



УДК 004.036

Р.Т. Бржанов

Ш. Есенов атындағы КМТЖИУ, Ақтау қ.

В.А. Лахно

Еуропалық университеті, Киев қ.

**АҚПАРАТТЫ ӨНДЕУДІҢ АВТОМАТТАНДЫРЫЛҒАН ЖҮЙЕЛЕРІНІҢ ОСАЛДЫҚТАРЫҢ МОДЕЛЬДЕУ
ҮШІН MATLAB ЖӘНЕ SIMULINK ПАКЕТІН ПАЙДАЛАНУ**

В статье содержатся результаты исследований, позволяющих повысить уровень защиты автоматизированных информационных систем предприятий. В работе предложена последовательность этапов моделирования работы систем защиты информации от сетевых атак, когда имеются конфликтные потоки передаваемых данных между сервером и автоматизированными рабочими местами.

Мақалада кәсіпорындардағы автоматтандырылған ақпараттық жүйелерін қорғау деңгейін көтерудегі зерттеу нәтижелері қарастырылған. Жұмыста сервер мен жұмыс станциялары арасында берілетін деректердің қайшылықтары болған жағдайда, желілік ақпараттарды қорғау жүйелерінің жұмысын модельдеу кезіндегі кезеңдерді ұсынылды.

The article contains the results of studies to improve the level of protection of automated information systems of enterprises. The paper proposes a sequence of stages of modeling of information security systems from network automated information systems in conditions where there are conflicting flows of transmitted data between the server and automated workstations.

Кілтті сөздер: автоматтандырылған ақпараттық жүйе, қорғау деңгейі, математикалық модельді қорғау.

Ақпараттық қауіпсіздіктің талдауға келмейтін оқиғалары мен инциденттерінің көп болуына әкеледі, олар технологиялық осалдықтардың алдын алуды, сонымен қатар АКТ-ны қылмыстар жасау үшін құрал ретінде пайдаланатын қылмыскерлермен күресті қиындатады [1]. Көптеген ресурстарға қол жеткізу автоматтандырылған ақпараттық жүйелерді (ААЖ) көздейді ресурстарды бөлу арна абоненттері арасында [2, 3]. Осы жұмыс жалғастыруда тақырыбы бойынша жарияланымдар зерттеумен байланысты даулы ағындарын ААЖ деректерді, және қалай салдары туындайтын проблемаларға кіріспе ақпаратты қорғау жүйесін (АҚЖ) осындай ААЖ.

Ықтималдығы трансформация қызмет көрсету рәсімдерін мәліметтер ағымының ба-стапқы жай-күйін түпкі желісі бойынша орналасқан шешімі негізінде жүйесін интегро-дифференциалдық тендеулер [4,5]:

$$\Phi_{ij}(h(tr_{ij}), \tau) = o_{ij} \cdot \int_0^{\tau} f_{ik}(h(tr_{ij}), \tau) \cdot \Phi_{kj}(h(tr_{ij}) - \tau) \cdot d\tau, \quad (1)$$

мұндағы $\{o_{ij}\}$ – матрица ықтималдықтар трансформация күйлердің ААЖ; $\{f_{ij}(\tau)\}$ – матрица тығыздық ықтималдық заманнан болу рәсімдерді қызмет көрсету деректер ағынын әрбір i -ші жағдайы.

Шешу үшін пайдаланылады математикалық аппараты желілерін Петри-Марков (ЖПМ). Ойлаймыз, өйткені желісі таңдалды (доғаның баған белгілі), онда балама нұсқалары бойынша орын ауыстыру дугам, қарауға болады.

$$\Phi_g(\tau) = \begin{cases} \prod_{h(tr) \in H} \Phi_{i+n_h(tr_{ij})}(h(tr_{ij}), \tau) - \text{for } \wedge; \\ 1 - \prod_{h(tr) \in H} [1 - \Phi_{i+n_h(tr_{ij})}(h(tr_{ij}), \tau)] - \text{for } \vee, \end{cases} \quad (2)$$

мұндағы – $\Phi_{i+n_h(tr_{ij})}(h(tr_{ij}), \tau)$ өтпелі ықтималдығы. Анықталып, нәтижесінде тендеулер жүйесін шешудің олар сипаттайды жүйені трансформациялау.

Егер $h(tr_{ij})$ желі астам емес анықталады өткелдер (немесе жоқ), онда ықтималдығы, бұл процесс есеп айырысу сәтіне уақыт жетеді түпкілікті көшу, яғни, тисе соңғы шыңына баған орналасқан, мысалы:

$$\Phi_{i,j(t)} = \prod_{k=1}^j o_{spm(s_a)}, k(s_a, j) \cdot \int_0^t \Phi_{s_a}(\tau) \cdot f_{spm(s_a),j}(t-\tau) \cdot d\tau, \quad (3)$$

мұндағы $spm(s_a)$ – позициясының нөмірі бойынша ЖПМ, ол керек бірден көше нөмірімен; $s_a; k(s_a, j)$ – нөмірі бойынша өту тәртібі бойынша ауыстыру кезінде СПМ жылғы s_a көшу нөмірімен көшу нөмірімен j .

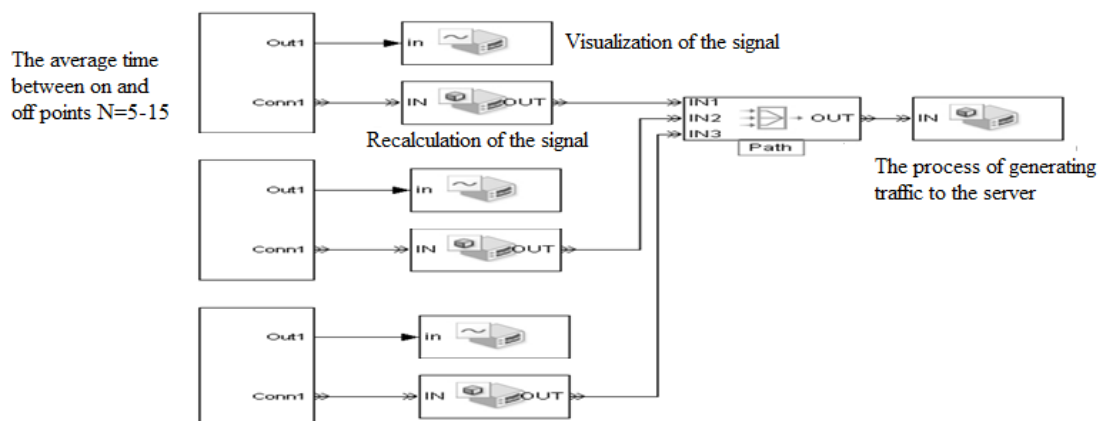
Ықтималдығы $\Phi_{i,j(t)}$ болып табылады, шын мәнінде, мүмкіндігі іске асыру ақпараттық қауіптер. Егер қарастырылып отырған желісі тағы бар өткелдерін отырып, логикалық критерийлері болса, онда есептеу қайталанады.

Үлкен жүйесін, оның ішінде ААЖ жарактандырылған кешендермен АҚЖ тұрады жүздеген, ал кейбір жағдайларда мың элементтерін және тағы көп санын олардың арасындағы байланыстар. Мұндай жүйе сипатталады неоднородностью элементтері. Сондай-ақ, гетерогенді арасындағы байланыс элементтері. Қарамастан дербес компоненттері немесе байланыс ААЖ барабар сипатталған модельдермен дискретті математика немесе теориясы жаппай қызмет көрсету жүйесі туралы тұтастай алғанда, бұл айтуға болмайды. Баламалы тәсіл қолдану болып табылады әдіснамасы имитациялық моделдеу (М). Имитациялық модельдеу біріктіруге мүмкіндік береді бір-бірімен әртекті математикалық модельдер элементтерінің құрамына кіретін ААЖ.

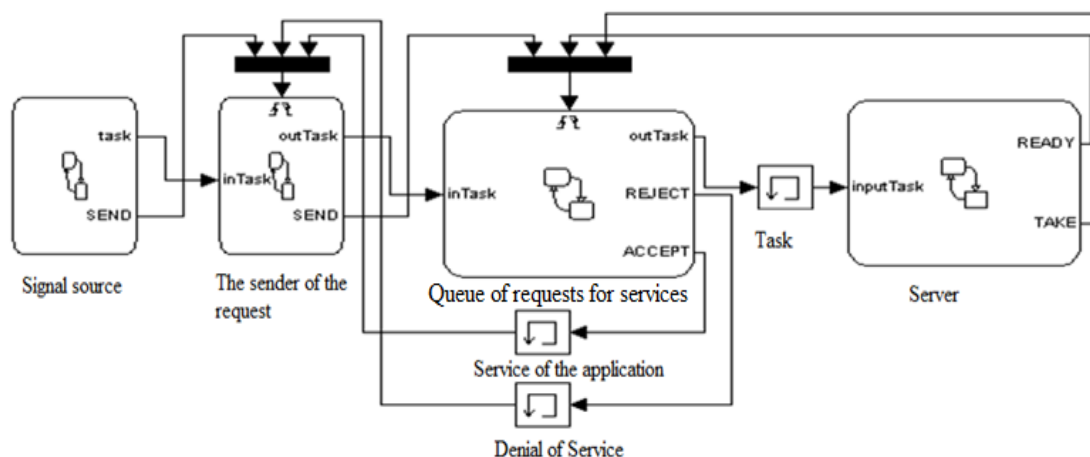
Үшін оларға ААЖ (құрылымдық компоненті ескеріледі және болуы АҚК) даулы ағынын таңдалған пакет Simulink. Мақсаты пакетін жүзеге асыру болып табылады имитациялық модельдеу жүйелері мен құрылғыларын [6].

Ескере отырып, рекурренттық білдірген келтірілген жұмыстарға [5], пайдаланып, құрал-саймандар имитациялық модельдеу пакетін MATLAB+Simulink, қарастырайық тәсілдерді талдау, динамикасын және өзара байланыстарын трафиктердің деректерді моделируемой ААЖ. Ескеріледі жанжалдық деректер ағымын [6]. Осы зерттеулерді жалғастыру [5, 6], деп ойлаймыз, желісі ААЖ дан тұрады бір деректерді беру желісі және станциялары (АЖО). АЖО мезгіл-мезгіл жіберуге талаптар деректерді беру желісі бойынша, суретті. 1, 2. Сұрау параметрлерін құрылды деректерге сәйкес [6]. (Автоматтандырылған жұмыс орны) АЖО1 – басым ағыны аз қарқындылығы; АЖО 2 – ағыны аз қарқындылығы; АЖО 3 – басым ағыны ең жоғары қарқындылығы түсетін сервер деректер).

Әуелде сондай-ақ, бұл уақытта дискретті және өзгеріп отырады 0-ден белгілі маңызы бар. АЖО бір бірінен тәуелсіз жұмыс істейді, және әрбір уақытта белгілі бір ықтималдықпен кез келген станция бола алады түсу талабы деректерді беру желісі бойынша немесе орын босату желісі.



Сур. 1 - Схемасы имитациялық модельдеу осалдықтар ААЖ шабуыл үшін "типті қызмет көрсетуден бас тарту"

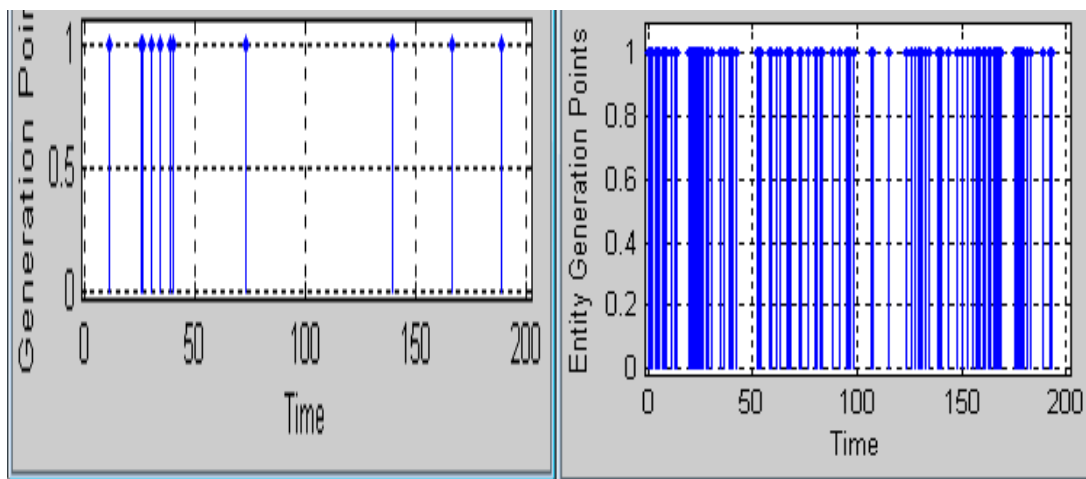


Сур. 2 - Жұмыс істеу схемасы сервер ААЖ қызмет көрсететін өтінім

Бағдарламалық іске асыру математикалық модельдер [6] визуализация үшін мүмкіндік берді процестер шығу әрбір АЖО. Т. е. болуы мүмкін, онда маңызды сипаттамалары ҚМК (время простоя деректерді беру желісі ААЖ-орташа уақыты деректер беру желісі бойынша әрбір АЖО және т. б.). Іс жүзінде ұсынады күрделілігі іске асыру бағдарламалық тест реттілігі кез келген таңдалған пайдаланушы немесе неғұрлым маңызды компоненті жай-күйін ААЖ (ойымызша, белгілі саны АЖО ААЖ). Мысалы, күріш. 3 көрсетілді тест трафиктерді желіде ААЖ.

1-нұсқа – барлық АЖО тең ықтималдығын алуы жолға беру. Шамамен бірдей және ықтималдық босату беру желілері кейін оның сабақтары тиісті АЖО-суретті қараңыз. 3 "а".

2-нұсқа – сервер ААЖ шабуылға ұшырағанда үлгідегі бас тарту қызмет көрсетуде. Содан кейін ықтималдығы жағдайларға сабақтар желісі әрбір АЖО ААЖ ұмтылу "0" немесе бірінші және екінші АЖО бар көп сұраныстарын деректерді және ұзақ алады сызығына қарағанда, бірінші АЖОЗ 3 -суретті қараңыз.



Сур. 3 - Модельдеу нәтижелерін жұмыс істеу режимдерінің ААЖ

ҚОРЫТЫНДЫ

Орындалған зерттеулер мүмкіндік берді әзірлеу кезендерінің реттілігі келесі модельдеу жұмыс АҚҚ-ның құрамында ААЖ:

- 1) көп фазалы ҚМК салу схемасы есептеу үшін сыйымдылық және уақытша сипаттамаларын АҚҚ ААЖ пайдалана отырып, MATLAB + Simulink;
- 2) орындау заңдастыруды барлық ықтимал жағдайлар модельдеу ААЖ және оның АҚҚ желілік шабуылдардан.
- 3) бөлу қауіпті жай-күйін; анықтау қарқындылығы мен көшу ықтималдығын тораптарын АҚҚ мен ААЖ жалпы жай-күйі;
- 4) сәйкес граф өтуілер жүйесін ескере отырып, дифференциалдық немесе рекурренттік тендеулер анықтау ықтималдық сипаттамалары күйлердің ААЖ уақыт функциясы сияқты және қарқындылығы кіріс ағынының талаптар.

Қолданған әдебиеттер

1. Киберқауіпсіздік тұжырымдамасы. Қазақстан Республикасы Үкіметінің 2017 жылғы «30» маусым № 407 [қаулысымен](#) бекітілген.
2. Petrov, B. Borowik, M. Karpinsky, O. Korchenko, V. Lakhno (2016). Immune and defensive corporate systems with intellectual identification of threats, Psczyna :ŚląskaOficynaDrukarska, 222 p. ISBN: 978-83-62674-68-8.
3. Lakhno, V. (2014). Ensuring of information processes' reliability and security in critical application data processing systems. MEST Journal, Vol. 2, pp. 71–79. doi: 10.12709/mest.02.02.01.07
4. Воробьев А.А. Анализ моделей процессов защиты информации от несанкционированного доступа в автоматизированных системах // Информатика-машиностроение. - 1999. - № 2. - С. 32-34.
5. Лакно В.А. Обеспечение защищенности автоматизированных информационных систем транспортных предприятий в условиях роста транзитных перевозок: Збірник наукових праць Військового інституту Київського національного університету ім. Тараса Шевченка / В.А. Лакно, А.С. Петров. - № 21. - 2009. - С. 110-120.
6. Лакно В.А. Исследование конфликтных потоков заявок в системах защиты информации / В.А. Лакно, А.С. Петров, Н.Н. Чертунина // Вісник СНУ ім. В. Даля. - № 6 (136). - 2009. - С. 200-209.

Получено 13.08.2018

МРНТИ 681.518:339.138(045)

Б.Е. Ибадильдин, А.А. Исмаилова

Казахский агротехнический университет им. С. Сейфуллина, г. Астана

РОЛЬ ЧАТ-БОТОВ В АВТОМАТИЗАЦИИ СЕРВИСОВ

В данной статье рассматривается развитие чат-ботов как инструмента автоматизации различных сервисов. Приведены и рассмотрены виды ботов и процессы взаимодействия с пользователями, а также базовые примеры разработки. Рассмотрены главные преимущества и недостатки использования данного вида автоматизации в различных сферах.

Бұл мақалада әртүрлі қызметтерді автоматтандыру құралы ретінде чат боттардың дамуы талқыланады. Пайдаланушылар мен боттар және өзара мәлімет алмасу процестері, сондай-ақ негізгі даму мысалдары ұсынылып, қарастырылады. Әртүрлі салаларда автоматтандырудың осы түрін пайдаланудың негізгі артықшылықтары мен кемшіліктері қарастырылады.

This article discusses the development of chat bots as a tool for automating various services. Bots and interaction processes with users, as well as basic development examples, are presented and considered. The main advantages and disadvantages of using this type of automation in various fields are considered.

Ключевые слова: чат-бот, сервисы, автоматизация, мессенджер, социальная сеть, Telegram, медиатренды, маркетинг.

Мессенджер как платформа развития для автоматизации. Актуальность данной темы определена тем, что в данный момент мессенджеры становятся более популярными. Как и мобильные телефоны, плавно перешедшие в облик смартфонов, мессенджеры обрастают новыми функциями, которые положительно влияют на их популярность в современном обществе, соответствуя современным медиатрендам.

Согласно данным аналитической компании eMarketer, в 2016 году более 1,4 миллиарда человек использовали приложения для обмена сообщениями. К 2019 году более 25% населения мира (примерно 1,75 миллиарда человек) будут постоянно использовать мессенджеры для связи друг с другом [1]. На сегодняшний день мессенджеры начинают частично вытеснять социальные сети, новостные ленты, а также различные сервисы, которые были доступны и ранее, но разнесены по отдельным классам. Одним из инструментов, позволяющих прогрессировать в направлении автоматизации, являются чат-боты.

Если обратиться к определению данного термина, чат-бот - это некая компьютерная программа, имитирующая поведение человека при общении, иначе говоря, программа-собеседник. Но и это определение уже тоже можно признать устаревшим, поскольку нынешнее поколение чат-ботов, обитающее в мессенджерах, уже не используется только для имитации общения. На сегодняшний день чат-боты - это целый программный кластер, объединяющий в себе различные сервисы в одном месте [2].

Основные векторы использования чат-ботов. Согласно докладу Humanity in the Machine, подавляющее большинство людей готовы к тому, что бизнес будет предоставлять свои услуги с помощью чат-ботов. Ради быстрых и точных ответов люди готовы общаться с ботами-продавцами, консультантами, секретарями. Онлайн-чат и приложения для обмена сообщениями являются наиболее предпочтительным способом связи с продавцом для 29% американцев. В других странах этот показатель пока ниже, но постепенно всё больше людей начинают отдавать предпочтение чат-ботам. В данный момент существует очень много вариаций их использования. Основные чат-боты:

- автоматизация рутинной работы - позволяет выполнять определенные функции, не привлекая людей, а работа будет выполнена моментально и безупречно;

- поиск и агрегация новостей, аналитики, данных (data-driven collaboration) - данные доступны в месте принятия решений - мессенджерах и всем участникам, которым они нужны;

- электронная коммерция - для спонтанных покупок без долгого поиска, mobile ecommerce + visual search + chatbots, для общения с клиентами;

Первая линия работы с клиентами, помощники, консультанты, типовые вопросы, телефония [3]. Ярким примером использования данной технологии является чат-бот, разработанный АО «Казпочта», взявший за основу платформу мессенджера Telegram. Несмотря на то что сервисы данной компании доступны как в веб-приложении, так и в мобильном приложении, все основные и часто используемые сервисы есть и в чат-боте (рис. 1).

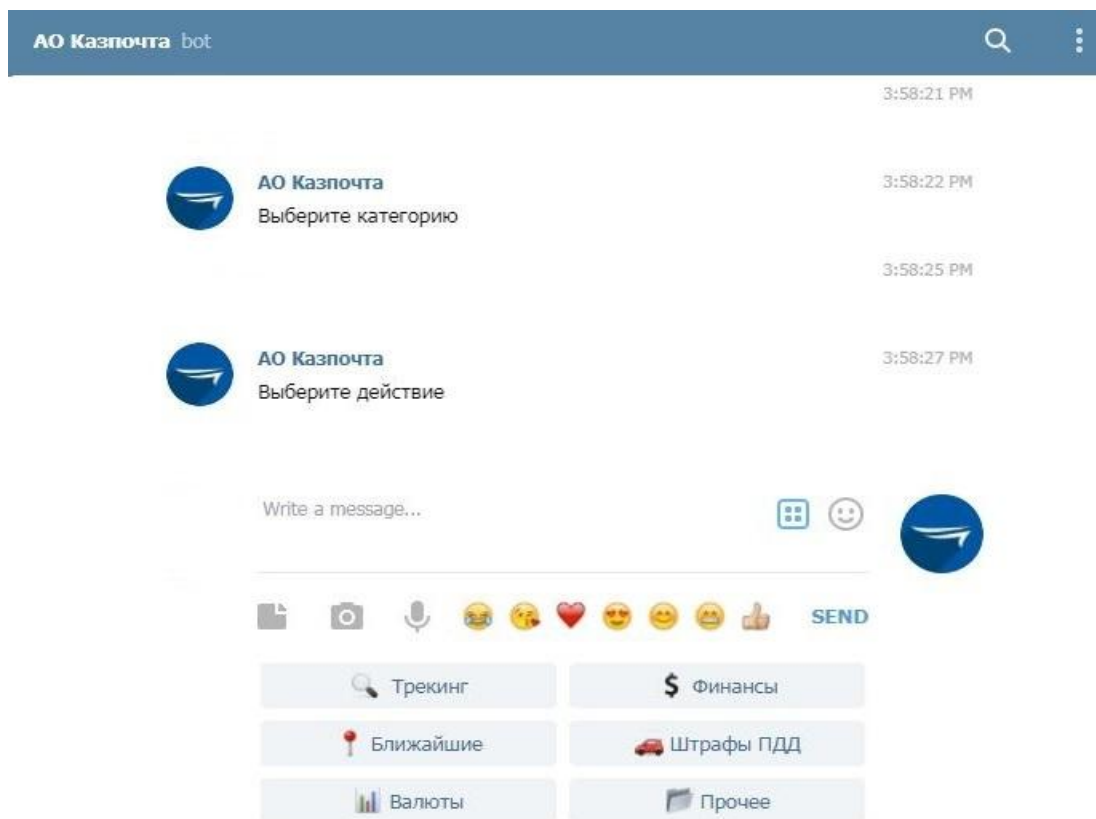


Рисунок 1 - Вид бота из веб-версии Telegram

Примеры реализации. Разработка и реализация чат-ботов может быть абсолютно разнообразной, в основном зависящей от выбранной платформы (мессенджера). В основе каждого бота лежит некий обработчик событий, который срабатывает при определенном запросе со стороны клиента (рис. 2). Это может быть как заранее определенная четкая логика или сценарий, так и искусственный интеллект, который работает на основе машинного обучения. Второй вариант используется для создания личных виртуальных помощников или виртуальных собеседников, предоставляющих некую информацию или набор сервисов. Первый вариант используется более часто, поскольку в данный момент чат-боты используют в первую очередь компании для продвижения и оптимизации своих сервисов и услуг.

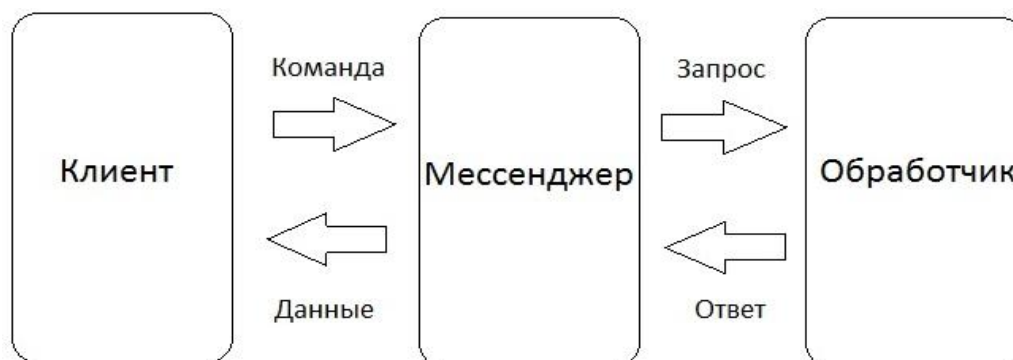


Рисунок 2 - Схема работы чат-ботов

Как можно заметить, мессенджер, как и говорилось ранее, выступает в роли платформы для связи между пользователем и обработчиком. К примеру, если брать мессенджер Telegram, в нем все команды, подаваемые обработчику, указываются через символ «/», тем самым дав понять, что это команда. Но есть и более удобные реализации общения с ботом, которые сводятся к нажатию специальных кнопок на клавиатуре, заранее подготовленной разработчиками, которая интерпретирует нажатие определенной кнопки в определенную команду.

Выбор платформы и базовые особенности разработки. Если обратиться к статистике, то большинство чат-ботов разрабатываются для таких платформ, как Facebook Messenger, Slack, Telegram, WeChat, Kik (см. табл).

Сводная статистика популярных мессенджеров [6]

Мессенджер	Аудитория в РФ (MAU)	Аудитория в мире (MAU)	В скольких странах лидирует?	Боты
WhatsApp	30 млн	1 млрд	109	2017-2018
Viber	20 млн	249 млн	15	Скоро
Skype	10 млн	300 млн	Нет данных	Да
Facebook Messenger	4 млн	1 млрд	49	Да, платный аппрув (\$99)
ICQ	6,7 млн	11 млн	-	Возможно
Telegram	3,5 млн	120 млн	Иран Узбекистан	Да
WeChat	Нет данных	800 млн	3 (Китай)	Да
Line	Нет данных	218 млн	4 (Япония, Таиланд, Тайвань, Индонезия)	Да
Kik	Нет данных	Нет данных	-	Да

Для примера мы решили выбрать платформу мессенджера Telegram, поскольку его разработчики сами открыто распространяют API для написания обработчика, тем самым делая разработку чат-бота для сторонних разработчиков более удобной. К примеру, тот же WhatsApp категорически запрещает вмешиваться в процесс работы своего продукта сторонним разработчикам, тем самым ограничивая свою платформу, несмотря на свои лидирующие позиции в отношении аудитории [2].

Перейдем непосредственно к части разработки. Для создания и регистрации непосредственно самого бота необходимо обратиться к боту BotFather с командой /newbot, будет

получен токен для коммуникации с будущим ботом, а также заданы базовые параметры в виде имени, никнейма, аватара и т.п. [4]. После этого необходимо определиться с методом общения мессенджера с обработчиком для нашего бота. Telegram предлагает целых два - это GetUpdates и Webhook. Разница в них в том, что в первом случае наш сервер будет опрашивать серверы мессенджера на наличие запросов к нашему боту, что не очень стабильно, поскольку такие запросы могут быть отклонены по тем или иным причинам (недоступность серверов, технические ошибки при запросе). Во втором же случае мы задаем конкретный адрес, к которому серверы Telegram будут обращаться сами при поступлении запросов. Но для этого нужен хостинг или выделенный виртуальный сервер, имеющий домен любого уровня, с установленным SSL-сертификатом для шифрации соединения. Поскольку второй является более рациональным и отказоустойчивым, при разработке лучше использовать его. Далее мы подвязываем при помощи метода setWebhook бота к нашему домену и пишем обработчик, который будет возвращать некоторые данные по запросам, что будут приходить из мессенджера [5].

Благодаря API мессенджер Telegram уже предлагает готовые методы «из коробки», в связи с чем обработчик можно будет написать вручную, создав для удобства небольшую библиотеку для регулярных выражений, функций и методов. Но также можно воспользоваться и готовыми библиотеками, написанными и дорабатываемыми сообществом разработчиков.

Для платформы Telegram есть множество библиотек на таких языках программирования, как PHP, Python, Javascript, C# и даже Haskell. После того как мы настроили базовый ввод и вывод запросов через обработчик (рис. 3), можно уже приступить к интеграции с нашими сторонними сервисами, которые в свою очередь, как и боты, работают через API или связаны напрямую с обработчиком. Создавая логику или сценарий обработчика, надо и не забывать о пользователях, для которых и создается конечный продукт [6].

```
// обязательное. Запуск бота
$bot->command('start', function ($message) use ($bot) {
    $answer = 'Добро пожаловать!';
    $bot->sendMessage($message->getChat()->getId(), $answer);
});

$bot->command('help', function ($message) use ($bot) {
    $answer = 'Команды:
/help - помощь';
    $bot->sendMessage($message->getChat()->getId(), $answer);
});

// запускаем обработку
$bot->run();
```

Рисунок 3 – Базовый пример реализации ввода-вывода запросов на языке PHP

Процесс взаимодействия с ботом должен быть максимально удобным, а также интуитивно понятным. Для этого существуют кастомные клавиатуры, всплывающие сообщения-ссылки и другие инструменты, которые также можно использовать в ходе разработки для улучшения чат-бота.

Из всего вышесказанного можно сделать вывод, что чат-боты действительно становятся очень популярным инструментом для получения разнообразной информации в одном месте. Однако в то, что они могут полностью вытеснить традиционные способы предоставления сервисов (мобильные приложения, веб-приложения), верится с трудом. Поскольку, несмотря на все свое удобство и гибкость, все равно есть привязка к платформе, которая является сторонней, а именно к самому мессенджеру. Несмотря на хорошие возможности кастомизации, у ботов есть предел, который обусловлен самой платформой. К примеру, если пользователь будет получать какие-то конфиденциальные данные, бот не сможет обеспечить должный уровень безопасности, даже если реализовать что-то похожее на процесс аутентификации. Также некоторые приложения используют специфичные методы при разработке сервисов, которые изначально являются непертируемыми. Но, несмотря на все эти минусы, хотелось бы отметить и положительные стороны. Боты намного проще в установке и использовании, чем приложения, легче в распространении, дешевле в разработке и поддержке. Также сама платформа (мессенджеры) используется намного чаще, чем другие приложения, в связи с этим чат-боты отлично подходят для сервисов, которые используют очень часто. Ну и конечно не стоит забывать и про человеческий фактор, то, что люди с рождения предрасположены к разговорам.

Суммируя вышеизложенное, можно с уверенностью сказать, что чат-боты - очень интересный вариант автоматизации с некоторыми ограничениями. При использовании которого нужно учитывать нюансы использования платформы и наборов сервисов.

Список литературы

1. ASUS Russia. Чат-боты: где, как и когда заменить человека [Электронный ресурс]. – 2017. – URL: <https://geektimes.ru/company/asus/blog/289997> (дата обращения 13.01.2018).
2. БИТ. Чат-боты для бизнеса. Рубрика: Технологии бизнеса [Электронный ресурс]. – 2017. – URL: <http://bit.samag.ru/archive/article/1855> (дата обращения 22.01.2018).
3. Документация по REST. Чат-боты [Электронный ресурс]. – 2017. – URL: https://dev.1c-bitrix.ru/rest_help/imbot/index.php?print=Y (дата обращения 25.01.2018).
4. Telegram Bots. Telegram Bot API [Электронный ресурс]. – 2017. – URL: <https://core.telegram.org/bots/api#setwebhook> (дата обращения 31.01.2018).
5. Telegram Bots. Bots: An introduction for developers [Электронный ресурс]. – 2017. – URL: <https://core.telegram.org/bots#inline-mode> (дата обращения 01.02.2018).
6. Любимцева М. Мессенджеры и боты: возможности для продвижения [Электронный ресурс]. – 2016. – URL: <https://spark.ru/startup/546dae36cfa4a/blog/17709/messendzheri-i-boti-vozmozhnosti-dlya-prodvizheniya> (дата обращения 02.02.2018).

Получено 13.08.2018

УДК 004.42

С.К. Кумаргажанова, А.А. Серикпаев

Восточно-Казахстанский государственный технический университет им. Д. Серикбаева,
г. Усть-Каменогорск

МИГРАЦИЯ ДАННЫХ SAP – ЗАГРУЗКА ПОЛЬЗОВАТЕЛЬСКИХ ПОЛЕЙ С ПОМОЩЬЮ СТАНДАРТНОГО ИНСТРУМЕНТА МИГРАЦИИ LSMW

В данной статье рассмотрена задача миграции полей, созданных в рамках конкретного решения системы SAP, с помощью стандартного инструмента загрузки данных Legacy System Migration

Workbench. Возможность переноса данных полей реализуется с помощью использования специальной структуры и использования фрагмента кода на языке программирования ABAP.

The article discusses the task of migrating fields created within a specific SAP system solution using the standard Legacy System Migration Workbench data loader. The ability to transfer field data is realized by using a special structure and using a code fragment in the ABAP programming language.

Бұл мақалада стандартты Legacy System Migration Workbench деректерді енгізу құралы көмегімен SAP жүйесіндегі белгілі бір шешімі негізінде жасалған өрістерді миграциялау тапсырмасы қарастырылады. Арнайы құрылымды қолдану арқылы және ABAP бағдарламалау тіліндегі код фрагменті арқылы өріс деректерін ауыстыру мүмкіндігі жүзеге асырылады.

Ключевые слова: ERP-системы, бизнес-процессы, SAP, миграция данных.

В настоящее время многие предприятия переходят на новые стандарты, усложняются требования бизнес-процессов, увеличивается количество документации и отчетности. Это подразумевает всё больший уровень автоматизации бизнес-процессов посредством ERP-системы. Одним из лидирующих продуктов на рынке систем управления ресурсами предприятия является SAP ERP.

Внедрение SAP – это целый проект с отдельно выделенной командой. В данной статье рассматривается процесс внедрения системы SAP ERP в компании ТОО «Казцинк», в частности задача миграции данных.

Общий план проекта по внедрению выглядит следующим образом (рис 1.):

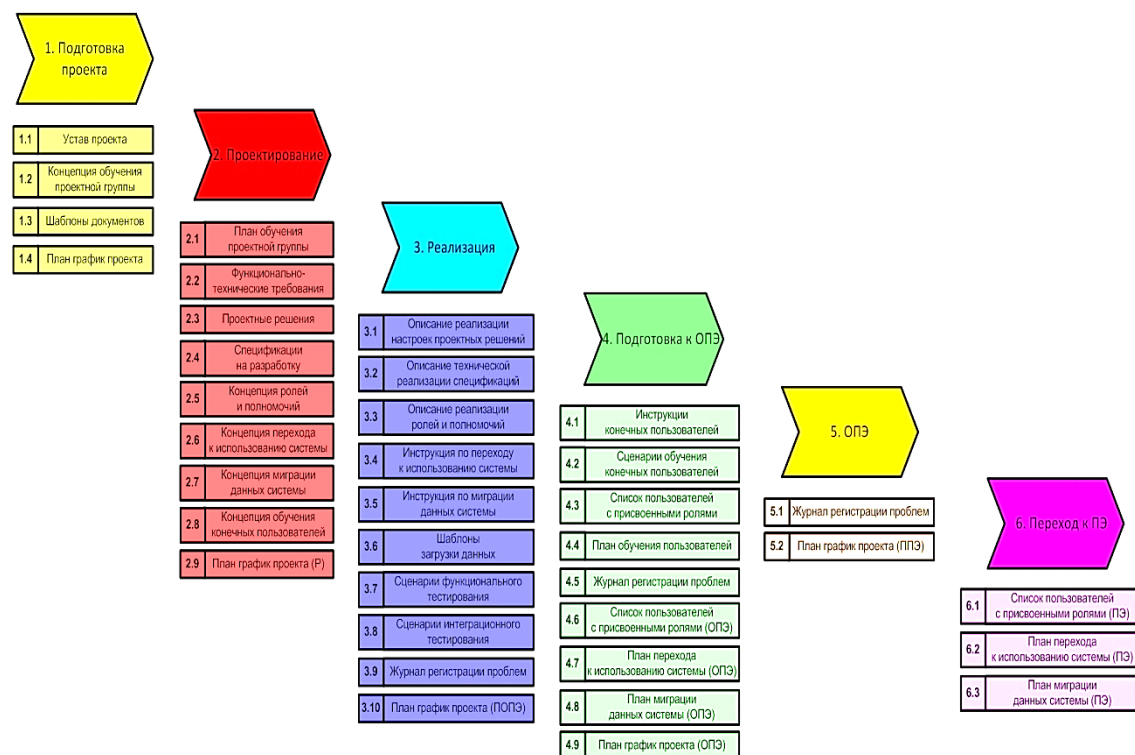


Рисунок 1 - Общая схема внедрения системы SAP

Часть процесса по миграции данных входит в этап реализации (рис. 2, п. 3.5).

Алгоритм миграции данных:

- выявление объектов по каждому направлению, необходимому для миграции;

- составление шаблонов для загрузки с обязательными для бизнес-процесса полями;
- разработка инструментов выгрузки из исторической системы на основе данных шаблонов;
- разработка инструментов загрузки в систему SAP (рассматривается в данной статье);
- тестирование и реализация процесса миграции:
- выгрузка данных из исторической системы;
- проверка выгруженных данных;
- загрузка проверенных данных в систему SAP;
- сверка загруженных данных непосредственно в системе;
- согласование загруженных данных.



Рисунок 2 - Этап реализации

Стандартных решений SAP не всегда хватает для полной реализации бизнес-процессов предприятий. С этой целью система SAP позволяет создать пользовательские поля (далее – Z-поля) в стандартных таблицах, а также полностью пользовательские таблицы с полями в случае необходимости. Стандартные инструменты миграции данных не предусматривают миграцию Z-полей напрямую, однако существует специальная структура,

позволяющая передавать необходимые данные и затем мигрировать их значения в пользовательские поля. В данной статье рассмотрена возможность включения в стандартные инструменты загрузки Z-полей.

В основе предлагаемого метода лежит использование специальной структуры EXTENSIONIN. Связь внутри программы загрузки осуществляется через ссылочную структуру E1BPPAREX. Данная структура имеет 5 текстовых полей: 4 по 240 символов и одно поле 30 символов. Таким образом, всего структура содержит строку из 1000 символов. Данная строка представляет собой последовательную комбинацию значений полей из структуры расширения любой таблицы. Также строка содержит поле с именем структуры, для которой выполнено расширение.

Для работы со стандартным средством существует ограничение – дополнительные (новые) поля таблицы не должны иметь тип P. Однако, даже если бизнес-процесс требует использования именно этого типа данных, предусмотрены расширения обработки данных структуры EXTENSIONIN с помощью BADI. В расширении можно произвести перекодировку цифрового поля в текстовое.

Структура E1BPPAREX представлена следующим образом (табл. 1):

Таблица 1

Структура E1BPPAREX

Поле	Тип	Длина
STRUCTURE	C	30
VALUEPART1	C	240
VALUEPART2	C	240
VALUEPART3	C	240
VALUEPART4	C	240

Поле STRUCTURE содержит имя структуры BAPI, которая в процессе обработки дополняется новыми полями. В зависимости от документа в системе это может быть и структура заголовка, и структура позиций. Также, как правило, необходима передача индикаторов изменяемых полей с помощью дополнительной структуры индикаторов, имеющий символ «X» в конце названия.

Далее рассмотрим конкретный пример заполнения пользовательских полей для заказа клиента в рамках внедрения системы SAP в компании ТОО «Казцинк».

Таблица позиций заказа на поставку расширена следующей структурой полей (табл. 2).

Таблица 2

Структура ZZ_VBAP_ESF

Поле	Тип	Длина	Название
ZZGTDNUMBER	C	20	Номер ГТД
ZZGTDPOS	C	3	Позиция ГТД
ZZSTAWN	C	17	Код ТНВЭД
ZZATTORIGIN	C	2	Признак происхождения

Задача: загрузить в позиции заказов клиента значения данных полей.

В созданном проекте LSMW использован метод загрузки IDOC: SALESORDER_CREATEFROMDAT2.

В качестве структуры передачи значений дополнительных полей используется структура VAPE_VBAP. К ней обязательно заполнение полей структуры VAPE_VBAPX, кото-

рая содержит индикаторы изменения дополнительных полей. Названия данных структур должны быть переданы в поле STRUCTURE структуры E1BPPAREX. Значения полей передаются в соответствующие поля VALUEPART* последовательно.

Для решения представленной задачи в проект необходимо передать следующую структуру данных (табл. 3).

Таблица 3

Значения структуры E1BPPAREX

STRUCTURE	VALUEPART1
BAPE_VBAP	000010009483746578 0019408237543 22
BAPE_VBAPX	000010XXXX

Таким образом, для структуры BAPE_VBAP в поле VALUEPART1 первые 10 символов представляют собой номер заказа, так как мы только создаем заказ, номер не указывается, поскольку заказ ещё не создан. Затем следуют 6 символов, указывающие номер позиции заказа (в данном случае указана первая позиция 000010), затем указывается номер ГТД длиной 20 символов (указывается номер 009483746578 и недостающие пробелы). Далее идут по порядку: позиция ГТД, код ТНВЭД и признак происхождения.

Для структуры BAPE_VBAPX первые шесть символов поля VALUEPART1 также содержат номер позиции заявки, затем следуют пять символов «X», что означает, что заполняются все поля данной структуры.

На рис. 3 показано, как выглядит заполнение полей непосредственно в проекте LSMW.

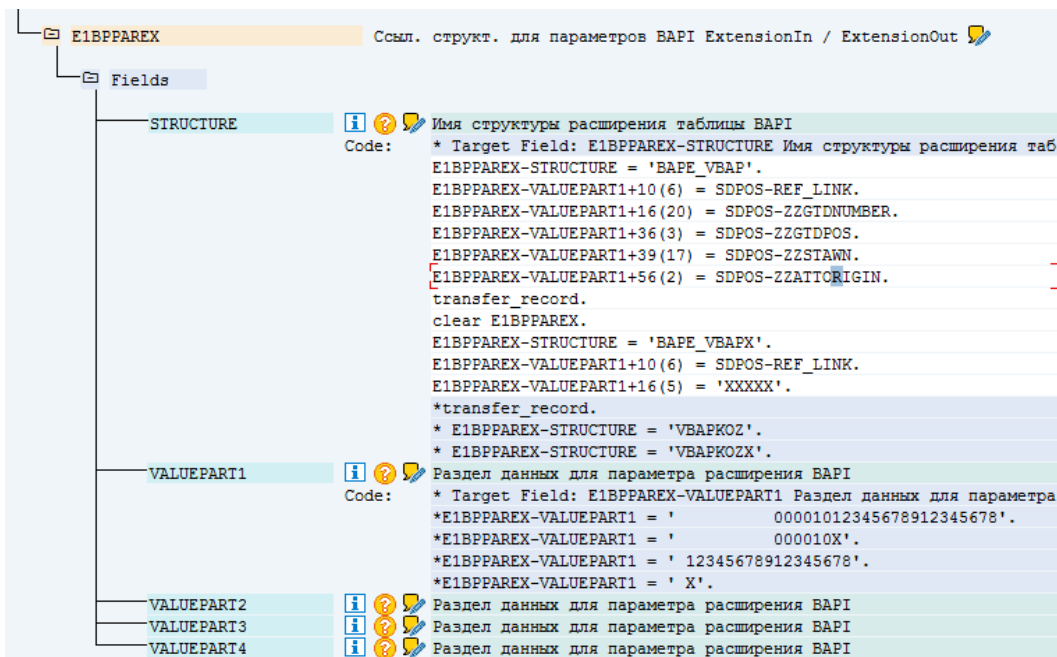


Рисунок 3 - Структура E1BPPAREX в проекте LSMW

Важный момент. Так как структура у нас одна (E1BPPAREX), а нам нужно подать данные для двух структур (BAPE_VBAP и BAPE_VBAPX), используем следующий под-

ход: сначала подаём данные для структуры BAPE_VBAP, затем передаём данные с помощью метода transfer_record, далее очищаем полностью структуру оператором clear и после этого подаём данные для второй структуры BAPE_VBAPX.

В коде, сформированном автоматически при запуске данного инструмента, данный фрагмент выглядит в соответствии с рис. 4 (второй раз данные передаются автоматически).

```
809 |
810 | + --- E1BPPAREX-STRUCTURE
811 | * Target Field: E1BPPAREX-STRUCTURE Имя структуры расширения таблицы BAF
812 | E1BPPAREX-STRUCTURE = 'BAPE_VBAP'.
813 | E1BPPAREX-VALUEPART1+10(6) = SDPOS-REF_LINK.
814 | E1BPPAREX-VALUEPART1+16(20) = SDPOS-ZZGTDNUMBER.
815 | E1BPPAREX-VALUEPART1+36(3) = SDPOS-ZZGTDPOS.
816 | E1BPPAREX-VALUEPART1+39(17) = SDPOS-ZZSTAWN.
817 | E1BPPAREX-VALUEPART1+56(2) = SDPOS-ZZATTORIGIN.
818 | transfer_record.
819 | clear E1BPPAREX.
820 | E1BPPAREX-STRUCTURE = 'BAPE_VBAPX'.
821 | E1BPPAREX-VALUEPART1+10(6) = SDPOS-REF_LINK.
822 | E1BPPAREX-VALUEPART1+16(5) = 'XXXXX'.
823 | *transfer_record.
824 | * E1BPPAREX-STRUCTURE = 'VBAPKOZ'.
825 | * E1BPPAREX-STRUCTURE = 'VBAPKOZX'.
826 |
827 | + --- E1BPPAREX-VALUEPART1
828 | * Target Field: E1BPPAREX-VALUEPART1 Раздел данных для параметра расшире
829 |
830 | *E1BPPAREX-VALUEPART1 = '          00001012345678912345678'.
831 | *E1BPPAREX-VALUEPART1 = '          000010X'.
832 | *E1BPPAREX-VALUEPART1 = ' 12345678912345678'.
833 | *E1BPPAREX-VALUEPART1 = ' X'.
834 |
835 | * --- __END_OF_RECORD__
836 | transfer_record.
```

Рисунок 4 - Фрагмент кода программы конвертации в проекте LSMW

Описанная выше техника позволяет с помощью стандартных средств загрузки загружать значения нестандартных полей различных документов, не прибегая к средствам АВАР-разработки.

Так, например, если требуется заполнить дополнительные поля заголовка и позиций заказа на поставку, следует пользоваться значениями поля STRUCTURE структуры E1BPPAREX: BAPI_TE_MERONEHEADER, BAPI_TE_MERONEHEADERX и BAPI_TE_MEROTEEM, BAPI_TE_MEROTEEMX.

Список литературы

1. Форум консультантов SAP www.sapland.ru.
2. Полезные ресурсы для SAP-консультанта www.saptex.blogspot.ru.
3. Курс BC100 – Введение в программирование (на основе АВАР). - 2000.
4. Курс BC400 – АВАР-инструментальные средства: основы. - 2005.
5. Курс BC402 – Расширенные средства АВАР. - 2009.
6. Статья по загрузке данных с помощью LSMW www.habrahabr.ru/post/215655/.

Получено 13.08.2018

УДК 004.93

Е.В. Рыжкова, О.Е. Бакланова, А.Е. БаклановВосточно-Казахстанский государственный технический университет им. Д. Серикбаева,
г. Усть-Каменогорск**РАЗРАБОТКА И ИССЛЕДОВАНИЕ АЛГОРИТМОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ ВИЗУАЛЬНОГО КОНТРОЛЯ ГЕОМЕТРИЧЕСКИХ ПАРАМЕТРОВ ОБЪЕКТОВ (ОПРЕДЕЛЕНИЯ ГРАНИЦ КОНТУРА ДЕТАЛИ)**

В данной статье рассматриваются методы обработки изображения с применением машинного зрения, возможность замены визуального контроля качества технологических изделий в сферах промышленности, требующих высокоточных измерений.

Бұл мақалада машиналық керуді қолданып суреттерді өңдеу әдістері, өнеркәсіп салаларында жоғары дәлдікті өлшеуді қажет ететін технологиялық бұйымдардың сапасын көзбен бақылауды ауыстыру мүмкіндігі қарастырылады.

This article deals with image processing techniques using machine vision, the ability to replace visual quality control of technological products in industries requiring high-precision measurements.

Ключевые слова: цифровизация, машинное зрение, распознавание изображения, методы обработки изображения, контроль качества.

Государственная программа «Цифровой Казахстан» на 2017-2020 гг. [1] направлена на развитие и цифровые преобразования в отраслях экономики, развитие цифровой индустрии путем автоматизации транспортно-логической системы страны, внедрение цифровых технологий в сфере промышленности; обеспечение доступности цифровой информации, реализацию технологий для создания умных городов. Большое значение цифровизации придает президент РК Н.А. Назарбаев: «Я объявил в своем послании народу Казахстана о третьей модернизации, стержнем которой является цифровизация» [2].

В связи с развитием промышленности в наше время стал актуальным вопрос высокоточной проверки качества продукции с помощью новых технологий. Машинное зрение позволяет сохранить, записать изображения, форму, размеры, местоположение, а также текстуру изделия. С помощью машинного зрения системы визуального контроля имеют высокую скорость работы, возможность 24-часовой работы, точность и надежность измерений. Преимущество машин над человеком заключается в отсутствии усталости, болезней или невнимательности.

Машинное зрение - это непосредственно инженерное направление, к нему относят цифровые устройства ввода-вывода и компьютерные сети, предназначенные для контроля производственного оборудования, роботы-манипуляторы или аппараты для извлечения бракованной продукции. Машинное зрение - это направление в инженерии, связанное с вычислительной техникой, оптикой, машиностроением и промышленной автоматизацией. Одним из самых распространенных применений машинного зрения является инспекция товаров, таких как полупроводниковые чипы [3], автомобили [4], продукты питания [5] и лекарства [6]. Системы контроля машинного зрения для проверки продукции используют цифровые и интеллектуальные камеры, программное обеспечение для обработки изображения и выполнения аналогичных проверок.

Перспективным является использование машинного зрения в медицине, травматологии и стоматологии. Эффективным способом лечения и восстановления целостности костной ткани является эндопротезирование. Имплантаты позволяют сократить сроки лечения при тяжелых заболеваниях и исключить отторжение последних. При изготовлении имплантата непременно важная роль отводится как качеству исходного материала, так и качеству напыления при эндопротезировании. В статье рассмотрено применение системы

контроля качества машинного зрения на примере зубных имплантатов, в данном случае основной частью для контроля является штифт или шуруп, представляющий собой корневую часть, внедряемую в челюстную ткань пациента, от его целостности зависит установка, вживление и дальнейшая эксплуатация. Использование машинного зрения дает следующие возможности.

Система контроля машинного зрения позволит исключить человеческий фактор при визуальном контроле качества изделий, что немаловажно при изготовлении точных изделий, а именно при получении медицинских материалов и изделий. На основе подобранных методов открывается возможность в создании автоматизированного комплекса с применением машинного зрения.

Система контроля машинного зрения может быть использована как технология для оценки геометрических параметров объекта, наличия дефектов на контурах детали, а именно на резьбе, что немаловажно для проведения работ в прикладных материаловедческих исследованиях, металлургии и получении медицинских материалов и изделий.

Применение машинного зрения позволит заменить визуальный контроль качества медицинских имплантатов на машинный, это позволит исключить человеческий фактор при появлении брака, ускорить процесс производства, снизить затраты на оплату труда, тем самым сократить стоимость готового изделия и самое главное, повысить качество продукции.

Нами рассмотрена система контроля качества машинного зрения, позволяющая распознавать дефекты, связанные с геометрическими характеристиками объекта. В частности, цилиндрических деталей и деталей сложной формы (в нашем случае с резьбой).

Существует множество подходов к выделению границ, но практически все можно разделить на две категории: методы, основанные на поиске максимумов, и методы, основанные на поиске нулей. Методы, основанные на поиске максимумов, выделяют границы с помощью вычисления силы края, обычно выражающейся первой производной, такой, как величина градиента, и затем поиска локальных максимумов силы края, используя предполагаемое направление границы, обычно это перпендикуляр к вектору градиента. Методы, основанные на поиске нулей, ищут пересечения оси абсцисс выражения второй производной, обычно нули лапласиана [7] или нули нелинейного дифференциального выражения, как будет описано далее. В качестве шага предобработки к выделению границ практически всегда применяется сглаживание изображения, обычно фильтром Гаусса [8]. Опубликованные методы выделения границ отличаются применяемыми фильтрами сглаживания и способами вычисления силы края. Хотя многие методы выделения границ основываются на вычислении градиента изображения, они отличаются типами фильтров, применяемых для вычисления градиентов в x - и y -направлениях.

В результате проведения наблюдений с точки зрения распознавания изображений наиболее информативным является очертание объектов, то есть их граница.носителем информации является не яркость, а граница объектов на изображении.

Оператор Собеля [9] — дискретный дифференциальный оператор, вычисляющий приближенное значение градиента яркости изображения. Результатом применения оператора Собеля в каждой точке изображения является либо вектор градиента яркости в этой точке, либо его норма. Используется в области обработки изображений, в частности часто применяется в алгоритмах выделения границ. Оператор Собеля основан на свертке изображения небольшими сепарабельными целочисленными фильтрами в вертикальном и горизонтальном направлениях, поэтому его относительно легко вычислять. С другой стороны, используемая им аппроксимация градиента достаточно грубая, особенно это сказывается на высокочастотных колебаниях изображения. Оператор вычисляет градиент яркости изображения в каждой точке. Так находится направление наибольшего увеличе-

ния яркости и величина её изменения в этом направлении. Результат показывает, насколько резко или плавно меняется яркость изображения в каждой точке, а значит, вероятность нахождения точки на грани, а также ориентация границы. На практике вычисление величины изменения яркости (вероятности принадлежности к грани) надёжнее и проще в интерпретации, чем расчёт направления. Математически градиент функции двух переменных для каждой точки изображения (которой и является функция яркости) — двумерный вектор, компонентами которого являются производные яркости изображения по горизонтали и вертикали. В каждой точке изображения градиентный вектор ориентирован в направлении наибольшего увеличения яркости, а его длина соответствует величине изменения яркости. Это означает, что результатом оператора Собеля в точке, лежащей в области постоянной яркости, будет нулевой вектор, а в точке, лежащей на границе областей различной яркости, вектор, пересекающий границу в направлении увеличения яркости. Оператор использует ядра 3×3 , с которыми сворачивают исходное изображение для вычисления приближённых значений производных по горизонтали и по вертикали. Пусть A - это исходное изображение, а G_x и G_y - два изображения, на которых каждая точка содержит приближённые производные по x и по y . Они вычисляются следующим образом:

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * A$$

и

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A,$$

где * обозначает двумерную операцию свертки.

Координата x здесь возрастает направо, а y вниз. В каждой точке изображения приближённое значение величины градиента можно вычислить путём использования полученных приближённых значений производных:

$$G = \sqrt{G_x^2 + G_y^2} \text{ (имеется в виду поэлементно).}$$

Используя эту информацию, мы можем также вычислить направление градиента:

$$\theta = \arctan \frac{G_y}{G_x},$$

где, к примеру, угол θ равен нулю для вертикальной границы, у которой тёмная сторона слева.

Поскольку функция яркости известна только в дискретных точках, мы не можем определить производные до тех пор, пока не положим яркость дифференцируемой функцией, которая проходит через эти точки. С этой дополнительной предпосылкой производную дифференцируемой функции яркости можно вычислить как от функции, с которой взяты замеры - точки изображения. Оказывается, что производные в любой отдельной точке есть функции яркости от всех точек изображения. Однако приближения их производных можно определить с большей или меньшей степенью точности. Оператор Собеля представляет собой более неточное приближение градиента изображения, но он достаточно качественен для практического применения во многих задачах. Точнее, оператор использует значения интенсивности только в окрестности 3×3 каждого пиксела для получения приближения соответствующего градиента изображения и только целочисленные значения весовых коэффициентов яркости для оценки градиента.

Программная реализация оператора Собеля может эффективно использовать SIMD-расширения системы команд современных процессоров (т. н. векторизация кода), при этом выигрыш в скорости вычисления оператора может составлять до пяти раз по сравнению с высокоуровневой реализацией. Ручное кодирование на языке ассемблера позволяет обогнать по скорости такие компиляторы, как Microsoft Visual C++ и Intel C++ Compiler. [10]

Вычисление оператора Собеля элементарно распараллеливается на произвольное число потоков (в пределе каждую точку результирующего изображения можно вычислять независимо от соседних). Например, при наличии двух процессоров (ядер) верхний полукадр изображения может быть обработан одним из них, а нижний - другим.

Практическая реализация использования метода Собеля была произведена в программе Matlab. В качестве объекта был выбран имплантат, представляющий собой шуруп, изготовленный из титана, изображение имплантата было загружено в программу Matlab. После этого осуществилось преобразование изображения в полутоновое изображение (рис. 1). С помощью возможностей библиотеки Matlab был написан код использования преобразования Собеля. В результате работы программы был выделен контур имплантата, по которому можно в дальнейшем определить дефекты, возникающие в результате его изготовления. Полученное изображение контура имплантата показано на рис. 2.

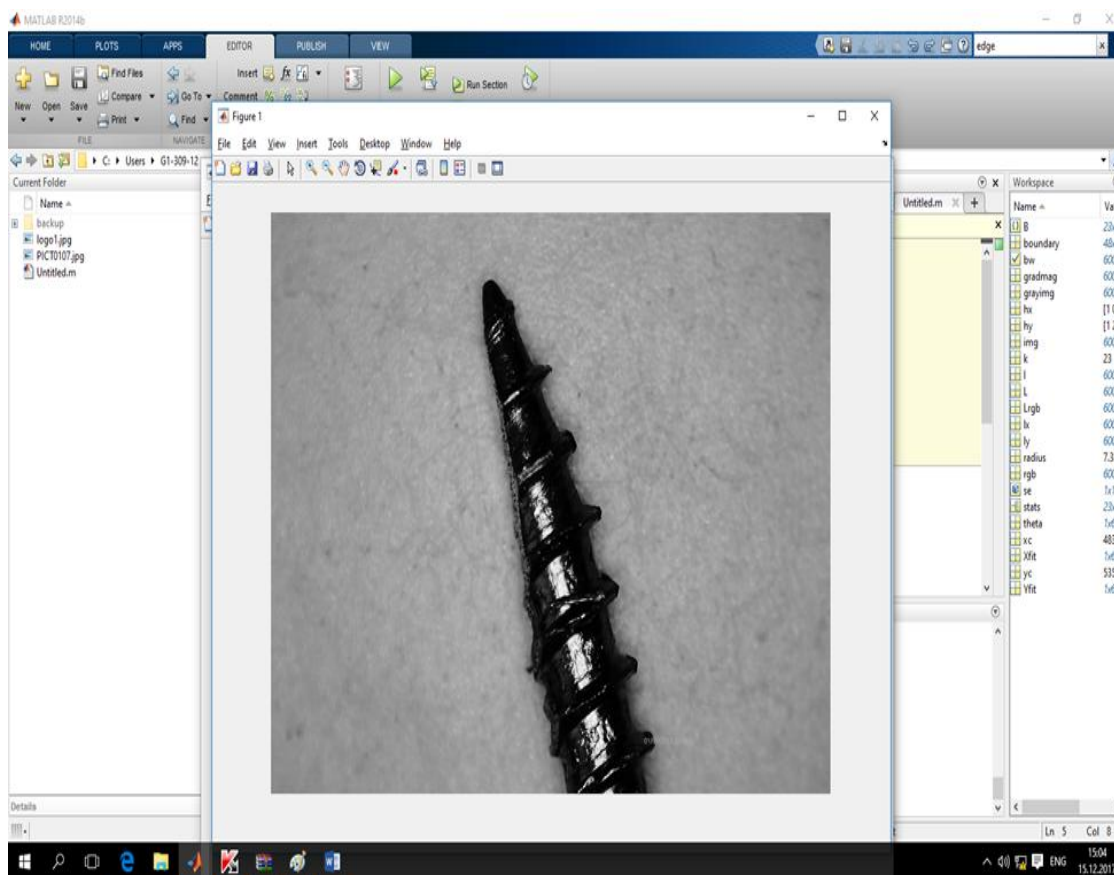


Рисунок 1 - Полутоновое изображение в программе Matlab

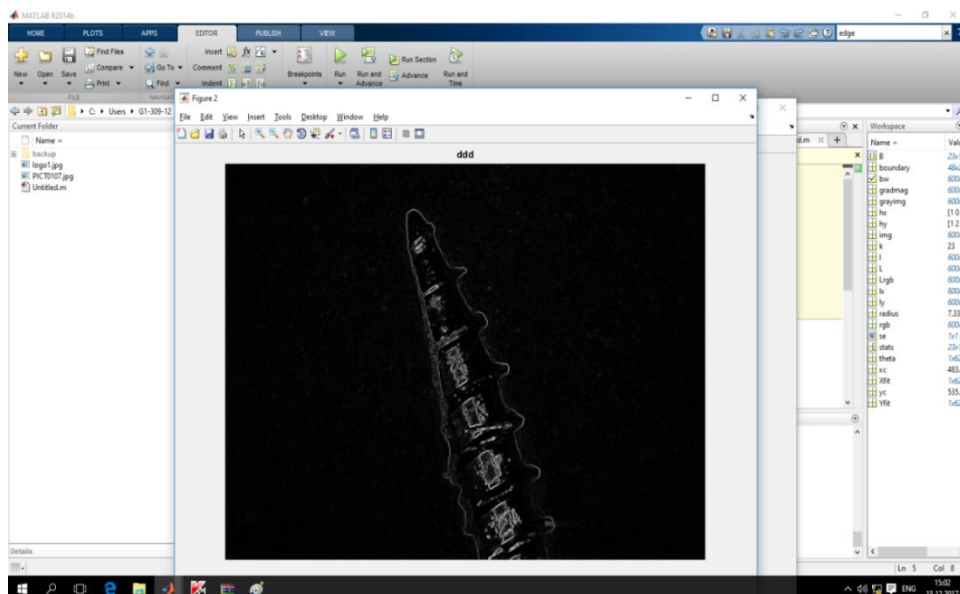


Рисунок 2 - Нормализованный Собелев градиент изображения шурупа в программе Matlab

В нашей статье рассмотрена обработка изображения на базе программы Matlab с использованием компьютера, программы Matlab, системы подобранных методов преобразования изображений. С помощью этих инструментов можно создать библиотеку данных с содержанием эталона продукции, в данном случае штифта, при этом программа будет автоматически отсортировать продукцию, что позволит автоматизировать контроль качества продукции с помощью машинного зрения. Все это позволит автоматизировать процесс контроля качества продукции на стадиях производства.

Список литературы

1. Государственная программа «Цифровой Казахстан» на 2017-2020 гг. [Электронный ресурс]. – Режим доступа: <https://zerde.gov.kz/images>.
2. Выступление главы государства Н.А. Назарбаева [Электронный ресурс]. – Режим доступа: <https://ru.sputniknews.kz/>.
3. Машинное зрение [Электронный ресурс]. – Режим доступа: <http://neuronus.com/>.
4. Машинное зрение: что и как видят автомобили [Электронный ресурс]. – Режим доступа: <https://geektimes.ru/>.
5. Компьютерное зрение в сельском хозяйстве [Электронный ресурс]. – Режим доступа: <http://cosmoport.club/post/>.
6. Обзор задач компьютерного зрения в медицине [Электронный ресурс]. – Режим доступа: <https://habrahabr.ru/>.
7. Дуда Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт; Пер. с англ. Вайнштейна Г.Г., Васильковского А.М. — М.: Мир, 1976. – С. 271-272.
8. Ватутин Э.И. Программная оптимизация оператора Собела с использованием SIMD-расширений процессоров семейства x86 / Э.И. Ватутин, С.Ю. Мирошниченко, В.С. Титов. - Телекоммуникации. - 2006. - № 6. - С. 12—16.
9. Дубков А.А. Преобразование Лапласа: Учеб.-метод. пособие / А.А. Дубков, Н.В. Агудов. - Нижегородский гос. ун-т им. Н.И. Лобачевского. - Нижний Новгород, 2016.
10. Mark S. Nixon and Alberto S. Aguado. Feature Extraction and Image Processing. — Academic Press, 2008. — С. 88.

Получено 13.08.2018

ЭОЖ 004.8

Н.Е. Сауырқанова, Ә.Б. Нұғыманова

С. Аманжолов атындағы Шығыс Қазақстан мемлекеттік университеті, Өскемен қ.

С.С. СмаиловаД. Серікбаев атындағы Шығыс Қазақстан мемлекеттік техникалық университеті,
Өскемен қ.**R ОРТАСЫНЫҢ ВИЗУАЛИЗАЦИЯЛАУ ҚҰРАЛДАРЫ КӨМЕГІМЕН
TF ЖӘНЕ TF-IDF ӨЛШЕМДЕРІН КӨРНЕКІ САЛЫСТЫРУ**

В статье рассматриваются две фундаментальные метрики взвешивания ключевых слов: TF и TF-IDF. На примере корпуса, содержащего 11 документов, анализируются особенности каждой из метрик. Чтобы отличия между метриками были более наглядны, используются инструменты визуализации языка R.

Мақалада кілттік сөздер салмағын анықтаудың екі фундаменталды TF және TF-Idf өлшемдері қарастырылған. 11 құжаттан тұратын корпус мысалында әрбір өлшем ерекшеліктері талданды. Өлшемдер арасындағы артықшылықтарды көрнекі түрде көрсету үшін R тілінің визуализациялау құралдары қолданылды.

This article describes comparison of two fundamental keywords weighting metrics: TF and TF-IDF. Features of each metric are analyzed on the example of a corpus containing 11 documents. To make differences between these metrics more obvious, visualization tools of R language are used.

Пәндік аймақтың мәгіндерінен термин сөздерді алу көптеген қолданбалы есептеулерде, ең алдымен тезаурустар мен онтологиялар сияқты түрлі терминологиялық ресурстарды әзірлеуде және оларды толықтыруда аса маңызға ие. Мұндай ресурстарды қолмен әзірлеу біршама ауқымды жұмыс болғандықтан, соңғы жылдары бұл үрдістерді автоматтандыру бойынша көптеген зерттеулер жүргізілуде.

Термин сөздер белгілі бір пәндік аймақ материалдың мазмұнын ашатын маңызды кілттік сөздерді қамтиды. Кілттік сөздерді шығарудың заманауи әдістері сөздердің түрлі статистикалық және лингвистикалық қасиеттерін қолдануға негізделген. Ұсынылған қасиеттерді 3 топқа жіктеуге болады [1]:

- үміткер сөздердің кездесу жиілігіне негізделген қасиеттер. Бұл топқа TF, DF, TF IDF әдістерін жатқызуға болады;
- контрасты, яғни ортақ тақырыптағы жиынтықтарды қолданатын қасиеттер. Аталмыш топқа қарастырылатын және контрасты мәтіндік жиынтықтағы сөздердің қатыстық жиіліктерін салыстыруға негізделген қатыстық жиілік қасиеті жатқызылады;
- үміткер сөздердің жиілігі мен олардың қолданылу контексті туралы мәліметтерді қамтитын контекстті қасиеттер. Бұл топты сипаттаушылар C-Value және NC-Value қасиеттері болып табылады.

Жиілікке негізделген қасиеттер	Контрасты жиынтықты қолданатын қасиеттер	Контекстті қасиеттер
<ul style="list-style-type: none">• <i>Идеясы:</i> терминдер басқа сөздерге қарағанда жиі кездеседі.• Tf, Df, Tfidf, Domain Consensus, т.б. жатады.	<ul style="list-style-type: none">• <i>Идеясы:</i> мақсатты және контрасты жиынтықтағы терминдер жиілігі ерекшеленеді.• Weiridness, KFIDF, Loglikelihood, т.б. жатады.	<ul style="list-style-type: none">• <i>Идеясы:</i> сөздердің жиілігі мен олардың контексте қолданылуы туралы мәліметтерді байланыстырады.• C-Value, NC-Value, Sum 3, Insideness және т.б. жатады.

1-сурет – Кілттік сөздерді автоматты шығару әдістері негізделген қасиеттер

Эксперименттік жұмыста кілттік сөздердің кездесу жиіліктеріне негізделген TF және TF-IDF өлшемдеріне салыстырмалы талдау жүргізілді. TF өлшемі арқылы құжаттағы кез келген сөздің үлес салмағын, яғни маңыздылығын анықтауға болады. TF (term frequency - сөздер жиілігі) өлшемі — сөздің кездесу санының құжаттағы барлық сөз санына қатынасы:

$$TF\ t, d = \frac{n_t}{\sum_k n_k},$$

мұндағы n_t - құжаттағы t терминінің кездесу саны.

TF-IDF (TF және IDF – inverse document frequency) – құжат немесе корпус жиынтығының бір бөлігі болып табылатын құжат контекстінде сөздің маңыздылығын анықтауда қолданылатын статистикалық өлшем. Бұл өлшем мәтінді талдау мен ақпараттық іздеу мәселелерін шешуде кеңінен қолданылады. Өлшем формуласы төмендегідей:

$$TF - IDF = TF\ t, d \times IDF\ t, D = \frac{n_t}{\sum_k n_k} \times \log \frac{D}{d_i \in D\ t \in d_i},$$

мұндағы n_t - сөздің кездесу саны, $\sum_k n_k$ - берілген құжаттағы жалпы сөз саны, $|D|$ - жиынтықтағы құжат саны, $\{|d_i \in D | t \in d_i\}$ - t кездесетін D жиынтығындағы құжаттар саны (егер $n_t \neq 0$ болса).

TF-IDF өлшемі бойынша белгілі бір құжатта жоғары жиілікпен кездесетін және өзге құжаттарда қолданылу жиілігі төмен термин маңыздылыққа ие.

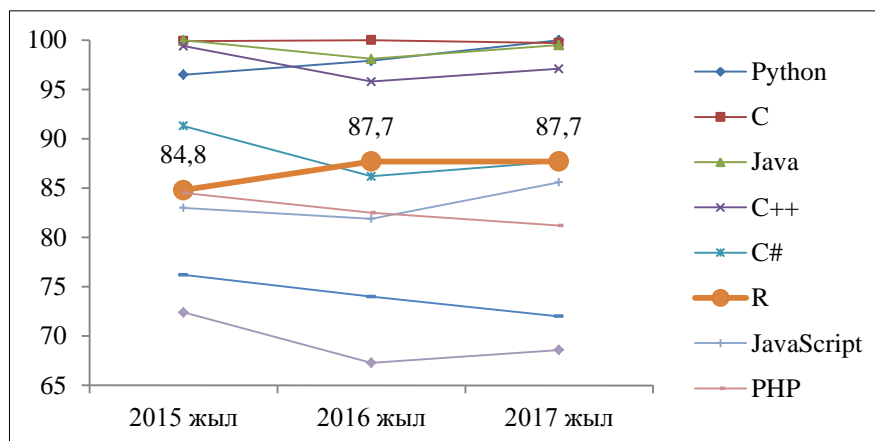
Эксперименттік жұмыста TF және TF-IDF өлшемдері бойынша құжаттың кілттік сөздерді шығару және нәтижелерін R ортасының визуализациялау әдістері көмегімен көрнекі түрде салыстыру әрекеттері жасалынды. Зерттеуге 11 тараудан тұратын Жозеф Новак пен Альберто Канастың ағылшын тіліндегі «The Theory Underlying Concept Maps and How to Construct and Use Them» атты жұмыстары алынды [2]. Зерттеуге алынған құжат концепт-карта теориясының негіздерін, сонымен қатар оны қалай құру туралы және қолдану аясы жайлы мәліметтерді қамтиды. Құжат тарауларының мағыналық тақырыптары бойынша 11 құжатқа бөлініп, жеке *.txt файлдар түрінде сақталды.

Эксперименттік жұмысты жүзеге асыру үшін статистикалық мәліметтерді талдауда, нәтижелерді анықтауда көптеген мүмкіндіктерге ие R статистикалық платформасы таңдалды, себебі R бағдарламалық ортасы ақысыз таратылады, 7000-нан аса пакеттері бола отырып, пайдаланушының жаңа пакеттерді құруына мүмкіндік береді және сапалы визуализациялау құралдарымен жабдықталған, сонымен қатар құжаттық, техникалық жабдыкталуы үнемі жаңартылып отырады.

Бүгінгі таңда R ортасы еркін таратылатын статистикалық талдау жүйелері арасында алдыңғы орын алады. IEEE Spectrum журналы жыл сайын танымал бағдарламалау тілдерінің рейтингісін жариялап отырады [3]. Төмендегі суретте соңғы үш жыл нәтижесінде анықталған бағдарламалау тілдерінің үздік ондығы берілген.

Жоғарыда берілген статистикаға көз жүгіртсек, R тілі IEEE Spectrum журналы жыл сайын жүргізетін танымал бағдарламалау тілдерінің рейтингісінде 2015-2016 жылдары 5-ші орында болса, өткен 2017 жылы үздік алтылықты аяқтады.

R ортасы талдау жүргізу таңдап алынған соң, қарастырылатын құжаттар «корпусы» тұрғызылды, яғни алынған 11 құжат бір объектіге біріктірілді. Зерттеуде R тілінің мәтіндік мәліметтерімен жұмыс жасауға арналған `library(tm)`, `library(RTextTools)`, құжат деңгейіндегі айнымалылары бар `library(quanteda)` пакеті және графикалық объектілермен жұмыс жасауға арналған `library(ggplot2)` пакеті қолданылды. Quanteda пакеті корпус мәтіндерімен жеңіл әрі жылдам орындалатын табиғи тілді өңдеу амалдарын жүргізуге мүмкіндік беретін көптеген құралдармен жабдықталған [4].



2-сурет - IEEE Spectrum журналы жүргізген танымал бағдарламалау тілдерінің 2015-2017 жылғы рейтингісі

Алдымен корпустағы тыныс белгілерді алып тастау, артық бос орындар мен сандарды жою, мәтінді төменгі регистрге көшіру, ағылшын тіліндегі жалпы стоп-сөздерді алып тастау сияқты жұмыстар жүргізіледі. Корпусты өңдеу жұмыстарынан кейін TF-IDF өлшемі арқылы құжаттың кілттік сөздері алынды.

R ортасында TF-IDF мәнін есептеу үшін `weightTfidf(tdm)` функциясы қолданылады. Ең алдымен тұрғызылған корпус `TermDocumentMatrix` функциясы арқылы термин-құжат матрицасына түрлендіріп алынды. Содан кейін «термин жиілігі-кері құжат жиілігі» бойынша термин-құжат матрицасының кілттік сөздер салмағын анықтайтын `WeightTfidf(tdm)` функциясы есептелді. Есептеу орындалған бағдарлама коды:

```
tdm <- TermDocumentMatrix(corpus)
x<-weightTfidf(tdm)
y <- as.data.frame(as.matrix(x[,1]))
y<-cbind(rownames(y),y)
colnames(y)<- c("term","tfidf")
z<-y[order(-y$tfidf),]
head(z,10)
write.csv2(z, file = "D:/Dis/TFIDF10.csv")
```

Жұмыстың нәтижесінде әрбір құжаттың кілттік сөздері алынды. Мақалада 11 құжаттың үшеуі бойынша алынған термин сөздер ұсынылған. Алғашқы «SmartTools бағдарламалық қамсыздандыру құралы» атты құжаттың қысқаша мазмұны келесідегідей: SmartTools бағдарламасының құралдар жиынтығы, кез келген ресурсты интернет желісі арқылы іздеу және оған қатынай алу, бірлесіп топта оқуға және қашықтықтан оқуға мүмкіндік алу, әзірлеген карталарды SmartServer-де сақтау және т.б туралы ақпаратты қамтиды. Осы құжаттың термин сөздерінің тізімі 1-кестеде берілген.

Ал келесі «Бағалауға арналған концептуалды карта» атты құжаттың кілттік сөздері 2-кестеде берілген. Аталмыш құжатта концептуалды картаны қандай да бір білім жүйесінің тұтас бөлімін немесе оның бөлігін меңгергендігін тексеру құралы ретінде пайдаланудың маңыздылығы, алғаш әзірленген карта мен соңғы картаны салыстыра отырып, студенттердің білімдерін бақылай алу жолдары жайлы баяндалады.

1-кесте

Термин (қазақша, ағылшынша)		TF-IDF мәні
интернет	internet	0,030807
smtpservers	smtpservers	0,028889
пайдаланушы	user	0,027393
ресурс	resource	0,027393
байланыс	link	0,024375
топ	group	0,021328
жариялау	publish	0,020538
canas	canas	0,016623
бірлескен	collaborative	0,015653

2-кесте

Термин (қазақша, ағылшынша)		TF-IDF мәні
емтихан	exam	0,076311
ұлттық	national	0,054252
онжылдық	decade	0,050874
баға	evaluation	0,033456
жетістік	achievement	0,027566
тарау	chapter	0,025437
салыстыру	comparison	0,025437
ынталандыру	incentive	0,025437
соңғы	late	0,025437
өту	pass	0,025437

Үшінші «Концептуалды карталар және оқу жоспарлары» атты құжатта оқу бағдарламаларын жоспарлауда концептуалды карталардың маңыздылығы мен оның иерархиялық ұйымдастырылуы оқу материалының бірізділігін анықтауға мүмкіндік беретіндігі баяндалған. Үшінші құжат бойынша анықталған кілттік сөздер тізімін төмендегі 3-кестеден көруге болады.

3-кесте

Термин (қазақша, ағылшынша)		TF-IDF мәні
оқу жоспары	curriculum	0,085298
жоспарлау	planning	0,05999
жоспар	plan	0,056865
есте сақтау	memorize	0,028433
арнайы	specific	0,026301
нұсқаулық	instruction	0,026301
студент	student	0,022615
нұсқау	instructional	0,02167
факультет	faculty	0,02167
бұлыңғыр	blur	0,019997

TF-IDF өлшемі нәтижесінде алынған кілттік сөздер әрбір құжаттың жеке мазмұнын жақсы ашты. Алынған нәтижелерді талдап, саралау мақсатында Quantada пакетінің

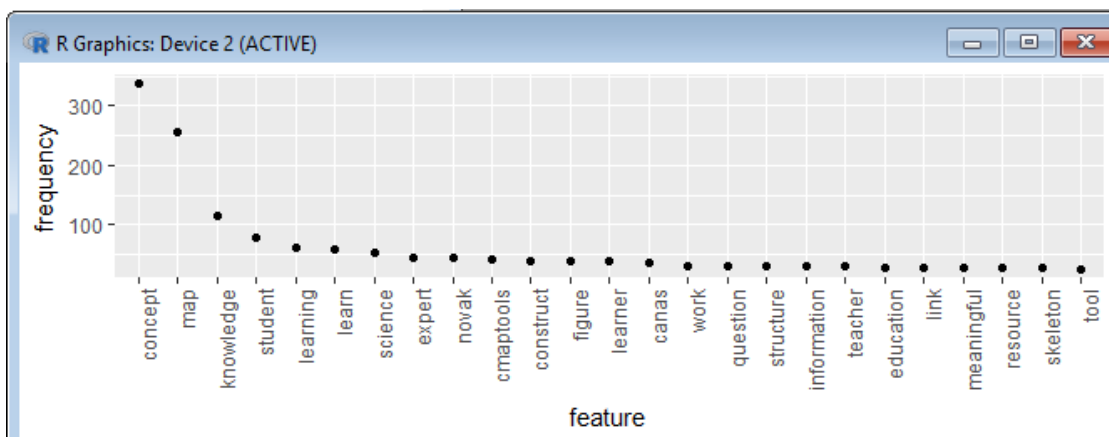
кілттік сөздерді алуда кеңінен қолданылатын функциялары пайдаланылды. Алдымен құжат-қасиет матрицасы `dfm()` тұрғызылды, содан кейін `quanteda` пакетінің `textstat_frequency()` функциясы арқылы барлық 11 құжаттың мазмұнына сәйкес маңызды 25 кілттік сөзі алынды:

```
dfm <- dfm(tokens.Ref)
features_dfm <- textstat_frequency(dfm, n = 25)
```

Алынған кілттік сөздер графикалық объект түрінде визуализациялау құралы `ggplot()` функциясын қолдану арқылы көрсетілді (3-сурет). Бұл функция көмегімен кез келген мәліметтер жайлы статистиканы түсінікті әрі көрнекі түрде көрсетуге болады [5]. Бағдарламалау коды:

```
ggplot(features_dfm, aes(x = feature, y = frequency)) + geom_point() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Жалпы құжаттың маңызды термин сөздерін төмендегі 3-суреттен көруге болады.

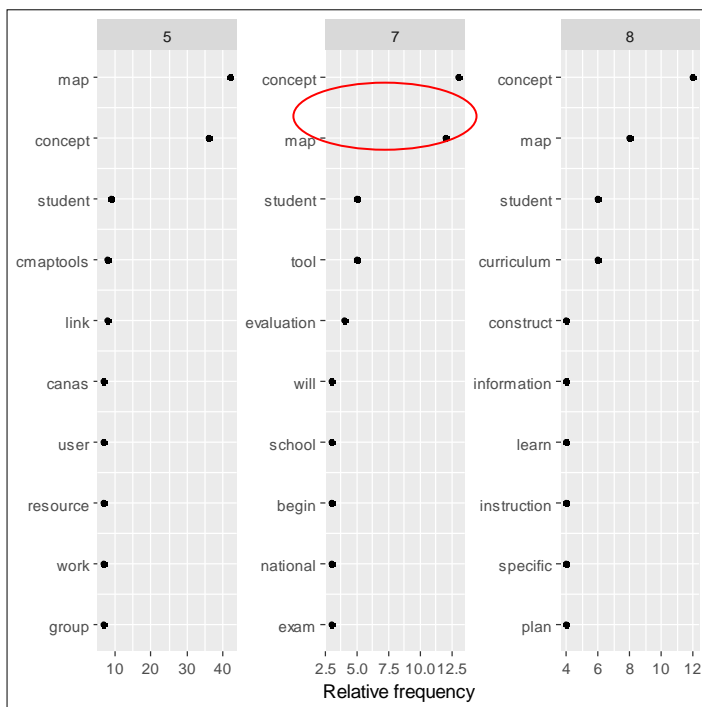


3-сурет – Жалпы корпусстың анықталған кілттік сөздер графигі

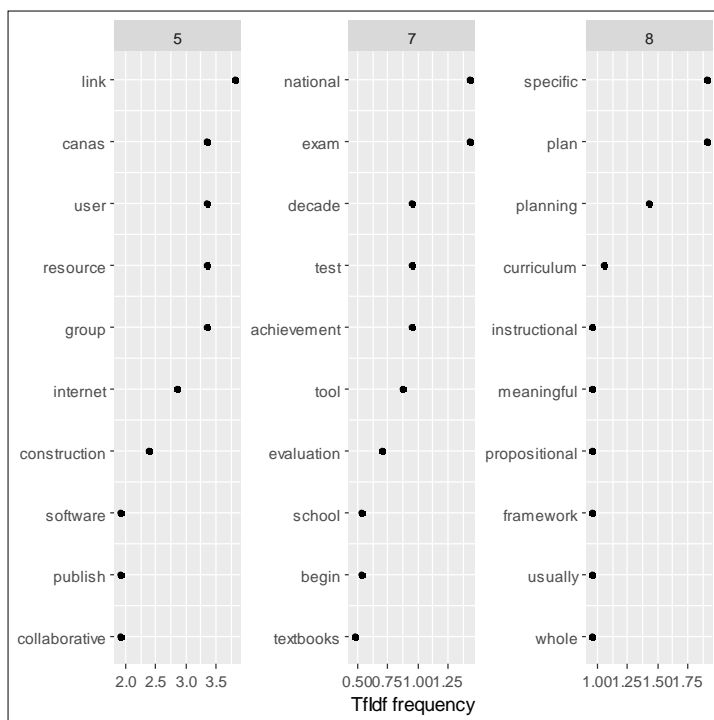
Графиктен жалпы «The Theory Underlying Concept Maps and How to Construct and Use Them» атты жұмыстың қысқаша мазмұнын анықталған кілттік сөздері арқылы байқауымызға болады. Көріп отырғанымыздай, бұл жұмыстың мазмұны концепт, карта, білім, студент, оқу, оқыту, ғылым және т.б. маңызды кілттік сөздерді қамтитын ақпараттан тұрады.

Қарастырылған өлшемдер арқылы алынған кілттік сөздер `ggplot()` функциясы арқылы визуализацияланды (4-сурет және 5-сурет). Екі өлшем нәтижелеріне талдау жасасақ, айырмашылықтардың бар екендігін байқауға болады. 5-суретте TF өлшемі арқылы үш құжаттың жеке-жеке кілттік сөздер тізімі алынған. Байқағанымыздай, 3 құжатта да концепт, карта, студент сөздерінің кездесу жиілігі жоғары болғандықтан, бұл сөздер кілттік сөздер деп алынғандығын байқаймыз, яғни қарастырылған құжаттар мазмұны бір-бірінен аса ерекшеленбеді.

Ал TF-IDF өлшемімен анықталған кілттік сөздер (5-сурет) әрбір құжаттың жеке мазмұнын өте жақсы ашқандай. Мысалы, үшінші құжат тақырыбы «Концептуалды карталар және оқу жоспарлары» екендігін анықтайтын жоспар, жоспарлау, оқу жоспары, нұсқаулық сияқты кілттік сөздер бұл құжатты өзге құжаттардан ерекшелейді.



4-сурет – TF өлшемі арқылы алынған кілттік сөздер



5-сурет – TF-IDF өлшемі арқылы алынған кілттік сөздер

Сонымен талдау нәтижелеріне көз жүгіртсек, кез келген мәтіннің кілттік сөздерін анықтауда TF-IDF статистикалық өлшемін қолдану жақсы нәтиже беретіндігі анықталды. Мәтіннен кілттік сөздерді алуда, ең алдымен, бұл сөздерді не үшін, яғни қандай мақсатта алатындығымызды анықтап алған жөн. Егер қандай да бір нақты мазмұнды құжатты анықтау қажет болса, сөздердің кездесу жиілігіне негізделген TF-IDF өлшемін пайдалануды ұсынуға болады.

Пайдаланылған әдебиеттер

1. Нокель М.А. Использование тематических моделей в извлечении однословных терминов / М.А. Нокель, Н.В. Лукашевич // Программная инженерия. — 2014. — № 3. — С. 34–40.
2. Novak, J. D. , Sañas A. J., The Theory Underlying Concept Maps and
3. How to Construct and Use Them, Technical Report IHMC CmapTools // Institute for Human and Machine Cognition, 2008”, available at: <http://cmap.ihmc.us/docs/pdf/TheoryUnderlyingConceptMaps.pdf>.
4. Cass S. The 2017 Top Programming Languages [Электронды ресурс] // IEEE SPECTRUM: электрон. ғыл. Журнал. - 2017. URL: <https://spectrum.ieee.org/computing/software/the-2017-top-programming-languages>.
5. Welbers K., Van Atteveldt W., Benoit K. Text Analysis in R //Communication Methods and Measures. – 2017. – Т. 11. – №. 4. – С. 245-265.
6. Wickham H. ggplot2: Elegant Graphics for Data Analysis (Use R) [Электронды ресурс] //Second Edition. – 2016. p. 260. ISBN 978-3-319-24277-4 (ebook). URL: <https://books.google.kz/books?id=XgFkDAAAQBAJ&lpg=PR6&hl=ru&pg=PR6#v=onepage&q&f=false>.

Получено 13.08.2018

по страницам



БИОЛОГИЧЕСКИЕ ПЬЕЗОКРИСТАЛЛЫ

Сотрудники университета города Лимерик (Ирландия) показали, что аминокислота бета-глицин, довольно широко распространённая в живых организмах, в кристаллическом виде обладает пьезоэлектрическими свойствами. Если получать её из органических отходов, скажем из древесных опилок, она обойдётся в сто раз дешевле других пьезоэлектрических материалов (кварц, титанат бария), используемых в электронике. При том напряжение, даваемое кристаллами глицина при механической деформации, в 10 раз выше, чем у используемых сейчас пьезоэлектриков.

«Наука и жизнь» № 3, 2018