категорий, отталкиваясь от поведенческого аспекта объектов. Однако часто поведение объектов оказывается более сложным, чем бинарное взаимодействие.

Цель — математически обосновать категоризацию для отношений, арности большей двух, а также обосновать другие случаи высокой арности глаголов в языке. Сформулирован метод представления соответствия Галуа для арности, большей двух. Доказана теорема о сохранении информации об отношении в этом случае. Показано, что расширение соответствия Галуа на n -арный случай категоризации с точностью до пустой категории создает в точности такие же категории, что и бинарный случай. Определены математические условия возникновения псевдовысоких арностей глаголов в языках. Показано, что они сводятся к многозначным зависимостям, действующим между атрибутом подлежащего и дополнениями. В заключении сформулированы некоторые соображения относительно арностей глаголов в естественных языках.

Ключевые слова: R-лингвистика, многозначная зависимость, категоризация, глаголы высокой арности, косвенные дополнения.

ВВЕДЕНИЕ

Предсказательный характер деятельности сознания очевиден. Предсказание является главным эволюционным преимуществом человека, позволившим ему занять высшую ступень в иерархии живых существ на Земле. Предсказание невозможно без модели и, следовательно, определение принципов построения такой модели является важнейшим направлением в рамках развития когнитивных и кибернетических наук. Язык, как система, обслуживающая мышление, является в значительной степени производной от процесса моделирования мира. R-лингвистика пытается построить математически обоснованную модель деятельности сознания по структурированию знаний о мире. Это новый вид модели данных, обладающий способностью к прогнозу. Начальная аксиоматика посвящена категоризации, причем главным ее направлением является глагольная категоризация. К сожалению, соответствие Галуа, лежащее в основаниях такой категоризации, не распространяется на арность глаголов выше второй. Эта статья призвана восполнить указанный пробел. Кроме того, здесь будет математически описан широко представленный в языке случай псевдовысоких арностей глаголов.

АКСИОМАТИКА КАТЕГОРИЗАЦИИ

В основе моделирования мира и формирования естественных языков лежит процесс формирования категорий. В [1] определена следующая аксиоматика формирования категорий. Пусть U — универсум объектов и θ — оператор, который по любому множеству X объектов из U (примеров категории) формирует категорию в виде множества объектов, удовлетворяющих двум аксиомам категоризации:

A1 $X \subseteq \theta(X)$ (аксиома экстенсивности);

A2 если K — категория, то из $X \subseteq K$ следует $\theta(X) \subseteq K$ (аксиома корректности).

Аксиома экстенсивности означает, что множество примеров, формирующих категорию, всегда входит в категорию на том простом основании, что это примеры данной категории. В частности, в силу максимальности U и аксиомы A1 получаем $U = \theta(U)$ (универсум является категорией). Аксиома корректности в свою очередь требует, чтобы никакое множество примеров категории в процессе категоризации не могло выйти за пределы этой категории. Другими словами, если K — категория «деревья», то никакое множество примеров деревьев не позволяет сформировать категорию «кусты» или «древкусты», хотя при некотором наборе примеров может сформироваться более узкая категория, скажем, «фруктовые деревья».

Как показано в [1], данная система аксиом эквивалентна системе аксиом замыкания:

A1 $X \subseteq \theta(X)$ (аксиома экстенсивности);

C1 $\theta(X) = \theta(\theta(X))$ (аксиома идемпотентности);

C2 из $X \subseteq Y$ следует $\theta(X) \subseteq \theta(Y)$ (аксиома монотонности).

Кроме того, система замкнутых множеств (категорий) определяет муровское семейство [1, 2] относительно операции пересечения замкнутых множеств, являющееся полной решеткой [3], в которой роль \sup выполняет точная нижняя грань.

Разумеется, оператор замыкания может быть задан различными способами [1], из которых важнейшим для лингвистики является глагольный, основанный на соответствии Галуа [3, 4], определенном для бинарных отношений (бинарных глаголов: подлежащее — глагол — прямое дополнение) или предикатов от двух переменных. В лингвистике глагольная категоризация практически всегда предшествует признаковой, а иногда, когда необходимые признаки найти не удастся, глагольной категоризации.

ей завершается процесс определения категории. Хорошо известен пример Л. Витгенштейна [5] с категорией «игра», которая не может быть выражена в признаках. Аналогично, например, категория «хищники» столь разнообразна, что не имеет общих признаков, характеризующих ее элементы. Рослянка, лев, окунь и богомол являются хищниками, но не имеют общих признаков: их объединяет одинаковое поведение по трофическому отношению.

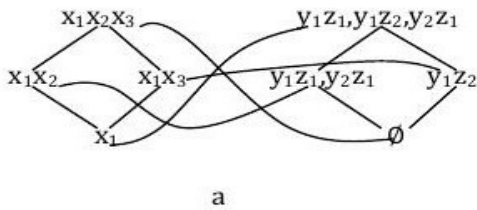
В [1] соответствие Галуа задается следующим образом.

Пусть $S(U, V)$ — бинарное отношение с атрибутами U и V ($S \subseteq D_U \times D_V$, где D_U и D_V — домены атрибутов U и V соответственно). Пусть также $x \in D_U$. Обозначим через x^Δ множество всех элементов $y \in D_V$, для которых $(x, y) \in S$. Аналогично символом y^∇ ($y \in D_V$) обозначим множество элементов $x \in D_U$, для которых $(x, y) \in S$. Операторы Δ и ∇ распространяются на произвольные подмножества следующим образом:

$$X^\Delta = \bigcap_{x \in X} x^\Delta \quad (X \subseteq D_U); \quad Y^\nabla = \bigcap_{y \in Y} y^\nabla \quad (Y \subseteq D_V).$$

Для пустого подмножества полагаем $\emptyset^\Delta = D_V$ и $\emptyset^\nabla = D_U$.

Нетрудно показать [1], что операторы Δ и ∇ являются операторами замыкания на D_U и D_V соответственно. При этом между муровскими множествами, определенными таким образом на D_U и D_V , операторы Δ и ∇ задают дуальный изоморфизм. В дальнейшем по аналогии с оператором топологического замыкания [6] систему муровских множеств, удовлетворяющую аксиомам $A1$ и $A2$ будем называть лингвистическим пространством (или просто пространством). В [1] определена также операция сведения пространств, которая состоит в том, что сведённое пространство образуется из всех категорий пространств-операндов, плюс все недостающие пересечения этих категорий. Сведённое пространство является наименьшим пространством, содержащим пространства-операнды.



РАСШИРЕНИЕ СООТВЕТСТВИЯ ГАЛУА НА ПРОИЗВОЛЬНУЮ КОНЕЧНУЮ АРНСТЬ

В языке существуют не только унарные и бинарные глаголы. Как минимум широко распространены и тернарные глаголы, которые предполагают обязательное участие не только подлежащего, прямого, но и косвенного дополнения, например, глаголы передавать, дарить, просить, обвинять и т.д.

Пример 1. В таблице 1 приведено тернарное отношение «передать» с атрибутами: КТО (передает) — 1, КОМУ (передает) — 2, ЧТО (передает) — 3.

В качестве домена (универсума) для каждого атрибута отношения будем использовать проекцию $S[i]$ ($i = 1, 2, 3$) отношения на соответствующий атрибут (атрибуты). Отношение (табл. 1) можно рассматривать как бинарное между двумя проекциями, например, на $S[1] \times S[2, 3]$. Пространство и дуальное пространство для этого отношения изображено на рисунке 1, а.

В дуальном пространстве для $S[2, 3]$ каждое понятие представляется бинарным отношением (рис. 1, а, правое пространство), поэтому для каждой такой категории можно также построить пространство и дуальное пространство. Результат этого перехода от бинарных отношений-категорий к соответствующим пространствам изображен на рисунке 1, б.

Таблица 1

Отношение для глагола «передать»

Кто (1)	Кому (2)	Что (3)
x1	y1	z1
x2	y1	z1
x1	y2	z1
x2	y2	z1
x1	y1	z2
x3	y1	z2

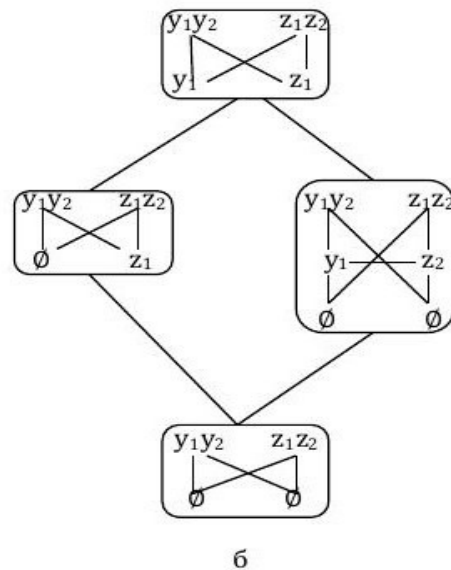


Рис. 1. Лингвистические пространства тернарного отношения

Подобную вложенную систему лингвистических пространств удобно представить в виде дерева (рис. 2). Здесь корень дерева соответствует пространству 1, первый уровень — пространству 2 и конечные вершины — простран-

ству 3. Ветви дерева одного уровня слева направо соответствуют категориям соответствующего пространства, выбираемым слева направо и сверху вниз.

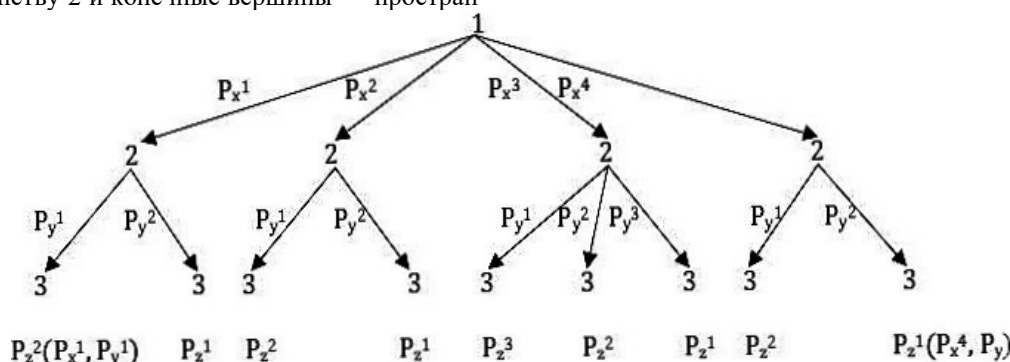


Рис. 2. Понятийное иерархическое представление глагола «передать»

На рисунке 2 дуги дерева отмечены именами категорий соответствующего пространства. Для примера: самый левый лист дерева кроме отметки номера категории имеет отметку пути от корня к листу. Например, запись $P_z^2(P_x^1, P_y^1)$ означает, что тернарный глагол соединяет три категории: КТО передает — $P_x^1 = \{x_1, x_2, x_3\}$; КОМУ передает — $P_y^1 = \{y_1, y_2\}$; ЧТО передает — $P_z^2 = \emptyset$. Итак, каждый путь от корня дерева к конечной вершине содержит некоторую возможную (пока смысловую, а не языковую) фразу из категорий, связанных глаголом «передать».

Определение 2. Дерево на рисунке 3 назовем **понятийным иерархическим представлением (ПИП)** глагола (отношения S).

Понятно, что для отношения любой арности можно построить такое дерево, причем арность определит его высоту. Кроме того, надо заметить, что в примере 1 первое пространство находилось в привилегированном положении, а именно: атрибут 1 был выбран первым в n-арном отношении, и ему было сопоставлено (n - 1)-арное отношение. Впрочем, в качестве первого атрибута можно было бы выбрать любой другой.

Определение 3. Пространство атрибута, выбранного в качестве первого, называется **безусловным пространством** этого атрибута. Собственно, в языке ему соответствует подлежащее. Все остальные пространства называются **условными**. Условность каждого из условных пространств определяется категориями (условиями) в ПИП, лежащими на пути от корня дерева до атрибута условного пространства.

Лемма 4. По любому ПИП можно восстановить исходное n-арное отношение.

Доказательство.

Для $n = 2$ пусть $(x, y) \in S$, тогда $(x, y) \in (x^{\Delta V}) \times (x^{\Delta})$, так что в ПИП это соответствует пути из корня дерева по дуге, отмеченной категорией $x^{\Delta V}$, в конечную вершину, отмеченную категорией x^{Δ} . Обратное, если (x, y) входит в декартово произведение $P^1 \times P^2$ категорий, лежащих на пути из корня в некоторый лист, то по построению ПИП $P^{1\Delta} = P^2$ и, значит, $(x, y) \in S$. Пусть лемма справедлива для $n = k$ и пусть отношение S имеет арность $k+1$. Рассмотрим

некоторое ПИП отношения S, у которого корень помечен номером i. Каждая дочерняя вершина корня ПИП является корнем поддерева ПИП k-арного отношения $S(P_i^j)$ для некоторой категории (P_i^j) пространства i. По нашему предположению $S(P_i^j)$ совпадает с объединением произведений понятий, помечающих ветви поддерева от корня поддерева до листьев поддерева, тогда, в соответствии с бинарным случаем, объединение всех произведений $P_i^j \times S(P_i^j)$ будет равно S.

Замечание 5. В ПИП отношения S на дугах могут встречаться пустые категории соответствующих пространств. В этом случае удаление из ПИП путей из корня в лист, содержащих такие понятия, с учетом доказательства леммы 4 не влияет на восстанавливающую способность ПИП. В связи в этом будем говорить, что два пространства почти равны, если они отличаются не более, чем на пустую категорию. Аналогично, одно пространство почти принадлежит другому, если все непустые категории одного пространства входят в другое пространство.

Определение 6. \cap -образующими лингвистического пространства называются категории, которые нельзя представить в виде пересечения других категорий.

Из определения 6 следует, что все категории лингвистического пространства либо являются \cap -образующими, либо могут быть получены пересечением некоторых \cap -образующих.

Предложение 7. Если $D_U^{\Delta} \neq \emptyset$ ($D_V^{\nabla} \neq \emptyset$), то для любой \cap -образующей найдется элемент $y(x)$ такой, что $y^{\nabla}(x^{\Delta})$ совпадает с этой образующей. Если же $D_U^{\Delta} = \emptyset$ ($D_V^{\nabla} = \emptyset$), то предложение справедливо для всех \cap -образующих, исключая D_U (D_V).

Доказательство.

Пусть X — \cap -образующая, для которой $X^{\Delta} = Y \neq \emptyset$. Тогда $X = \bigcap_{y \in Y} y^{\nabla}$. Если в Y найдутся такие элементы y, что $X \subset y^{\nabla}$ и их пересечение равно X, то X не является \cap -образующей. Следовательно, в Y нет таких элементов и X совпадает с некоторой категорией y^{∇} ($y \in Y$). Аналогично для \cap -образующих на D_V .

Теорема 8. Пусть S — тернарное отношение. Для каждого x первого атрибута определим бинарное отношение $S^x \subseteq S[2] \times S[3]$ следующим образом: $(y, z) \in S^x$ тогда и только тогда, когда $(x, y, z) \in S$. Обозначим через \prod операцию сведения пространств, а сами пространства обозначим через \mathbb{P} с соответствующими индексами. Тогда $\mathbb{P}_3 \cong \prod_{x \in S[1]} \mathbb{P}_3^x$, где отношение \cong означает отношение почти равенства из замечания 5, а \mathbb{P}_3^x обозначает пространство по третьему атрибуту отношения S^x . Аналогично для атрибута 2.

Доказательство.

В соответствии с предложением 7 выберем некоторые x, y . Тогда $(x, y)^\Delta$ — категория безусловного пространства \mathbb{P}_3 . Но тогда для S^x имеем $y^\Delta = (x, y)^\Delta$ и $(x, y)^\Delta \in \mathbb{P}_3^x$, а значит, в соответствии с определением операции сведения, $(x, y)^\Delta \in \prod_{x \in S[1]} \mathbb{P}_3^x$. В соответствии с предложением 7 получаем $\mathbb{P}_3 \subseteq \prod_{x \in S[1]} \mathbb{P}_3^x$. Обратно, пусть $Z \neq \emptyset$ и $Z \in \prod_{x \in S[1]} \mathbb{P}_3^x$. Тогда найдутся такие $x_i \in S[1]$, ($i = 1, \dots, k$), что для каждого i и отношения S^{x_i} в условном

пространстве на $S[2]$ найдется категория Y^{x_i} , которой по условному глаголу соответствует категория Z^{x_i} , такая, что $Z = \bigcap_{x_i} Z^{x_i}$. Но $Z^{x_i} = \bigcap_{y \in Y^{x_i}} y^{\Delta x_i}$ по соответствующему условному глаголу. То есть $Z^{x_i} = \bigcap_{y \in Y^{x_i}} (x_i, y)^\Delta$, а значит, $Z \in \mathbb{P}_3$.

Итак, для тернарных отношений сведение условных пространств с точностью до пустой категории равно безусловному пространству. Например, для отношения в таблице 1 безусловное пространство для второго атрибута (КОМУ) показано на рисунке 3, а. На рисунке 3, б изображены условные пространства для этого атрибута, а на рисунке 3, в показано сведенное пространство для всех условных пространств, которое с точностью до пустой категории совпадает с безусловным пространством. Заметим, что, например, после фиксации категорий первого атрибута, условные пространства второго и третьего атрибутов находятся в равном положении независимо от их перестановки (рис. 1, б).

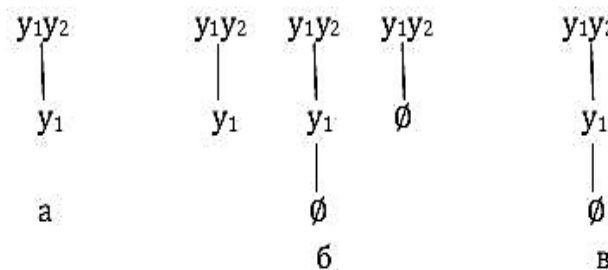


Рис. 3. Условные лингвистические пространства

Разумеется, если в качестве начального взять, допустим, третий атрибут, то условные пространства по второму атрибуту изменятся, но их сведение по-прежнему с точностью до пустой категории даст безусловное пространство. Почему это так важно? Потому что безусловное пространство атрибута соответствует случаю, когда данный атрибут выступает в качестве подлежащего, а условные пространства характеризуют понятия, которые занимают место прямых и косвенных дополнений. Полученный результат позволяет нам быть уверенными, что изменение статуса атрибута в предложении не может породить новых категорий в сравнении с положением этого атрибута в качестве подлежащего. Случай арности большей трех получается индукцией по арности.

Теорема 9. Для любого n -арного отношения S ($n > 2$) и любого атрибута A категории условных пространств по этому атрибуту с точностью до пустой категории совпадают с безусловными категориями.

Доказательство.

Случай $n = 3$ соответствует теореме 8. Пусть теорема верна при любом $m \leq n$ и пусть S имеет арность $n + 1$. Пусть также для $(n + 1)$ -арного отношения S построено ПИП, имеющее $n + 1$ ярус. Выберем в ПИП произвольный атрибут, расположенный на i -ом ярусе. Если $i = 1$, то формируемое пространство для этого атрибута по определению является безусловным. Если $i = 2$, то в каждом поддереве, начинающемся со второго уровня, согласно индукционному предположению лингвистическое пространство

для каждой вершины второго уровня с точностью до пустой категории совпадает со сведением условных лингвистических пространств, на каком бы уровне от 3-го до $n + 1$ в другом ПИП они ни находились. Это в свою очередь означает, что, если бы $3 \leq i \leq n + 1$, то, перестроив каждое поддерево второго уровня ПИП так, чтобы атрибут i занял место в вершине поддерева второго уровня, мы в силу индукционного предположения не породили и не потеряли бы категории в лингвистических пространствах этого атрибута. Итак, без потери общности достаточно рассмотреть только случай $i = 2$.

Каждой вершине второго уровня соответствует n -арное отношение $S(P_1^j)$ без атрибута 1, которое получается выделением из отношения S кортежей, у которых первый атрибут принимает значения из категории P_1^j с последующим проектированием этого отношения на атрибуты $2 \dots n + 1$. Для отношения $S(P_1^j)$ строится лингвистическое пространство для каждой вершины второго уровня, причем при этом построении рассматривается соответствие Галуа между значениями второго атрибута и остальных атрибутов $3 \dots n + 1$ как единого атрибута, значениями которого являются кортежи длины $n - 1$ со значениями атрибутов $3 \dots n + 1$. Другими словами, мы можем рассматривать отношение S как тернарное отношение на атрибутах 1, 2 и $(3 \dots n + 1)$. Тогда по теореме 8 безусловное лингвистическое пространство атрибута 2 получается сведением (с точностью до пустой категории) лингвистических пространств, получаемых по отношениям $S(P_1^j)$, которые в

свою очередь получают сведением условных пространств, когда атрибут 2 занимает более низкое положение в ПИП. Но операция сведения является ассоциативной, что и завершает доказательство теоремы.

ПСЕВДОВЫСОКИЕ АРНОСТИ ГЛАГОЛОВ

Итак, категории сохраняют стабильность независимо от их места в предложении и арности глагола. Надо иметь в виду, что до сих пор мы рассматривали глаголы (предикаты), арность которых определяется природой отношений, соответствующих этим глаголам. Но в языке сплошь и рядом встречаются высокие арности (множественность косвенных дополнений), которые, однако, не связаны с природой глаголов. Эта ситуация часто вводит в заблуждение логиков [7], занимающихся языком, поскольку создается впечатление, что в языке может произвольным образом меняться размерность предикатов. Здесь мы рассмотрим механизм появления таких высоких арностей.

Может ли арность отношения возражать без увеличения арности глагола? Почему глаголы большей частью бинарные? Эти, казалось бы, разные вопросы имеют в качестве источника один ответ: многозначные зависимости данных [8, 9]. Напомним, что в отношении $S(ABC)$ выполняется многозначная зависимость (МЗ) $A \rightarrow B$, если и только если из того, что $(A, B, C) \in S$, $(A, B', C') \in S$ следует $(A, B, C') \in S$. Здесь A, B и C — некоторые наборы атрибутов, причем $B \cap C = \emptyset$. Такое «кортежное» симметричное определение позволяет сформулировать правило дополнения: МЗ $A \rightarrow B$ в отношении $S(ABC)$ выполняется тогда и только тогда, когда в этом отношении выполняется МЗ $A \rightarrow C$. Главное свойство МЗ состоит в том, что она обеспечивает декомпозицию отношения на две части без потерь: в отношении $S(ABC)$ выполняется (МЗ) $A \rightarrow B$, если и только если из проекций этого отношения $S[AB]$ и $S[AC]$ с помощью операции соединения можно восстановить исходное отношение $S(ABC)$. Если C пусто, как, например, в $S[AB]$, то по определению МЗ всегда выполняется $A \rightarrow \emptyset$ и, значит, по правилу дополнения в $S[AB]$ всегда выполняется МЗ $A \rightarrow B$. Обилие в языке бинарных глаголов как раз и объясняется тем, что они вычленились с помощью МЗ из более общих наблюдаемых взаимодействий.

Что будет происходить, если к атрибутам $A \rightarrow B$ в отношении будут добавляться дополнительные атрибуты с сохранением МЗ $A \rightarrow B$? Это будет означать, что новое отношение может быть без потери информации разложено на два отношения с атрибутами AB и атрибутами AC , где C — добавленные атрибуты. Итак, по отношению AB сохраняются уже существующие категории с глаголом между ними, а отношение AC может определить новые категории на A . Поскольку в новом общем отношении ABC одновременно действуют сразу два отношения $S[AB]$ и $S[AC]$, то категории на B и C будут определяться, исходя из бинарных отношений, а категории на A фактически будут представлять из себя пересечение категорий по одному и другому отношению.

Обычно такой случай представляется двумя предложениями с глаголами отношений AB и AC . В тех случаях, когда можно использовать одно название глагола для обоих атрибутов B и C , объекты по атрибуту C превращаются в косвенные дополнения («девочка была мальчика пал-

кой»). Теперь надо выяснить условия стабильности многозначных зависимостей при добавлении или изъятии некоторых атрибутов отношения. Без понимания этих условий говорить о косвенных дополнениях невозможно.

Обычно зависимости данных рассматриваются для фиксированных схем отношений, поэтому необходимо уточнить ситуацию, когда схема отношения меняется. Рассмотрим сначала случай уменьшения числа атрибутов в отношении, который, разумеется, не затрагивает МЗ $A \rightarrow B$.

Предложение 10. Пусть в отношении S на атрибутах $ABCD$ выполняется МЗ $A \rightarrow B$. И пусть S' — проекция отношения S на атрибуты ABC . Тогда в S' справедлива МЗ $A \rightarrow B$.

Доказательство.

Итак, пусть $(x, y, z) \in S'$ и $(x, y', z') \in S'$, где x, y, z — значения атрибутов A, B, C соответственно. Поскольку S' — проекция отношения S на ABC , найдутся такие значения атрибута D v и v' , что $(x, y, z, v) \in S$ и $(x, y', z', v') \in S$. Поскольку в S выполняется МЗ $A \rightarrow B$, то $(x, y, z', v') \in S$, а значит, $(x, y, z') \in S'$, и в S' также выполняется МЗ $A \rightarrow B$.

Итак, уменьшение числа атрибутов в отношении не отменяет справедливость многозначных зависимостей. При этом сохраняется не только МЗ $A \rightarrow B$, но и дополнительная к ней МЗ $A \rightarrow CD$. В самом деле, поскольку в проекции согласно предложению 10 выполняется МЗ $A \rightarrow B$, то по правилу дополнения выполняется и МЗ $A \rightarrow C$.

Рассмотрим теперь случай увеличения числа атрибутов в отношении. Здесь все не так очевидно, поскольку существуют встроенные многозначные зависимости, то есть такие МЗ, которые выполняются в проекции, но не выполняются в самом отношении. Пример такого отношения представлен в таблице 2.

Таблица 2

Встроенные МЗ

A	B	C	D	A	B	C
a	b	c	d	a	b	c
a	b'	c	d	a	b'	c
a	b	c'	d'	a	b	c'
a	b'	c'	d	a	b'	c'
a'	b'	c'	d'	a'	b'	c'

В правом отношении на атрибутах ABC (проекция) выполняется МЗ $A \rightarrow B$, но она не выполняется в левом отношении на атрибутах $ABCD$. Такие МЗ как раз и называют встроенными. Итак, мы видим, что произвольное приписывание значений атрибута D к кортежам отношения на ABC может «разрушить» существующую МЗ.

Теорема 11. Пусть в отношении S' на атрибутах ABC выполняется МЗ $A \rightarrow B$. И пусть к отношению S' добавляется атрибут D и кортежи отношения S' дополняются значениями атрибута D для получения отношения S . Если это добавление осуществляется так, что в S выполняется МЗ $A \rightarrow D$, то в S также выполняется и МЗ $A \rightarrow B$.

Доказательство.

Пусть $(x, y, z, v) \in S$ и $(x, y', z', v') \in S$. Для того, чтобы в S выполнялась МЗ $A \rightarrow B$ надо показать, что $(x, y, z', v') \in S$. Поскольку $(x, y, z) \in S'$, $(x, y', z') \in S'$ и в S' выполняется МЗ $A \rightarrow B$, то $(x, y, z') \in S'$, а значит, существует v'' такое, что $(x, y, z', v'') \in S$.

Поскольку $(x, y, z', v'') \in S$, $(x, y', z', v') \in S$ и в S выполняется МЗ $A \rightarrow D$, то $(x, y, z', v') \in S$, а значит, в S выполняется МЗ $A \rightarrow B$.

Если $C = \emptyset$, то условия теоремы 11 вырождаются в аксиому вывода «дополнение».

Теорема 11 говорит о том, что если косвенные дополнения добавлять аккуратно, то сохраняется не только бинарность начального глагола, но и те пространства, которые он образует, поскольку сохранение МЗ $A \rightarrow B$ означает возможность выделить АВ без потери информации в отдельное первоначальное отношение. Это в свою очередь означает, что новые дополнения должны быть независимы от уже находящихся в отношении.

Рассмотрим пример с глаголом «бить» в предложении «девочка била мальчика». Здесь «бить» выступает в качестве бинарного глагола. Понятно, что имеется в виду конкретная девочка и конкретный мальчик. Категории «девочка» и «мальчик» в этом предложении фактически задают области определения переменных, которые принимают значения конкретных участников события. Если взять отношение между людьми и орудиями избиения и добавить его к этому предложению, то, например, получим: «девочка била мальчика палкой». На самом деле, новое отношение соответствует своему глаголу, и правильнее было бы сформулировать предложение так: «девочка била мальчика, используя палку». Однако, поскольку из контекста понятно, о каком глаголе идет речь, а падежи компенсируют пропажу глагола, то и первый вариант оказывается вполне допустимым. Область переменной «мальчика» и орудия избиения не изменились в новом предложении. Но вот область переменной «девочка» сузилась, поскольку далеко не всякая девочка использует в драке палку. Переменная орудия избиения не зависит от прямого дополнения. В самом деле, на вопрос: «Чем била девочка мальчика?» вы скорее всего пожмете плечами и ответите: «Да чем угодно». Это как раз и означает наличие многозначной зависимости. Косвенные дополнения в своем действии на подлежащие очень напоминают прилагательные, поскольку сужают область действия категории. Вместе с тем, есть и различие в действии прилагательного и косвенного дополнения, но его можно обсуждать лишь после исследования признаков пространств.

Мы можем продолжить наращивать кажущуюся арность глагола «бить», добавляя новые дополнения: «на улице девочка била мальчика палкой по голове». И опять действующая в предложении область определения «девочка» сузилась, поскольку совсем немного девочек столь кровожадны, что готовы лупить на улице прилюдно беззащитного ребенка палкой по голове. Появление в языке таких псевдовысоких арностей довольно частое явление («я защищал дерево веслом от бобра»), и всегда в этих случаях за ширмой прячутся зависимости данных, а первоначальная арность глагола сохраняется. В этом принципиальное отличие таких косвенных дополнений от

«настоящих» косвенных дополнений, порожденных арностью глаголов.

Описанные выше примеры изменения арности глагола не надо путать с одним из семантических способов определения объектов с помощью уточнения его места в иерархии, которая, конечно, также связана с разложением, но только не глаголов, а самих объектов [10]. Например, фраза «деньги лежат в спальне в шкафу на второй полке сверху под полотенцами» совсем не характеризует арность глагола «лежать», а относится к разложению самих объектов: квартиры на комнаты и т.д.

ЗАКЛЮЧЕНИЕ

Мы рассмотрели два случая глаголов арности большей двух. При этом было показано, что часто высокие арности связаны просто с независимостью дополнений, поскольку многозначные зависимости скорее характеризуют не зависимость правой части от левой, а независимость левых частей МЗ [11], то есть дополнений. Выше было приведено несколько примеров «настоящих» тернарных глаголов. Возникает вопрос: «Существуют ли в языке глаголы арности большей трех?» Хотя теорема 9 снимает проблемы в математическом смысле, но вопрос остается открытым для самой лингвистики. Очевидно, что положение категории в кортеже некоторого отношения имеет принципиальное значение. Например, сказать, что «щука проглотила карася», или что «карась проглотил щуку» — вещи существенно различные. Поэтому для «настоящих» n -арных глаголов важно фиксировать место каждой категории в кортеже, а значит, необходимо как-то отражать эту информацию и в предложении. В различных языках проблема решается по-разному: эту роль либо выполняют предлоги с падежами, либо жесткая привязка к месту в предложении. Например, в выражениях «царь направил посла королю» и «царь направил королю посла» идентификация атрибута в кортеже определяется с помощью падежа. Если предположить, что выразительные способности английского языка в смысле отражения глаголов высокой арности, не уступают другим языкам, следует ожидать ограничения арности глаголов цифрой три. Это утверждение, разумеется, является лингвистической гипотезой. Кроме того, далеко не все глаголы языка образуют категории. Большинство глаголов языка используют категории, созданные другими глаголами. Существуют ли тернарные (и выше) глаголы, образующие категории? Этим открытым вопросом к лингвистике мы завершим рассмотрение данной темы.

ЛИТЕРАТУРА

1. Polyakov O. M. Linguistic Data Model for Natural Languages and Artificial Intelligence. Part 1. Categorization // ДИСКУРС. 2019. Т. 5, № 4. С. 102–114. DOI: 10.32603/2412-8562-2019-5-4-102-114.
2. Blyth T. S. Lattices and Ordered Algebraic Structures. London, Springer-Verlag, 2005, 315 p.
3. Биркгоф Г. Теория решеток / Пер. с англ. В. Н. Салей; под ред. Л. А. Скорнякова. — М.: Наука. Главная редакция физико-математической литературы, 1984. — 568 с.
4. Гретцер Г. Общая теория решеток / Пер. с англ. А. Д. Большота и др.; под ред. Д. М. Смирнова. —

М.: Мир. Редакция литературы по математическим наукам, 1982. — 456 с.

5. Витгенштейн Л. Философские исследования. — М: Русский гуманитарный Интернет-университет, 2008. — 142 с.

6. Куратовский К. Топология: в 2-х т. / Пер. М. Я. Антоновского. — М.: Мир, 1966–1969.

Т. 1. — 1966. — 594 с.

7. Андреев А. В. Введение в формальную семантику: Учебное пособие. / А. В. Андреев, О. А. Митрофанова, К. В. Соколов; Санкт-Петербургский гос. ун-т, Филологический фак. — СПб.: СПбГУ, 2014. — 88 с.

8. Fagin R. Multivalued Dependencies and a New Normal Form for Relational Databases, *ASM Transactions on Database Systems*, 1977, Vol. 2, No. 3, Pp. 262–278.

9. Дейт К. Дж. Введение в системы баз данных. Восьмое издание = An Introduction to Database Systems. 8th Edition / Пер. с англ. и ред. К. А. Птицына. — М.: Издательский дом «Вильямс», 2005. — 1328 р.

10. Блюм В. С. Лингвистическое пространство языка клинической медицины // Актуальные проблемы экономики и управления. 2018. Вып. 3 (19). С. 27–30.

11. Мейер Д. Теория реляционных баз данных / Пер. с англ. М. К. Валиева и др.; под ред. М. Ш. Цаленко. — М.: Мир. Редакция литературы по математическим наукам, 1987. — 608 с.

Categorization in Natural Languages for Verbs of Increased Arity

PhD, Assoc. Prof. O. M. Polyakov
Saint Petersburg State University of Aerospace Instrumentation
Saint Petersburg, Russia
Road.Dust.SPb@gmail.com

Abstract. R-linguistics is a new axiomatic direction in cognitive Sciences, which deals with the study of modeling the world by consciousness. The basis of axiomatics are the axioms of categorization. The axioms of categorization allow us to build systems of interrelated categories, starting from the behavioral aspect of objects. However, often the behavior of objects is more complex than binary interaction. To mathematically substantiate categorization for relations, arity greater than two, as well as to substantiate other cases of high arity of verbs in the language. A method for representing the Galois correspondence for an arity greater than two is formulated. The theorem on the conservation of information about the relation in this case is proved. It is shown that the extension of the Galois correspondence to the n-ary case of categorization up to an empty category creates exactly the same categories as the binary case. The mathematical conditions for the emergence of pseudo-high arities of verbs in languages are determined. It is shown that they are reduced to multivalued dependencies acting between the subject attribute and additions. In conclusion, some considerations are formulated regarding the varieties of verbs in natural languages.

Keywords: R-linguistics, multivalued dependence, categorization, verbs of high arity, indirect additions

REFERENCES

1. Polyakov O. M. Linguistic Data Model for Natural Languages and Artificial Intelligence. Part 1. Categorization, *DISCOURSE [Diskurs]*, 2019, Vol. 5, No. 4, Pp. 102–114. DOI: 10.32603/2412-8562-2019-5-4-102-114.
2. Blyth T. S. Lattices and Ordered Algebraic Structures. London, Springer-Verlag, 2005, 315 p.

3. Birkhoff G. Lattice Theory [Teoriya reshetok], Moscow, Nauka Publishers, 1984, 568 p.
4. Grätzer G. General lattice theory [Obschchaya teoriya reshetok], Moscow, Mir Publishers, 1982, 456 p.
5. Wittgenstein L. Philosophical studies [Filosofskie issledovaniya], Moscow, Russian Humanities Internet University, 2008, 142 p.
6. Kuratowsky K. Topology [Topologiya], Vol. 1, Moscow, Mir Publishers, 1966, 594 p.
7. Andreyev A. V., Mitrofanova O. A., Sokolov K. V. Introduction to formal semantics: Study guide [Vvedenie v formal'nyu semantiku: Uchebnoe posobie], Saint Petersburg, St. Petersburg University, 2014, 88 p.
8. Fagin R. Multivalued Dependencies and a New Normal Form for Relational Databases, *ASM Transactions on Database Systems*, 1977, Vol. 2, No. 3, Pp. 262–278.
9. Date C. J. An Introduction to Database Systems. 8th Edition [Vvedenie v sistemy baz dannykh. Vos'moe izdanie], Moscow, Willams Publishing House, 2005, 1328 p.
10. Blum V. S. Linguistic Space of the Language of Clinical Medicine [Lingvisticheskoe prostranstvo yazyka klinicheskoy meditsiny], *Actual Problems of Economics and Management [Aktual'nye problemy ekonomiki i upravleniya]*, 2018, Is. 3(19), Pp. 27–30.
11. Maier D. The Theory of Relational Databases [Teoriya relyatsionnykh baz dannykh], Moscow, Mir Publishers, 1987, 608 p.

Интеллектуальный анализ данных для задач идентификации состояний объектов контроля в условиях неопределенности

А. В. Панков, Я. В. Дашкевич

Военно-космическая академия имени А. Ф. Можайского
Санкт-Петербург, Россия
yvdspsb55@yandex.ru

Аннотация. Рассмотрена возможность реализации методов интеллектуального анализа данных для совершенствования информационно-аналитической работы в деятельности органов государственного и регионального управления. Основная идея предлагаемого подхода заключается в реализации нейросетевых моделей обработки информации для идентификации кризисных состояний при решении задач, связанных с обеспечением общественной безопасности, в условиях неполноты и частичной недостоверности исходных данных. Представленный имитационный эксперимент описывает особенности подготовки исходных данных для обучения нейросетевой модели на основе частичной рандомизации эталонных векторов признаков ситуаций. Обоснованность применения нейросетевых методов для идентификации состояний объектов контроля определена на основе сравнительного анализа результативности их применения по отношению к традиционным используемым методам, основанным на установлении метрического расстояния между выявленным по результатам оперативного мониторинга и эталонным состоянием объекта контроля.

Ключевые слова: нейронная сеть, распознавание ситуации, обработка информации, расстояния Хемминга, Big Data, OLAP, аналитические структуры ведомств.

ВВЕДЕНИЕ

Многофакторность и высокая неопределенность возникновения угроз безопасности определяют ситуационные условия функционирования информационно-аналитических структур ведомств, деятельность которых направлена на предотвращение кризисных ситуаций (МЧС, ГИБДД, органы здравоохранения, полиция и т. д.). Сложнопрогнозируемые ситуационные условия, связанные с неполнотой и, возможно, недостоверностью исходных данных, формируют дополнительные требования к процессам обработки информации, обеспечение которых может быть реализовано на основе активно развивающихся информационных технологий, реализующих новые методы интеллектуального анализа данных.

В соответствии с реализуемой Стратегией развития информационного общества в Российской Федерации на 2017–2030 годы, информационные и коммуникационные технологии становятся частью современных управленческих систем во всех сферах государственного управления, обороны страны, безопасности государства и обеспечения правопорядка [1].

В настоящее время в различных сферах государственного и регионального управления все более остро проявляются проблемы, связанные с необходимостью постоян-

ного мониторинга ключевых параметров функционирования сложных объектов. Специфика решаемых задач связана с обеспечением общественной безопасности, оперативным выявлением возможных угроз и принятием эффективных управленческих решений по разрешению возникших кризисных ситуаций.

Так, в интересах обеспечения комплексной безопасности Санкт-Петербурга с 2004 года функционирует государственное казенное учреждение «Городской мониторинговый центр» (ГКУ ГМЦ), деятельность которого направлена на повышение уровня обеспечения общественной безопасности. Одним из основных направлений его деятельности является круглосуточный мониторинг экстренных событий и нестандартных ситуаций на объектах города и в местах проведения массовых мероприятий. Для решения задач идентификации состояний объектов контроля во всех 18 районах Санкт-Петербурга установлено более 30 тысяч камер. В поле зрения которых включены наиболее сложные перекрестки автомобильных дорог и оживленные магистрали, городские парки и дворы спальных районов [2].

Активное развитие технических средств автоматизированного сбора информации формирует дополнительные возможности, связанные с непрерывной актуализацией оперативной информации, используемой для принятия управленческих решений. Уже сейчас СПб ГКУ «ГМЦ» совместно с Комитетом по информатизации завершил реализацию первого этапа проекта «Умный город», в рамках которого создана интеллектуальная система обработки данных с городских видеокамер [3]. Реализовано специальное мобильное приложение, обеспечивающее оперативный сбор видеоинформации с мест происшествий.

В настоящий момент на улицах Петербурга установлено более 150 камер с функционалом интеллектуальной видеоаналитики, реализующих нейросетевые технологии, для обработки видеоизображений в реальном времени. Реализованные программно-аппаратные решения обеспечивают решение задач идентификации лиц, с программированием их биометрических моделей, распознавания государственных номерных знаков автотранспорта, фиксации оставленных предметов, выявления фактов возникновения различных чрезвычайных ситуаций, в том числе пожаров и дорожно-транспортных происшествий. В ближайшей перспективе данные системы будут способны вести учет интенсивности движения транспорта и пешеходов в интересах эффективного и безопасного управле-

ния дорожным движением и общественным правопорядком.

Аналогичные способы интеллектуализации процессов обработки информации востребованы и в сфере обороны государства. В интересах концепции развития системы управления Вооруженными силами Российской Федерации (ВС РФ) до 2025 года, одной из главных целей реализации которой является гарантированное управление войсками (силами) и оружием в едином информационном пространстве (ЕИП), осуществляется разработка и модернизация автоматизированных систем военного и специального назначения. Создание ЕИП ВС должно способствовать эффективному использованию информационных ресурсов государства для решения задач в сфере военного строительства и управления войсками (силами). Так, с 1 декабря 2014 года на боевое дежурство заступил Национальный центр управления обороной Российской Федерации (НЦУО). Сформированная структура предназначена для обеспечения централизованного управления боевой и повседневной деятельностью ВС РФ. НЦУО реализует функции оперативной координации деятельности Министерства обороны РФ и 73 федеральных органов исполнительной власти, органов власти всех субъектов РФ, 1 320 государственных корпораций и предприятий оборонно-промышленного комплекса России [4]. Функции контроля и управления основаны на существенном увеличении объемов данных (включая видеоматериалы), используемых в ходе взаимодействия указанных элементов информационной системы.

Процессы, происходящие в сфере военного управления, в полной мере соотносятся с изменениями, происходящими и в других сферах информатизации общества, как в России, так и за рубежом. Статистика объема данных, создаваемых на протяжении последних лет, поражает воображение. По оценкам западных компаний, специализирующихся в сфере индустрии информационных технологий (EMC, Cisco, IBM, Google) в 2012 году в мире было сгенерировано 2 зетабайта (2 тысячи миллиардов гигабайтов) информации, а в 2020 году эта величина достигнет 35-40 зетабайтов [4]. При этом большая часть данной информации будет сгенерирована различного рода техническими устройствами регистрации (видеопотоки с камер наблюдения, данные от устройств радиочастотной идентификации, спутниковых систем навигации, различного рода сенсоров и многое другое).

Обработать такие объемы информации без использования автоматизированных систем становится практически невозможным. Резкое увеличение объемов информации требует поиска новых методов их обработки. Ответом на этот вызов стали технологии, получившие в технической и информационной среде обобщенное название «Большие данные» (Big Data).

Стратегия развития отрасли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года определяет технологии обработки Big Data в числе «прорывных для мировой индустрии, на основе которых в перспективе 10-15 лет с высокой вероятностью может быть обеспечена глобальная технологическая конкурентоспособность России» [5].

Обработка больших объемов данных — совокупность подходов, инструментов и методов автоматической обра-

ботки структурированной и неструктурированной информации, поступающей из большого количества различных, в том числе разрозненных или слабосвязанных, источников информации, в объемах, которые невозможно обработать вручную за разумное время [1].

Необходимость оперативной аналитической обработки значительного объема информации, характеризующегося существенной неопределенностью, вызванной неполнотой и возможной недостоверностью исходных данных, вследствие преднамеренных и непреднамеренных искажений, определяет потребность в использовании методов искусственного интеллекта в процессах обнаружения новых знаний.

На фоне активного внедрения интеллектуальных информационных систем, ориентированных на распознавание образов ситуаций, регистрируемых техническими устройствами сбора данных, интеллектуализация процессов, связанных с поддержкой принятия решений при оценивании состояний сложных объектов контроля, существенно отстает. В качестве причин данного отставания чаще всего выступает высокая ответственность за принимаемые информационные решения и недостаточная формализованность используемых моделей. Недостаточная формализованность моделей принятия решений приводит к тому, что зачастую отдельные эксперты по-разному оценивают ситуацию, сформированную на основе одинаковых исходных данных. Проблемы агрегирования знаний из нескольких источников различной типологии (эксперты, естественно-языковые тексты, базы данных), а также вопросы создания эффективных технологий автоматизированного приобретения знаний находятся в центре внимания разработчиков современных интеллектуальных систем [6].

В соответствии с традиционно используемым подходом к организации информационно-аналитической работы для идентификации состояний объектов контроля используются детерминированные модели эталонных состояний объектов контроля. В общем виде модель, описывающая возможные варианты состояний объекта контроля, в большинстве случаев определяется конечным множеством состояний и идентифицирующих их признаков, и может быть представлена в виде матрицы, общий вид которой представлен в таблице 1.

Таблица 1

Модель состояний объекта контроля в матричном виде

		Признаки, определяющие возможное состояние (p_i)					
		p_1	p_2	...	p_i	...	p_l
Состояния	S_1	b_{11}	b_{21}	...	b_{i1}	...	b_{l1}
	S_2	b_{12}	b_{22}	...	b_{i2}	...	b_{l2}
	
	S_r	b_{1r}	b_{2r}	...	b_{ir}	...	b_{lr}
	
	S_R	b_{1R}	b_{2R}	...	b_{3R}	...	b_{lR}

Здесь b_{ir} — весовой коэффициент, определяющий значение признака для идентификации состояния.

В ходе решения задач идентификации состояний объектов контроля по результатам оперативного мониторинга

параметров, производится соотнесение значений выявленных системой сбора параметров с эталонными образами возможных состояний. Сходство между выявленным и эталонным состоянием объекта контроля устанавливается в зависимости от выбранного метрического расстояния между ними. Если возможное состояние объекта контроля описывается r -м числом признаков, то он может быть описан как некоторая точка в N -мерном пространстве, а его сходство с эталонными состояниями будет определяться в соответствии с используемой мерой близости. Наиболее известными мерами близости, образованных N -мерных векторов, являются следующие расстояния: Евклидово, Хемминга, Минковского, Махаланобиса [7].

Выражение для расчета показателя близости образов состояний, с использованием расстояния Хемминга (city-block, или Манхэттенское) может быть представлено следующим выражением:

$$d_{ij} = \sum_{r=1}^m |x_{ir} - x_{jr}|,$$

где d_{ij} — расстояние между i -м и j -м образом состояний объектов;

x_{ir} , x_{jr} — значение r -го признака соответственно у i - и j -го образа состояний объектов.

Описанный подход долгое время обеспечивал требуемое качество решения информационно-аналитических задач. Но традиционная математическая статистика, претендовавшая на роль основного инструмента анализа данных и создавшая основу оперативной аналитической обработки данных (online analytical processing, OLAP), не обеспечивает решение задач связанных с необходимостью оперативной адаптации к изменяемым условиям ведения информационно-аналитической работы, связанных со значительным объемом подлежащей обработке информации в условиях ее высокой изменчивости и неопределенности.

Анализ моделей, используемых специалистами аналитических органов при решении задач, связанных с идентификацией состояний объектов контроля, свидетельствует о том, что в большинстве случаев эталонные образы возможных состояний исследуемых объектов основаны на усредненных (среднестатистических) значениях учитываемых параметров. В силу чего формируемые на их основе модели эталонных состояний содержат не фактические, а «обобщенные» образы реальных процессов, очищенные от специфики реальных возможных ситуаций в предметной области. Данные «усредненные» значения в значительной степени опираются на экспертные знания об исследуемых процессах, которые нередко являются субъективными. Кроме того, сформированные значения зачастую нуждаются в своевременной адаптации к изменениям, происходящим в исследуемой предметной области, что представляет серьезную практическую проблему.

Предполагается, что в ближайшем будущем в основу современных подходов к организации и ведению информационно-аналитической работы будут положены методы интеллектуального анализа данных (discovery-driven data mining, Data Mining), ориентированные на выявление и отражение многоаспектных взаимоотношений в данных.

Ожидается, что работы в области искусственного интеллекта и моделирования когнитивных процессов приведут к созданию принципиально нового поколения интеллектуальных технологий обнаружения новых фактов, закономерностей и связей в данных [6].

Для демонстрации потенциальных возможностей реализации интеллектуальных методов в интересах повышения качества решения задач информационно-аналитической работы, связанных с достоверной идентификацией состояний объектов контроля в условиях неполноты и неопределенности исходных данных, проведен имитационный эксперимент. При его реализации осуществлено сравнение результативности решения задачи распознавания образа состояния объекта контроля с использованием традиционных методов расчета, описанных выше мер близости и реализации нейросетевых методов распознавания.

ФОРМИРОВАНИЕ ИСХОДНЫХ ДАННЫХ ДЛЯ ЭКСПЕРИМЕНТА

В ходе эксперимента для подготовленной модели эталонных состояний объектов контроля, представленной в матричном виде (см. таблицу), содержащей пять возможных эталонных состояний ($R = 5$), представленных 16 признаками ($I = 16$). Каждый признак дискретен и бинарен. Основываясь на предположении о том, что доля возможных искажений исходных данных, поступающих в систему обработки, может составлять до 30 % ($\delta \leq 30$) от истинных исходных данных, была осуществлена генерация тестовой выборки данных $\{P_{(12)}, S_i\}_{100}$ с использованием генератора случайных чисел.

Сформированная выборка данных представляет собой модель «допустимых мутаций» ($\delta \leq 30$) относительно эталонных значений вектора признаков состояния S_i . Она отражает возможные значения векторов, описывающих текущие значения параметров, поступающих в систему обработки информации. Предположение о менее 30 % доли возможных искажений исходных данных основано на исследовании особенностей функционирования существующих информационно-аналитических систем.

Схема формирования тестовой выборки требуемого объема N , иллюстрирующая реализацию рандомизации эталонных значений признаков, представлена на рисунке 1.

На первом этапе эксперимента осуществлялось распознавание образов моделируемых ситуаций на основе традиционно реализуемых методов распознавания. После чего для каждого из сформированного множества векторов тестовой выборки данных $\{P_{(12)}\}_{100}$ посредством последовательного сравнения с каждым из эталонных наборов векторов $\{P_{(12)}, S_i\}_5$ осуществлен расчет показателя близости образов состояний с использованием нормированного расстояния Хемминга, которое может быть представлено следующим выражением:

$$d_{ij} = \sum_{r=1}^m |x_{ir} - x_{jr}| / m.$$

В качестве наиболее вероятного вывода о наблюдаемом состоянии формируемого в результате оценивания значений показателя близости рассматривалось значение S_i , соответствующее максимальному значению показателя близости ($\max(d_{ij})$).

После чего производилось оценивание соответствия распознанного (случайного) состояния \tilde{S}_i и эталонного значения состояния S_i , используемого при мутации r -го

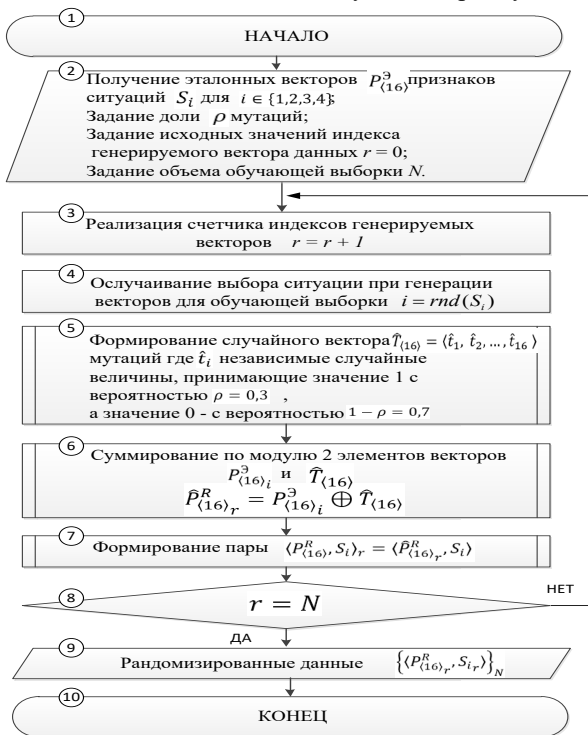


Рис. 1. Схема рандомизации эталонных векторов признаков ситуаций для формирования обучающей выборки потребного объема N .

вектора тестовой выборки данных $\{P_{(16)}, S_i\}_r$. Результаты реализации предложенного метода распознавания образа состояния объекта контроля, представленного описанной моделью, представлены на рисунке 2.



Рис. 2. Результаты реализации традиционных методов распознавания образа состояния объекта контроля

На втором этапе эксперимента осуществлялось распознавание образов моделируемых ситуаций на основе реализации нейросетевых технологии обработки данных.

Исходные данные для обучения нейросети были представлены сформированным множеством векторов признаков $P_{(R)} = \langle p_1, p_2, \dots, p_R \rangle$, описывающих возможные состояния S_i исследуемых объектов. При реализации нейросеть формирует зависимости для отображения $P_{(R)} \rightarrow \tilde{S}$: $\tilde{S} = f(P_{(R)}; A_{\diamond})$, где \tilde{S} — прогнозируемое нейросетью состояние объекта, A_{\diamond} — параметры модели нейросети, $f(P_{(R)}; A_{\diamond})$ — функция, символизирующая отображение аргумента $P_{(R)}$ при заданных параметрах A_{\diamond} в значение S_i . В процессе обучения используемые

в нейросети алгоритмы адаптируют параметры A_{\diamond} для снижения среднеквадратического расхождения между \tilde{S}_i и S_i состояниями объектов.

В случае недостаточного объема обучающей выборки возникает задача увеличения ее объема. Возможным методом увеличения объема обучающей выборки является рандомизация значений векторов $P_{(R)_j}$ из имеющихся пар $\langle P_{(R)_j}, S_{ij} \rangle$. Суть метода заключается в присвоении случайных значений заданной доли элементов $P_{(R)}$. Реализация данного подхода основана на учете возможных искажений в регистрации данных значений системой наблюдения.

При проведении эксперимента объем сформированной обучающей выборки составил 1 000 вариантов. Обучение проводилось методом Ф. Розенблата: проводилась итерационная подстройка матрицы весов для уменьшения ошибки выходных значений. Для оценивания качества распознавания состояний объектов с использованием обученной нейросети сформированная выборка разделена на обучающую и тестовую выборки в соотношении 80/20. Данная выборка использовалась для обучения 4 нейросетей с 1, 2, 3 и 10 слоями по 16 нейронов в каждом.

Полученные вероятности правильного распознавания ситуаций S_i на тестовой выборке в зависимости от числа эпох обучения и используемых параметров модели нейросети представлены на рисунке 3.

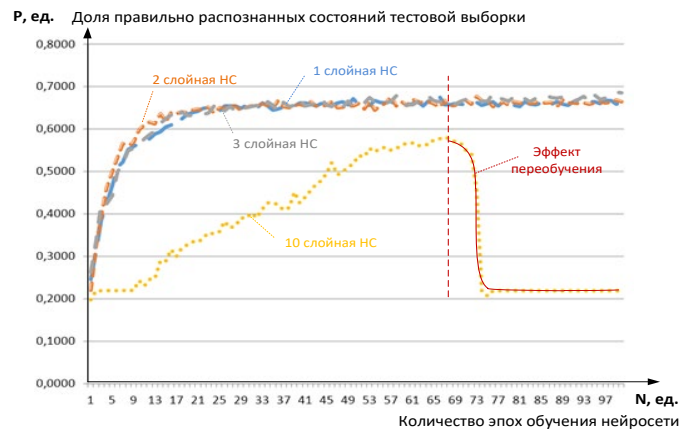


Рис. 3. Графики вероятностей правильного распознавания ситуаций S_i в зависимости от числа эпох обучения

Полученные графические зависимости определяют практическую пригодность выбранной структуры нейросети для решения описанной в эксперименте задачи. В соответствии с полученными результатами, реализованные на основе 1-3-уровневой структуры нейросети модели демонстрируют более высокую скорость обучения и более точное качество идентификации состояний объектов. Представленный на графике, демонстрирующем возможность реализации 10-уровневой нейросети, характерный спад показателя правильности идентификации состояний определяется эффектом переобучения нейросети [8]. Учет данного эффекта является важным фактором, подлежащим учету при выборе структуры нейросети.

Результаты имитационного моделирования демонстрируют достаточно высокий практический потенциал, связанный с использованием нейросетевых технологий

для решения задач идентификации состояний объектов мониторинга в условиях неполноты и частичной недостоверности исходных данных. Представленные значения (~0,67) вероятности правильного распознавания состояний объекта контроля, демонстрируют преимущества использования описанного подхода для совершенствования информационно-аналитической работы в деятельности органов государственного и регионального управления.

ЗАКЛЮЧЕНИЕ

Результаты проведенного анализа современных ситуационных условий функционирования информационно-аналитических структур ведомств, деятельность которых направлена на предотвращение кризисных ситуаций, демонстрируют потребность в формировании и реализации новых интеллектуальных методов обработки данных, призванных обеспечить помощь специалистам информационно-аналитических подразделений органов государственного и регионального управления в обработке существенно увеличившегося объема информации, характеризующегося высокой степенью неопределенности.

Результаты проведенного имитационного эксперимента свидетельствуют о реализуемости и потенциальной эффективности реализации предложенного подхода к совершенствованию информационно-аналитической работы посредством применения методов интеллектуального анализа данных для решения задач, связанных с идентификацией состояний объектов контроля в условиях неопределенности исходных данных. Общий прирост результативности решения рассмотренной задачи составил более 14 %.

ЛИТЕРАТУРА

1. Стратегия экономического и социального развития Санкт-Петербурга на период до 2030 года. URL: <http://spbstrategy2030.ru> (дата обращения 15.07.2019).
2. В систему видеонаблюдения Петербурга внедрили технологию распознавания лиц // Официальный сайт СПб ГКУ «ГМЦ». URL: <http://spb112.ru/news/193/detail/> (дата обращения 15.07.2019).
3. Савельев А. И. Проблемы применения законодательства о персональных данных в эпоху «Больших данных» (Big Data) // Право. Журнал Высшей школы экономики. 2015. № 1. С. 43–66.
4. Об утверждении Стратегии развития отрасли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года: распоряжение Правительства Российской Федерации от 01.11.2013 № 2036-р. URL: <http://government.ru/docs/8024>. (дата обращения 15.07.2019).
5. Поллак Г. А. Интеллектуальные информационные системы: Учебное пособие. — Челябинск: Издательский центр ЮУрГУ, 2011. — 141 с.
6. Николенко С. Глубокое обучение. Погружение в мир нейронных сетей / С. Николенко, А. Кадурын, Е. Архангельская — СПб.: Питер, 2018. — 480 с. — (Библиотека программиста).
7. Google Colaboratory. URL: <http://colab.research.google.com> (дата обращения 25.05.2019).
8. Теория и практика машинного обучения: Учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. — Ульяновск: УлГТУ, 2017. — 290 с.

Data Mining for the Tasks of Identifying the States of Objects Control in Conditions of Uncertainty

A. V. Pankov, J. V. Dashkevich
A. F. Mozhaisky Military Space Academy
Saint Petersburg, Russia
yvdspp55@gmail.ru

Abstract. The article proposes the ways to improve information analysis conducted by national and regional government bodies for public security tasks. The proposed approach is based on the idea of possible adaptation of artificial neural networks for identifying critical states of subject areas in conditions of incomplete or partially invalid input data. The presented simulation experiment describes the particular steps for the preparation of initial data to study a neural network model based on the situation of partial randomization of reference vectors of signs. The validity of using neural network methods for identifying the control objects state is determined by a comparative analysis between the effectiveness of their use and traditionally used methods that are based on the establishment of a metric distance between the results of operational monitoring and the reference state of the control object.

Keywords: neural network, situation recognition, information processing, Hamming distances, Big Data, OLAP, analytical structures of departments.

REFERENCES

1. The Strategy of Economic and Social Development of St. Petersburg for the Period Until 2030 [Strategiya ekonomicheskogo i sotsial'nogo razvitiya Sankt-Peterburga na period do 2030 goda]. Available at: <http://spbstrategy2030.ru> (accessed 15 July 2019).
2. Face Recognition Technology Implemented in St. Petersburg Video Surveillance System [V sistemu videonablyudeniya Peterburga vnedrili tekhnologiyu raspoznavaniya lits], *Official website St. Petersburg State-Funded Institution «City monitoring center» [Ofitsial'nyy sayt Sankt-Peterburgskogo gosudarstvennogo kazennogo uchrezhdeniya «Gorodskoy monitoringovyy tsentr»]*. Available at: <http://spb112.ru/news/193/detail/> (accessed 15 July 2019).
3. Savelyev A. I. The Issues of Implementing Legislation on Personal Data in the Era of Big Data [Problemy primeneniya zakonodatel'stva o personal'nykh dannykh v epokhu «Bol'shikh dannykh» (Big Data)], *Law. Journal of the Higher School of Economics [Pravo. Zhurnal Vysshey shkoly ekonomiki]*, 2015, No. 1. Pp. 43–66.
4. On Approval of the Strategy for the Development of the Information Technology Industry in the Russian Federation for 2014–2020 and for the Future Until 2025: Order of the Government of the Russian Federation [Ob utverzhdenii Strategii razvitiya otrasli informatsionnykh tekhnologiy v Rossiyskoy Federatsii na 2014–2020 gody i na perspektivu do 2025 goda: rasporyazhenie Pravitel'stva Rossiyskoy Federatsii] from 01.11.2013 No. 2036-r. Available at: <http://government.ru/docs/8024>. (accessed 15 July 2019).
5. Pollak G. A. Intelligent information systems: Study guide [Intellektualnye informatsionnye sistemy: Uchebnoe posobie], Chelyabinsk, Publishing center of South Ural State University, 2011, 141 p.
6. Nikolenko S., Kadurin A., Arkhangelskaya E. Deep learning. Immersion in the world of neural networks [Glubokoe obuchenie. Pogruzhenie v mir neyronnykh setey], St. Petersburg, Piter Publishing House, 2018, 480 p.
7. Google Colaboratory. Available at: <http://colab.research.google.com> (accessed 25 May 2019).
8. Voronina V. V., Mikheev A. V., Yarushkina N. G., Svyatov K. V. Theory and practice of machine learning: Study guide [Teoriya i praktika mashinnogo obucheniya: Uchebnoe posobie], Ulyanovsk, Ulyanovsk State Technical University, 2017, 290 p.

Метод распознавания радиосигналов на основе теории образов

к.т.н. Е. Г. Воробьев, аспирант Д. А. Олаоде

Санкт-Петербургский государственный электротехнический университет (ЛЭТИ) им. В.И. Ульянова (Ленина)

Санкт-Петербург, Россия

vrbyug@mail.ru, ola.john34@yahoo.com

Аннотация. Рассматриваются вопросы обеспечения доступности спутникового интернет-сервиса в регионах Республики Нигерия при мешающем воздействии зарубежных радиостанций на основе формирования и распознавания образа радиостанции по параметрам поляризационно-временного кодирования. Показано, что наиболее конструктивны методы защиты от помех, включая поляризационно-временные, которые обладают высокой эффективностью и структурной устойчивостью в условиях значительных изменений параметров сигналов и помех. Предлагаемый подход позволяет гарантированно селективировать сигнал спутника в любой точке наблюдения и создать ретрансляционный центр для внутренних районов Нигерии.

Ключевые слова: доступность информации, поляризация, кодирование, образ радиостанции.

ВВЕДЕНИЕ

Актуальность задачи распознавания радиосигналов спутника-ретранслятора интернет-сервиса в Республике Нигерия обусловлена наличием мешающих станций на территории сопредельных государств в условиях гористой местности и низкой высоты спутника над горизонтом из-за прохождения трассы спутника достаточно далеко от государственной границы Нигерии.

ОСНОВНЫЕ ПОДХОДЫ К РЕШЕНИЮ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБЪЕКТОВ И ЯВЛЕНИЙ

Процесс распознавания в общем случае состоит в том, что система распознавания на основании сопоставления апостериорной информации относительно каждого поступившего на вход системы объекта или явления с априорным описанием классов принимает решение о принадлежности этого объекта (явления) к одному из классов. Правило, которое каждому объекту ставит в соответствие определенное наименование класса, называют *решающим правилом*. В литературе, посвященной распознаванию образов, утвердилось мнение, что суть проблемы распознавания заключается в определении решающих правил, нахождении в признаковом пространстве таких границ (решающих границ), придерживаясь которых признаки пространства оптимальным образом, например, с точки зрения минимизации ошибок распознавания, подразделяются на области, соответствующие классам. Так, в [1] сказано, что в отыскании таких решающих правил на основании заданных описаний классов и заключается проблема распознавания.

При определении решающих правил (решающих границ в признаковом пространстве) в зависимости от объема исходной априорной информации рассматриваются следующие ситуации:

1. Количество исходной информации достаточно для того, чтобы путем ее анализа и непосредственной обработки определить решающие правила (системы распознавания без обучения).

2. Количество исходной информации недостаточно для определения решающих правил на основе ее непосредственной обработки, в связи с чем реализуется процедура обучения (обучающиеся системы распознавания).

В ситуациях 1 и 2 задача отыскания решающих правил базируется на том, что алфавит классов объектов и априорный словарь признаков, предназначенных для их описаний, известны. Рассматривается также и такая ситуация, когда словарь признаков известен, но неизвестен алфавит классов. При этом, однако, определен некоторый набор правил, в соответствии с которыми на основании процедуры самообучения находится искомым алфавит классов. Затем определяются решающие правила (самообучающиеся системы).

Исторически сложилось так, что первые теоретические исследования и прикладные работы в области распознавания базировались на том, что признаковое пространство известно, известен также и алфавит классов [1]. В этих условиях проблема распознавания действительно может трактоваться как проблема определения в некотором смысле наилучших решающих границ (решающих правил). В настоящее время часто при построении распознающих устройств имеет место ситуация, когда известны и алфавит классов, и словарь признаков [2–4]. Однако в общем случае при построении реальных систем распознавания, требующих разработки специальных измерительных средств и целых измерительных комплексов, исходить из того, что алфавит классов и словарь признаков априорно известны, к сожалению, не приходится, поэтому системы распознавания, являясь частью системы управления (автоматической или автоматизированной), должны строиться с учетом обеспечения наиболее эффективного использования всего набора допустимых решений. Этот факт накладывает на построение систем распознавания следующие ограничения.

1. При прочих равных условиях повышение эффективности принимаемых решений следует связывать со степенью детализации определения или назначения либо характера распознаваемого объекта или явления. Степень детализации определяется количеством классов, на которые подразделено множество объектов или явлений. Так, если система управления располагает m различными решениями, то в алфавите классов системы распознавания, учитывая сказанное, целесообразно предусмотреть $m + 1$ классов. Тогда, если распознанный объект относится к классу Ω_1 ,

принимается решение l_1 , если к классу Ω_2 — решение l_2 и т. д., если же объект относится к классу Ω_{m+1} , решение не принимается.

2. Эффективность принимаемых системой управления решений при прочих равных условиях (в том числе, естественно, при заданном алфавите классов) зависит от точности определения принадлежности распознаваемого объекта или явления к соответствующему классу. Точность же определения или ошибка распознавания при заданном по точности априорном описании классов определяется размерностью и информативностью признакового пространства, объемом и качеством апостериорной информации о значениях признаков (параметров), которыми характеризуется распознаваемый объект. Иначе говоря, расширение алфавита классов, увеличивающее степень детализации определения назначения либо характера распознаваемого объекта (явления), при неизменном словаре признаков увеличивает ошибку распознавания.

Пусть заданы три класса Ω_1, Ω_2 , и Ω_3 объектов распределениями $f_1(x), f_2(x), f_3(x)$, априорными вероятностями появления объектов соответствующих классов $P(\Omega_1) = P(\Omega_2) = P(\Omega_3) = P$, а также потерями $c_{11} = c_{22} = c_{33} = 0, c_{12} = c_{21} = c_{13} = c_{31} = c_{23} = c_{32} = c$.

На рисунке 1 представлены законы распределений. Средний (байесовский) риск

$$\begin{aligned} \dot{R}_1 &= Pc \left[\int_{-\infty}^a f_2(x) dx + \int_{-\infty}^a f_3(x) dx + \int_a^b f_1(x) dx + \right. \\ &\quad \left. + \int_a^b f_3(x) dx + \int_b^{\infty} f_1(x) dx + \int_b^{\infty} f_2(x) dx \right] = \\ &= Pc \left[\int_{-\infty}^a f_2(x) dx + \int_{-\infty}^b f_3(x) dx + \int_a^{\infty} f_1(x) dx + \int_b^{\infty} f_2(x) dx \right]. \end{aligned}$$

Положим теперь, что объекты, относящиеся к классам Ω_1 и Ω_2 , решено объединить в один класс Ω_4 , описание которого

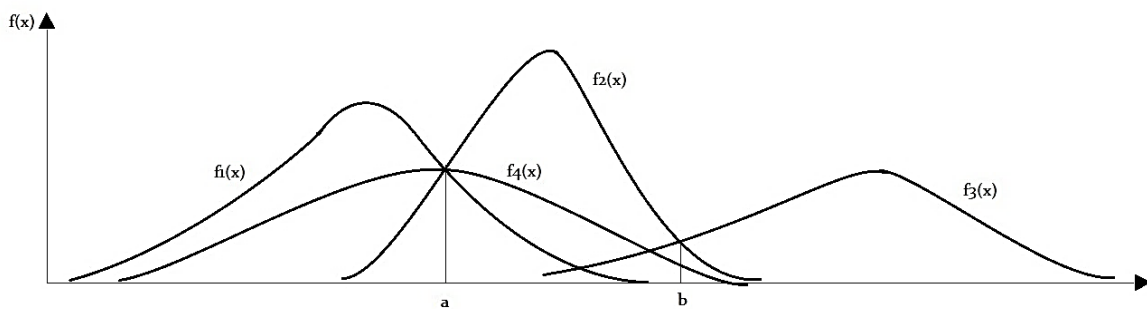


Рис. 1. Законы распределений появления объектов соответствующих классов

Таким образом, стремление по возможности наиболее эффективно использовать набор возможных решений системы управления приводит к необходимости увеличения алфавита классов до $m + 1$. Однако естественная ограниченность ресурсов, ассигнованных на построение измерительных средств системы распознавания или системы распознавания в целом, приводит к тому, что по мере увеличения алфавитов классов ошибки распознавания растут, а это уменьшает эффективность использования возможных решений. Только некоторый компромисс между размерами алфавита классов и объемом рабочего словаря признаков системы, базирующийся на исходных данных

$$f_4(x) = \frac{P(\Omega_1)f_1(x) + P(\Omega_2)f_2(x)}{P(\Omega_1) + P(\Omega_2)} = \frac{f_1(x) + f_2(x)}{2};$$

$$P(\Omega_4) = 2P, C_{43} = C_{34} = C.$$

Средний риск в данном случае в предположении неизменности границы b составит

$$\begin{aligned} \dot{R}_2 &= c \left[P \int_{-\infty}^b f_3(x) dx + 2P \int_b^{\infty} f_4(x) dx \right] = \\ &= Pc \left[\int_{-\infty}^b f_3(x) dx + \int_b^{\infty} f_1(x) dx + \int_b^{\infty} f_2(x) dx \right]. \end{aligned}$$

Из сравнения величин \dot{R}_1 и \dot{R}_2 видно, что $\dot{R}_1 > \dot{R}_2$ на величину

$$\Delta R = \dot{R}_1 - \dot{R}_2 = Pc \left[\int_{-\infty}^a f_2(x) dx + \int_a^b f_1(x) dx \right].$$

Следовательно, при заданном признаковом пространстве и прочих равных условиях уменьшение числа классов приводит к уменьшению ошибок распознавания, и наоборот, при увеличении числа классов системы распознавания в целях поддержания на заданном уровне или даже уменьшения среднего риска (вероятности ошибочных решений) надо расширять словарь признаков (естественно, при прочих равных условиях). В то же время расширение признакового пространства в целях уменьшения ошибок распознавания сопряжено с увеличением числа технических измерительных средств, каждое из которых обеспечивает определение соответствующего признака или группы признаков. Это, в свою очередь, требует увеличение затрат на построение системы распознавания. На величину затрат в реальных условиях, как правило, накладываются те или другие ограничения.

относительно набора возможных решений и величины ресурсов, отпущенных на создание измерительной аппаратуры, реализующей словарь признаков, позволяет обеспечить решение задачи построения системы распознавания оптимальным образом.

Исходя из вышеизложенного, суть проблемы распознавания состоит в разработке таких алфавита классов и словаря признаков, которые в условиях ограниченных ресурсов на построение системы распознавания обеспечивают максимальную эффективность системы управления, принимающей соответствующее решение в зависимости от результатов решения задачи распознавания. При этом,

безусловно, выбирая словарь признаков и определяя алфавит классов, следует находить наилучшие решающие правила, решающие границы между классами. Однако в общем случае не в этом состоит проблема распознавания, как ни важна и как подчас ни сложна задача определения оптимальных решающих правил, обеспечивающих в условиях заданных алфавита классов и словаря признаков наибольшую точность распознавания [4].

ПОСТАНОВКА ЗАДАЧИ РАСПОЗНАВАНИЯ СИГНАЛА СПУТНИКОВОЙ РАДИОСТАНЦИИ НА ОСНОВЕ ПОЛЯРИМЕТРИИ

Пусть задано множество признаков радиосигналов $\Omega = \{\omega_1, \dots, \omega_z\}$, а также множество возможных решений $L = \{l_1, \dots, l_k\}$, которые могут быть приняты системой управления по результатам решения задачи распознавания. Введем в рассмотрение множество возможных вариантов разбиения объектов на классы $A = \{A_1, \dots, A_r\}$. Будем полагать, что если выбран вариант разбиения $A_a, a = 1, \dots, r$, то множество Ω подразделяется на m_a классов, т. е. $A_a: \Omega_q^{A_a} \cap \Omega_g^{A_a} = \emptyset, \cup_{i=1}^{m_a} \Omega_i^{A_a} = \Omega, q, g = 1, \dots, m_a, q \neq g$.

Пусть первоначальная информация позволяет построить априорное признаковое пространство (составить априорный словарь признаков), описываемое многомерным вектором $x_a = \{x_1, \dots, x_N\}$. Информация относительно множества решений $L = \{l_1, \dots, l_k\}$ позволяет произвести исходное разбиение множества объектов на классы, т. е. составить априорный алфавит классов. В первом варианте подразделения объектов на классы ($\alpha = 1$), т. е. когда $A_a = A_1$, их число равно $m_a = m_1 = k + 1$. Исходное множество объектов $\Omega = \{\omega_1, \dots, \omega_z\}$ (обучающую выборку) подразделим на подмножества — классы $\Omega_1^{A_1}, \Omega_2^{A_1}, \dots, \Omega_{m_1}^{A_1}$.

Если обучающая выборка достаточно представительна, то непосредственной обработкой исходной информации можно определить описания классов. При статистическом подходе к задаче распознавания такими описаниями являются априорные вероятности $P(\Omega_i^{A_1})$ появления объектов соответствующих классов, а также условные плотности распределения значений признаков по классам, т. е. функции $f_{\Omega_i}^{A_1}(x_1, \dots, x_N), i = 1, \dots, m_1$. Если объем исходной априорной информации недостаточен для непосредственного описания классов, то они могут быть получены с помощью процедуры обучения.

Наличие описаний классов параметров радиосигналов в принципе позволяет определить решающие правила (решающие границы), использование которых обеспечивает минимизацию ошибок при распознавании неизвестных сигналов. Обозначим через $P_{\text{пр}}(\Omega_i^{A_1} | x_a)$ оценку апостериорной вероятности правильного решения задачи распознавания, усредненную по всем возможным значениям признаков априорного словаря, описываемого вектором x_a . Эта оценка может быть получена проведением статистических испытаний (метод Монте–Карло) математической модели системы распознавания.

Если бы не было ограничений на величину ресурсов, ассигнованных на построение измерительных устройств, предназначенных для определения поляризационных признаков x_1, x_2, \dots , то можно было бы полагать, что основные

характеристики системы распознавания — алфавит классов и словарь признаков — определены, и можно приступить к построению системы распознавания [1, 4]. В условиях ограничений, когда реализовать априорное признаковое пространство $x_a = \{x_1, \dots, x_N\}$ в полном объеме не представляется возможным, приходится его сокращать по сравнению с априорным, т. е. переходить от априорного словаря признаков к рабочему.

Рассмотрим вектор $\lambda = \{\lambda_1, \dots, \lambda_N\}$, компоненты которого $\lambda = \frac{1}{0}$ (в зависимости от того, используется ли данный признак априорного словаря в рабочем словаре или нет). Кроме того, введем обозначение для рабочего словаря

$x_p = \{x_{j_1}, \dots, x_{j_n}\}$, где $j_1, \dots, j_n \in 1, \dots, N$, т. е. множество признаков рабочего словаря состоит из элементов множества признаков априорного словаря (рабочий словарь представляет собой подмножество множества признаков априорного словаря).

Обозначим через C_j стоимость создания измерительно-го устройства, обеспечивающего определение x_j -го признака, $j=1, \dots, N$, а через C_0 — общую величину ресурсов, ассигнованных на создание всех измерителей. Если $\sum_{j=1}^N C_j = C_0$, то в качестве рабочего словаря системы распознавания может быть использован априорный словарь. Однако в общем случае, как правило, суммарная стоимость создания комплекса технических средств, обеспечивающих измерение всех признаков априорного словаря, превышает величину C_0 , т. е. $\sum_{j=1}^N C_j > C_0$. Затраты на создание комплекса технических средств, обеспечивающих измерение признаков рабочего словаря, определяются величинами $C = \sum_{j=1}^N \lambda_j C_j = \sum_{\beta=1}^N C_{j\beta} \leq C_0$.

Обозначим через $G(\Omega_i^{A_1})$ выигрыш, связанный с реализацией возможных решений при распознавании объекта ω , отнесенного к классу $\Omega_i^{A_1}$ в варианте классификации A_1 . Тогда математическое ожидание выигрыша от выбора варианта A_1 при использовании априорного словаря признаков

$$R = M[G(\Omega_i^{A_1})] = \sum_{i=1}^{m_1} P_{\text{пр}}(\Omega_i^{A_1} | x_a) G(\Omega_i^{A_1}).$$

Величину R уместно рассматривать в качестве критерия эффективности системы распознавания, и, следовательно, с его максимизацией нужно связывать увеличение эффективности ее функционирования.

В условиях ограничений, определяемых величиной C_0 , возникает следующая экстремальная задача: необходимо в пределах C_0 найти такой вариант разбиения объектов на классы и такое пространство признаков, при которых обеспечивается максимальное значение критерия эффективности системы R . Другими словами, необходимо определить $A_\alpha = A^0$ из множества $A = \{A^1, \dots, A_\alpha, \dots, A_r\}$ и вектор $\lambda = \lambda^0$, которые при наилучшем решающем правиле доставляют экстремальное (максимальное) значение величины R при соблюдении ограничений на величину $C \leq C_0$, т. е.

$$R = R(A^0, \lambda^0) = \max_{A_\alpha} \max_{\lambda} R = \\ = \max_{A_\alpha} \max_{\lambda} \sum_{i=1}^{m_\alpha} P_{\text{пр}}(\Omega_i^{A_\alpha} | \lambda) G(\Omega_i^{A_\alpha}),$$

с учетом $C = \sum_{j=1}^N \lambda_j C_j \leq C_0$. При этом A^0 определяет алфавит классов, а λ^0 — оптимальный рабочий словарь признаков.

Итак, *общая постановка проблемы распознавания спутниковых радиосигналов по их поляризационным параметрам может быть сформулирована следующим образом: в условиях первоначального (априорного) описания исходного множества объектов на языке априорного словаря признаков необходимо в пределах выделенных ресурсов на построение измерительной аппаратуры определить оптимальный алфавит классов и оптимальный рабочий словарь признаков, которые при наилучшем решающем правиле обеспечивают наиболее эффективное использование решений, принимаемых по результатам распознавания неизвестных объектов или явлений системой управления.*

МЕТОД РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ

Рассмотренная постановка проблемы распознавания позволяет определить последовательность задач, возникающих при разработке системы распознавания, предложить их формулировки и возможные методы решения. Наиболее экономный метод решения проблемы построения системы распознавания — метод математического или физико-математического моделирования. Основная идея работы предлагаемой модели разрабатываемой системы распознавания — реализация итеративной процедуры, обеспечивающей путем последовательных приближений синтез системы, эффективность работы которой достаточно близко приближается к потенциально достижимой.

Для построения модели необходимы:

1. Множество возможных решений, которые могут быть приняты системой управления на основании результатов распознавания неизвестных объектов или явлений $L = \{l_1, \dots, l_k\}$.

2. Априорный словарь признаков $x_\alpha = \{x_1, \dots, x_N\}$.

3. Исходное множество объектов $\Omega = \{\omega_1, \dots, \omega_z\}$.

4. Величина ресурсов C_0 , ассигнованных на построение измерительной аппаратуры системы.

5. Значение выигрышей, получаемых системой управления от конкретных решений из множества возможных решений $L = \{l_1, \dots, l_k\}$, принимаемых по результатам решения задачи распознавания, т. е. величин $G(\Omega_i^{A_\alpha}), i = 1, \dots, m; \alpha = 1, r$.

Последовательность построения и работы модели состоит из таких этапов.

Первый этап предназначен для построения модели системы распознавания в первом приближении ($\alpha = 1$). Алгоритм реализации этого этапа следующий.

1. Определяется первый вариант разбиения множества объектов на классы A_1 , в соответствии с которым количество классов $m_1 = k + 1$. При этом к классу $\Omega_1^{A_1}$ относятся объекты, применительно к которым следует принимать решение l_1 , к классу $\Omega_2^{A_1}$ — объекты, к которым надо при-

нимать решение l_2 , и т. д., к классу $\Omega_k^{A_1}$ — объекты, к которым надо принимать решение l_k , к объектам класса $\Omega_{m_1}^{A_1}$ решение не принимается.

2. Определяется непосредственно либо подмножество множества каждого класса:

$$\Omega_1^{A_1} = (\omega_{\gamma_1}^1, \omega_{\gamma_2}^1, \dots, \omega_{\gamma_\psi}^1), \dots, \Omega_k^{A_1} = (\omega_{\gamma_\varphi}^1, \dots, \omega_{\gamma_r}^1),$$

где $\{(\gamma_1, \dots, \gamma_\psi), \dots, (\gamma_\varphi, \dots, \gamma_r)\} \in 1, \dots, z$, либо разрабатывается некоторый набор правил относительно значений признаков, содержащихся в априорном словаре, в соответствии с которым на основе методов самообучения при известном числе классов определяются объекты исходного множества, относящихся к каждому классу.

3. Производится описание каждого класса на языке признаков априорного словаря, а затем находятся наилучшие решающие границы между классами. Эта задача проблемы распознавания — традиционная, методы ее решения достаточно подробно рассмотрены, например, в [1, 5–8].

4. Проверяется, достаточна ли величина C_0 для построения измерителей, обеспечивающих всех признаков $x_\alpha = \{x_1, \dots, x_N\}$ априорного словаря. Если $C_0 \geq \sum_{j=1}^n C_j$, то в рабочий словарь системы включаются все признаки априорного словаря. Если $C_0 < \sum_{j=1}^n C_j$, для определения первого приближения *оптимального рабочего словаря системы* (словаря, который при заданных ограничениях на величину C_0 обеспечивает, например, либо минимум величины среднего риска, либо максимум среднеквадратического расстояния между классами, либо экстремальное значение какого-нибудь другого критерия) могут быть использованы другие методы.

5. Производится описание классов $\Omega_i^{A_1}, i = 1, \dots, m$, на языке рабочего словаря признаков первого приближения и определяются наилучшие решающие границы между ними.

6. Оценивается вероятность правильного решения задачи распознавания $P_{\text{пр}}[\Omega_i^{A_1} | x_{j_1}, \dots, x_{j_m}]$. Для этого строится математическая модель и проводятся ее статистические испытания.

7. Вычисляется первое приближение значения критерия эффективности системы $R^{(1)}$.

На этом завершается первый этап построения системы распознавания. Второй и последующие этапы предназначены для уточнения модели системы. Их цель — определить такой вариант разбиения объектов на классы A^0 и такой словарь признаков, при которых критерий R достигает наибольшего значения. Алгоритм реализации этих этапов таков.

1. В алфавите классов первого приближения определяется такой класс $\Omega_v^{A_1}, v = 1, \dots, m$ (либо, исходя из практических соображений, 2-3 класса), для которого величина $G(\Omega_v^{A_1}) = \min_i G(\Omega_i^{A_1})$. Это означает, что к классу Ω_v относятся такие объекты, распознавание которых обеспечивает по сравнению с распознаванием объектов других классов наименьший выигрыш.

2. Из алфавита классов первого приближения исключается класс $\Omega_v^{A_1}$, а объекты этого класса надлежит отне-

сти к такому классу $\Omega_{\mu}^{A_1}, \mu = 1, \dots, m$, для которого уменьшение величины $G(\Omega_{\mu}^{A_1})$ по сравнению с уменьшением величины $G(\Omega_{\mu}^{A_1}), i = 1, \dots, m$, минимально, т. е.

$$\Delta G(\Omega_{\mu}^{A_1}) = \min_i \Delta G(\Omega_{\mu}^{A_1}), \mu, i = 1, \dots, m; \mu \neq i.$$

Таким образом, определяется второй вариант разбиения объектов на классы A_2 , применительно к которому вновь повторяются операции 1–7, и определяется второе приближение значения критерия R^2 эффективности системы. Практически несколько итераций достаточно для определения такого варианта построения системы, при котором критерий R эффективности системы достигает наибольшего значения.

В заключение следует заметить, что только системный подход к проблеме распознавания объектов и явлений позволяет в реальных условиях при наличии неизбежных ограничений добиться наибольшей эффективности комплекса «система распознавания + система управления».

ЗАКЛЮЧЕНИЕ

Сигнально-помеховая ситуация, складывающаяся в различных условиях, достаточно разнообразна. Можно, однако, утверждать, что вероятность обеспечения связи с заданной надежностью возрастает с увеличением помехозащищенности средств и систем связи. Наиболее конструктивны при этом те методы помехозащиты, включая поляризационно-временные, которые обладают высокой эффективностью и структурной устойчивостью в условиях значительных изменений параметров сигналов и помех. Выбор структуры алгоритмов поляризационно-временной обработки с оценкой вектора весовых коэффициентов, удовлетворяют этим требованиям, поскольку в достаточно широком диапазоне изменений параметров объекта управления и регулятора сохраняется внутренняя устойчивость системы и регулируемость ее выходного сигнала.

С незначительными модернизациями разработанные алгоритмы оказались пригодны и эффективны для селекции как непрерывных, так и дискретных (в том числе импульсных) помех. Их можно использовать даже в неопределенных статистических ситуациях и непосредственно реализовать с помощью элементов ЭВМ.

В статье рассмотрены особенности применения распознавания образов сигналов спутника, обеспечивающего интернет-сервис. На основании записи паспорта (образа) сигнала в каждой точке зоны наблюдения спутника возможна пространственная селекция в условиях массовой работы мешающих станций.

ЛИТЕРАТУРА

1. Лекции по теории образов: в 3-х т. Т. 2: Анализ образов / У. Гренандер; Пер. с англ. И. Гуревича; под ред. Ю. Журавлева. — М.: Мир. Редакция литературы по математическим наукам, 1981. — 448 с.
2. Воробьев Е. Г. Управляемая поляризация электромагнитных волн как средство повышения скрытности передачи информации // Известия СПбГЭТУ «ЛЭТИ». 2014. № 9. С. 44–49.
3. Маткасым Н. Н. Распознавание образов с помощью нейронных сетей / Н. Н. Маткасым; науч. рук. С. В. Аксёнов // Технологии Microsoft в теории и практике программирования: сборник трудов XIII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых, г. Томск, 22–23 марта 2016 г. — Томск: Издательство ТПУ, 2016. — С. 23–25.
4. Обрывалин С. Г. Влияние поляризации радиоволн на характеристики обнаружения радиолокационных станций // Научный Вестник МГТУ ГА. 2011. № 168. С.122-125.
5. Фомин Я. А. Распознавание образов: теория и применения. — 2-е изд., доп. — М.: ФАЗИС, 2012. — 429 с.
6. Фомин Я. А. Статистическая теория распознавания образов / Я. А. Фомин, Г. Р. Тарловский. — М.: Радио и связь, 1986. — 264 с.
7. Форсайт Д. А., Понс Д. Компьютерное зрение. Современный подход = Computer Vision. A Modern Approach / Пер. с англ. под ред. А. В. Назаренко. — М.: Издательский дом «Вильямс», 2004. — 928 с.
8. Черногорова Ю. В. Методы распознавания образов // Молодой ученый. 2016. № 28 (132). С. 40–42.
URL: <http://moluch.ru/archive/132/36964> (дата обращения 10.10.2019)

Method Radio Signals Detection on the Image Recognize Theory

PhD E. G. Vorobiev, PhD Student J. A. Olaode
Saint Petersburg Electrotechnical University «LETI»
Saint Petersburg, Russia
vrbyug@mail.ru, ola.john34@yahoo.com

Abstract. The issues of ensuring the availability of satellite Internet service in the regions of the Republic of Nigeria under the interfering effect of foreign radio centers based on the formation and recognition of the image of a radio station by the parameters of polarization-temporal coding are considered. It is shown that the most constructive methods of protection against interference, including polarization-temporal, which are highly efficient and structurally stable under conditions of significant changes in signal and interference parameters. The proposed approach makes it possible to reliably select a satellite signal at any observation point and create a relay center for the interior of Nigeria.

Keywords: accessibility of information, polarization, coding, image of a radio station

REFERENCES

1. Grenander U. Pattern Analysis. Lectures in Pattern Theory. Volume II [Lektsii po teorii obrazov. Tom 2. Analiz obrazov], Moscow, Mir Publishers, 1981, 448 p.
2. Vorobiev E. G. Use of Electromagnetic Waves Operated Polarization for Increase of Information Transfer Reserve [Upravlyayemaya polarizatsiya elektromagnitnykh voln kak sredstvo povysheniya skrytnosti peredachi informatsii] *Proceedings of Saint Petersburg Electrotechnical University [Izvestiya SPbGETU «LETI»]*, 2014, No. 9, P. 44–49.
3. Matkashim N. N. Recognition of Images Using Neural Networks [Raspoznavanie obrazov s pomoshch'yu neyronnykh

setey] *In: Microsoft Technologies in theory and programming practice. A collection of works XIII All-Russian Scientific and Practical Conference of students, graduate students and young scientists [Tekhnologii Microsoft v teorii i praktike programirovaniya: sbornik trudov XIII Vserossiyskoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh]*, Tomsk, March 22–23, 2016, Tomsk, Tomsk Polytechnic University, 2016, Pp. 23–25.

4. Obryvalin S. G. Influence to Polarizations Radiowave on the Radar Finding Characteristics [Vliyaniye polarizatsii radiovoln na kharakteristiki obnaruzheniya radiolokatsionnykh stantsiy], *Civil Aviation High Technologies [Nauchnyy Bestnik MGTU GA]*, 2011, No. 168, Pp. 122–125.

5. Fomin Ya. A. Image recognition: theory and application [Raspoznavanie obrazov: teoriya i primeneniya], Moscow, FAZIS Publishers, 2012, 429 p.

6. Fomin Ya. A., Tarlovsky G. R. Statistical Theory of Pattern Recognition [Statisticheskaya teoriya raspoznavaniya obrazov], Moscow, Radio and Communication Publishers, 1986, 624 p.

7. Forsyth D. A., Ponce J. Computer vision. A Modern Approach [Komp'yuternoe zrenie. Sovremennyy podkhod], Moscow, Williams Publishing House, 2004, 928 p.

8. Chernogorova Y.V. Image Recognition Techniques [Metody raspoznavaniya obrazov], *Young Scientist [Molodoy uchenyy]*, 2016. No. 28. Pp. 40–42.

Available at: <http://moluch.ru/archive/132/36964> (accessed 10 Oct. 2019).

Характеристики алгоритма сжатия данных без потерь на основе псевдорегулярных чисел

магистр И. А. Молодкин, к.т.н. Е. М. Шурыгин

Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
imolodkin@gmail.com

Аннотация. Рассматривается алгоритм сжатия данных на основе псевдорегулярных чисел. В основе алгоритма заложен принцип использования специальных кодов, названных псевдорегулярными числами, по аналогии с широко известными псевдослучайными последовательностями. ПРЧ выступают в качестве «осей симметрии» при математическом преобразовании двоичного кода. Проводится ряд экспериментов для его сравнения с классическим алгоритмом LZW и алгоритмом Хаффмана. При этом оцениваются характеристики степени сжатия и быстродействия. Показано, что для псевдорегулярных чисел использование указанного алгоритма обеспечивает сжатие целочисленных данных без потерь с существенно лучшими характеристиками по степени сжатия.

Ключевые слова: сжатие, кодирование, архиваторы, алгоритм LZW, алгоритм Хаффмана.

ВВЕДЕНИЕ

Процесс цифровизации, который происходит во всех областях мировой деятельности, приводит к чрезвычайно быстрому росту объема данных. Компания IDC в своем исследовании выдвигает гипотезу о том, что глобальный объем данных вырастет с 33 зеттабайт в 2018 до 175 зеттабайт к 2025 году (рис. 1). Из них 49 % будет храниться в облачных хранилищах [1].

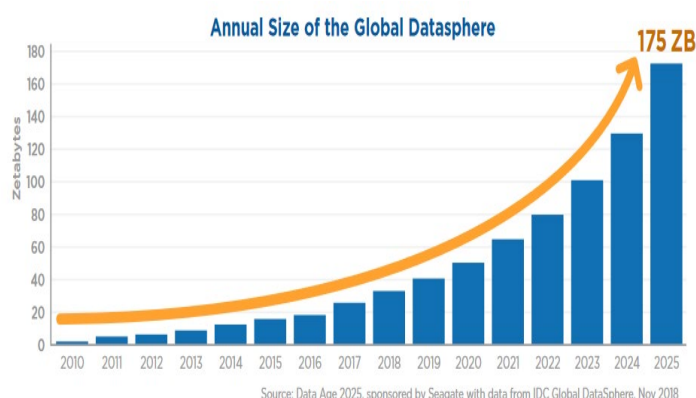


Рис. 1. Динамика роста объема информации в мире

Проблемы хранения и эффективного сжатия данных являются одними из важнейших в области информационных технологий [2]. Существующие алгоритмы сжатия данных без потерь плохо обрабатывают длинные двоичные коды, а также ограничения, связанные с применением модели избыточности, введенной в практику К. Шенноном в 40-х годах прошлого века [3].

На сегодняшний день наиболее распространенными методами сжатия данных без потерь являются следующие алгоритмы [4]:

- LZ77;
- LZ78;
- LZW;
- RLE;
- DEFLATE;
- RAR;
- Алгоритм Хаффмана.

Эти алгоритмы, как правило, используют статистические свойства отдельных битов или байтов файлов [3].

В алгоритме, описанном в статье [5], предлагается использование нового подхода к сжатию данных — на основе применения так называемых псевдорегулярных чисел (ПРЧ). Для псевдорегулярных чисел использование указанного алгоритма обеспечивает сжатие целочисленных данных без потерь с существенно лучшими характеристиками по степени сжатия.

Рассмотрим этот алгоритм подробнее и проведем его сравнение с классическим методом сжатия LZW и алгоритмом Хаффмана.

АЛГОРИТМ LZW

Алгоритм LZW (Lempel–Ziv–Welch) был опубликован Велчем в 1984 году в качестве улучшенной реализации алгоритма LZ78, созданного Лемпелем и Зивом в 1978 году [6].

Как было отмечено ранее, LZW является универсальным алгоритмом сжатия без потерь и использует технологию адаптивного словаря. Словарь создается в процессе сжатия, что позволяет производить кодирование на лету. Стоит отметить, что сам словарь не передается и формируется при декодировании [7].

Основные достоинства алгоритма — отсутствие вычисления вероятностей встречаемости символов или кодов и, следовательно, высокая скорость работы.

Очевидно, что этот алгоритм наиболее эффективен для файлов, содержащих длинные повторяющиеся последовательности [8].

Этот метод позволяет достичь одной из наилучших степеней сжатия среди других существующих методов сжатия данных (при полном отсутствии потерь или искажений в исходных файлах). В настоящее время используется в файлах формата TIFF, PDF, GIF, PostScript и других, а также, отчасти, во многих популярных программах сжатия данных (ZIP, ARJ, LHA).

АЛГОРИТМ ХАФФМАНА

Еще один классический алгоритм сжатия данных — алгоритм Хаффмана, разработанный в 1952 году аспирантом Массачусетского технологического института Дэвидом Хаффманом.

Построение кода Хаффмана сводится к построению соответствующего бинарного дерева по следующему алгоритму:

1. Символы входного алфавита образуют список свободных узлов. Каждый лист имеет вес, который может быть равен либо вероятности, либо количеству вхождений символа в сжимаемое сообщение.

2. Выбираются два свободных узла дерева с наименьшими весами.

3. Создается их родитель с весом, равным их суммарному весу.

4. Родитель добавляется в список свободных узлов, а два его потомка удаляются из этого списка.

5. Одной дуге, выходящей из родителя, ставится в соответствие бит 1, другой — бит 0. Битовые значения ветвей, исходящих от корня, не зависят от весов потомков.

6. Шаги, начиная со второго, повторяются до тех пор, пока в списке свободных узлов не останется только один свободный узел. Он и будет считаться корнем дерева.

Суть алгоритма заключается в кодировании символов, которые имеют наибольшую вероятность появления, наиболее короткими кодами. Несмотря на то, что коды имеют переменную длину, они имеют уникальный префикс, который позволяет однозначно их декодировать [9].

Недостатком этого алгоритма является необходимость дважды производить чтение файла — первый раз для составления вероятностной модели, второй раз — непосредственно для кодирования. Это приводит к увеличению временных затрат.

АЛГОРИТМ НА ОСНОВЕ ПРЧ

В основе алгоритма заложен принцип использования специальных кодов, названных автором псевдрегулярными числами, по аналогии с широко известными псевдослучайными последовательностями. ПРЧ выступают в качестве «осей симметрии» при математическом преобразовании двоичного кода. Эти коды должны удовлетворять следующим условиям:

- быть вычислимыми;
- количество операций для генерации ПРЧ, независимо от его длины, не должно превышать числа основания кода.

Примеры ПРЧ приведены в таблице 1.

Существует пять дополнительных типов ПРЧ:

- 10-й тип — 10...01 — это 10 + ПРЧ тип 4 (заполненный справа);
- 11-й тип — 110...01 — это 11 + ПРЧ тип 4 (заполненный справа);
- 12-й тип — 101111...1 — это 10 + ПРЧ тип 7 (заполненный справа);
- 13-й тип — 01000...01 — это 01 + ПРЧ тип 4 (заполненный справа);
- 14-й тип — 111...101 — это ПРЧ тип 7 (заполненный слева) + 01 (т. е. отображение типа 12);
- 15-й тип - 000...000.

Кодирование данных происходит в несколько шагов:

1. Исходный файл представляется в памяти ЭВМ в виде двоичного числа длины n .

Таблица 1

Примеры ПРЧ

Тип ПРЧ	Вид	Структура, i — номер бита
1	10101010	0 для нечетных i 1 для четных i
2	01010101	1 для четных i 0 для нечетных i
3	01111111	1 для $i \in [1, n-1]$ 0 для $i=n$
4	10000000	0 для $i \in [1, n-1]$ 1 для $i=n$
5	00001111	1 для $i \in [1, n/2]$ 0 для $i \in [n/2+1, n]$
6	11110000	0 для $i \in [1, n/2]$ 1 для $i \in [n/2+1, n]$
7	11111111	1 для $i \in [1, n]$
8	00000001	1 для $i=1$ 0 для $i \in [2, n]$
9	11111110	0 для $i=1$ 1 для $i \in [2, n]$
10	00000000	0 для $i \in [1, n]$

2. Проводится проверка двоичного числа на псевдрегулярность путем сверки с основными типами ПРЧ длины исходного файла n : если «да», то производится запись в служебный (выходной) файл требуемых характеристик. Если «нет», то переходим к следующему шагу.

3. Производится деление двоичного числа на все типы ПРЧ. В каждом раунде цикла проверяется целочисленность результата деления: если «да», то производится запись в служебный (выходной) файл требуемых характеристик. Если «нет», то переходят к следующему шагу.

4. Выбирается самое меньшее по количеству разрядов частное и производится запись в служебный (выходной) файл требуемых характеристик.

В результате работы алгоритма в служебный файл помещается информация об имени и расширении файла, его исходный размер, тип ПРЧ и частное (сомножитель ПРЧ).

Рассмотрим последовательность шагов алгоритма декодирования:

1. Считывание параметров из служебного файла.
2. Подготовка пустого файла, размер которого соответствует исходному файлу.
3. Генерация ПРЧ необходимой длины.
4. Умножение подготовленного ПРЧ на сомножитель.
5. Запись результата в файл.

Благодаря тому, что файл архива содержит всего лишь параметры генерации исходного файла, удается достичь высоких коэффициентов сжатия.

СРАВНЕНИЕ АЛГОРИТМОВ

Для проведения сравнения рассмотренных выше алгоритмов был разработан комплекс программ для генерации тестовых файлов, а также программ кодирования и декодирования на основе ПРЧ. Для кодирования по алгоритмам LZW и Хаффмана применялось готовое решение — ПО с открытым исходным кодом FileCompressor.

При помощи программы генерации тестовых файлов был сформирован файл размером 4 КБ, содержащий последовательность нулей и единиц (рис. 2).

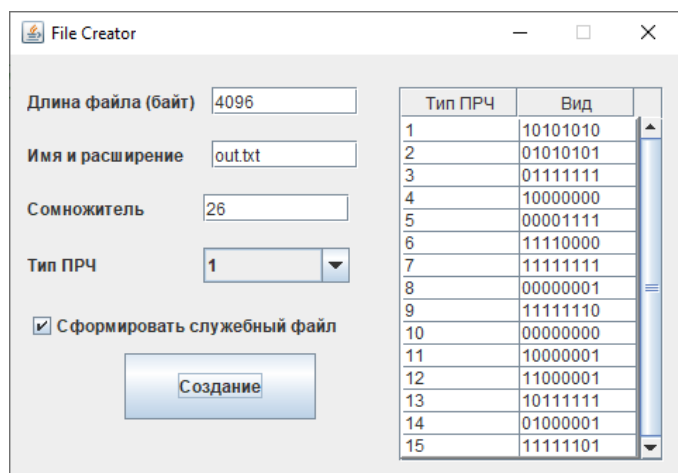


Рис. 2. Программа генерации файлов

Затем было произведено кодирование этого файла рассматриваемыми алгоритмами и сравнение размеров архивов.

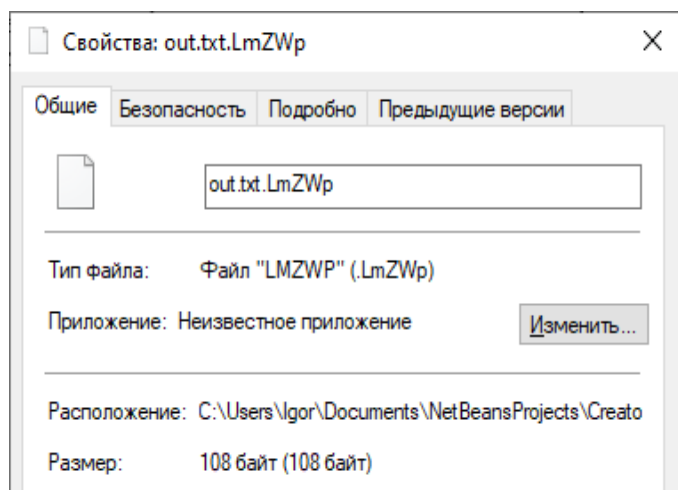


Рис. 3. Свойства архива по алгоритму LZW

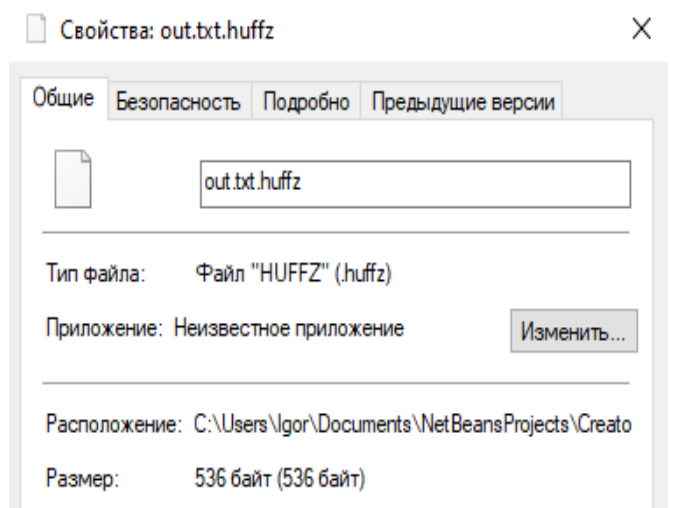


Рис. 4. Свойства архива по алгоритму Хаффмана

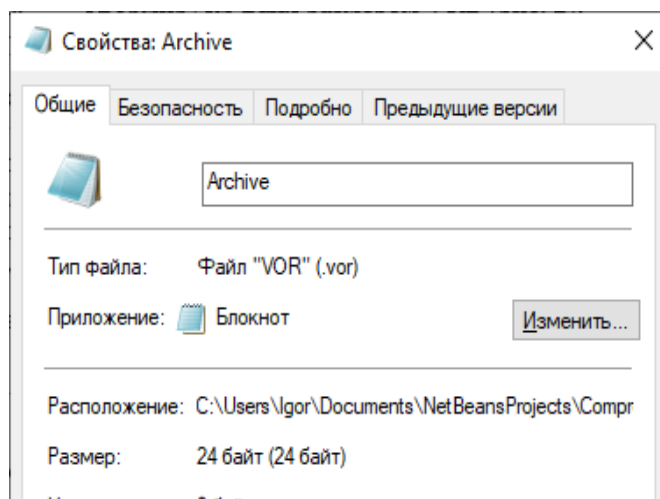


Рис. 5. Свойства архива по алгоритму на основе ПРЧ

Как видно из окна свойств (рис. 3–5) получившихся архивов, размер архива, сформированного по алгоритму Воробьева, в 4,5 раза меньше архива по алгоритму LZW и в 22 раза меньше, чем архив по алгоритму Хаффмана.

Проверим алгоритмы на файлах большего размера. Результаты приведены в таблице 2.

Таблица 2

Результаты сжатия файлов

Объем файла	На основе ПРЧ		Алгоритм LZW		Алгоритм Хаффмана	
	Размер, байт	Кэф. сжатия	Размер, байт	Кэф. сжатия	Размер, байт	Кэф. сжатия
10 КБ	25	400	165	61	1 274	7.9
50 КБ	26	1 923	402	124	6 274	8
500 КБ	26	19 231	566	883	62 524	8
1 МБ	27	37 037	3 083	324	125 024	8

Из таблицы 2 следует, что увеличение размера исходного файла приводит к незначительному увеличению размера архива при кодировании с использованием ПРЧ, в то время как другие исследуемые алгоритмы (LZW и алгоритм Хаффмана) заметно увеличивают размер выходных файлов.

ЗАКЛЮЧЕНИЕ

Разработка новых алгоритмов сжатия данных и усовершенствование старых — крайне актуальная задача на сегодняшний день. Оптимизация хранения данных в эпоху цифровизации позволит сократить расходы на IT-услуги и увеличить прибыльность бизнеса.

Алгоритм, основанный на использовании псевдорегулярных чисел, рассмотренный в статье, несомненно, позволяет добиться высокой степени сжатия файлов, однако конечный результат зависит от структуры сжимаемого файла. Помимо этого, алгоритм требует довольно больших вычислительных ресурсов. Это связано с операциями над числами большой разрядности.

Исходя из вышесказанного, необходимы дальнейшие исследования по развитию алгоритма на основе ПРЧ, что позволит увеличить его практическую применимость

и завоевать популярность как в домашних системах архивации, так и в промышленных.

ЛИТЕРАТУРА

1. Reinsel D., Gantz J., Rydning J. The Digitization of the World. From Edge to Core, *IDC White Paper No. US44413318*, November 2018, Pp. 2–3.

URL: <http://www.seagate.com/www-content/our-story/trends/files/idc-seagate-data-age-whitepaper.pdf> (дата обращения 30.07.2019).

2. Основы современных компьютерных технологий: Учебник / Б. Н. Артамонов [и др.]; под ред. А. Д. Хомоненко. — СПб.: Корона Принт, 2005. — 672 с.

3. Воробьев Е. Г., Хомоненко А. Д. Математическая модель и алгоритмы для сжатия информации без потерь // Защита информации. Инсайд. 2019. № 3 (87), С. 52–59.

4. History of Lossless Data Compression Algorithms // Engineering and Technology History Wiki (ETHW).

URL: http://ethw.org/History_of_Lossless_Data_Compression_Algorithms (дата обращения 30.07.2019).

5. Воробьев Е. Г. Модели и методы оперативного восстановления и обеспечения доступности данных автоматизированных информационных систем: автореф. дис. ... д-ра техн. наук: 05.13.18. — СПб.: ПГУПС, 2019. — 34 с.

6. Хомоненко А. Д. Методы сжатия изображений: Учебное пособие. — СПб.: ПГУПС, 2010. — 39 с.

7. Dheemanth H. N. LZW Data Compression, *American Journal of Engineering Research (AJER)*, 2014, Vol. 3, Is. 2. Pp. 22–26. URL: [http://ajer.org/papers/v3\(2\)/C0322226.pdf](http://ajer.org/papers/v3(2)/C0322226.pdf).

8. Алгоритм LZW // Викиконспекты. URL: http://neerc.ifmo.ru/wiki/index.php?title=Алгоритм_LZW (дата обращения: 30.07.2019).

9. Матрюков Д. Алгоритмы сжатия информации. Часть 1. Сжатие по Хаффмену // Монитор. 1993. № 7–8, С. 14–20.

Characteristics of a Lossless Data Compression Algorithm Based on Pseudo-Regular Numbers

M. Sc. I. A. Molodkin, PhD E. M. Shurygin
Emperor Alexander I St. Petersburg State Transport University
St. Petersburg, Russia
imolodkin@gmail.com

Abstract. An algorithm for data compression based on pseudo-regular numbers is considered. The algorithm is based on the principle of using special codes called pseudo-regular numbers, by analogy with the well-known pseudo-random sequences. HRP's act as "axes of symmetry" in the mathematical transformation of binary code. A series of experiments are carried out to compare it with the classical LZW algorithm and the Huffman algorithm. In this case, the characteristics of the degree of compression and speed are evaluated. It is shown that for pseudo-regular numbers, the use of this algorithm provides lossless compression of integer data with significantly better compression characteristics.

Keywords: data compression, encoding, archive, LZW algorithm, Huffman Encoding.

REFERENCES

1. Reinsel D., Gantz J., Rydning J. The Digitization of the World. From Edge to Core, IDC White Paper No. US44413318, November 2018, Pp. 2–3. Available at: <http://www.seagate.com/www-content/our-story/trends/files/idc-seagate-data-age-whitepaper.pdf> (accessed 30 July 2019).
2. Artamonov B. N., et al. Basics of Modern Computer Technologies: Textbook [Osnovy sovremennykh komp'yuternykh tekhnologiy: Uchebnik], edited by A. D. Khomonenko, St. Petersburg, Korona Print Publishers, 2005, 672 p.
3. Vorobiev E. G., Khomonenko A. D. Mathematical Model and Algorithms for Lossless Data Compression [Matematicheskaya model' i algoritmy dlya szhatiya informatsii bez poter'], *Zashita informatsii. Inside [Zashchita informatsii. Insayd]*, 2019, No. 3 (87), Pp. 52–59.
4. History of Lossless Data Compression Algorithms, *Engineering and Technology History Wiki (ETHW)*. Available at: http://ethw.org/History_of_Lossless_Data_Compression_Algorithms (accessed 30 July 2019).
5. Vorobiev E. G. Models and Methods of Operational Restoration and Making Accessibility of Data of Automated Information Systems [Modeli i metody operativnogo vostanovleniya i obespecheniya dostupnosti dannykh avtomatizirovannykh informatsionnykh sistem]: abstract of dissertation on competition of scientific degree Grand PhD (Engin.), St. Petersburg, PSTU, 2019, 34 p.
6. Khomonenko A. D. Data compression methods: Study guide [Metody szhatiya izobrazheniy: Uchebnoe posobie], Saint Petersburg, PSTU, 2010, 39 p.
7. Dheemanth H. N. LZW Data Compression, *American Journal of Engineering Research (AJER)*, 2014, Vol. 3, Is. 2. Pp. 22–26. Available at: [http://ajer.org/papers/v3\(2\)/C0322226.pdf](http://ajer.org/papers/v3(2)/C0322226.pdf)
8. Algorithm LZW [Algoritm LZW], *WikiNotes [Wikikon-spekty]*. Available at: http://neerc.ifmo.ru/wiki/index.php?title=Алгоритм_LZW (accessed 30 July 2019).
9. Mastrykov D. Data Compression Algorithms. Chapter 1. Huffman Compression [Algoritmy szhatiya informatsii. Chast' 1. Szhatie po Haffmenu] Monitor [Monitor], 1993, No. 7–8, Pp. 14–20.

Логическая интеграция данных на примере организации запросов к базе данных авиаперевозок

магистрант Ю. С. Фоменко,
магистрант Р. Абу Хасан,
Петербургский государственный университет путей
сообщения Императора Александра I
Санкт-Петербург, Россия
kosmonavt.98@mail.ru, ragheb1997@yandex.ru

д.т.н., проф. А. Д. Хомоненко
Петербургский государственный университет путей
сообщения Императора Александра I,
Военно-космическая академия имени А. Ф. Можайского
Санкт-Петербург, Россия
khomon@mail.ru

Аннотация. Процесс интеграции данных может быть рассмотрен как на глобальном уровне — интеграция данных из нескольких источников, так и на локальном — интеграция данных внутри одной системы для решения прикладной задачи. В статье представлены краткие теоретические основы интеграции приложений и интеграции данных. Описаны прикладная предметная область и «полигон» для интеграции. Обоснована актуальность решения задачи интеграции данных. Приведены маркетинговая и техническая постановки задачи. Приведена схема интеграции данных для поставленной задачи, выполнено ее решение.

Ключевые слова: интеграция данных, база данных, информационные системы, запрос, PostgreSQL.

ВВЕДЕНИЕ

В настоящее время большое число информационных источников нуждается в оперативной и качественной обработке, а также последующем представлении данных из этих источников или из структурных составляющих источников в унифицированном виде, либо в едином представлении для конкретной пользовательской системы или задачи. Стремительная скорость и уровень развития автоматизированных систем в современном мире заставляет задумываться о необходимости создания единой унифицированной системы, которая будет являться хранилищем интегрированных данных рабочего процесса предприятия. Источники информационных ресурсов и данных по способу представления информации принято подразделять на: базы данных, web-страницы, файловые серверы, телеконференции [1].

Неоднородность интегрируемых данных в разных источниках заключается в:

- представлении данных в разных форматах;
- существовании своей понятийной системы (онтологии) предметной области источника;
- различных моделях данных;
- различных схемах данных.

В зависимости от причин возникновения неоднородности выделяют три основные модели интеграции данных: физическая интеграция, логическая интеграция и семантическая интеграция [2, 3]. Выбор модели интеграции для разработки архитектуры интегрирующей системы зависит от совокупности интегрируемых данных, уровня доступа

к локальным системам, технического задания и функциональности прикладных программ [4].

КОНЦЕПТУАЛЬНЫЕ ОСОБЕННОСТИ ВИДОВ ИНТЕГРАЦИИ

Enterprise Application Integration (EAI) — интеграция корпоративных приложений подразумевает две основные концепции интеграции: интеграция приложений и интеграция данных.

Концепция *интеграции приложений* определена следующими факторами:

- 1) увеличение производительности разрабатываемой комплексной прикладной программы;
- 2) следствие использования разных платформ и поддерживающих их программных средств;
- 3) неполноценный функционал отдельных программных средств;
- 4) автоматизация бизнес-процесса.

Интеграция данных неразрывно связана с интеграцией приложений. Она является ведущим звеном в обеспечение корректной и слаженной работы предприятия. Интегрированные данные могут быть использованы как средство для выполнения определенной задачи, а также для проведения анализа, который может быть осуществлен после выполнения конкретной задачи.

В рамках логического уровня интеграции данные, хранящиеся в своих сущностях, доступ к которым осуществляется на основе некоторой глобальной схемы, в процессе осуществления пользовательской задачи нуждаются в объединении (данные из разных сущностей) для анализа и выборки результата, удовлетворяющему условию задачи.

Интеграция данных в рамках конкретной предметной области может быть полезна для выполнения актуальных задач по улучшению работы предприятия. В частности, рассматривая дальнейшую практическую задачу, интеграция данных поможет управляющему составу получить некоторую статистику, позволяющую скорректировать работу организации для поддержания высокого уровня компетентности или для улучшения проводимых мероприятий по повышению уровня компетентности.

ИНТЕГРИРУЕМАЯ ОБЛАСТЬ И «ПОЛИГОН» ДЛЯ ИНТЕГРАЦИИ

Постановка задачи выполнялась для демонстрационной базы данных под управлением СУБД PostgreSQL [5], предметной областью которой являются авиаперевозки по

России. Выбор этой предметной области для анализа и осуществления процесса интеграции обоснован следующими отличительными характеристиками:

- простая понятийная система предметной области;
- широкий спектр данных для анализа и интеграции;
- максимальное сходство данных с реально существующими.

Последняя отличительная характеристика обеспечивает следующее:

- возможность постановки максимально реалистичной задачи;
- разработку алгоритмов, применимых к базам данных, наполняемых в реальном масштабе времени;
- получение максимально реалистичных результатов в ответ на поставленную задачу.

«Полигоном» для демонстрации процесса интеграции является СУБД PostgreSQL. В совокупности с вышеуказанными преимуществами демонстрационной базы данных, СУБД обладает также рядом достоинств, прежде всего практических для пользователя системы [6–11]:

- объектно-реляционная структура;
- возможность работы со сложными структурами данных;
- возможность создания собственного типа данных;
- высокий уровень сохранения целостности данных за счет соответствия стандарту ANSI-SQL:2008.

ПОСТАНОВКА ЗАДАЧИ

Процесс интеграции данных сопровождается прежде всего анализом интегрируемых данных. Обобщенной целью анализа является в первую очередь – оценка текущего состояния некоторой области исследования. В соответствии с этим определением поставленной задаче было дано две формулировки: маркетинговая и техническая.

Маркетинговая постановка задачи. Авиаперевозки России – предметная область, деятельность структурных филиалов которой поддерживает высокий уровень конкуренции. Помимо выполнения своих основных обязанно-

стей, компании должны не забывать о маркетинговой составляющей своего бизнеса.

Маркетинговая постановка представляет задачу в том виде, в каком ее бы преподнес отдел маркетинга и продаж IT-подразделению предприятия. Определение поставленной задачи без ее технической адаптации звучит следующим образом: *«Мы хотим улучшить бизнес-залы ожидания и повысить компетенции персонала. Необходимо понять, в каких подразделениях следует начать процесс улучшения. Между какими тремя парами городов совершается больше всего бизнес-командировок?».*

Техническая постановка задачи. Процесс интеграции, как и любой автоматизированный процесс, нуждается в техническом описании, необходимом для корректности разработки алгоритма его реализации. Выполнение поставленной задачи подразумевает процесс логической интеграции данных. Поэтому перед ее выполнением определена строгая техническая постановка задачи: *«На какие рейсы покупается больше всего билетов «туда и обратно» со сроком «обратно» не более 5 дней».*

РЕШЕНИЕ ЗАДАЧИ

Вследствие того, что основная часть работы для получения ответа на поставленную задачу будет проводиться с базой данных авиаперевозок по России, то можно поэтапно описать алгоритм выполнения задачи:

1. Ознакомиться с демонстрационной базой данных (подробно изучить диаграмму схемы данных, структуру сущностей, список отношений, представления, функции и т.д.).
2. При отсутствии опыта работы с СУБД PostgreSQL изучить основные принципы работы системы.
3. Развернуть СУБД PostgreSQL.
4. Развернуть демонстрационную базу данных.
5. Разработать структуру запроса в соответствии с поставленной задачей.
6. Произвести оптимизацию запроса (по необходимости).

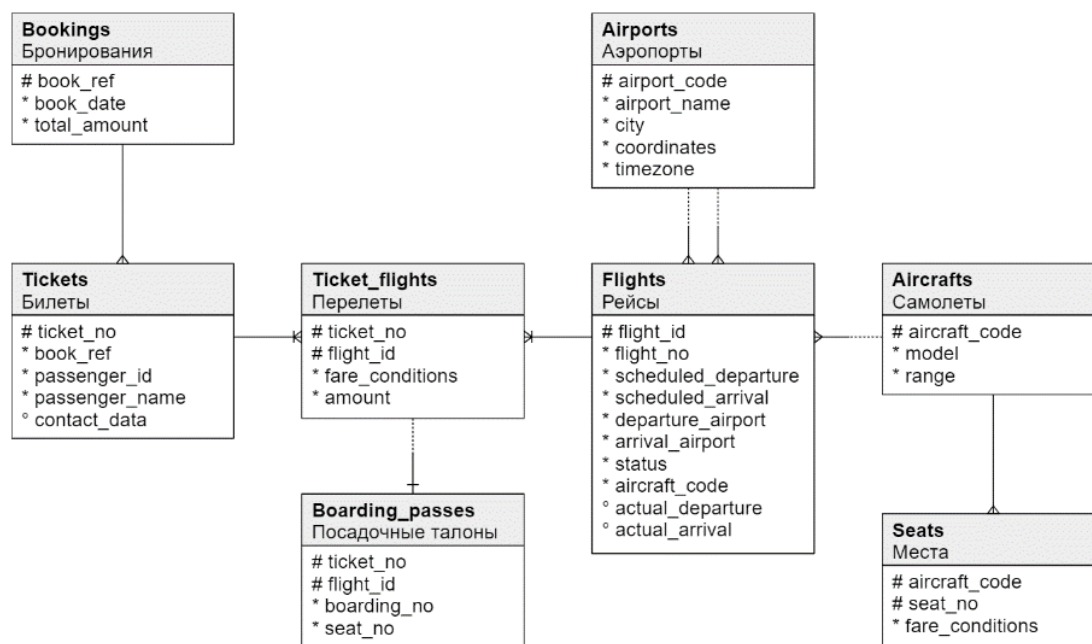


Рис. 1. Схема демонстрационной базы данных «Авиаперевозки по России»

ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА

I. Диаграмма схемы данных

Кроме сущностей, представленных на рис. 1, база данных имеет четыре представления:

- airports (характеристика аэропортов);
- flights_v (дополнительная информация о полетах);
- routes (содержит метаданные таблицы Flights);
- aircrafts (характеристики самолетов).

II. Развертывание базы данных

Демонстрационная база данных распространяется в виде файла, содержащего SQL-скрипт, который создает базу данных *demo* и наполняет ее данными по полетам. Следует обратить внимание на то, что при установке су-

ществующая база данных *demo* будет удалена и создана заново. Если имеется необходимость развернуть базу данных по полетам под другим именем, то это можно сделать, открыв в текстовом редакторе SQL-скрипт и подкорректировать сценарий скрипта, как показано на рисунке 2, где произведена замена *demo* на *AirTransportation*.

Для развертывания базы данных необходимо воспользоваться терминальным клиентом для PostgreSQL — psql [8]. Стандартное использование графической среды pgAdmin для выполнения SQL-скрипта будет бесполезным. Это вызвано большим набором хранимых данных.

```

16
17 DROP DATABASE AirTransportation;
18
19 -- Name: AirTransportation; Type: DATABASE; Schema: -; Owner: -
20
21
22 CREATE DATABASE AirTransportation;
23
24
25 \connect AirTransportation
26
27 SET statement_timeout = 0;
28 SET lock_timeout = 0;
29 SET idle_in_transaction_session_timeout = 0;
30 SET client_encoding = 'UTF8';
31 SET standard_conforming_strings = on;
32 SET check_function_bodies = false;
33 SET client_min_messages = warning;
34 SET row_security = off;
35
36 --
37 -- Name: bookings; Type: SCHEMA; Schema: -; Owner: -
38
39
40 CREATE SCHEMA bookings;
41
42
43 --
44 -- Name: SCHEMA bookings; Type: COMMENT; Schema: -; Owner: -
45
46
47 COMMENT ON SCHEMA bookings IS 'Airlines AirTransportation database schema';
2290678 ALTER DATABASE AirTransportation SET search_path = bookings, public;
2290679 ALTER DATABASE AirTransportation SET bookings.lang = ru;

```

Рис. 2. Замена названия базы данных

Процесс развертывания базы данных осуществляется в несколько действий:

1. Запуск командной строки и переход в папку, содержащую интерактивный терминал (рис. 3).

2. Настройка параметров psql и запуск процесса развертывания базы данных (рис. 4).

3. Просмотр базы данных через графический интерфейс pgAdmin4 (рис. 5).

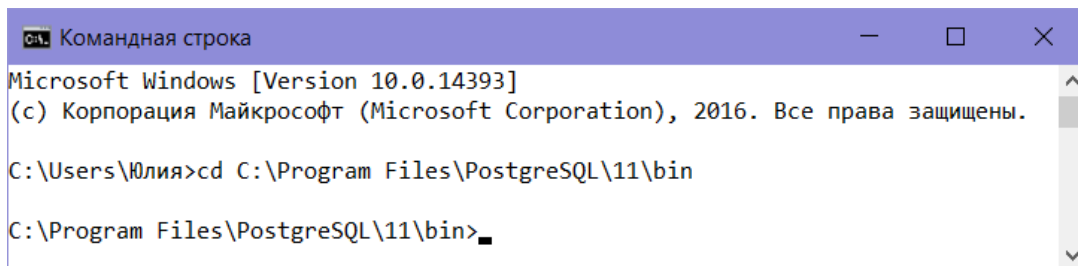


Рис. 3. Переход к терминалу psql

```

Командная строка
Microsoft Windows [Version 10.0.14393]
(c) Корпорация Майкрософт (Microsoft Corporation), 2016. Все права защищены.

C:\Users\Юлия>cd C:\Program Files\PostgreSQL\11\bin

C:\Program Files\PostgreSQL\11\bin>psql.exe -h 127.0.0.1 -U postgres -f C:\demo-small.sql
SET
SET
SET
SET
SET
SET
SET
SET
SET
DROP DATABASE
CREATE DATABASE
Вы подключены к базе данных "demo" как пользователь "postgres".
    
```

Рис. 4. Процесс развертывания базы данных

	flight_id [PK] integer	flight_no character (6)	scheduled_departure timestamp with time zone	scheduled_arrival timestamp with time zone	departure_airport character (3)	arrival_airport character (3)	status character varying (20)	aircraft_code character (3)
1	1185	PG0134	2017-09-10 09:50:00+03	2017-09-10 14:55:00+03	DME	BTK	Scheduled	319
2	3979	PG0052	2017-08-25 14:50:00+03	2017-08-25 17:35:00+03	VKO	HMA	Scheduled	CR2
3	4739	PG0561	2017-09-05 12:30:00+03	2017-09-05 14:15:00+03	VKO	AER	Scheduled	763
4	5502	PG0529	2017-09-12 09:50:00+03	2017-09-12 11:20:00+03	SVO	UFA	Scheduled	763
5	6938	PG0461	2017-09-04 12:25:00+03	2017-09-04 13:20:00+03	SVO	ULV	Scheduled	SU9
6	7784	PG0667	2017-09-10 15:00:00+03	2017-09-10 17:30:00+03	SVO	KRO	Scheduled	CR2
7	9478	PG0360	2017-08-28 09:00:00+03	2017-08-28 11:35:00+03	LED	REN	Scheduled	CR2
8	11085	PG0569	2017-08-24 15:05:00+03	2017-08-24 16:10:00+03	SVX	SCW	Scheduled	733
9	11847	PG0498	2017-09-12 10:15:00+03	2017-09-12 14:55:00+03	KZN	IKT	Scheduled	319

Рис. 5. Развернутая база данных

III. Разработка алгоритма выполнения задачи

1 этап. *Определение вспомогательных данных.* Для получения результата, удовлетворяющего запрос, работа будет проводиться со следующими данными:

- Аэропорт отправления (departure_airport);
- Аэропорт прибытия (arrival_airport);
- Актуальное время отправления (actual_departure);
- Уникальный идентификатор пассажира (passenger_id; выбран именно идентификатор, а не ФИО на тот случай, если у пассажиров одинаковое ФИО);
- Номер бронирования (book_ref).

2 этап. *Интеграция данных в результирующую таблицу.* Ответ на запрос будет представлен в дополнительно сформированной таблице. Соответственно, перед добавлением запрошенных данных необходимо удалить эту таблицу на тот случай, если данный запрос уже был реализован относительно конкретной базы данных. После чего необходимо объявить создание этой таблицы. Определенная для запроса таблица представлена на рисунке 6.

Тип полей таблицы *arr_depart* и *arr_arival* определен как текстовый массив. В каждой ячейке этих полей будет интегрироваться по два значения — аэропорта отправления и аэропорта прибытия соответственно. Поле *dat* определено как массив дат. В ячейках этого поля будут инте-

грироваться даты отправления из двух городов. Тип поля *lim* определен как числовой. В ячейках этого поля будет рассчитываться разница между датой отправления из одного аэропорта и датой отправления из другого аэропорта.

out_table
arr_depart text[]
arr_arival text[]
dat date[]
lim integer

Рис. 6. Выходная таблица запроса

Интеграция данных из нескольких таблиц реализуется по схеме, приведенной на рисунке 7.

Рассмотрим две ситуации: бронирование билетов «туда-обратно» и бронирование билетов до пункта назначения через пересадочный пункт.

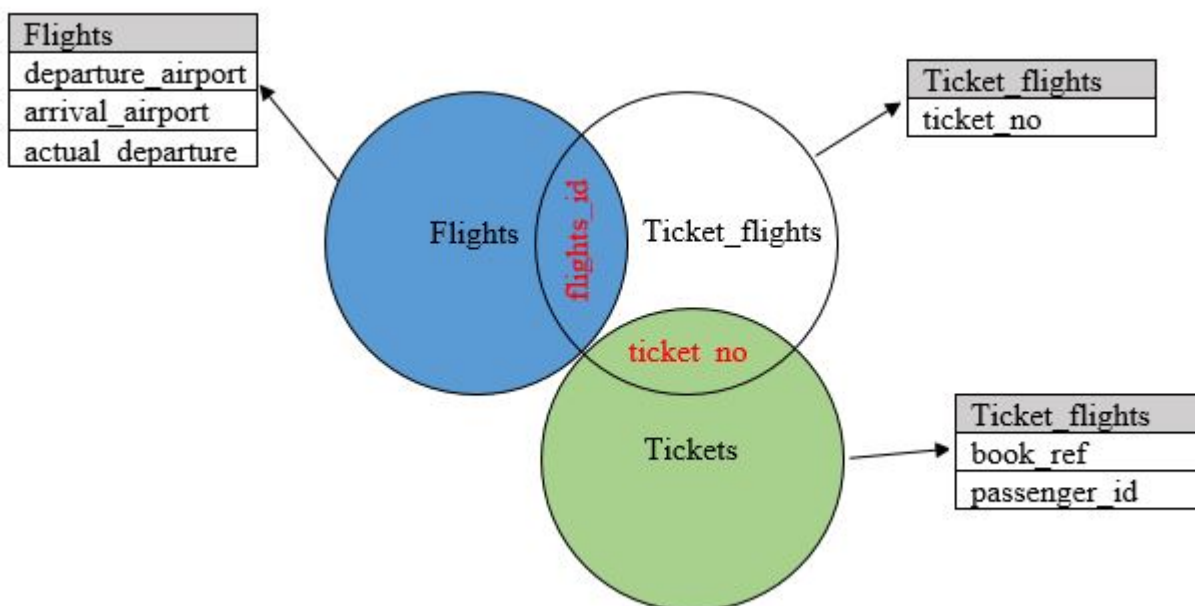


Рис. 7. Схема объединения данных из нескольких таблиц

Ситуация 1. Пассажиры, совершающие бизнес-поездки, одним разом бронируют два билета. Номер брони для билета с маршрутом из пункта А в пункт В и билета с маршрутом из пункта В в пункт А будет одинаковый. Результат выборки будет группироваться (GROUP BY) относительно уникального идентификатора пассажира с условием (HAVING COUNT), что в оформленных им билетах количество одинакового номера бронирования равно 2. Соответственно, каждому такому пассажиру в базе данных будут присущи два аэропорта отправления, два аэропорта прибытия и две даты отправления (табл. 1).

Таблица 1

Представление данных при бронировании билетов «туда-обратно»

	Аэропорт отправления	Аэропорт прибытия	Дата отправления
Пассажир	Пункт А	Пункт В	Дата отправления из аэропорта А
	Пункт В	Пункт А	Дата отправления из аэропорта В

Ситуация 2. Рассмотрим вторую ситуацию, когда пассажир за один раз бронирует два билета на рейсы: из пункта А в пункт В и из пункта В в пункт С. В такой ситуации пункт В будет являться пересадочным пунктом по пути в пункт С. В таком случае результат выборки (в соответствии с таким же алгоритмом, описанным для первой ситуации) будет иметь общий вид данных, представленных в таблице 2.

Все данные при первой и второй ситуациях заносятся в заранее определенные поля результирующей таблицы как элементы массива.

Таблица 2

Представление данных при бронировании транзитных билетов

	Аэропорт отправления	Аэропорт прибытия	Дата отправления
Пассажир	Пункт А	Пункт В	Дата отправления из аэропорта А
	Пункт В	Пункт С	Дата отправления из аэропорта В

3 этап. Выборка требуемых по условию данных. После заполнения выходной таблицы сгруппированными и отобранными согласно условия данными стоит задача подсчитать разницу между датами отправления из двух пунктов (данные, хранящиеся в массивах поля *dat*) для каждого пассажира. Эта разница заносится в поле *lim*. После чего выходная таблица становится полностью заполненной и содержит все необходимые данные для получения ответа на поставленную задачу.

Таблицу 3 можно назвать фрагментом истинного результата, удовлетворяющего запросу. Именно такая совокупность фрагментов будет подвергаться подсчету с последующей сортировкой и выделением тройки пар аэропортов. Выбор данного фрагмента основывается на том, что в соответствии с указанными аэропортами отправления, аэропорты прибытия должны иметь обратный порядок (т. е. ситуация 2 не подходит) при условии разницы между датами отправления не более 5 дней.

В скрипте это реализовано проверкой элементов массива *arr_depart* и *arr_arival* и проверкой полученного результата в ячейках поля *lim*.

Таблица 3

Данные, удовлетворяющие условию задачи

Аэропорты отправления	Аэропорты прибытия	Разница между датами отправления
[Пункт А, Пункт В]	[Пункт В, Пункт А]	<=5

После получения фрагментов, подобных представленным в таблице 3, необходимо сгруппировать данные, например, по аэропорту отправления и определить огра-

ничения вывода. По условию задачи необходимо определить тройку пар городов.

IV. Код запроса

Код запроса для решаемой задачи имеет вид, представленный на рисунке 8.

Последовательность действий, выполняемых при запуске кода запроса, представлена на рисунке 9.

```
drop table out_table

create table out_table (
    arr_depart text[],
    arr_arrival text[],
    dat date[],
    lim integer
);

insert into out_table (arr_depart, arr_arrival, dat)
select
array_agg(departure_airport) as arr_depart,
array_agg(arrival_airport) as arr_arrival,
array_agg(cast(actual_departure as date)) as dat
FROM flights
right join ticket_flights on ticket_flights.flight_id = flights.flight_id
right join tickets on tickets.ticket_no = ticket_flights.ticket_no
group by tickets.passenger_id having count(tickets.book_ref) = 2
order by dat

update out_table
set lim = abs(out_table.dat[2] - out_table.dat[1])

select count (out_table.arr_depart) as num, out_table.arr_depart
from out_table
where lim <= '5' and ((arr_depart[1]=arr_arrival[2] and arr_depart[2]=arr_arrival[1]))
group by out_table.arr_depart
order by num desc
limit 3
```

Рис. 8. Код запроса для решаемой задачи

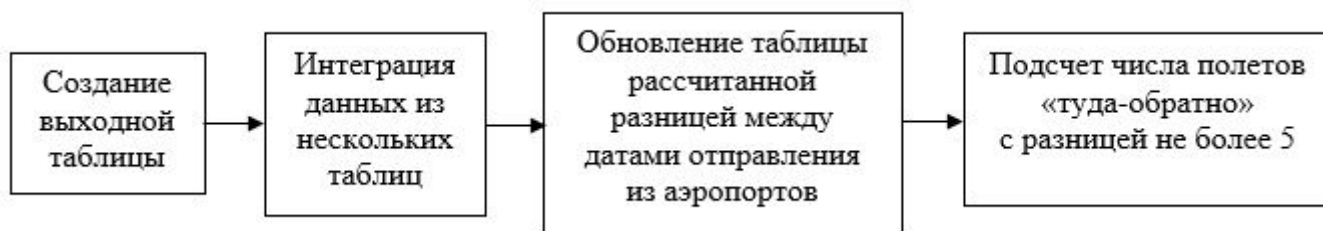


Рис. 9. Последовательность действий, выполняемых при запуске кода запроса

V. Результат выполнения

Городами, между которыми чаще совершаются поездки «туда-обратно» со сроком обратно не более 5 суток являются:

1. Москва (код аэропорта: SVO; аэропорт Шереметьево) – Екатеринбург (код аэропорта: SVX; аэропорт Кольцово);
2. Москва (код аэропорта: SVO; аэропорт Шереметьево) – Новосибирск (код аэропорта: OVB; аэропорт Толмачево);
3. Новосибирск (код аэропорта: OVB, аэропорт Толмачево) – Москва (код аэропорта: DME; аэропорт Домодедово).

Выходная таблица содержит коды аэропортов, которые нужно расшифровать и предоставить информацию не только о названии аэропорта, но и о городе, в котором аэропорт расположен. Для этого необходимо сделать запрос к сущности `airports_data`, которая хранит данные об аэропортах и городах их расположения. Код запроса представлен на рисунке 10.

Результат выполнения кода запроса показан на рисунке 11, расшифровка кодов аэропортов приведена на рисунке 12.

Как видим, решение приведенной задачи поможет сотрудникам авиакомпании правильно расставить приоритеты между аэропортами городов, между которыми выполняются перевозки.

```
select airports_data.city, airports_data.airport_name, airports_data.airport_code
from airports_data
where airports_data.airport_code = 'SVO' or airports_data.airport_code = 'SVX'
or airports_data.airport_code = 'OVB' or airports_data.airport_code = 'DME'
```

Рис. 10. Код запроса данных об аэропортах

	num bigint	arr_depart text[]
1	298	{SVO,SVX}
2	282	{SVO,OVB}
3	271	{OVB,DME}

Рис. 11. Выходная таблица

	city jsonb	airport_name jsonb	airport_code [PK] character (3)
1	{'en': 'Yekaterinburg', 'ru': 'Екатеринбург'}	{'en': 'Koltsovo Airport', 'ru': 'Кольцово'}	SVX
2	{'en': 'Moscow', 'ru': 'Москва'}	{'en': 'Sheremetyevo International Airport', 'ru': 'Шереметьево'}	SVO
3	{'en': 'Moscow', 'ru': 'Москва'}	{'en': 'Domodedovo International Airport', 'ru': 'Домодедово'}	DME
4	{'en': 'Novosibirsk', 'ru': 'Новосибирск'}	{'en': 'Tolmachevo Airport', 'ru': 'Толмачёво'}	OVB

Рис. 12. Расшифровка кодов аэропортов

ЗАКЛЮЧЕНИЕ

В статье поставлена задача, корректное решение которой с использованием интеграции может быть полезным для структурных подразделений. В частности, решение конкретной прикладной задачи поможет маркетологам-аналитикам авиакомпании правильно расставить приоритеты между аэропортами городов, которые в первую очередь нуждаются в поддержании или повышении высокого уровня компетентности организации. Полученный результат отвечает на поставленную задачу, доказывает корректность и правильность предложенного способа реализации.

ЛИТЕРАТУРА

1. Шуремов Е. Л. Информационные ресурсы: классификация, источники, поставщики. Коротко о главном. – М.: Издательские решения, 2017. — 160 с.

2. Когаловский М. Р. Методы интеграции данных в информационных системах [Электронный ресурс]. — М.: б/и, 2010. — 9 с.

URL: <http://www.ipr-ras.ru/articles/kogalov10-05.pdf> (дата обращения 10.11.2019).

3. Черняк Л. С. Интеграция данных: синтаксис и семантика // Открытые системы. СУБД. 2009. № 10. С. 24–29.

4. What is Data Integration? // Big Data Articles / Datamation. URL: <http://www.datamation.com/big-data/what-is-data-integration.html> (дата обращения 10.11.2019).

5. Демонстрационная база данных // Компания PostgreSQL Professional. URL: <http://postgrespro.ru/education/demodb> (дата обращения 05.11.2019).

6. What PostgreSQL has over other open source SQL databases: Part I // Compose Articles. URL: <http://www.compose.com/articles/what-postgresql-has>

over-other-open-source-sql-databases (дата обращения 01.11.2019).

7. Бьюли А. Изучаем SQL = Learning SQL / Пер. с англ. Н. Шатохиной. — СПб.: Символ-Плюс, 2007. — 312 с.

8. Моргунов Е. П. PostgreSQL. Основы языка SQL: Учебное пособие / Е. П. Моргунов; Компания Postgres Professional; под ред. Е. В. Рогова, П. В. Лузанова. — СПб.: БХВ-Петербург, 2018. — 336 с.

9. Базы данных: Учебник для высших учебных заведений / А. Д. Хомоненко, В. М. Цыганков, М. Г. Мальцев;

под ред. А.Д. Хомоненко. — 6-е изд., доп. — СПб.: КОРОНА-Век, 2009. — 736 с.

10. Основы технологий баз данных: Учебное пособие / Б. А. Новиков, Е. А. Горшкова; под ред. Е. В. Рогова. — М.: ДМК Пресс, 2019. — 240 с.

11. Уорсли Дж. PostgreSQL для профессионалов / Дж. Уорсли, Дж. Дрейк; пер. с англ. Е. Матвеева. — СПб.: Питер, 2003. — 496 с. — (Для профессионалов).

Logical Data Integration on the Example of Organizing Queries to the Database Air Transportation

Master's Degree student Yu. S. Fomenko,
Master's Degree student R. Abou Hasan
Emperor Alexander I Petersburg State Transport University
St. Petersburg, Russia
kosmonavt.98@mail.ru, ragheb1997@yandex.ru

Grand PhD, Prof. A. D. Khomonenko
Emperor Alexander I Petersburg State Transport University
A. F. Mozhaisky Military Space Academy
St. Petersburg, Russia
khomon@mail.ru

Abstract. The process of data integration can be considered both at the global level – data integration from some sources and at the local level – data integration inside the one system handling of applied problems. The article presents a brief theoretical science for application integration and data integration. The applied subject area and the "testing ground" for integration are described. The relevance of solving the problem of data integration is justified. The marketing and technical statements of the problem are given. The scheme of data integration for the problem is presented, its solution is completed.

Keywords: data integration, database, information systems, query, PostgreSQL.

REFERENCES

1. Shuremov Ye. L. Information resources: classification, sources, providers. Briefly about the main [Informatsionnye resursy: klassifikatsiya, istochniki, postavshchiki. Korotko o glavnom], Moscow, Publishing Solution Publishing House, 2017, 160 p.
2. Kogalovsky M. R. Data integration methods in information systems [Metody integratsii dannykh v informatsionnykh sistemakh], Moscow, 2010, 9 p. Available at: <http://www.ipr-ras.ru/articles/kogalov10-05.pdf> (accessed 10 Nov 2019).
3. Chernyak L. S. Data integration: syntax and semantics [Integratsiya dannykh: sintaksis i semantika], *Open systems. DBMS [Otkrytye sistemy. SUBD]*, 2009, No. 10, Pp. 24–29.
4. What is Data Integration? *Datamation*. Available at: <http://www.datamation.com/big-data/what-is-data-integration.html> (accessed 16 Nov 2019).
5. Demonstration database [Demonstratsionnaya baza dannykh], *Postgres Professional [Kompaniya Postgres Professional]*. Available at: <http://postgrespro.ru/education/demodb> (accessed 1 Nov 2019).
6. What PostgreSQL has over other open source SQL databases: Part I, *Compose Articles*. Available at: <http://www.compose.com/articles/what-postgresql-has-over-other-open-source-sql-databases> (accessed 10 Nov 2019).
7. Beaulieu A. Learning SQL [Izuchaem SQL], St. Petersburg, Symbol Plus Publishers, 2007, 312 p.
8. Morgunov E. P. PostgreSQL. Base SQL language [PostgreSQL. Osnovy yazyka SQL: Uchenoe posobie], St. Petersburg, BHV-Petersburg Publishing House, 2018, 336 p.
9. Khomonenko A. D., Tsygankov V. M., Maltzev M. G. Database: Textbook for universities [Bazy dannykh: Uchebnik dlya vysshykh uchebnykh zavedeniy], St. Petersburg, Korona-Vek Publishing House, 2009, 736 p.
10. Novikov B. A., Gorshkova E. A., Database technology basics: Study guide [Osnovy tekhnologiy baz dannykh: Uchebnoe posobie], Moscow, DMK Press Publishers, 2019, 240 p.
11. Worsley J. C., Drake D. J. Practical PostgreSQL [PostgreSQL dlya professionalov], St. Petersburg, Piter Publishing House, 2003, 496 p.

Directions of Informatization of Education at the University

PhD, Assoc. Prof. A. I. Dergachev, PhD S. A. Dergachev,
PhD O. N. Kuranova, PhD, Assoc. Prof. A. M. Perepechenov
Emperor Alexander I Petersburg State Transport University
St. Petersburg, Russia

d_ader@mail.ru, deburg@mail.ru, olga_kuranova@mail.ru, ddamper@yandex.ru

Abstract. The most important aspects of mastering a teacher by teaching methods using modern technical teaching aids and based on them new information technologies are considered. In particular, the application features of computer-aided training aids are described. The study strategies of the academic discipline are described, which includes various forms of training and a reasonable ratio of training with the use of a computer and without it.

Keywords: computer technologies of training, methods of training, creation and applications of the automated educational and methodical complex.

The intensification of training in the world practice is considered as a process of its Informatization and industrialization on the basis of the introduction of new information technologies, integrated introduction of computer technology and telecommunications in educational activities. Therefore, the creation and introduction of new information technologies in the educational process of our University is now considered one of the most important tasks of its development [1, 2].

One of the most important aspects of pedagogical technology is the methodological aspect — the teacher's mastery of teaching methods with the use of modern technical means of education (TME) and based on them new information technologies [3, 4].

Teaching methods are chosen by the teacher. The teaching method is based on figure 1 on the basis of philosophy, specialty and psychological processes. It depends on age physiology, pedagogical psychology, professional training and the use of technical means of teaching TME.

TME are distinguished by the nature of their direct impact, by methods of absorption by the organs of vision and hearing, or both at the same time. To implement the functions of information presentation, various recording, storage and playback devices can be introduced into the TME complex. Simulators, simulators of real objects, laboratory equipment are used to increase the activity of students in various forms of training sessions. In addition, to improve the performance of students in the complex TME may include means of implementation of functional music, hypnopedia and relaxation.

TME can be used for different purposes — both autonomously and in complexes, the composition of which is determined by specific tasks.

According to one of the most common classifications, all existing TSO are divided into three categories: information, training and control. Category educational TME is a compound

of information and regulatory categories of TME [1, 5, 6].

The term «informatization of training» will be understood as the process of production, distribution and use of information and information services in training based on the introduction of computer methods and means of collecting, processing, searching, issuing and storing information. material; sending training materials on computer communications; video broadcasts of training programs, etc. [4, 5, 7].

Continuous computerization and application of computer technology in all areas was initially taken as an axiom. Is this the case in education? Note the following main features of the use of computer tools in training.

– Computerization of the educational process covers only part of the cognitive activity.

– When particular importance is attached to learning through practice and experience that cannot be expressed in a formal description, attempts transfer them into the language of machines is ineffective.

– When teaching practical knowledge, the teacher conveys the culture of any professional activity, i.e. provides a link between generations. This is where the superiority of human knowledge over machine knowledge is manifested.

For such a traditionally basic form of education as a lecture, the computer cannot replace live communication with the teacher.

In the world practice, modern information technologies should provide:

– shipping to the trainees of the basic amount of educational material;

– interactive interaction of students and teachers

– in the learning process;

– providing the trainees opportunities for independent work;

– democratic assessment of knowledge and skills acquired.

To achieve these goals, the following computer information technologies have been used: work in the environment of automated training systems (ATS); presentation support for the supply of educational material; work in the environment of virtual laboratories; testing and automated control of the level of assimilation of educational. The computer may not be able to focus on the main idea of the learning process as effectively as the teacher, timbre coloring and voice modulation will not be able to make the learner get carried away with the idea as with live learning.

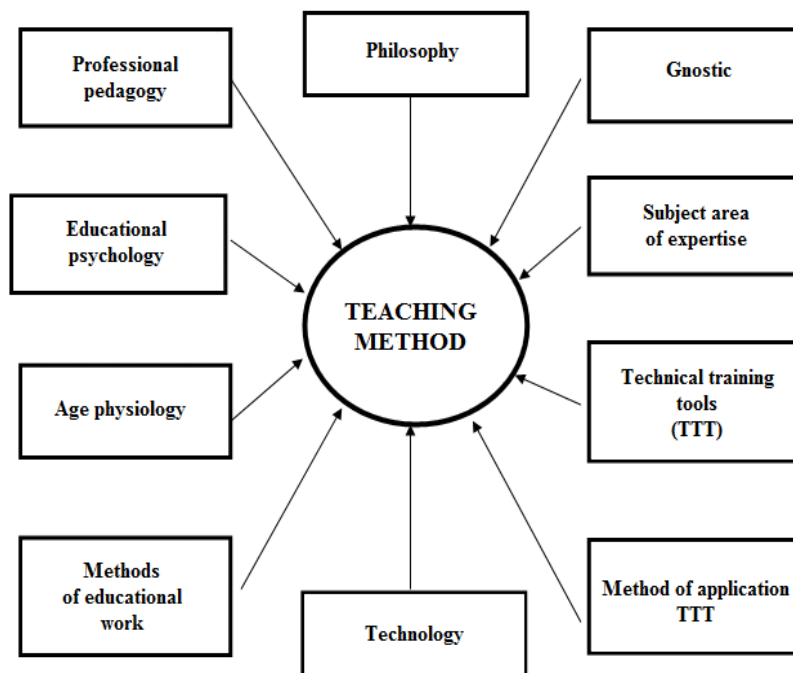


Fig. 1. Sources of formation of teaching methods

However, as illustrating the control and TSO computer more effective than slides, charts, etc. static TSO. What is the secret of this efficiency?

Based on the analysis of the above features of the use of computer TCO, the strategy of studying the discipline, which includes various forms of training, should provide a reasonable ratio of training using a computer and without it [4, 7]. This gives a clear idea of the true place of computers in the learning process:

– Computers by their very nature are unable to cancel any creative pedagogical work, the responsibilities of the teacher to educate the students.

– Computers do not eliminate the need for persistent learning and deep understanding of educational material by students.

– Computer and software and information products act as a means of improving the efficiency and quality of training by automating the management of joint activities of the teacher and trainees [1, 5, 7, 8].

– The inclusion of software and information products and methods of their application in the standard educational-methodical complex of the discipline allows us to talk about the creation of an automated educational-methodical complex (AEMC) of this discipline, the typical composition of which is shown in figure 2.

The main principle proposed as a basis for the development and application of AEMC in the educational process: not a replacement car of the teacher that combines in its activities the functions of training and education, and a reasonable synthesis of the possibilities of man and computer in the interests of solving the problems of education [1].

The teacher acts not only as an expert in the subject area, but also organizes and mainly monitors and makes decisions on the implementation of a set of activities to achieve General and private educational and educational goals of the training session:

- General management of the learner's cognitive activity;
- presentation of the topic, training questions and literature used;
- transfer of new knowledge and skills in the subject area;
- acquisition of new knowledge about themselves;
- motivation, monitoring and evaluation, gap filling;
- repetition, generalization, consolidation;
- information on the course, conduct and results of the lesson;
- registration of achievements and development of educational material.

As can be seen, the ATS, in addition to the transfer of knowledge, also takes over the performance of part of the management and technical non-creative functions of the teacher, which is especially important in the independent study of educational material.

The use of AUM should provide each trainee with:

- ability to work at your own pace in accordance with their psychological and physiological characteristics and knowledge;
- democratic control and the objectivity of the evaluation of his knowledge;
- reduction in wastage of teaching time to support work to find educational information and getting the references;
- developing a sustainable habit of self-study;
- gaining experience in working with computers and knowledge of its application in solving practical problems.

The use of AUM in the classroom on the curriculum should allow the teacher:

- to achieve increase of productivity of the work at the expense of unloading from some types of technical non-creative work connected with issue of tasks, check of their performance, registration of results;

- receive statistical information on the course of the educational process, the work of individual students and training groups for analysis and informed pedagogical decisions;
- to provide a productive independent work of trainees;
- throughout the allocated teaching time at the expense of individualization of a dialogue with a computer program;

- to reduce the number of control and inspection of works carried out not in the specialized audience for the purpose of obtaining information about the results of the students work on the training material.

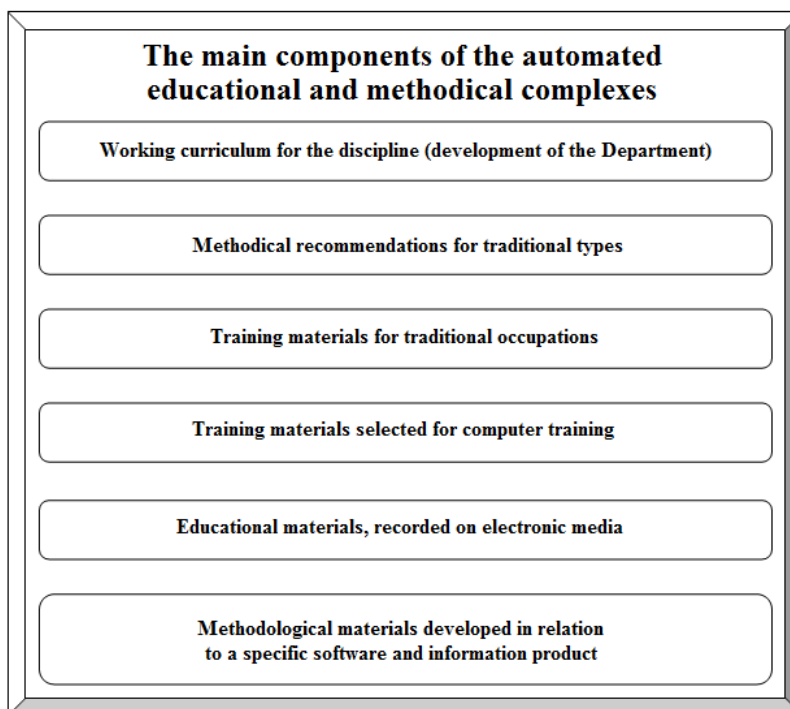


Fig. 2. The structure of typical automated educational systems

The development of new (high) information technologies leads to a rapid increase in the amount of knowledge that a modern specialist must possess in order to carry out active professional activities.

REFERENCES

1. Bryn M. Ya., Dergachev A. I., Perepechenov A. M. Multimedia Technologies in Teaching Students [Multimediyne tekhnologii pri obuchenii studentov], *Proceedings of the IV International Scientific and Methodological Conference «Problems of Mathematical and Natural Science Training in Engineering Education» [Sbornik trudov IV Mezhdunarodnoy nauchno-metodicheskoy konferentsii «Problemy matematicheskoy i estestvennonauchnoy podgotovki v inzhenernom obrazovanii»]*, November 03, 2016, St. Petersburg, St. Petersburg, PSTU, 2016, P. 30–34.

2. Kuranova O. N., Dergachev S. A., Lebedeva N. A. The Influence of the Level of Training of Personnel of Railway and Air Transport on the Efficiency of Transportation [Vliyanie urovnya podgotovki personala zheleznodorozhnogo i vozdushnogo transporta na effektivnost' vypolneniya perevozok], *Proceedings of the IV International Scientific and Methodological Conference «Problems of Mathematical and Natural Science Training in Engineering Education» [Sbornik trudov IV Mezhdunarodnoy nauchno-metodicheskoy konferentsii «Problemy matematicheskoy i estestvennonauchnoy podgotovki v inzhenernom obrazovanii»]*, November 03, 2016, St. Petersburg, Volume 2, St. Petersburg, PSTU, 2017, P. 92–95.

3. Kuranova O. N., Dergachev A. I., Vatulin Ya. S., Lebedeva N. A. Justification of Efficiency of Remote Additional Training of the Personnel of Railway Transport [Obosnovanie effektivnosti distantsionnoy doppodgotovki personala zheleznodorozhnogo transporta], *Proceedings of the VII International Scientific and Practical Conference of Students, Postgraduates and Young Scientists «Computer-Aided Design Systems in Transport» [Sbornik trudov VII Mezhdunarodnoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh "Sistemy avtomatizirovannogo proektirovaniya na transporte"]*, April 17–19, 2017, St. Petersburg, St. Petersburg, PSTU, 2017, P. 119–124.

4. Dergachev A. I., Kuranova O. N., Perepechenov A. M. Modern Multimedia Technologies in Engineering Education [Sovremennye multimediyne tekhnologii v inzhenernom obrazovanii], *In: Augustine Betancourt: from Traditions to the Future of Engineering Education. Proceedings of the International Scientific and Practical Conference [Avgustin Betankur: ot traditsiy k budushchemu inzhenernogo obrazovaniya. Sbornik trudov mezhdunarodnoy nauchno-prakticheskoy konferentsii]*, February 01–02, 2018, St. Petersburg, St. Petersburg, PSTU, 2018, P. 65–69.

5. Bryn M. Ya., Dergachev S. A., Perepechenov A. M. Computer Technologies in Teaching Students in the Disciplines of «Informatics» and «Engineering geodesy» [Komp'yuternye tekhnologii pri obuchenii studentov po distsiplinam «Informatika» i «Inzhenernaya geodeziya»], *Proceedings of the IV International Scientific and Methodological*

Conference «Problems of Mathematical and Natural Science Training in Engineering Education» [Sbornik trudov IV Mezhdunarodnoy nauchno-metodicheskoy konferentsii «Problemy matematicheskoy i estestvennonauchnoy podgotovki v inzhenernom obrazovanii»], November 03, 2016, St. Petersburg, Volume 2, St. Petersburg, PSTU, 2017, P. 12–16.

6. Dergachev A. I., Dergachev S. A., Perepechenov A. M. Modern Information Technology in the University [Sovremennye informatsionnye tekhnologii v PGUPS], *Proceedings of the XII St. Petersburg Congress «Professional Education, Science and Innovation in the XXI Century»* [Sbornik trudov XII Sankt-Peterburgskogo kongressa «Professional'noe obrazovanie, nauka i innovatsii v XXI veke»], November 12–30, 2018, St. Petersburg, St. Petersburg, PSTU, 2018, P. 73–74.

7. Dergachev A. I., Dergachev S. A., Perepechenov A. M. Automated Program Control As One of the Ways of Knowledge Assessment [Avtomatizirovanny programmy

kontrol' kak odin iz sposobov otsenki znaniy], *Proceedings of the IV International Scientific and Methodological Conference «Problems of Mathematical and Natural Science Training in Engineering Education»* [Sbornik trudov IV Mezhdunarodnoy nauchno-metodicheskoy konferentsii «Problemy matematicheskoy i estestvennonauchnoy podgotovki v inzhenernom obrazovanii»], November 03, 2016, St. Petersburg, St. Petersburg, PSTU, 2016, P. 68–70.

8. Bulavsky P. E., Dergachev A. I., Dergachev S. A., Kuranova O.N. Creating a Database in MS Excel on the topic «Analysis of the Economic Work of the Company for the Sale of Equipment»: Study guide on the discipline «Informatics in Economics» [Sozdanie bazy dannykh v MS Excel na temu «Analiz ekonomicheskoy raboty kompanii po prodazhe oborudovaniya»: Uchebnoe posobie po distsipline «Informatika v ekonomike»], St. Petersburg, PSTU, 2018, 98 p.

Направления информатизации обучения в вузе

к.в.н., доцент А. И. Дергачёв, к.э.н. С. А. Дергачёв,
к.т.н. О. Н. Куранова, к.т.н., доцент А. М. Перепеченов
Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
d_ader@mail.ru, debug@mail.ru, olga_kuranova@mail.ru, ddamper@yandex.ru

Аннотация. Рассмотрены наиболее важные аспекты овладения педагогом методами преподавания с применением современных технических средств обучения и на их основе новыми информационными технологиями. В частности, описаны особенности применения компьютерных технических средств обучения. Охарактеризованы стратегии изучения учебной дисциплины, включающей в себя различные формы обучения и разумное соотношение обучения с использованием компьютера и без него.

Ключевые слова: компьютерные технологии обучения, методика обучения, создание и применение автоматизированного учебно-методического комплекса.

ЛИТЕРАТУРА

1. Мультимедийные технологии при обучении студентов / М. Я. Брынь, А. И. Дергачёв., А. М. Перепеченов // Сб. трудов IV Международной научно-методической конференции «Проблемы математической и естественнонаучной подготовки в инженерном образовании», 03 ноября 2016 г., Санкт-Петербург. — СПб.: ПГУПС, 2016. — С. 30–34.
2. Влияние уровня подготовки персонала железнодорожного и воздушного транспорта на эффективность выполнения перевозок / О. Н. Куранова, С. А. Дергачёв, Н. А. Лебедева // Сб. трудов IV Международной научно-методической конференции «Проблемы математической и естественнонаучной подготовки в инженерном образовании», 03 ноября 2016 г., Санкт-Петербург. — Т. 2. — СПб.: ПГУПС, 2017. — С. 92–95.
3. Обоснование эффективности дистанционной подготовки персонала железнодорожного транспорта / О. Н. Куранова, А. И. Дергачёв, Я. С. Ватулин, Н. А. Лебедева // Системы автоматизированного проектирования на транспорте: Сб. трудов VII Международной научно-практической конференции студентов, аспирантов и молодых ученых, 17–19 апреля 2017 г., Санкт-Петербург. — СПб.: ПГУПС, 2017. — С. 119–124.
4. Современные мультимедийные технологии в инженерном образовании / А. И. Дергачёв, О. Н. Куранова, А. М. Перепеченов // Августин Бетанкур: от традиций к будущему инженерного образования: Сб. трудов международной научно-практической конференции, 01–02 февраля 2018 г., Санкт-Петербург. — СПб.: ПГУПС, 2018. — С. 65–69.
5. Компьютерные технологии при обучении студентов по дисциплинам «Информатика» и «Инженерная геодезия» / М. Я. Брынь, С. А. Дергачёв, А. М. Перепеченов // Сб. трудов IV Международной научно-методической конференции «Проблемы математической и естественнонаучной подготовки в инженерном образовании», 03 ноября 2016 г., Санкт-Петербург. — Т. 2. — СПб.: ПГУПС, 2017. — С. 12–16.
6. Современные информационные технологии в ПГУПС / А. И. Дергачёв, С. А. Дергачёв, А. М. Перепеченов // Профессиональное образование, наука и инновации в XXI веке: Сб. трудов XII Санкт-Петербургского конгресса, 12–30 ноября 2018 года, Санкт-Петербург. — СПб., ПГУПС, 2018. — С. 73–74.
7. Автоматизированный программный контроль как один из способов оценки знаний / А. И. Дергачев, С. А. Дергачев, А. М. Перепеченов // Сб. трудов IV Международной научно-методической конференции «Проблемы математической и естественнонаучной подготовки в инженерном образовании», 03 ноября 2016 г., Санкт-Петербург. — СПб.: ПГУПС, 2016. — С. 68–70.
8. Создание базы данных в MS Excel на тему "Анализ экономической работы компании по продаже оборудования": учебное пособие по дисциплине "Информатика в экономике" / П. Е. Булавский, А. И. Дергачев, С. А. Дергачев, О. Н. Куранова. — СПб.: ПГУПС, 2018. — 98 с.

Transboundary Trust Space as a Component of an International E-Commerce Soft-Infrastructure

Grand PhD, Prof. V. N. Kustov, PhD T. L. Stankevich
Emperor Alexander I Petersburg State Transport University
St. Petersburg, Russia
kvnvika@mail.ru, stankevich-t@gaz-is.ru

Abstract. The global economy development logic dictates necessity of barriers removal on a way of acceleration and simplification of business processes and administrative procedures the majority from which by tradition of conducting business relations are documented. Within the limits of the offered approach it is possible to draw a conclusion on that problem "Global Value Chains" widely considered now has the decision by its consideration, as the big transboundary information system uniting set of registration systems (control, business, bank and others) in the various countries. The report suggests ways of organizing transboundary trust space to implement the international e-Commerce.

Keywords: transboundary trust space, soft-infrastructure, e-commerce, INCOTERMS rules, electronic transferable records, Buy–Ship–Pay, Public Key Infrastructure, Global Value Chains, EDI-messages, Chock-Point.

INTRODUCTION

The development logic global economy dictates necessity of barriers removal on a way of acceleration and simplification of business processes and administrative procedures the majority from which by tradition of conducting business relations are documented.

There is no necessity, and possibilities, to list all variety of documents, which are used at fulfillment of information transactions in spheres of manufacture and sale, such as:

- researches and workings out;
- design of products;
- manufacture;
- marketing;
- sales;
- distribution;
- service.

Besides, documents fix various stages of complex procedures of e-commerce Buy–Ship–Pay to which it is possible to carry:

- information-marketing services;
- procedures of the contracts conclusion;
- control operations on check points and tracing of cargoes passage;
- bank and insurance operations, including possibility of arbitration protection of the rights of counterparts.

So, has developed that documenting of such information transactions is carried out mainly in the paper form. Such state of affairs in the conditions of rapid development of information technology, and first of all, networks the Internet, is

represented an anachronism. Thus, it is possible to notice that basic possibility of documenting of the information in the electronic form is provided for a long time already. In the trading custom developed by the International chamber of commerce and known, as Rule INCOTERMS, in 1990 was following position is entered:

«The main reason for successive revisions of Incoterms has been the need to adapt them to contemporary commercial practice...».

Further, in the 1990 revision of Incoterms, the clauses dealing with the seller's obligation to provide proof of delivery permitted a replacement of paper documentation by EDI-messages provided the parties had agreed to communicate electronically...».

However, it is possible to ascertain that for expired more than 20 years which have expired more the given norms haven't received wide practical application. Though the problem is in attention focus, both the international expert community, and working groups.

There is a question, and whether it is necessary to solve the given problem if it costs decades on the agenda and does not find while a practical embodiment? The answer is in our opinion obvious. The logistical costs connected with physical moving of paper documents, for a measure exponential growth of transboundary trading operations become too big. Besides, the negative influence renders also the factor of superfluous time for documenting of operations in a paper kind. These two problems become present Chock-Point (barriers) on a way of development of global world economy.

THE GENERAL APPROACHES TO A SOLUTION OF A PROBLEM

Then it is necessary to answer the following question, and in what the reasons of such state of affairs?

First, in our opinion problem roots consist in inertness of thinking. Business in that that information documenting in the physical world was carried out from time immemorial. It is known that the first case of a ceremony of real estate transfer is described in the Old Testament (the Bible. Life. Chapter 23). Prosecuted documenting subjects also in Babylon, the Sumer Kingdom, Ancient Egypt, Rome and Greece. Various types of data carriers, such as the papyrus, clay vessels copper tablets, mortgage stones and others in which number the paper is certainly presented were thus used.

Unites all it that technologies of use of such carriers concern the world of subjects and consequently the document in a general view is considered «as a unit», according to the inter-

national standard ISO 15489). The cyberspace nature essentially another, already other, expanded approaches that, however, should keep continuity in relation to tradition here again are required.

Necessity of change of a foreshortening speaks some simple reasons. In the physical world of operation on substantial drawing up of the document and documenting (registration) of a total content are carried out by the same persons or bodies, for example notaries or customs officers. In a cyberspace there is a possibility, (and it is necessary to underline especially it - economically defensible expediency connected with convenience of users) to generate the general infrastructures for documenting. A vivid example of such infrastructure is the well-known Public Key Infrastructure (PKI).

However, it conducts to differentiation of legal responsibility for granting of final service between various operators who can be under jurisdiction of the different countries. Accordingly, additional audit of activity of all operators is required. Functioning of such distributed information system should be carried out based on the unified rules and requirements. All it concerns a problem of interoperability of information systems that should be carried out at three levels:

- technological,
- legal,
- organizational.

In Addition to, at personal interaction in the course of an exchange of documents many questions, for example, a mutual cognizance (authentication), are carried out by default. Use of networks of the general access demands working out of new technologies of the information security which correct application in the conditions of mass use involves also new problems.

It conducts all too sharp complication of system, system, instead of separate documents as it was considered earlier. In addition, in Russia attempts to describe such model are undertaken, using some already turned out in UNCITRAL approaches and terms. Thus, the basic components of such model can be:

- set of Registry Systems (RS),
- set of managing directors infrastructural a component such, for example, as Identity Management System (IMS),
- set not so isolated subject («as unit») "documents", but Electronic Transferable Records (ETR).

Thus, ETR are stored in electronic databases (RS) with use of services of distribution of access and the powers provided IMS.

Besides, from the point of view of validity ETR possess doubtless advantage as they are supported in constantly the actual status and to them access to mode On-line can be organized. E-documents can be made based on ETR, be fixed on separate material carriers and further there is possible their use already in mode Off-line in cases when there is no access to a network the Internet. It is clear that in due course necessity for such e-documents can be reduced to a minimum; they will be replaced with a personal "electronic" office of the citizen or the legal person.

It is possible to add that the technology differentiated On-line access to ETR gives good possibility for reengineering of separate business processes and administrative procedures. It is actual, for example, within the limits of «one window» technology practical realization - the themes which problematics actively develops CEFACT.

Thus, for a number of information technology such, for example, as the e-government, essentially important is system consideration of a problem where the e-document is only a part of larger category that can be designated as the Transboundary space of trust formed as an information-security superstructure in the Internet.

SUGGESTED SOLUTIONS

Within the limits of the offered approach it is possible to draw a conclusion on that that problem widely considered now "Global Value Chains" has the decision by its consideration, as the big transboundary information system uniting set RS (control, business, bank and others) in the various countries. These RS should use the general security infrastructure, for example, with a view of realization of the international e-commerce.

Practical implementation of such a transboundary system for Global Value Chains participants can be carried out by a specific set of access and documenting services.

The documenting services should include:

- signature service, designed to confirm the will of individuals - participants of information exchange;
- the service of time intended for acknowledgement of time of fulfillment of electronic transferred record to which it will be organized differentiated on-line access, or editions on the basis of this record of the electronic document having the status off-line;
- the service of a place intended for assurance of a place of entering of electronic transferred record (a place of the edition of the electronic document);

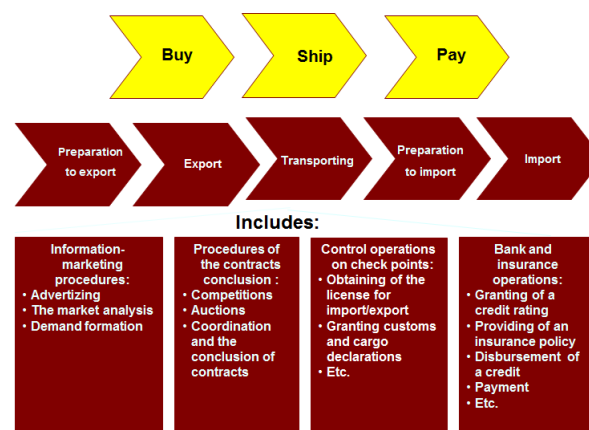


Fig. 1. Global Value Chains Procedural steps

- notarial service, is intended for assurance of the data (a content, the signature, the ciphered content, etc.) for legally significant application within one jurisdiction;
- apostille service, is intended for assurance of the data (a content, the signature, the ciphered content, others) for legally significant application within various jurisdictions;
- service of gathering payment, is intended for realization of payment of gathering, tariffs, taxes, etc., legally significant actions connected with fulfillment;
- set of services of monitoring of legal statuses of subjects of the information interaction intended for registration, maintenance in an actual condition and cancellation of legal statuses of subjects of information interaction - legal bodies, and also competences, powers and the right to sign of physical persons;

- service of the entrusted data storage, is intended for realization original «a bank cell» in which it is possible to store, for example, in the ciphered kind, including having validity the information concerning separate physical and legal bodies (under the contract with them), in a case when expenses for independent storage of such data are represented for users excessive or unsafe.

It is possible to carry to access services:

- the base service of authentication intended for check and acknowledgement of an accessory to the subject of access (the subject of information interaction) the identifier shown to it;
- the auxiliary services (options) connected with the organization of remote access; use protected (cryptographically protected) access; use of those or other accessors and others.

Some elements of such a system are described in INCOTERMS 2010 [1].

When INCOTERMS 2000 describes trading customs in all regulations, excluding Regulation EXW (EX Works), it always refers to Item 8, which defines an order of presentation proofs of delivery, transport documents or equivalent electronic messages. Unfortunately, INCOTERMS describes this order for application only within the limits of one-domain EDI.

This point specifically for the seller says, ‘...Where the seller and the buyer have agreed to communicate electronically, the document referred to in the preceding paragraphs may be replaced by an equivalent electronic data interchange (EDI) message’.

However, in case of the transboundary EDI, for example, in a two-domain system there arises a problem related to using different, incompatible cryptography by each party. It becomes a next Chock Point in an attempt to ensure legal force of such electronic documents and makes it necessary to use a Trusted Third Party (TTP).

In this case, options of TTP construction can be as follows (Figure 2): TTP A; TTP B; Double TTP; Bridge TTP.

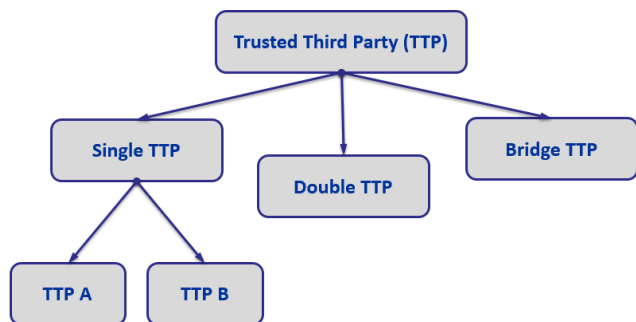


Fig 2. TTP Classification

Let us describe TTP B in more detail (Figure 3) because it is this construction that can provide a transfer of electronic documents from the seller to the buyer according to INCOTERMS rules.

The buyer sends the transport document in electronic form, signed by a digital signature according to the cryptographic standard of party A. The electronic document arrives through transboundary space to the buyer on party B's side.

The buyer, staying in another jurisdiction and in another cryptography zone, is unable to confirm a legal force of the electronic document. He has to use TTP services on his side for this purpose. TTP functions include a check of the digital signature validity, issuance of the receipt and its sending to the

buyer. The buyer, having received acknowledgement in the form of a positive receipt from TTP on digital signature validity, is obliged to accept the transport document.

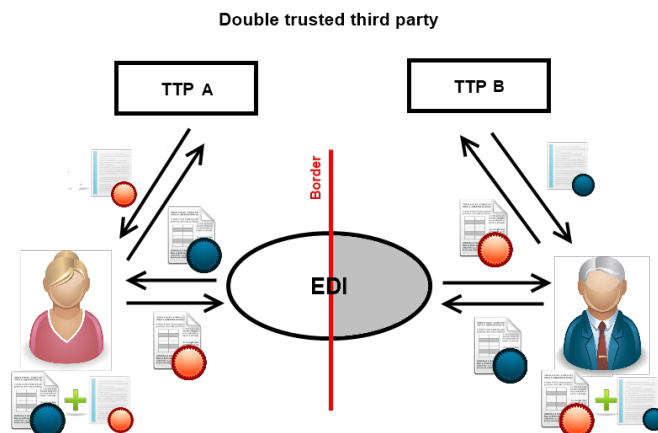


Fig 3. Double TTP

DISCUSSION

Work in the proposed direction is already underway in a number of international formats. Thus, within the framework of the Commonwealth of Independent States, a Model [2] for the formation and operation of a transboundary trust space of the CIS member States on the Internet has been developed, discussed and approved. To date, the methodology of [3] system description of this Model has been developed.

Similar actions are undertaken in other international formats, such as the Commission of the Customs Union and the Shanghai Cooperation Organization. The initiative WEB 3.0 is offered in a format of Dialogue Russia – EU on an information society, elaboration of specific aspects is carried out. Offers on the basis of the stated approaches are presented also in CEFACT within the framework of discussion of a draft for 37th Recommendation on mutual recognition of digital signatures.

CONCLUSION

Thus, this paper deals with the problems of creating a transboundary trust space as an integral part of the soft-infrastructure of international e-Commerce. Significant problems of multiple barriers removal on acceleration and simplification way of various business processes and administrative procedures providing cross-border e-Commerce systems are noted.

Within the framework of the proposed approach, it is concluded that the widely considered problem of "Global Value Chains" has a solution when it is considered, only as a large transboundary information system that combines the set of RS (control, business, banking and others) in different countries. These RS must use a common soft infrastructure.

Practical implementation of such transboundary system for Global Value Chains participants can be carried out by a specific set of access and documentation services. The full list of these services is given. On the example of the rules of trade customs described in INCOTERMS 2010, the ways of realization of transboundary trust space are given.

It is noted that with transboundary EDI, for example, in a two-domain system, there is a problem associated with the use of different cryptography by each party, which is not compati-

ble with each other, which necessitates the use of TTP. The classification and options for building a cross-border trust space based on the use of TTP are given. The procedure of document exchange between buyer and seller in two-domain EDI using different cryptographic standards is described.

The results of the work on the creation of a cross-border trust space in various international formats are presented.

As additional sources of literature on the issue we have examined, standards [4–5] and articles [6–8] are recommended.

REFERENCES

1. Incoterms 2010. ICC Rules for the use of domestic and international trade terms [Inkoterm 2010. Pravila ICC po ispol'zovaniyu natsional'nykh i mezhdunarodnykh torgovykh terminov], Moscow, ICC Russia, 2010, 33 p.

2. Methodology of formation and functioning of cross-border trust space in the Internet. Version 3.0 [Metodologiya formirovaniya i funkcionirovaniya v seti Internet transgranichnogo prostranstva doveriya. Versiya 3.0], Moscow, 2012, 72 p. Available at: <http://docplayer.ru/33957887-Metodologiya-formirovaniya-i-funkcionirovaniya-v-seti-internet-transgranichnogo-prostranstva-doveriya-versiya-3-0.html>.

3. Anikin S. N., Domrachev A. A., Dralo M. P., Dupan A. S., Kiriushkin S. A., Furgel I. A. Methodology of Formation of

Transboundary Trust Environment [Metodologiya formirovaniya transgranichnogo prostranstva doveriya], *Zashita informacii. Inside [Zashchita informatsii. Insayd]*, 2017, No. 2 (74), P. 13–17 (Part 1); No. 3 (75), P. 6–13 (Part 2).

4. ITU-T Recommendation X.842 (2000) | ISO/IEC TR 14516:2002(E): Information technology — Security techniques — Guidelines for the use and management of Trusted Third Party services. Available at: <http://www.itu.int/rec/T-REC-X.842-200010-I/en>.

5. RFC 3029 — Internet X.509 Public Key Infrastructure Data Validation and Certification Server Protocols. Available at: <http://tools.ietf.org/html/rfc3029>.

6. Miller R., Cross F. B. The Legal and E-Commerce Environment Today: Business in Its Ethical, Regulatory, E-Commerce, and Global Setting, Thomson Learning, 2002, 741 p.

7. Thompson J. A Matter of Trust. Infoconomy: Business Briefing, No 16, 2001. Pp. B27–B29.

8. Baldwin A., Beres Y., Casassa Mont M., Shiu S. Trust Services: A Trust Infrastructure for E-Commerce. HPL-2001-198, 15 August 2001, 23 p.

Available at: <http://www.hpl.hp.com/techreports/2001/HPL-2001-198.html>.

Трансграничное пространство доверия как составная часть софт-инфраструктуры международной электронной коммерции

д.т.н., проф. В. Н. Кустов, к.т.н. Т. Л. Станкевич
Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
kvnvika@mail.ru, stankevich-t@gaz-is.ru

Аннотация. Логика развития мировой экономики диктует необходимость снятия барьеров на пути ускорения и упрощения бизнес-процессов и административных процедур, большинство из которых по традиции ведения деловых отношений документированы. В рамках предложенного подхода можно сделать вывод о том, что широко рассматриваемая в настоящее время проблема "Глобальных Цепочек Поставок" имеет решение путем её рассмотрения как большой трансграничной информационной системы, объединяющей множество учётных систем (контрольных, деловых, банковских и других) в различных странах. В докладе предлагаются пути организации трансграничного пространства доверия для осуществления международной электронной торговли.

Ключевые слова: трансграничное пространство доверия, софт-инфраструктура, электронная коммерция, правила ИНКОТЕРМС, электронные передаваемые записи, Покупка-Отгрузка-Оплата, инфраструктура открытых ключей, глобальные цепочки поставок, ЭДО-сообщения, препятствие.

ЛИТЕРАТУРА

1. Инкотермс 2010. Правила ICC (International Chamber of Commerce) по использованию национальных и международных торговых терминов (публикация Международной Торговой Палаты № 715 от 16.09.2010). — М.: Международная торговая палата, 2010. — 33 с.
2. Методология формирования и функционирования в сети Интернет трансграничного пространства доверия. Версия 3.0 — Москва, 2012. — 72 с.

URL: <http://docplayer.ru/33957887-Metodologiya-formirovaniya-i-funkcionirovaniya-v-seti-internet-transgranichnogo-prostranstva-doveriya-versiya-3-0.html>

3. Методология формирования трансграничного пространства доверия / С. Н. Аникин, А. А. Домрачев, М. П. Драло, А. С. Дупан, С. А. Кирюшкин, И. А. Фургель // Защита информации. Инсайд. 2017. № 2 (74). С. 13–17 (часть 1); № 3 (75). С. 6–13 (часть 2).

4. ITU-T Recommendation X.842 (2000) | ISO/IEC TR 14516:2002(E): Information technology — Security techniques — Guidelines for the use and management of Trusted Third Party services. URL: <http://www.itu.int/rec/T-REC-X.842-200010-I/en>.

5. RFC 3029 — Internet X.509 Public Key Infrastructure Data Validation and Certification Server Protocols. URL: <http://tools.ietf.org/html/rfc3029>.

6. Miller R., Cross F. B. The Legal and E-Commerce Environment Today: Business in Its Ethical, Regulatory, E-Commerce, and Global Setting, Thomson Learning, 2002, 741 p.

7. Thompson J. A Matter of Trust. Infoconomy: Business Briefing, No 16, 2001. Pp. B27–B29.

8. Baldwin A., Beres Y., Casassa Mont M., Shiu S. Trust Services: A Trust Infrastructure for E-Commerce. HPL-2001-198, 15 August 2001, 23 p.
URL: <http://www.hpl.hp.com/techreports/2001/HPL-2001-198.html>.

ПАМЯТИ КОЛЛЕГИ



Добрую память о себе в умах преподавателей и ученых оставил профессор, доктор технических наук, профессор кафедры «*Информационные и вычислительные системы*», почетный профессор Петербургского государственного университета путей сообщения императора Александра I, почетный железнодорожник, член-корреспондент Международной академии наук высшей школы, Заслуженный деятель науки Российской Федерации ЯКОВЛЕВ Валентин Васильевич (06.01.1939–01.12.2019).

Выпускник Электротехнического факультета Пражского института транспорта. Специалист в области стохастических методов и средств вычислений и моделирования, архитектур локальных вычислительных сетей ЭВМ. Проработал в университете более пятидесяти шести лет. Прошел путь от лекционного ассистента до заведующего кафедрой, проректора ПГУПС по информатизации.

Один из разработчиков отраслевой программы «Информатизация высших и средних специальных учебных заведений МПС РФ», председатель научно-методического совета «Информационные системы» учебно-методического объединения железнодорожных специальностей. Профессор Яковлев В. В. внес существенный вклад в становление и развитие кафедры ИВС, в развитие университета и информатизации. Основатель научной школы «Стохастические методы и средства вычислений».

По инициативе и при активном участии Валентина Васильевича с 1995 по 2009 гг. организовывалась и проводилась международная научно-практическая конференция «Информационные технологии на железнодорожном транспорте» — ИНФОТРАНС.

Яковлев В. В. активно участвовал в проведении лекционных занятий, руководил работой студентов бакалавриата, магистратуры и аспирантов. Подготовил 3 докторов технических наук и 12 кандидатов технических наук. Автор более 300 научных работ.

Валентин Васильевич Яковлев пользовался высоким и заслуженным авторитетом среди коллег и студентов, представителей других ВУЗов и железнодорожной отрасли. Светлая память об этом замечательном, мягком, умном, интеллигентном человеке навсегда останется в наших сердцах.

In Memory of Colleague

Good memories in the minds of teachers and scientists was left by the doctor of technical sciences, professor, professor of department of Information and Computing Systems, Honorary professor of Emperor Alexander I St. Petersburg State Transport University, Honorary railway worker, corresponding member of the International Academy of Sciences of Higher Education, Honored Scientist of the Russian Federation Valentin V. YAKOVLEV (01/06/1939—12/01/2019).

Graduate of the Faculty of Electrical Engineering at the Prague Institute of Transport. Specialist in the field of stochastic methods and means of computing and modeling, architectures of local computer networks. He worked at the university for more than fifty-six years. He has gone from a lecture assistant to the head of the department, vice-rector of PSTU for Informatization.

One of the developers of the industry program "Informatization of higher and secondary special educational institutions of the Ministry of Railways of the Russian Federation", Chairman of the Scientific and Methodological Council "Information Systems" of the Educational and Methodological Association of Railway Specialties.

Professor Yakovlev V. V. made a significant contribution to the formation and development of the department of temporary detention facilities, to the development of the university and informatization. Founder of the scientific school "Stochastic methods and means of computing".

On the initiative and active participation of Valentin Vasilievich, from 1995 to 2009, an international scientific-practical conference Information Technologies in Railway Transport — "INFOTRANS" was organized and held.

Yakovlev V. V. actively participated in lectures, led the work of undergraduate, graduate and postgraduate students. He prepared 3 doctors of technical sciences and 12 candidates of technical sciences. The author of more than 300 scientific papers.

Valentin Vasilievich Yakovlev enjoyed high and well-deserved authority among colleagues and students, representatives of other universities and the railway industry. The bright memory of this wonderful, gentle, wise, intelligent person will forever remain in our hearts.