CONFERENCE PROCEEDINGS

МАТЕРІАЛИ КОНФЕРЕНЦІЇ

winter infoCom 2019

VIII МІЖНАРОДНА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ З ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

Winter InfoCom Advanced Solutions 2019

8th INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGIES

2-3 грудня 2019 року
Україна, Київ

2-3 December 2019
Ukraine, Kyiv

ISBN 978-966-2344-67-7

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ІНСТИТУТ МОДЕРНІЗАЦІЇ ЗМІСТУ ОСВІТИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»

WINTER INFOCOM ADVANCED SOLUTIONS 2019

МАТЕРІАЛИ

VIII МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ

З ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

CONFERENCE PROCEEDINGS

8th SCIENTIFIC AND PRACTICAL CONFERENCE

КИЇВ, УКРАЇНА

2-3 грудня 2019 року

УДК 004

Winter InfoCom 2019: Матеріали VIII Міжнародної науково-практичної конференції з інформаційних систем та технологій, м. Київ, 2-3 грудня 2019 р. – К.: Вид-во ТОВ "Інжиніринг", 2019. – 50с. – Мови укр., рос., англ.

Конференція входить до Переліку міжнародних та всеукраїнських науково-практичних конференцій здобувачів вищої освіти та молодих учених у 2019 році.

Проведення конференції регламентоване наказом ректора КПІ ім. Ігоря Сікорського № 3/612 від 21 листопада 2019 р.

Редакційна колегія дотримується прийнятих міжнародною спільнотою принципів публікаційної етики, відображених, зокрема, в рекомендаціях Комітету з етики наукових публікацій (Committee on Publication Ethics, COPE), а також враховує досвід авторитетних міжнародних видавництв. Щоб уникнути недобросовісної практики в публікаційній діяльності (плагіат, виклад недостовірних відомостей та ін.), з метою забезпечення високої якості наукових публікацій, визнання громадськістю отриманих автором наукових результатів, кожен член редакційної колегії, автор, рецензент, видавець, а також установи, які беруть участь в видавничому процесі, зобов'язані дотримуватися етичних стандартів, норм і правил та вживати всіх можливих заходів для запобігання їх порушень. Дотримання правил етики наукових публікацій усіма учасниками цього процесу сприяє забезпеченню прав авторів на інтелектуальну власність, підвищенню якості видання і виключення можливості неправомірного використання авторських матеріалів в інтересах окремих осіб.

ISBN 978-966-2344-67-7

ПРОГРАМА КОНФЕРЕНЦІЇ

ПРОГРАМА/PROGRAM

2 грудня

Інформаційні системи та технології

Хлівненко М. **Писаренко А.**	Підсистема розпізнавання об'єктів для автомобільних систем автономного керування
Зубрицький А.	Проектування архітектури системи дослідження тексту
Покровський Є. **Савчук О.** **Моргаль О.** **Похиленко О.**	Про моделювання надійності та оцінювання в системі хмарних сервісів
Bodak B. **Doroshenko A.**	The impact and unforeseen challenges of E-procurement systems in Canada

Оброблення інформації у складних системах

Хололович К. **Букасов М.**	Автоматизація пошуку помилок у сирих даних та створення SDTM датасетів для медичних досліджень
Писаренко О. **Дорошенко А.**	Аналіз коментарів за допомогою машинного навчання
Poltorak V.	Analysis of the calculus basis boundary for redundant codes

3 грудня

Інформаційні системи та технології

Kharabet R. **Pysarenko A.**	The use of radio-frequency identification in information systems
Дяченко К. **Писаренко А.**	Інтегрована інформаційна система моніторингу та керування на основі інтернету речей для розумної ферми
Теленик А.	Автоматизована система віддаленої інсталяції програмного забезпечення
Alhawawsha M. **Anisimov A.**	Developing of the E-government System based on Java for Online Voting

Безпека та захист інформації

Романчук С.	Аналіз особливостей державних стандартів ЕЦП на властивостях еліптичних кривих
Калитюк Н.	Автентифікація зображень на основі методів цифрового підпису

ЗМІСТ/CONTENTS

ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ

INFORMATION SYSTEMS AND TECHNOLOGIES

Підсистема розпізнавання об'єктів для автомобільних систем автономного керування

Хлівненко Михайло
КПІ ім. Ігоря Сікорського
Київ, Україна
xlivnenko.michael@gmail.com

Писаренко Андрій
КПІ ім. Ігоря Сікорського
Київ, Україна
andrew.pisarenko@gmail.com

Анотація. **Наведено аналіз існуючих апаратних методів розпізнавання об'єктів для систем автономного керування. Представлено дослідження показників якості алгоритмів розпізнавання машинного навчання. На основі аналізу запропоновано структуру моделі системи розпізнавання об'єктів для автомобільних систем автономного керування.**

Ключові слова: системи автономного керування, алгоритми розпізнавання об'єктів, ACF, YOLO, лідар.

Вступ

Останнім часом, зростає кількість дорожньо-транспортних пригод пов'язаних з пішоходами та за таких обставин, що водій не має змоги вчасно зреагувати на виникнення небезпечних ситуацій [1]. Світові гіганти автомобілебудування приділяють багато уваги безпеці руху, кожного року представляють нові системи, вдосконалюють існуючі. Дані системи аналізують навколишній простір та у разі виникнення небезпечних ситуацій сповіщають про це водія, або самостійно втручаються у процес керування.

З огляду на такий стан речей, для розроблення систем розпізнавання об'єктів використовують сучасні апаратні рішення, що здатні забезпечити усі необхідні умови. Проте, не менш важливою частиною систем є програмне забезпечення. Кожного року представляють нові алгоритми, методи та підходи до машинного навчання, кожен з яких, має свої переваги та недоліки. Проблема, що виникає перед розробниками – це необхідність вибору певного підходу та алгоритму, що забезпечив би при цьому усі висунуті вимоги, при цьому, за час створення, відлагодження, тестування та сертифікації системи такі підходи зазвичай втрачають актуальність.

Отже, постає питання порівняння сучасних методів розпізнавання та на основі аналізу їх показників створення такого підходу, який би був легко інтегрованим, при цьому, мав високі показники якості роботи та мінімальний час навчання.

Огляд існуючих підходів та методів

Вирішуючи проблему виявлення об'єкта та його місцезнаходження, провідні виробники прийшли до розширення можливостей систем за рахунок використання різних апаратних засобів, що значно збільшують кількість необхідної інформації, проте, одразу ж виникає проблема з обробкою такого об'єму даних.

На сьогодні, існує три базових апаратних засоби, що використовуються в автомобільних системах – камери (включаючи тепловізори), лідари та радари.

Камери є універсальним засобом отримання інформації з порівняно невисокою ціною та високою ефективністю. До переваг камер можна віднести невеликий обсяг генерованої інформації та відносно невеликі вимоги до потужності апаратної платформи. Також, камери мають здатність розрізнення кольорів, що позитивно впливає на якість розпізнавання. Досить часто, виробники використовують камери, що мають теплову роздільну здатність – тепловізори. Такий підхід дозволяє ефективно орієнтуватися у темний час доби.

Лідар – це світовий радар, що дає змогу отримувати об'ємну модель простору навколо транспортного засобу. Крім цього, лідар дає змогу оцінювати відстань до об'єктів. Основною проблемою є великий об'єм генерованої інформації, що вимагає великої апаратної потужності для подальшої обробки, а також порівняно висока вартість. На роботу лідару можуть впливати погодні умови, наприклад, дощ та туман [2].

Радар – це сенсорна система, яка використовує радіохвилі для визначення швидкості, дальності та кута об'єкта. Радар використовує невеликий об'єм даних та потребує менше обчислювальної потужності ніж камера. За своєю точністю він поступається лідару, проте, має можливість працювати у несприятливих погодних умовах.

Огляд алгоритмів розпізнавання та їх аналіз

Для розпізнавання об'єктів застосовуються будь-які з доступних класифікаторів. Зазвичай, виробники використовують потужні алгоритми попередньої обробки та будь-який з доступних класифікаторі. Часто в якості класифікатора використовують нейронні мережі. В більшості розглянутих систем – це Support Vector Machine (SVM) [3], причому ні в одній роботі немає аргументації, чому слід застосовувати саме цей тип нейронних мереж. Порівняльний аналіз різних класифікаторів також відсутній.

Однак, такий підхід не є досконалим, тому постає

проблема створення універсального засобу для розпізнавання.

Для аналізу алгоритмів, було реалізовано програмні моделі детекторів за допомогою MATLAB. Для аналізу було обрано наступні алгоритми: ACF, R-CNN [4], Fast-RCNN [5], Faster-RCNN [6] та YOLO [7]. Такий вибір зумовлений можливістю порівняти двохетапні (використовує певний фрагмент зображення) та одноетапні детектори, а також легкістю їх реалізації та відносно високою ефективністю.

Для тренування було обрано найбільший з сьогодні існуючих датасетів – Caltech Pedestrian [8], що містить понад 10 годин реальних відео.

Експерименти використовують наступну процедуру оцінювання. Трекери ініціалізуються в першому кадрі відеопослідовності і відслідковують об'єкт, що нас цікавить до кінця. Потім отриману траєкторію порівнюють із еталонною використовуючи такі параметри: тривалість треку та точність. У таблиці 1 представлено показники якості для тренованих моделей з використанням різних алгоритмів.

Таблиця 1

Показники якості

Алгоритм	**Показник**	
	Тривалість треку	***Точність***
ACF	0.96	0.88
R-CNN	0.95	0.15
Fast R-CNN	1	0.29
Faster R-CNN	1	0.33
YOLO	1	0.77

Також, було оцінено швидкість обробки кадрів, результати показано на рис. 1.

Аналізуючи отримані результати, можна сказати, що кращим варіантом для використання в системах

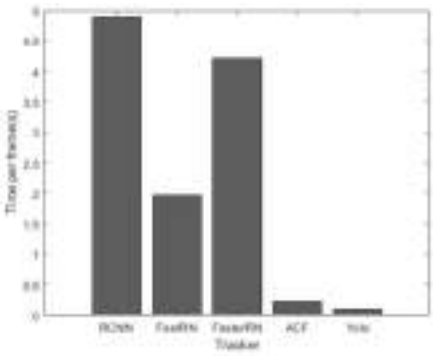

Рис. 1. Швидкість обробки у кадрах/секунду

розпізнавання реального часу є алгоритм YOLO(You only look once), адже він має найвищу швидкість роботи та досить високу точність.

Структура системи

З огляду на представлену інформацію, найкращим варіантом побудови системи є використання камери з лідаром. Такий підхід забезпечить високу ефективність розпізнавання об’єктів та дасть смогу побудувати об’ємну модель навколишнього простору, яка робить можливим більш точно визначати координати об’єктів, їх положення та швидкість відносно автомобіля. У якості алгоритму для розпізнавання об’єктів на зображеннях, отриманих з камери використовується YOLO.

Висновки

На основі аналізу існуючих методів та алгоритмів розпізнавання пішоходів було запропоновано модель системи розпізнавання об’єктів для автомобільних систем автономного керування в основі якої поєднано використання камери та лідару. У якості алгоритму розпізнавання використовується YOLO. Для передбачення траєкторії руху об’єкта використовується фільтр Калмана.

Література

1. Страховий інститут безпеки дорожнього руху [Електронний ресурс] // IIHS. – 2019. – Режим доступу до ресурсу: https://www.iihs.org.

2. Volz, B.; Behrendt, K.; Mielenz, H.; Gilitschenski, I.; Siegwart, R.; Nieto, J. A data-driven approach for pedestrian intention estimation. In Proceedings of the International Conference on Intelligent Transportation Systems. IEEE, Rio de Janeiro, Brazil, 1–4 November 2016; pp. 2607–2612, doi:10.1109/ITSC.2016.7795975.

3. Sun, W.; Zhu, S.; Ju, X.; Wang, D. Deep learning based pedestrian detection. In Proceedings of the Chinese Control And Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 1007–1011.

4. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the EEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 24–27 June 2014; pp. 580–587.

5. Li, X.; Li, L.; Flohr, F.; Wang, J.; Xiong, H.; Bernhard, M.; Pan, S.; Gavrila, D.M.; Li, K. A unified framework for concurrent pedestrian and cyclist detection. IEEE Trans. Intell. Transp. Syst. 2017, 18, 269–281, doi:10.1109/TITS.2016.2567418.

6. Brazil, G.; Yin, X.; Liu, X. Illuminating Pedestrians via Simultaneous Detection and Segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4960–4969, doi:10.1109/ICCV.2017.530.

7. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. European Conference on Computer Vision; Springer: Cham, Switzerland, 2016; pp. 21–37, doi:10.1007/978-3-319-46448-0_2.

8. Caltech Pedestrian Detection Benchmark [Електронний ресурс] // Caltech. – 2019. – Режим доступу до ресурсу: http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/

Проектування архітектури системи дослідження тексту

Зубрицький Андрій
КПІ ім. Ігоря Сікорського
Київ, Україна
andreizubritskiy@gmail.com

Анотація. **В роботі запропоновано рішення для *Анотація.* У даній статті запропоновано архітектуру системи для аналізу тексту дослідниками-лінгвістами. На основі порівняльного аналізу існуючих рішень визначено функціонал системи. Зроблені висновки щодо можливостей розширення функціоналу систему іншими видами аналізу.**

Ключові слова: архітектура систем, система для аналізу тексту, проектування, комп'ютерна лінгвістикаю

Вступ

На теперішній час на ринку програмних продуктів існує великий дефіцит систем для аналізу українського тексту дослідниками-лінгвістами. Для виконання своїх задач система повинна забезпечувати обробку тексту на різних рівнях з можливістю комбінування та порівняння результатів різних видів аналізу.

Серед багатьох систем, які було розглянуто можна виділити 3 основні групи аналізу:

- використання методів машинного навчання (прикладом є MonkeyLearn [1])
- статистичні методи (прикладом є NetXtract)
- лінгвістичні методи (прикладом є AOT [2])

Враховуючи проаналізовані програмні засоби спроектована система повинна бути модульною системою, де кожна з основних груп аналізу реалізується в окремому модулі.

Опис основних компонентів системи

Для того, щоб дослідники лінгвісти могли з легкістю використовувати систему вона була реалізована як веб додаток. Таким чином дослідникам лінгвістам не потрібно встановлювати її на своєму комп'ютері, а все що від них вимагається - це наявність веб браузера.

Однією з головних вимог, що висуваються до системи для використання дослідниками, є можливість розширення функціональності за рахунок введення нових методів аналізу та можливості зміни параметрів реалізованих методів.. Серед основних модулів, що реалізовано в системі є:

- модуль аналізу тексту за допомогою методів машинного навчання;
- модуль із статистичними методами аналізу (N-грами);
- модуль лінгвістичного аналізу

У свою чергу модуль лінгвістичного аналізу включає:

– графематичний аналіз
– морфологічний аналіз

При розробці додатку було використано розділення клієнтської та серверної частини, які взаємодіють між собою за допомогою REST API.

Для розробки веб-додатку була обрана мова програмування JavaScript як для клієнтської частини, так і для серверної, що дозволяє писати уніфікований код та робить контракти між сервером та клієнтом більш зручними.

Для розробки на сервері було використано Node JS із використанням фреймворку Express.

На Node JS сервері також було реалізовано server side rendering, що покращує продуктивність системи при її початковому завантаженні. Для реалізації server side rendering було використано фреймворк Nest JS, який містить велику кількість оптимізацій у своїй реалізації.

Також було реалізовано один мікросервіс з API на Python, на якому реалізовано аналіз тексту за допомогою машинного навчання.

У якості бази даних було обрано базу даних MySQL. При цьому робота з базою у системі виконується через ORM під назвою Sequelize. Це дозволяє працювати з даними на вищому абстрактному рівні.

Для клієнтської частини додатку була обрана архітектура Single Page Application. Вона широко використовується у додатках та дозволяє розділити розробку на сервері та на клієнті на окремі частини та уникнути таким чином сильної зв'язності клієнту та серверу [3].

Для роботи зі станом додатку на клієнтській частині була використана архітектура Redux. Однією з її головних переваг є легкість масштабування та знаходження помилок у додатку під час розробки.

Redux дозволяє зробити архітектуру більш прозорою та масштабованою, так як виключає складні перехресні зв'язки між модулями додатку, що призводять до непередбачуваних помилок у системі.

На серверній частині у основі системи лежить подійно-орієнтоване і асинхроне програмування з неблокуючи вводом/виводом [4].

Основні архітектурні компоненти системи представлені на рис. 1.

Сценарії використання системи представлені на рис. 2.

Висновки

Запропоновано архітектуру системи аналізу текстів для лінгвістів-дослідників, що включає основні класи видів аналізів. Система містить базу даних для зберігання корпусів текстів загального

Рис. 1. Основні архітектурні компоненти системи

користування. Розширення корпусу можливе шляхом додавання в особистому кабінеті необхідного тексту дослідником з окремою областю видимості. Розширення функціоналу системи можливе завдяки модульній структурі. Система розрахована насамперед на використання української мови в якості основної при формуванні корпусів текстів та при проведенні їх аналізу.

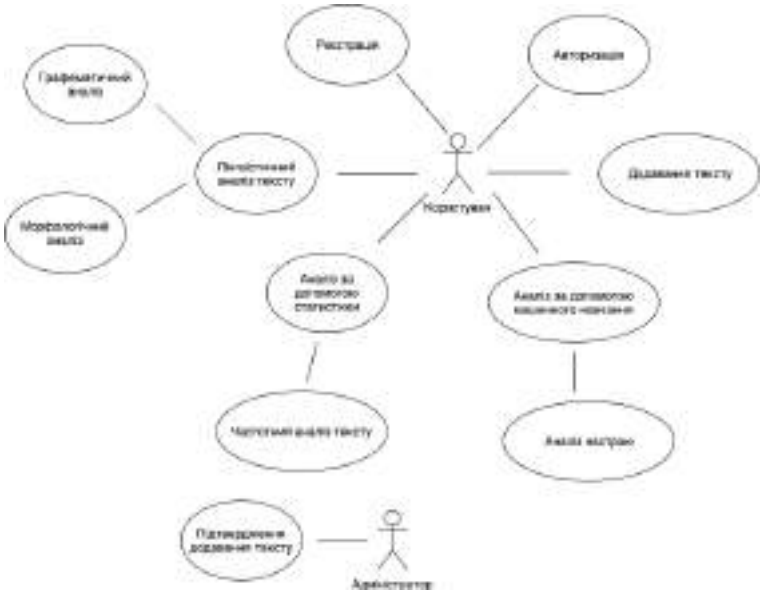

Рис. 2. Основні модулі системи

ЛІТЕРАТУРА

1. Monkey learn [Електронний ресурс]: (Стаття) / Sentiment analysis – Електрон. дан. (1 файл) – 2018. – Режим доступу: https://monkeylearn.com/sentiment-analysis - Назва з екрана

2. AOT [Електронний ресурс]: (Стаття) / Технологии автоматической обработки текста – Електрон. дан. (1 файл) – 2015. – Режим доступу: http://www.aot.ru/technology.html - Назва з екрана

3. Hacker Noon [Електронний ресурс]: (Стаття) / The 4 Layers of Single Page Applications You Need to Know – Електрон. дан. (1 файл) – 2018. – Режим доступу: https://hackernoon.com/architecting-single-page-applications-b842ea633c2e - Назва з екрана

4. Medium [Електронний ресурс]: (Стаття) / Patterns for designing flexible architecture in node.js – Електрон. дан. (1 файл) – 2018. – Режим доступу: https://medium.com/@domagojk/patterns-for-designing-flexible-architecture-in-node-js-cqrs-es-onion-7eb10bbefe17 - Назва з екрана

Рецензент: к.т.н., доц. каф. АСОІУ, КПІ ім. Ігоря Сікорського О.Д. Фіногенов

Про моделювання надійності та оцінювання в системі хмарних сервісів

Покровський Євген
КПІ ім. Ігоря Сікорського
Київ, Україна
ea.pokrovski@gmail.com

Моргаль Олег
КПІ ім. Ігоря Сікорського
Київ, Україна
m_olegm@ukr.net

Савчук Олена
КПІ ім. Ігоря Сікорського
Київ, Україна
savchuk_l1@ukr.net

Похиленко Олександр
КПІ ім. Ігоря Сікорського
Київ, Україна
pokhilenko.alex@gmail.com

Анотація. **Розглянуті деякі проблеми моделювання та оцінки надійності в системі хмарних сервісів. Досліджується модель Маркова для черги запитів з очікуванням. Пропонується підхід до забезпечення безпеки даних в хмарі, що гарантує конфіденційність та безпеку інформації.**

Ключові слова: хмарні сервиси, моделювання надійності, безпека даних у хмарних середовищах.

Вступ

Технологія хмарних обчислень стала популярною альтернативою традиційним обчислювальним технологіям. Ця технологія забезпечує нову концепцію плати за використання корисних обчислювальних ресурсів, що базується переважно на технологіях віртуалізації. Основними перевагами хмарних обчислювальних послуг є: самообслуговування, широкий доступ до мережі, об'єднання ресурсів, швидке масштабування. Незважаючи на ці переваги, широке використання цієї нової технології зіштовхуються з низкою перешкод, включаючи безпеку та конфіденційність. Хмарне обчислення може скоріше реалізувати спільне використання сервісу, ніж лише розподіл ресурсів, який створюється мережевими обчисленнями. Таким чином, хмарне обчислення більш сервісно орієнтоване, ніж орієнтоване на ресурси.

Проблеми

Дослідити модель надійності хмарних обчислень з метою забезпечення зручного, мережевого доступу по запиту до загального пулу обчислювальних ресурсів з мінімальними зусиллями управління з боку постачальника послуг.

Забезпечити захист інформації користувача перед доступом до загального пулу обчислювальних ресурсів.

Визначення вірогідності помилок: тайм-аут та переповнення

Розглядається архітектура типової системи хмарних сервісів [1], що є типовим поданням більшості сучасних або майбутніх систем хмарних сервісів, і надається проста класифікація вище зазначених помилок на дві групи за життєвим циклом:

1. Помилки на етапі запитів: переповнення та тайм-аут.
2. Помилки етапів виконання.

Ці дві групи відмов можна вважати незалежними. Проте, збій у кожній групі сильно корелює. Таким чином, моделювання надійності хмарних сервісів можна розділити на дві частини.

На першому етапі, якщо робочий запит не виконується планувальником до встановленого часу, він буде відкинутий з швидкістю відкидання $\mu_п$. Припускається, що прихід подання заявок на обробку підпорядковується процесу Пуассона з швидкістю прибуття $\lambda_п$.

Нехай загальна кількість S однорідних серверів-планувальників працює одночасно для виконання запитів. Час служби для завершення одного запиту для кожного сервера вважається експоненціально розподіленим з параметром $\mu_з$. Отже, маємо процес Маркова з очікуванням, в якому стан n $(n = 0,1, ..., N)$ являє собою кількість запитів у черзі (рис.1):

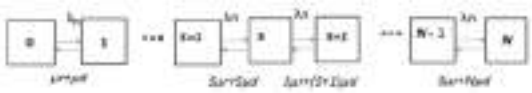

Рис.1. Модель Маркова з очікуванням для черги запитів

Позначимо через q_n стабільну ймовірність того, що система залишатиметься у стані n $(n = 0,1, ..., N)$, де q_n виводиться шляхом вирішення рівнянь Чепмена-Колмогорова. Звідки ймовірність $R_п$, що помилка через переповнення не відбудеться:

$$R_{\text{переповнення}} = \sum_{n=0}^{N-1} q_n \qquad (1)$$

$$\sum_{n=0}^{N} q_n = 1 \qquad (2)$$

де q_n (n = 0,1, ..., N). Якщо $n < S$, то новий запит може бути негайно обслуговий без будь-якого часу очікування. Тому імовірність помилок типу тайм-аут і переповнення не відбувається (тобто стадія запиту є надійною):

$$R_{\text{запиту}} = \sum_{n=0}^{S-1} q_n + \sum_{n=S}^{N-1} q_n \int_0^{T_3} f_n(t)dt \qquad (3)$$

де $f(t)$ щільність ймовірності часу очікування. Сума в (3) між *[0, N-1]* містить умову, що помилка через переповнення не відбувається, як проаналізовано в (1).

Якщо деякий програмний модуль містить помилку, ідентичні «запасні» модулі також будуть містити ту ж помилку. Наступний крок – виправлення помилок системою.

ОЦІНЮВАННЯ ПАРАМЕТРУ РОЗПОДІЛУ ПУАССОНА

Розрахункові вирази для ймовірностей станів системи хмарних сервісів надані згідно [2]. Введемо замість щільностей λ і ν «приведені» щільності:

$$\begin{cases} \lambda/\mu = \lambda m_{t_{ob}} = \alpha \\ \nu/\mu = \nu m_{t_{ob}} = \beta \end{cases} \qquad (4)$$

Параметри α і β означають відповідно середнє число заявок і середнє число доглядів заявки, яка стоїть у черзі, μ – середній час обслуговування однієї заявки.

Очевидно, що при $\beta \to \infty$ отримаємо систему Ерланга з відмовами (заявка миттєво йде з черги). Розглянемо інший крайній випадок: чисту систему з очікуванням ($\beta \to 0$, $t \to \infty$, $\alpha < n$). У такій системі $P_H=0$: кожна заявка рано чи пізно дочекається обслуговування [2].

Середнє число заявок, що знаходяться в черзі, визначається з формули (5) при $\beta \to 0$:

$$m_u = \frac{\frac{a^{n+1}}{nn!(1-a/n)^2}}{\sum_{k=0}^{n} \frac{a^k}{k!} + \frac{a^{n+1}}{n!(n-a)}}. \qquad (5)$$

МОДЕЛЮВАННЯ, ОЦІНЮВАННЯ ТА ЗАХИСТ ІНФОРМАЦІЇ СИСТЕМИ

На етапі виконання можливі відмови через відсутність ресурсів даних, недостатність обчислювальних ресурсів, збій програмного забезпечення, помилку бази даних, апаратні збої та помилки мережі. Застосовуємо логіко-ймовірнісний метод визначення безвідмовності обчислень. Мінімальний зв'язуючий граф підзавдань дає нижню границю імовірності безвідмовної роботи як послідовного з'єднання елементів за нормальним законом.

На етапі виконання сумуються імовірності виконання кожного завдання на попередньому етапі. Наступний крок генерує всі можливі комбінації виявлених критичних елементів за допомогою бінарного пошуку та обчислює ймовірність цих комбінацій. Їх підсумовування є $P(\overline{E_1}, \overline{E_2}, \dots, \overline{E_{j-1}} | E_j)$. Нарешті, якщо хмарне сервісне обслуговування має бути успішно завершено, етапи запиту та виконання повинні бути надійними одночасно.

$$P_{\text{Сервісу}} = P_{\text{запиту}} P_{\text{виконання}} \qquad (6)$$

Захист інформації авторизованого користувача використовує китайську теорему о залишках і реалізує криптосистему з публічним ключем для безпечного обміну даними захищених користувачів. Рішення розроблене на основі модулів, кожен з яких забезпечує набір сервісів, які в основному знаходяться в розпорядженні власника даних [3].

ВИСНОВОК

Розглянута модель хмарних обчислень для забезпечення зручного мережного доступу по запиту до загального пулу обчислювальних ресурсів, що конфігуруються. Пропонується модель Маркова для черги запитів з очікуванням та обчислювання параметрів відмов вирішуванням рівнянь Чепмена–Колмогорова для системи з необмеженим часом.

Запропоновані об'єктно-орієнтований підхід та заходи по забезпеченню безпеки даних в хмарі, що гарантують конфіденційність та безпеку інформації.

ЛІТЕРАТУРА

1. Telenyk S.F, Savchuk O.V., Pocrovskyi E.O., Morgal O.M., Pokhylenko O.A. On reability modeling and evaluating in cloud services system/Artificial Intelligence, 2018, vol.81 (2018'3), pp. 70-80.

2. Советов Б.Я., Яковлев С.А. Моделирование систем: — М.: Высш. шк., 2001. — 343 с.

3. Пирожков О.Ю., Савчук О.В. Інформаційно-орієнтована концепція забезпечення безпеки хмарних обчислень / Інфокомунікаційні системи та технології, вип. №2(2), 2018, с.32-36.

.

The impact and unforeseen challenges of E-procurement systems in Canada

Bodak Bohdan
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
bohdan.bodak@outlook.com

Doroshenko Anatoliy
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
a-y-doroshenko@urk.net

Abstract. **Automation in public procurement sector had become widespread across Europe and North America in mid-1990-s. Canada was among pioneers to implement the competitive process, which aims to create the best opportunities for Canadians, while enhancing transparency, competition, and fairness. Procurement data is available to view and download allowing any supplier to look for opportunities, access all listed tenders, bookmark and share searches. According to studies, the government was able to save up to 15% of budgetary funds on procurement services whilst reducing additional costs and time with the introduction of online system. Therefore, it is crucial to analyze the success and issues in development of an online procurement system in Canada in order to take over the experience and key concepts.**

Keywords: E-government, procurement systems, tenders, goods and services, E-procurement, fair competition, buy and sell.

INTRODUCTION

The implementation of E-procurement systems is a great way to overcome limitations and weaknesses of traditional public administration, such as bureaucracy, inefficiency, lack of flexibility, and corruption. The Government of Canada is among the largest public consumers of goods and services, purchasing goods approximately $22 billion worth every year [1].

The procurement process is regulated on behalf of the Public Works and Government Services Canada (PWGSC), which carries out a major role in helping federal agencies specify their needs and scope. Most contracts awarded to small and medium companies are conducted on a competitive basis in keeping with the Government Contracts Regulations [2], which ensures transparency and fair competition.

OVERVIEW OF THE SYSTEM

Based on a global trend, PWGSC is leading the charge of best practices in facilitating procurement through web with the new Government Electronic Tendering Service (GETS) website. Furthermore, to align with Open Government [3] initiative, any visitor can access the website without restrictions, giving them an unprecedented window into Government of Canada procurement. Potential suppliers can view business opportunities, and obtain all related procurement information.

First, the system utilizes Open Procurement Data Standard (OPDS), which makes it compatible with international procurement markets through Free Trade Agreements.

OPDS consists of five main stages (Fig. 1):

- Planning
- Initiation (Tender)
- Award
- Contract
- Implementation

Fig. 1. OPSD stages

It is worth to mention that Canada-Ukraine Trade Agreement exists between Trade Commissioner Service (TCS) [4] and Prozorro [5], in which former provides latter with assistance in navigating international markets.

Secondly, the E-procurement system using a distributed approach, whereby all information is stored in a central database. However, suppliers and government regulation committees can access and view the information from various platforms using Open-API standard [6]. Fig. 2

displays the standard's architecture including roles, modules, and services.

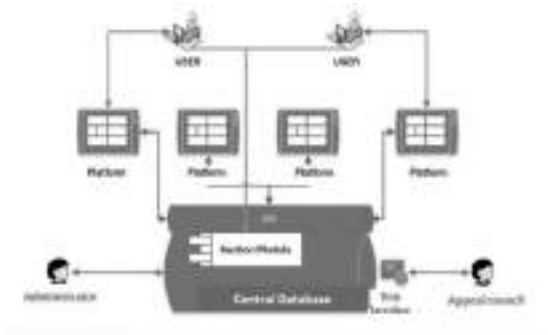

Fig. 2. Open-API architecture

Finally, main participants of the project play a unique role in a so-called 'golden partnership triangle' depicted below. Federal agencies represented by PWGCS are responsible for regulations, agreements, procedure, and guidelines. Meanwhile, enterprises provide government with goods and services. Additionally, a civil society takes care of monitoring and control of the procurement process, since the information is accessible online. This particular way of cooperation has enriched trust among all key stakeholders.

CONCLUSION

The E-procurement systems are a major step towards Open Government initiative and overall automation of public sector. This study focused on a great example of such kind of a system in Canada, describing key concepts and overall architecture. There is no doubt that the E-government paradigm has proved to be a tremendous improvement compared to a traditional model by providing transparency, fair competition, and flexibility. The GETS website serves as a bridge between suppliers and federal agencies, allowing former to look for opportunities and obtain procurement information.

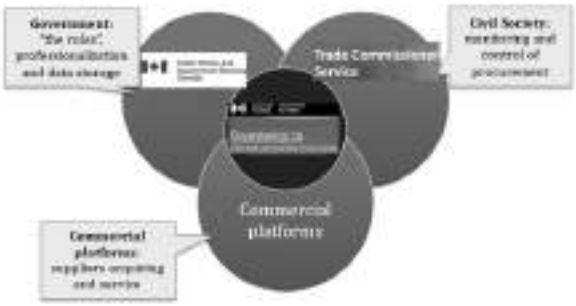

Fig. 3. 'Golder partnership triangle

On the other hand, the article discovers potential issues in E-procurement systems, such as data exchange, monitoring, and regulation process, while giving example solutions in terms of Canadian Government.

REFERENCES

[1] The Procurement Process [Electronic Resource] / Public Works and Government Services Canada – Access mode: https://buyandsell.gc.ca/for-businesses/selling-to-the-government-of-canada/the-procurement-process.

[2] Government Contracts Regulations [Electronic Resource] / Justice Laws Website Canada – Access Mode: https://laws-lois.justie.gc.ca/eng/regulations/SOR-87-402/index.html.

[3] Open Government Standard [Electronic Resource] / Open Canada – Access mode: https://open.canada.ca/en.

[4] Trade Commissioner Service [Electronic Resource] / Trade Comissioner – Acess mode: https://tradecommissioner.gc.ca/en.

[5] Prozorro Government Tenders [Electronic Resource] / Prozorro – Access mode: https://prozorro.gov.ua/en.

[6] Open Contracting Standard [Electronic Resource] / Open Contracting – Access mode: https://standard.open-contracting.org/latest/en/.

[7] Calista, Donald J., James Melitski. "E-government and egovernance: converging constructs of public sector information and communications technologies". University of Alaska. 2007.

[8] Cordella, Antonio. "E-government: towards the e-bureaucratic form?". Journal of Information Technology (2007) 22, 265–274.

[9] Alawneh, A., Al-Refai, H., Batiha, K., 2013. Measuring user satisfaction from E-Government service.

O'Toole, Laurence, and Kenneth I, Hanf. "American Public Administration and Impacts of International Governance". Public Administration Review. 62 (September 2002): 158-169.

The use of radio-frequency identification in information systems

Kharabet Rodion
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine

Andrii Pysarenko
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine

Abstract. **The article describes the use of radio-frequency identification in information systems. It covers the application of RFID in the housekeeping process, in particular, goods monitoring. The existing solutions and their pros and cons were reviewed. To increase the efficiency and convenience of shopping the software and hardware system based on radio-frequency identification was proposed. The article contains an explanation of its workflow and interaction with users and outlines the advantages of this system compared to existing systems.**

Keywords: radio-frequency identification, RFID, information systems, housekeeping, barcode, esp8266, automation.

INTRODUCTION

Information system is a communication system that provides the collection, processing and transmission of information [1].

The Law of Ukraine "On Information Protection in Information and Telecommunication Systems" defines an information (automated) system as an organizational and technical system in which information processing technology using technical and software tools is implemented [2].

Information systems have surrounded us more and more in the last decade. The needs of organizations and users are the main factors that influence the implementation of information systems in various industries and areas of life. Especially this is facilitated by advances in computer technology and telecommunication networks.Information systems are created with the aim of increasing the productivity of everyday processes or reducing the number of unnecessary iterations.

Consider the process of housekeeping in everyday life. Its important component is the provision of housing with necessary resources, such as food and non-food products. A critical factor in this is the awareness of the situation with the resources in the house at the moment. This allows rationalizing purchases, preventing unnecessary waste of money on unnecessary goods. Information systems are created with the purpose to increase the productivity of shopping. The simplest example is a shopping list.

OVERVIEW OF EXISTING SOLUTIONS

The shopping list can be created either manually or automatically using special devices. One such device is Hiku [3]. This is a smart device that can receive voice commands to create a shopping list. Hiku has the ability to scan product barcodes with a built-in scanner in addition to voice control. Once the barcode is read, the corresponding item will be added to the shopping list. This method of use is very convenient for everyday foods such as bread, milk, etc.

Similar in functionality is GeniCan [4] – a device that attaches to the trash can and scans the bar codes of the packaging that is going to be thrown into the trash. Like Hiku, GeniCan automatically adds all the scanned items to the shopping list.

The disadvantage of both of these devices is that using a barcode scanner is not convenient enough because the packaging of the item to be identified may be damaged or otherwise unreadable. There may also be difficulties in positioning the package properly. Additionally, in the case of GeniCan, there is limited space for handling the packaging. It caused by the GeniCan is located directly inside the bin.

An alternative solution for the identification of goods was proposed. The solution is to use radio frequency identification (RFID) jointly with barcode identification. This approach to product identification is more convenient because the speed of reading RFID tags is much faster than reading the barcode. Also, the RFID readers can reach up to 1000 of tag reads per second.

Radio Frequency Identification (RFID) is a form of wireless communication that uses radio waves to identify and track objects [5]. RFID is a generic term that covers identification technologies with different standards. These include NFC and RAIN, two technologies that are the most common among others.

The radio frequency identification subsystem consists of three main elements:

1. An item that has an RFID tag that uniquely identifies the item.
2. A device that provides wireless bidirectional communication between the items described above.
3. Software that collects and transforms data from transmitters, providing real-time information to the software above.

THE PROPOSED SOLUTION

The above-mentioned RFID technology has been applied to software and hardware system, which aims to automate the household process. A software part was implemented in the form of a web application that contains current information about the available goods in the house and their quantity.

Because the use of RFID imposes an additional cost on the user, the barcode identification option has been retained for ease of use. The user can use the web camera to add or remove a product by barcode using your smartphone camera.

The web application was written using the .NET Core 2.1 platform and the ASP.NET Core framework. These

technologies are cross-platform, which allows deploying the application on a server with any operating system: both Windows and Unix-based.

The device with a hardware barcode scanner clings to the recycle bin. It removes goods from the system by a scanned barcode on the package by analogy with existing solutions. The device built with a 1D barcode scanner, a diode distance sensor, and an ESP8266 Wi-Fi module.

The distance sensor is required to activate the scanner only when an obstacle (i.e. package with barcode) appears in front of it. This approach saves energy and does not create the discomfort of constantly active LEDs emitted by the scanner.

Once the barcode is read, the data is transmitted from the scanner to the ESP8266 where it is processed and transmitted to the server by the network.

The RFID subsystem includes two devices. With the first of these, the user adds an RFID tag to the system, while the other device automatically removes the item that was thrown into the trash. Both devices consist of a MFRC522 RFID reader and an ESP8266 WiFi module that transmits the read information to a web application server. The structure diagram of the whole system is shown in Fig 1.

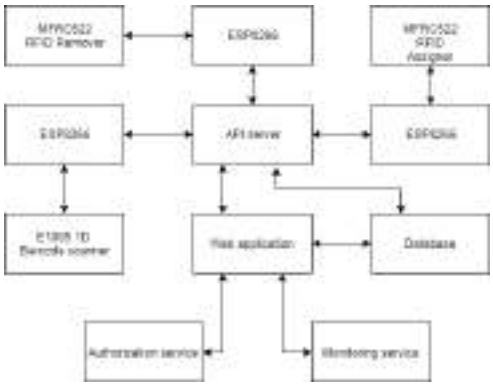

Fig. 1. The structure diagram of the software and hardware system

Let's look at the operation of each device separately.

The first device aims to add new products to the system. The RFID subsystem assumes that as RFID tags Mifare 1k standard tags are used. To add a product to the system, the user scans the tag by bringing it to the RFID scanner. The scanner sends the received tag ID to the server and marks the tag as free to associate with the product. The next step is to attach the tag to the product or its package and select the product from the web application. The user has to clicks the "bind RFID" button. This tag now identifies a specific product in the web application's product dashboard. The UML sequence diagram for adding new product using RFID is shown in Fig 2.

The second device is mounted to a trash bin and aims to reads RFID tags from packages or items that are thrown into bin. When the tag has been read, data is sent to the server that contains its identifier. Then the item associated with this identifier is marked as discarded. These changes are visible on the main screen of the web application, where all currently available products are displayed.

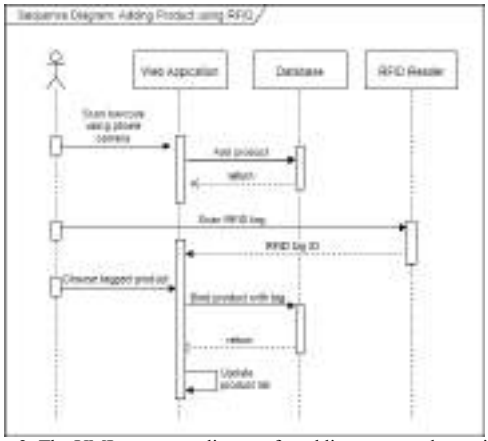

Fig. 2. The UML sequence diagram for adding new product using RFID

Since RFID technology supports quite a large tag reading distance, depending on the tags purchased by the user, it is possible to throw products as simple as it could be. The RFID reader will identify the tag on the fly. The larger the area of the antenna on the tag and the more powerful the emission of radio waves from the reader, the larger the possible distance of successful tag reading.

CONCLUSIONS

The proposed solution describes the use of radio frequency identification in information systems. In particular, the application of RFID in the case of housekeeping is considered. The advantages of using this technology in comparison with existing solutions are described. The hardware and software automation complex for housekeeping using radio frequency identification is considered.

The scheme and characteristics of the system are presented. The principle of action and the components of the system are described. In accordance with the description, a working prototype was created. Currently, the prototype is being tested. But general tests have confirmed the advantages of the proposed approach in comparison with existing solutions.

REFERENCES

1. DSTU ISO 5127-2007 Information and documentation. Base terms. – Kyiv: DSSU, 2010. – 240p.

2. Law of Ukraine “On Protection of Information in Automated Systems” from 05.07.994, № 80/94 [Electronic resource] – Access mode for resource: https://zakon3.rada.gov.ua/laws/show/80/94-%D0%B2%D1%80

3. Gearbrain editorial team. Review: Hiku, the smart tool for today’s smart kitchen [Electronic resource] / Gearbrain editorial team – Access mode for resource: https://mediafeed.org/review-hiku-the-smart-tool-for-todays-smart-kitchen/.

4. Losaw J. A Good Trash Talking with the GeniCan [Electronic resource] / Jeremy – Access mode for resource: https://www.inventorsdigest.com/articles/good-trash-talking/.

5. What is RFID and Why is it so Useful? [Electronic resource] – Access mode for resource: https://www.impinj.com/about-rfid/about-rfid/

Інтегрована інформаційна система моніторингу та керування на основі інтернету речей для розумної ферми

Дяченко Каріне
КПІ ім. Ігоря Сікорського
Київ, Україна
diachenkokarine@gmail.com

Писаренко Андрій
КПІ ім. Ігоря Сікорського
Київ, Україна
andrew.pisarenko@gmail.com

Анотація. **Запропоновано інтегровану інформаційну систему моніторингу та керування на основі інтернету речей для розумної ферми та алгоритми, які застосовуються для аналізу великих даних та прогнозування стану здовов'я тварини.**

Ключові слова: розумна ферма, датчики, інтернет речей, моніторинг, керування, алгоритми.

Вступ

Інтернет речей (IoT) почав відігравати головну роль у повсякденному житті, розширюючи наше сприйняття та здатність змінювати навколишнє середовище. З кожним днем все більше об'єктів мають якесь мережеве підключення. Немає жодної сфери нашого життя, яку IoT не торкнеться у наступному десятилітті. Зокрема, агропромислова та сільськогосподарські сфери застосовують IoT як у діагностиці, так і в контролі. Таким чином, дана стаття має на меті застосувати IoT для моніторингу та керування розумною фермою.

Використання IoT призводить до великомасштабних або великих даних, які надають цінну інформацію для користувачів цих технологій. IoT можна використовувати для підтримки та допомоги фермерам у будь-якому виді сільського господарства. Власники ферм можуть використати час, котрий заощаджується, на інші види діяльності, щоб збільшити свою продуктивність і дохід.

За даними журналу Sensors, в найближчому майбутньому ринок натільних девайсів виросте з нинішнього $ 1 млрд до $ 2,5 млрд до 2025 р. Ця тенденція відображає популярність руху точного землеробства, де технологія впроваджена в кожен аспект життя аграрія: трактори обладнані автономними GPS-модулями; логістика на полях розробляється за допомогою дронів; автоматизовані доїльні апарати та інші футуристичні речі перейшли зі сфери наукової фантастики в реальний світ промислового сільського господарства [1].

Тому було вирішено саме розробити інтегровану інформаційну систему моніторингу та керування на основі інтернету речей для розумної ферми. Так само як і для моніторингу врожаю, на розумній фермі є датчики IoT для сільського господарства, які можна застосувати для моніторингу та контролю за здоров'ям тварин. Технологія роботи аналогічна пристроям IoT для догляду за домашніми тваринами. Наприклад, SCR від Allflex та Cowlar використовують розумні сільськогосподарські датчики (нашийники) для надання інформації про температуру, здоров'я, активність та харчування кожної окремої корови, а також колективну інформацію про стадо.

Система складається з трьох основних компонентів – це апаратне забезпечення, система відображення моніторингу фізіологічних та біологічних параметрів, та мобільний додатки, як показано на рис. 1. Тобто архітектура система складається з трьох частин, а саме рівня збору даних про навколишнє середовище, рівня комунікації і даних та рівня додатків. Спочатку відбувається збір даних навколишнього середовища від датчиків. Потім вони транспортуються на сервер, звідки накопичені дані передаються прикладним рівнем для моніторингу та керування.

Посилаючись на рис. 1 перший компонент був розроблений у формі блоку управління. Він призначений для управління пристроями IoT та отримання даних з датчиків. Було вирішено використати саме такі датчики для моніторингу та керування: акселерометр, GPS, датчик сердечного ритму, датчик електропровідності шкіри, термометр, альтиметр та пульсометр.

Детальніше про їх функції [2]:

1. Основна функція акселерометру – підрахунок кількості зроблених кроків тварини. Він також надає дані про положення в просторі і швидкість пересування тварини.
2. GPS дозволяє визначити координати тварини з високою точністю, використовуючи сигнал, який надсилають супутники. GPS модуль дозволяє визначати швидкість пересування та висоту над рівнем моря.
3. Сучасні оптичні датчики сердечного ритму фіксують зміни рівня поглинання світла, які спричинені зміною кількості крові у сосудах. Спеціальний алгоритм на основі цих даних визначає частому сердечного ритму. Найбільш просунуті датчики сердечного ритму наближуються по точності до ЕКГ.
4. Модулі датчику електропровідності шкіри призначені для вимірювання провідності шкіри. Чим більше вологи на шкірі, тим краща її провідність.
5. Навіть елементарний термометр може надати точну оцінку температури шкіри. Інформація про температуру шкіри порівнюється з показами інших датчиків, після чого в систему відображення надаються дані про активність та стан тварини.
6. Альтиметр фіксує підняття та спуски.
7. Пульсометр працює по електрокардіосигналу. Надає інформацію про пульс тварини.

Другий компонент – це система відображення моніторингу фізіологічних та біологічних параметрів. Вона застосовується в системі для управління інформацією в реальному часі з пристроїв IoT. Веб-додаток дозволяє фермеру відслідковувати стан здоров'я тварини та контролювати її потреби. Ці дані будуть проаналізовані для прогнозування потреб тварин у майбутньому.

Третій компонент – це мобільний додаток. Він застосовується в системі для віддаленого моніторингу та контролю стану здоров'я тварини.

даних у вигляді діаграм або анімації для спрощення інтерпретації полегшення розуміння отриманих результатів.

Висновки

IoT було застосовано для розумної ферми, щоб відстежити та прогнозувати стан здоров'я тварини. Було розроблено інтегровану інформаційну систему моніторингу та керування на основі інтернету речей для розумної ферми. Тематика системи є актуальною адже найближчим часом мільярди одиниць

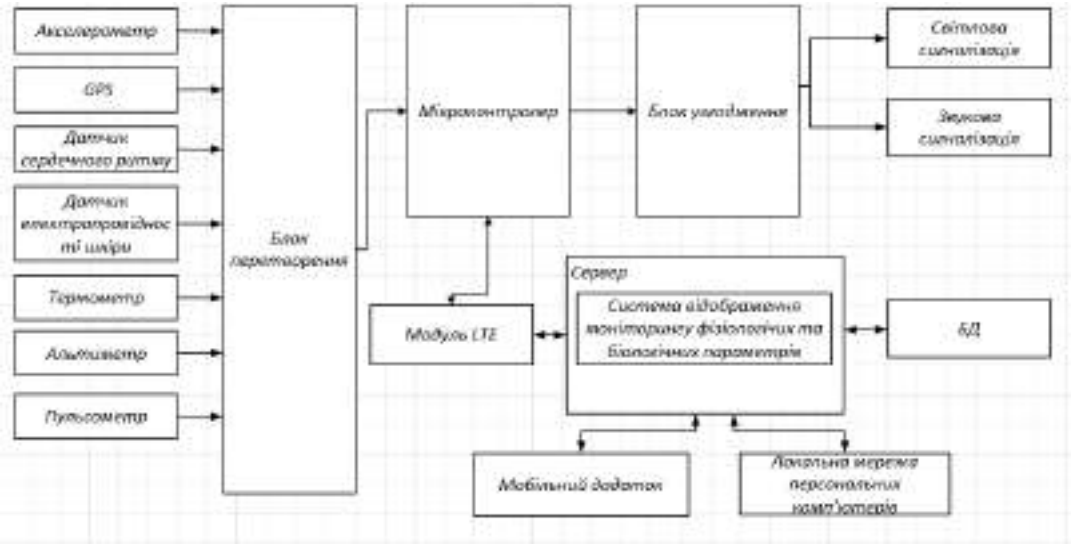

Рис. 1. Схема електрична структурна

Для моніторингу даних використовуються такі алгоритми [3]:

1. Cluster analysis. Статистичний метод класифікації об'єктів по групах за рахунок виявлення наперед невідомих загальних ознак.
2. Crowdsourcing. Методика збору даних з великої кількості джерел.
3. Genetic algorithms. У цій методиці можливі рішення представляють у вигляді «хромосом», які можуть комбінуватися і мутувати. Як і в процесі природної еволюції, виживає найбільш пристосована особина.
4. Візуалізація. Методи графічного представлення результатів аналізу великих

обладнання будуть підключені один до одного - все від найменших побутових речей до розумного міста, де дані будуть зібрані та проаналізовані в режимі реального часу.

Література

1. https://www.business.ua/strategies/4264-vrozhainyi-hadzhet-suchasni-tekhnolohii-na-sluzhbi-u-ahrariiv (дата звернення 05.11.2019)
2. https://ru.mouser.com/applications/smart-agriculture-sensors/ (дата звернення 05.11.2019)
3. https://www.science-education.ru/ru/article/view?id=10922 (дата звернення 06.11.2019)

Автоматизована система віддаленої інсталяції програмного забезпечення

Теленик Андрій
КПІ ім. Ігоря Сікорського
Київ, Україна
andrzej.telenik@gmail.com

***Аннотація*: Мета статті – переглянути наявні способи вирішення проблеми віддаленої інсталяції програм у локальних мережах, їх розгортання та описати бачення автора щодо вирішення проблеми. Найкращим рішенням проблеми буде впровадження системи віддаленого розгортання, кодованої мовою C#, заснованої на принципах об'єктно-орієнтованого програмування та архітектури клієнт-сервер, що дозволить легко встановити програмне забезпечення в комп'ютерній мережі.**

Ключові слова: розгортання програмного забезпечення; IT-інфраструктура; локальна мережа; апгрейд програм; інтерфейс; тиха інсталяція.

З розгортанням нового етапу Науково-Технічної Революції (НТР) відбувається постійне зростання кількості необхідних у науковій, виробничій, навчальній та інших сферах програмних застосунків (продуктів) та різноманіття їх форм. Від файлових менеджерів до графічних редакторів, усе більше і більше програмних застосунків допомогають людям практично усіх професій, у діяльності яких використовується цифрова техніка. Особливо це різноманіття програм стає у пригоді студентам та працівникам IT-сфери, значно полегшуючи процес навчання або роботи та прискорюючи їх. Накладання цих двох факторів і зумовило необхідність виникнення засобу, який б уможливив швидке забезпечення усіх абонентів комп'ютерних мереж різноманітними програмними додатками. Більшість осіб, що використовують комп'ютерні мережі, відчувають потребу в економії часу та сил, які витрачуються на встановлення програмного забезпечення. Ще більшою проблемою є питання синхронізації версій програмного забезпечення, що використовується на підприємстві. Різниця у версіях однієї і тієї ж програми, яку використовують працівники підприємства може призвести до невиправних наслідків через несумісність продукту та розбіжність у можливостях програмного застосунку.

Отже, усі ці чинники підводять користувачів, як корпоративних так і приватних, до необхідності узгоджувати ці дві тенденції – збільшення різноманіття використовуваних програмних додатків та консолідацію комп'ютерів у мережі. Припускаючи існування у користувача окремо двух компонент майбутньої системи – корпоративної мережі з певною топологією, деякою кількістю елементів та комплекту необхідного для роботи програмного забезпечення (ПЗ) - для досягнення цієї мети у користувача фактично є три можливих шляхи дій по встановленню необхідних для роботи програм на комп'ютери-абоненти.

Перший, найпростіший, шлях полягає у встановленні ПЗ на комп'ютери-абоненти вручну з фізичного носія – гнучкого диску, flash-накопичувача, ін. У такому випадку виключена будь-яка помилка мережі через фактичне незастосування мережевого підключення між комп'ютерами для процесу інсталяції. Разом з тим, переваги цього методу не можуть переважити численні недоліки – інсталяція при такому варіанті займає значний час, що збільшується пропорційно кількості комп'ютерів у мережі.

Другий варіант вирішення проблеми полягає у створенні хмарного сховища, у якому адміністратор розміщає стиснуті дані необхідних для інсталяції програм та їх компонентів. Після розміщення усіх матеріалів у «хмарі» користувачі-абоненти мережі завантажують інсталяційні файли та самостійно встановлюють програмне забезпечення на свої машини. Переваги даного методу прямо протилежні недолікам попереднього, але навіть у цьому випадку є «підводне каміння» - такий метод містить вразливості відразу на мережевому, прикладному та транспортному рівнях взаємодії членів мережі. Передача даних з хмарного сховища може постраждати через дію різноманітних завад, від білого шуму до електромагнітних хвиль, розриву з'єднання або помилки пакетування TCP\IP.

Нарешті, третій шлях, наскільки це можливо, компенсує недоліки першого та другого методів. Він полягає у використанні автоматизованої системи мережевої інсталяції – спеціальної програми, яка у режимі прямого підключення до клієнтів по локальній мережі буде проводити встановлення необхідного програмного забезпечення незалежно від наявності чи відсутності контролю з боку адміністратора.

Перш за все потрібно визначити вимоги, яким має відповідати система автоматизованої віддаленої інсталяції. Без сумнівів, така система повинна забезпечувати стійку роботу з усіма абонентами локальної мережі, допускаючи можливість працювати як з окремими комп'ютерами так і з цілими виділеними робочими групами комп'ютерів. По-друге, гіпотетична система інсталяції повинна мати можливість доступу до встановлених на клієнтських комп'ютерах програмних додатків щоб уникнути повторної інсталяції та, як її наслідок, нераціонального використання пам'яті комп'ютерів-абонентів, у випадку якщо необхідний софт все встановлений. Також очевидно, що необхідність швидкої встановки та оновлення програмного забезпечення для всіх абонентів мережі підприємства диктує потребу у можливості інсталяції з готових

пакетів софту, які будуть завантажені по мережі за допомогою протоколу TCP/IP у сумісному з операційною системою клієнтів вигляді. Усе це буде потребувати тісної роботи з мережевим, транспортним та презентаційним рівнями мережевої моделі.

Одним з найбільш досконалих з наявних систем автоматизованої дистанційної інсталяції є програма *Total Software Deployment (TSD).*

Ця програма призначена для управління та розгортання ПЗ в корпоративних мережах. Програмний додаток забезпечує виконання поставлених цілей шляхом сканування локальної мережі на предмет підключених до неї абонентів, перевірки наявності на них необхідного корпоративного програмного забезпечення, перевірки версії цього ПЗ та проведення апгрейду (або встановлення «з нуля») забезпечення згідно з результатом сканування. Потрібно відзначити наявність у *TSD* можливості проводити сканування не лише за конкретними IP-адресами окремих комп'ютерів, а й за робочими групами абонентів.

Автоматичний режим дозволяє визначати політики контролю софту для абонентів та особисто формувати пакети розгортування для тієї чи іншої політики. У даному варіанті програма підготовлює інсталяційний пакет на основі доступу до операційної системи клієнта та реєстрів цієї ОС. Одночасно з швидкодією та високою точністю, даний метод має свої недоліки – так, доступ до реєстрів системи може призвести до небажаних наслідків для безпеки та конфіденційності. На додачу, такий спосіб можна використовувати лише для інсталяції відносно примітивного програмного забезпечення, яке не вимагає встановки драйверів, бібліотек та ін.

Чи не найзручнішим аспектом *TSD* є використання так званої «тихої інсталяції» - способу розгортання (deployment) програмних файлів, які потрібно інсталювати на комп'ютері не перетинаючись з поточною діяльністю оператора комп'ютера-абонента, тобто процес під'єднання та інсталяції йде, керований виключно з комп'ютера-сервера без виникнення зайвих вікон, що можуть відволікти клієнта. *TSD* дає можливість обирати режим «тихої інсталяції» через командну строку за допомогою набору потрібних параметрів, для чого не потрібно навіть запитувати дозволу від користувача клієнтського комп'ютера.

Іншим наявним рішенням є програма *Maestro AutoInstaller.* На відміну від попереднього прикладу, *Maestro AutoInstaller* є програмою що від початку розрахована на дії в рамках однієї ОС та працює по принципу «запам'ятовування» дій користувача у ході інсталяції, створюючи сценарії встановлення програмного забезпечення за його діями та виконуючи потім автоматичне встановлення програм по даним сценаріям. Це означає, що для використання програми як автоматизованої системи дистанційної інсталяції потрібно проводити налаштовування доступу до файлів програми як на сервері, так і на клієнті, та, відповідно, наявності встановленої *Maestro AutoInstaller* однакової версії на обох комп'ютерах. Через це програму не можна повною мірою розглядати як чисту автоматизовану систему дистанційної інсталяції – для використання *Maestro AutoInstaller* потрібно підготовити усі комп'ютери мережі, встановивши на них одинакові версії цієї програми. Суттєвим недоліком цієї програми також є схильність програми вступати в конфлікт з технологію User Access Control (UAC), що використовується в операційних системах Microsoft Windows починаючи з ОС Windows Vista – забезпечення нормальної роботи *Maestro AutoInstaller* потребує спеціального відключення контролю облікових засобів користувачів через панель керування безпекою операційної системи, що тягне за собою загрозу потенційного порушення безпеки доступу до ресурсів Windows.

Таким чином, з огляду на недоліки та переваги розглянутих програм, найкращим варіантом вирішення проблеми синхронізації версій ПЗ буде розробка нової автоматизованої системи дистанційної інсталяції. Цей програмний продукт планується розробити за допомогою стека технологій, що включає у себе засоби програмування .NET реалізовані мовою програмування C#. Даний продукт буде мати клієнт-серверну структуру зі зв'язком між елементами системи, незалежним від топології (працювати як в мережах з P2P-зв'язком, так і в мережах, побудованих за принципом «зірки») та підтримувати «тиху інсталяцію» без конфлікту з основними потоками роботи операційної системи абоненту мережі, що планується реалізувати за допомогою засобів роботи з потоками бібліотеки System.Threading мови програмування C#.

Висновки

У результаті проведених досліджень була визначена проблема синхронізації та централізації управління ПЗ у комп'ютерних мережах, проаналізовані наявні методи її вирішення, описані їх недоліки та переваги. З урахуванням цієї інформації автором були сформовані вимоги до оптимальної системи автоматичної інсталяції та намічений шлях подальшої розробки даної системи.

Література

1. https://uk.wikipedia.org/wiki/Інформаційна_інфраструктура (дата звернення 01.11.2019).
2. https://uk.wikipedia.org/wiki/Інсталяція_ПЗ (дата звернення 01.11.2019).
3. http://maestro-kit.ucoz.ru/ (дата звернення 01.11.2019).
4. http://computerologia.ru/obzor-programmy-total-software-deployment/ (дата звернення 01.11.2019).

Developing of the E-government System based on Java for Online Voting

Alhawawsha Mohammad
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
mhawawsha@gmail.com

Anisimov Anatoly
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine,

Abstract. **The current article is focused on proposing a new electronic voting system that will assist the government in conducting the online voting for the elections. The system will be secure with the end to end encryption to ensure that no data theft occurs during the transfer. The key demerit of the system is secure from the end of the government. It has to be ensured that all the external threats are kept at bay when using the online voting system. The secondary research approach is considered to mine the information required to conduct the study and develop the system. Overall, it can be stated that the new system will allow the government and the public to have improved voting experience if implemented successfully.**

Keywords: Java-based E-government system, online voting, privacy theft, technological adoption, internet, encryption, hackers.

INTRODUCTION

The current paper is concerned with understanding the various aspects of the e-government system. In addition, a new e-government system has been proposed that is likely to support the government in handling the election polling. E-government system refers to the use of electronic mediums to facilitate the services to the public of the country [1].

The e-government system allows the citizen to engage with the government at all levels and along with gaining benefit from the actions of the government, it contributes to the action of the government. With the use of the e-government system, the citizens involved in local and national governance. The information and communication technology are used to facilitate these actions. As per United Nations, the e-government is referred to as the actions of the government that lead to the use of information technology to facilitate the delivery of government services to the citizens [2].

At present, the United Nations conducts an e-government survey twice a year to assess the readiness of the governments of different countries regarding the use of information and technology to conduct their business and cater to the masses. As a model, the e-government system should be able to allow any individual visiting the city or country website to engage in communication with the employees of the local government using the Internet, GUI, and IM. The current paper is concerned with understanding the current e-government system and proposing an additional system that can aid in the current system. The current e-government system is assessed and then an additional online system is proposed. Also, the potential merits and limitations of the proposed online system are identified.

The aim of the research is to come up with a suitable e-government system that can assist the government to improve its voting. The objectives of the research are:

1. To understand the existing solutions available for voting.
2. To propose a new e-government voting system.

RESEARCH PROBLEM

One can witness numerous researches that have been conducted regarding the usage of technologies to facilitate in the elections [3]. Some of these studies have warned against the rapid adaptation of the technology as it can be challenging for the stakeholders to get used to it. However, with improved security measures, the technologies can be a beneficial tool. The elections conducted in various countries use technologies in different ways. For instance, the election conducted in Florida (United States) in 2000 used a punch-card voting system which was later adopted by wide population. This new electronic voting system allows the voters to have a government-issued token or smartcard which they can take to the nearest voting terminal in their locality and then swipe the card. The system also allows them to make any changes if they want by entering pin. This has been considered appreciable in comparison to the traditional method where the voters approach the voting booth to cast their vote by first showing their voter id and proving who they are. On the other hand, the direct-recording electronic (DRE) voting system allows the users to have quick voting experience with faster service. There are only few instances of adoption of this system, as there are still countries where the method of voting is traditional and cumbersome [4].

One of the important aspects to understand here is that the parties in the election can get the benefit of the system if they identify any loopholes to utilize the system flaw. The recent development has found that the governments of few countries have adopted the usage of the DRE without considering or questioning the security of the system which can be a fatal mistake. Here, it is necessary that a voting system that is robust and secure is required to be developed the introduction of end to end encryption can also improve the security of the system. The end to end encryption has become a necessary tool for the current systems due to the increasing security threats [5].

METHODS OF RESEARCH

The method of research used for the current paper is secondary. The researcher referred to various secondary

resources for the collection of information that was relevant to the study. The secondary research allows a desk-based study of the required research area. The study could have been taken further by conducting primary research to gain more input for the required system, however due to the time constraint, secondary research was preferred.

RESEARCH RESULTS

The current chapter assesses the current e-governance system keeping in view the current position of various countries such as the United States, the United Kingdom, and India. In general, there are four models of service delivery from the government. They are from a government to the citizens, from a government to businesses operating in the country, from one department of government to another, and from the government to its employees. The adoption rate of e-government system is different in the different parts of the world. The developed economies such as The United States and the United Kingdom have an appreciable rate of adoption in comparison to developing nations like India. The developed nations also witness improved usage rate of the system by the citizens whereas same can be found far below in the developing nations [6].

The improved usage of the e-government system requires that the citizens of the country are well aware of modern information and communication technology and how to use them. If a government is willing to use the e-government system, then it is important that an education drive should be facilitated that can educate the citizens about the same. This will help them get the intended benefit from the system.

Governments can use information and communication technologies to improve their governance capability. IT can be implemented in almost all aspects of the government activities and all can be coordinated together to provide better service to the masses. Also, the Internet can help the government in establishing better coordination among the departments and improving the internal and external relations.

In general, when the discussion of e-government comes up, people think about the Internet which is appreciable as it is the only thing at present that should be guiding all the technological adoption by the government. However, there also exists e-government which is not based on the Internet. The use of SMS, telephone, wireless networks, smart cards, CCTV, and others are also on the premise of information and technology and thus are also an integral part of the e-government system. However, the definition of the e-government has not changed to something that allows the inclusion of citizens with the government and this can only be facilitated by the use of the Internet [7]. One of the important things, however, to consider is that the use of the e-government system is different in various countries and it impacts the level of equality. The countries, where fewer people are aware of the Internet and technology, such as India face a lack of equality. On one hand where the citizens who are aware of IT, gain the benefit from improved government communication, however, the other part of the citizens who are still unaware of this aspect face challenges.

There are many such people who are homeless, have lower income and cannot learn or purchase IT devices, or are living in the remotest of the locations where there is no IT infrastructure. In such scenarios, it is difficult for the government to provide benefits to these people directly with the use of e-government modules. Here, the elementary step of the government would be to train these people about the new technology and install the right infrastructure [8].

In addition to the above, it is important that the citizens trust the e-government model for whether it will perform as expected. Interestingly, the e-government system is still in developing phase from the most developed economies to the underdeveloped economies. Even countries like the US and UK have not fully adopted e-government mechanism due to various reasons.

PROPOSED E-GOVERNMENT SYSTEM

The e-government system consists of various modules and infrastructures that cater to different aspects of government functioning. The current chapter will propose a new module to the current e-government system that is likely to assist the government in getting more digitalized. The proposed system is an online voting system for the elections. This system is still in a debate that whether a government should adopt this method of voting considering various challenges, such as identity manipulation, data theft, and others.

The information will be stored in the government database and will not be shared with any third party. The current system will store the name of the voter, his unique voterID, password, phone number of the voter, address of the voter, date of birth, and nationality. The aforementioned information will be used to identify particular voters when they will cast their votes. The registration will generate a unique username and password that will be used by the voter to log in to the system using the window.

When users will enter their name (username) and password, then the system will take them to the voting page where the users will see the list of candidates who prefer to be elected in the particular election. The users will select a particular candidate and submit their vote. The system will ask for confirmation from the users regarding their vote and the voting will be completed. The users can then log out of the system after casting their votes. This will be the whole procedure of the new online voting system from the users' (voters) end. In addition to the above, the new system will also allow the admin to view the votes cast in a graphical manner. This will allow easy viewing of the votes. The system will allow the admin to view the data in three human-readable formats for the total casted votes.

Here, the admin will be able to view the parties for whom the vote was cast, a total number of votes cast for each of the parties, and the same value in the form of the percentage. This will allow the admin to understand the winner and the loser. Another format that will be used to show the data is a pie chart and the bar chart.

BENEFITS OF PROPOSED E-GOVERNMENT SYSTEM

The proposed system will allow the government to get votes from the citizens without arranging booths to cast

their votes. This will significantly reduce the cost incurred in conducting voting across the country every few years. Moreover, this system will allow the voters to cast their vote any time of the day they want and that too without lining up at the voting booths.

The secure login id of the voters will be used by them to log into the system and cast their vote as per their desire and comfort level. The new system will securely transfer all the votes to the central database instantly with the help of end to end encryption.

LIMITATIONS OF THE PROPOSED E-GOVERNMENT SYSTEM

The new system is still in the development stage and there are various aspects that are required to be developed to make it more robust and usable to the government. The current system is less secure and usable to the government and the public. There are also some demerits to the current system. The use of the current system requires that the voters know how to cast their votes through this system. In addition to this, there are other prerequisites such as the users should be able to operate computers, access the Internet, open a browser, and establish the connection.

PRINCIPLE OF OPERATION OF ALGORITHMS OF THE SYSTEM

The current system will be installed in the computer of the users as normal software. The system will have access to the MAC address of the computer and the IP. The system will open in a dedicated window which consists of user login and password. The system will first establish a secure connection over the Internet to the server. The entry of the data from the user to access the system will initiate the verification process and it will be approved in the back end. The approval process will also have one-time password (OTP) usage which will be sent to the users' phone. The process will be taken further with the voting. The second screen will have list of candidates and option to cast votes. The casting of votes will be allowed for 100 seconds. After that, the system will close automatically. Whether the voter has cast their vote or not, the system will close on its own after 100 seconds. The users can log in again to cast their votes if they did not.

CONCLUSION

1. The use of the e-government system is likely to improve if the governments of various countries make an effort to educate their population. The use of e-government system will help in the easy management of most of the governmental activities and also facilitate the successful interaction with the citizens.

2. The current article proposes an online voting system that is expected to be used by the government for conducting voting. The voters will cast votes by logging into the system which will be then summarized by the system automatically in the form of a table, pie chart and bar chart to know the result of the election.

REFERENCES

[1] Al-khamayseh, S. Towards Understanding Success Factors in Interactive Mobile Government [Text] / S. Al-khamayseh, E. Lawrence, A. Zmijewska // Proceedings of Euro mGov. – December 2006. – Available at: https://pdfs.semanticscholar.org/70eb/66b62cf8a5e67590849b182fb97ef39a4ef9.pdf

[2] Altameem, T. Critical Success Factors of E-Government: A Proposed Model for E-Government Implementation [Text] / T. Altameem, M. Zairi, S. Alshawi // 2006 Innovations in Information Technology. – IEEE, 2006. – P. 1–5.

[3] Bannet, J. Hack-a-vote: security issues with electronic voting systems [Text] / J. Bannet, D. W. Price, A. Rudys, J. Singer, D. Walach // IEEE Security & Privacy Magazine. – 2004. – Vol. 2, № 1. – P. 32–37.

[4] Chircu, A. M. E-government: key success factors for value discovery and realisation [Text] / A. M. Chircu, D. H.-D. Lee // Electronic Government, an International Journal. – 2005. – Vol. 2, № 1. – P. 11–25. doi:10.1504/eg.2005.006645

[5] Clarkson, M. R. Civitas: Toward a Secure Voting System [Text] / M. R. Clarkson, S. Chong, A. C. Myers // 2008 IEEE Symposium on Security and Privacy (sp 2008). – IEEE, 2008. – P. 354–368. doi:10.1109/sp.2008.32

[6] Electronic voting system [Electronic resource]: United States Patent 6250548 / Mcclure N., Lohry K. – Appl. №08/953003, Filed 16.10.1997, Publ. 26.06.2001. – Available at: http://www.freepatentsonline.com/ 6250548.html

[7] Kohno, T. Analysis of an Electronic Voting System [Text] / T. Kohno, A. Stubblefield, A. D. Rubin, D. S. Wallach // IEEE Symposium on Security and Privacy 2004. – IEEE Computer Society Press, May 2004. – P. 27–40.

[8] Mercuri, R. A better ballot box? [Text] / R. Mercuri // IEEE Spectrum. – 2002. – Vol. 39, № 10. – P. 46–50. doi:10.1109/mspec.2002.1038569.

ОБРОБЛЕННЯ ІНФОРМАЦІЇ У СКЛАДНИХ СИСТЕМАХ

INFORMATION PROCESSING IN COMPLEX SYSTEMS

Автоматизація пошуку помилок у сирих даних та створення SDTM датасетів для медичних досліджень

Хололович Катерина Вікторівна
КПІ ім. Ігоря Сікорського
Київ, Україна
k.9609.khololovich@gmail.com

Букасов Максим Михайлович
КПІ ім. Ігоря Сікорського
Київ, Україна
bukasov@gmail.com

***Анотація*. Розглянута проблема автоматизації пошуку помилок у сирих даних та створення SDTM (Standard Data Tabulation Model) датасетів для медичних досліджень, та запропоновано рішення з використанням системи SAS версії 9.4, в інтегрованому середовищі розробки SAS Enterprise Guide 7.1. Запропоноване рішення дозволить значно підвищити ефективність та покращити якість обробки даних медичних досліджень.**

Ключові слова: обробка даних медичних досліджень, CDISC, SDTM, SAS, SAS Enterprise Guide 7.1.

Вступ

За останні десятиліття проведення медичних досліджень зробило величезний крок вперед. Світовою спільнотою було розроблено десятки стандартів, що регулюють процес перевірки якості та ефективності ліків. Медичні дослідження можуть насправді зробити життя кращим, адже кожної хвилини проводиться розробка або аналіз препаратів, які дозволять боротися із до сих пір невиліковними хворобами.

Проведення медичних досліджень є дуже регульованим та стандартизованим процесом. Важливою умовою затвердження препарату є позитивні результати проведення дослідження. Лише 10% усіх препаратів, над якими проводяться медичні дослідження, стають затвердженими препаратами [1]. Обробка даних медичних досліджень полягає у проведенні статистичного аналізу даних та їх форматуванні згідно до стандартів [2]. Обробка даних також є дуже стандартизованою та регульованою. Такі організації як FDA (Food and Drug Administration) та PMDA (Pharmaceuticals and Medical Devices Agency) приймають тільки дані, оброблені згідно до стандартів CDISC (Clinical Data Interchange Standards Consortium) – некомерційної спільноти, що займається розробкою стандартів для медичних досліджень [3, 4].

Процес обробки даних медичних досліджень містить декілька етапів:

- створення SDTM (Standard Data Tabulation Model) датасетів на базі сирих даних;
- створення ADaM (Analysis Data Model) датасетів [5];
- створення звітів TLF (Tables, Listings, Figures).

Медичні дослідження можуть обходитись фармакологічним компаніям у мільярди доларів. Штраф який стягується з фармакологічної компанії у випадку недостовірності даних або наявності помилок у даних може обійтись майже в ту саму вартість, що й проведення дослідження. Тому виявлення помилок у даних є однією із головних задач, із якими стикаються спеціалісти IT у медичних дослідженнях. Проте перевірка даних часто децентралізована, що робить пошук помилок менш ефективним та створює загрозу невчасного виявлення помилок чи невиявлення помилок взагалі, що може зробити результат дослідження недійсним. Тому постає задача пошуку помилок у сирих даних для їх завчасного виявлення та виправлення.

Сирі дані надходять у вигляді SAS7BDAT файлів або датасетів. Кожен такий датасет містить у собі певну інформацію, що відповідає певній тематиці. Перший етап обробки сирих даних – створення SDTM датасетів, які проходять перевірку на відповідність стандартам CDISC (OpenCDISC validation) [6]. Після валідації SDTM датасетів, на їх основі створюються ADaM датасети, що певним чином модифіковані для підтримки ефективної генерації, реплікації та огляду результатів аналізу. На основі ADaM датасетів створюються звіти TLF, що є основними репрезентативними результатами обробки даних дослідження та дозволяють оцінювати ефективність або безпечність препарату.

Процес створення SDTM датасетів є приблизно однаковим для різних досліджень, адже по суті є первиною обробкою даних, не включає в себе ніяку статистичну обробку даних, а полягає у форматуванні даних згідно до стандартів. Тому доцільно автоматизувати цей етап.

У даній статті запропоноване рішення задачі автоматизації пошуку помилок у сирих даних та автоматизація етапу створення SDTM датасетів з використанням системи SAS.

Особливості системи SAS

Система SAS (Statistical Analysis System) – це набір статистичних програм, розроблений Інститутом SAS для розширеної аналітики, багатофакторного аналізу, управління даними та прогнозованої аналітики. SAS надає графічний інтерфейс користувача для нетехнічних користувачів та більш розширені можливості за допомогою мови програмування SAS [7].

Мова програмування SAS – це процедурна мова програмування високого рівня. Мова програмування SAS розвивалася майже одночасно з розвитком регулювання медичних досліджень, тому засоби розробки SAS максимально задовольняють потреби фармакологічних компаній у обробці даних [8]. На

сьогодні багато фармакологічних компаній звертають увагу на інші безкоштовні мови програмування (Python, R), але, незважаючи на це, SAS залишається лідером у сфері медичних досліджень.

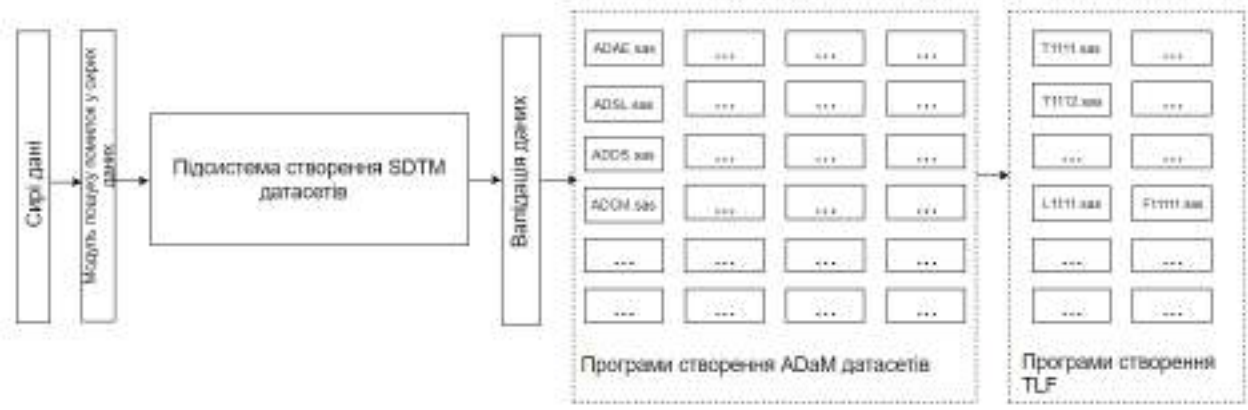

Рис. 1. Процес обробки даних у медичних дослідженях

Система SAS містить багато компонентів, але у сфері медичних досліджень найбільш широко застосованими є компоненти Base SAS, SAS/GRAPH та SAS/STAT (табл. 1).

Таблиця 1

Компоненти системи SAS

Назва компоненту	Опис
Base SAS	Це основний компонент, який містить засоби управління даними та мову програмування для аналізу даних. Він також є найбільш широко використовуваним.
SAS/GRAPH	Містить інструменти для створення графіків та презентацій для демонстрації результату у відповідному форматі.
SAS/STAT	Містить засоби проведення статистичного аналізу за допомогою дисперсійного аналізу, регресії, багатофакторного аналізу, аналізу виживання та психометричного аналізу, змішаного модельного аналізу.

Запропоноване рішення

Запропоноване рішення поставленої задачі реалізовано за допомогою системи SAS 9.4 у середовищі розробки SAS Enterprise Guide 7.1.

Запропонований процес обробки даних медичних досліджень представлений на рис. 1.

Модуль пошуку помилок у сирих даних інтегрований у підсистему створення SDTM датасетів таким чином, що пошук помилок у даних виконується до створення SDTM датасетів. Модуль містить у собі декілька під-модулів – макросів, таким чином, у залежності від особливостей проекту, цей модуль у інтерактивному режимі можна переналаштувати на виконання певних перевірок. Правила для перевірки змінних формуються на базі документу ALS (Architecture Load Specification), який містить повний список сирих датасетів, змінних та їх атрибутів. В результаті виконання модулю створюється XLSX файл, з вкладками для кожного сирого датасету та вкладкою, що містить загальну інформацію про знайдені помилки у сирих даних. Якщо помилки знайдені, то автоматично буде відправлено електронний лист до відділу управління даними для перевірки.

Підсистема створення SDTM датасетів містить декілька модулів та базується на специфікації SDTM датасетів – документу, який розробляється програмістами, та містить у собі правила мапінгу змінних, словники допустимих значень змінних та основні атрибути. В результаті виконання модулю створюються пакет SDTM датасетів у форматі SAS7BDAT (для зручності подальшого використання у системі SAS) та пакет SDTM датасетів у форматі XPT (для відправлення замовнику та подання у служби FDA або PMDA).

Висновки

Запропоноване рішення можна легко впровадити для більшості проектів, які підтримують стандарти CDISC та використовують систему SAS 9.4. Воно допоможе збільшити ефективність та покращити якість обробки даних медичних досліджень.

Література

1. en.wikipedia.org/wiki/Clinical_trial (дата звернення 28.10.2019)

2. Wang D. Clinical Trials. A Practical Guide to Design, Analysis, and Reporting / D. Wang, A. Bakhai. – London: Remedica, 2006. – 498 с.

3. Stemplinger R. T. Considerations for CDISC Implementation [Електронний ресурс] / R. T. Stemplinger, J. Lane // PhUSE. – 2007. – Режим доступу до ресурсу: www.phusewiki.org/docs/2007/PAPERS/_RA10.pdf.

4. en.wikipedia.org/wiki/Clinical_Data_Interchange_Standards_Consortium (дата звернення 30.10.2019)

5. www.cdisc.org/standards (дата звернення 30.10.2019)

6. www.pinnacle21.com/news/opencdisc-validator-15-available-now

7. en.wikipedia.org/wiki/SAS_(software) (дата звернення 01.11.2019)

8. www.sas.com/ru_ua/solutions/analytics.html (дата звернення 05.11.2019)

Аналіз коментарів за допомогою машинного навчання

Писаренко Олег Анатолійович
КПІ ім. Ігоря Сікорського
Київ, Україна
oa.pisarenko@gmail.com

Дорошенко Анатолій Юхимович
КПІ ім. Ігоря Сікорського
Київ, Україна
a-y-doroshenko@ukr.net

Анотація. **В роботі демонструється результат реалізації сервісу для аналізу текстових коментарів за допомогою машинного навчання. Для цього був обраний підхід ансамблювання алгоритмів машинного навчання, до якого війшли алгоритми Gradient Boosting, Random Forest та логістична регресія. Була навчена прогнозуюча модель та оцінена її точність на основі тестових даних.**

Ключові слова: машинне навчання, Tensorflow, Gradient Boosting, Random Forest, логістична регресія.

Вступ

В час швидкого розвитку інтернету, соціальних мереж та стрімінгових сервісів виникла необхідність в реалізації інструментів та платформ для обговорення та спілкування між людьми. Не завжди таке спілкування є конструктивним та по суті, а тим більше коли спільнота стає досить великою або популярною, там з'являються люди із зовсім інакшою метою – не конструктивна критика авторів, агресивна та нецензурна дискусія з іншими коментаторами, приниження та образи певних груп людей за різними ознаками. Таким чином виникла ідея в створенні інтелектуальної системи, завдяки якій автори та модератори різних інтернет-майданчиків зможуть автоматизувати стеження за своєю спільнотою коментаторів та завчасно фільтрувати негативні коментарі. Частиною такої системи є сервіс оцінки коментарів за різними негативними ознаками.

В якості негативних ознак для класифікації коментарів було обрано ознаки: токсичність, висока токсичність, нецензурна лексика, погроза, образа та персональна ненависть.

Ансамбль методів

Для створення інтелектуальної системи фільтрації коментарів, а саме в частині сервісу аналізу коментарів було вирішено використати машинне навчання, як один із найбільш сучасних та ефективних інструментів аналізу текстових та графічних даних і знаходження закономірностей в проаналізованих даних для подальшого їх використання на реальних даних. Чим більш різноманітні вхідні дані, тим простіше алгоритмам знайти закономірності та завдяки використанню різних алгоритмів класифікації даних машинного навчання можна досягти швидкодії та дуже високої кінцевої точності необхідного результату.

Для побудови прогнозуючої моделі було вирішено обрати один із підходів в машинному навчанні – ансамблювання алгоритмів машинного навчання [1]. Ансамблі є більш точними в прогнозуванні кінцевого результату за окремо взяті методи навчання, так як такий підхід дозволяє взяти різні методи класифікації та навчити їх виправляти помилки один одного. Якість такої системи буде набагато вище, чим використання кожного із методів окремо.

Алгоритм Gradient Boosting

Класифікатор Gradient Boosting – це алгоритм машинного навчання, що відноситься до вирішення проблем регресії та класифікації, який продукує прогнозуючу модель у вигляді дерев рішень [2].

Алгоритми класифікації часто використовують логарифмічні втрати. Класифікатору Gradient Boosting не потрібно отримувати нову функцію втрат щоразу, коли додається алгоритм бустингу, скоріше будь-яка диференційована функція втрат може бути застосована до системи.

Gradient Boosting складається з двох ключових елементів: дерева рішень та адитивного компоненту. Для мінімізації похибки між заданими параметрами використовується процедура, схожа на спуск градієнта [3]. Це робиться шляхом взяття обчисленої втрати та виконання градієнтного спуску для зменшення цих втрат. Потім параметри дерева змінюються для зменшення залишкових втрат.

Вихід нового дерева потім додається до виходу попередніх дерев, використовуваних у моделі. Цей процес повторюється, поки не буде досягнуто раніше визначеної кількості дерев, або втрати не зменшаться нижче певного порогу.

В цьому алгоритмі машинного навчання, процес навчання послідовно пристосовує створені ним моделі, щоб забезпечити більш точну оцінку для шуканої ознаки. Під час кожної конкретної ітерації нове дерево рішень, яке отримується під час навчання на попередніх помилках, які були отримані на попередніх ітераціях роботи алгоритму.

Алгоритм Random Forest

Класифікатор Random Forest – алгоритм машинного навчання для класифікації, регресії та інших завдань, які можна вирішити за допомогою

побудови численних (комітетів) дерев рішень під час тренування. Недоліком є схильність до перенавчання. Основна ідея алгоритму полягає в використанні великої кількості дерев рішень, кожне з яких саме по собі дає дуже невисоку якість класифікації, але за рахунок їх великої кількості результат виходить точним [4].

Алгоритм Random Forest вибирає випадково n записів з навчального датасету, а потім будує дерево рішень, що базується на цих записах. Задається кількість дерев, що будуть виконувати таку дію. Для проблеми класифікації кожне дерево рішень передбачає категорію, до якої належить новий запис.

Класифікація об'єктів проводиться шляхом голосування: кожне дерево категорії відносить об'єкт класифікації до одного з класів, і перемагає той клас, за який проголосувала найбільша кількість дерев.

Оптимальне число дерев підбирається таким чином, щоб мінімізувати помилку класифікатора на тестовій вибірці. У разі її відсутності, мінімізується оцінка помилки: тих зразків, які не потрапили в навчальну підвибірку за рахунок повторень.

Алгоритм логістична регресія

Логістична регресія – це статистичний алгоритм, що використовується для прогнозування ймовірності виникнення деякої події шляхом підгонки даних до логістичної кривої. Логістична регресія застосовується для прогнозування ймовірності виникнення деякої події по значеннях множини ознак. Для цього вводиться так звана залежна змінна y, що приймає лише одне з двох значень - як правило, це числа 0 (подія не відбулося) і 1 (подія відбулася), і безліч незалежних змінних (так звані ознаки) - на основі значень яких потрібно обчислити ймовірність прийняття того чи іншого значення залежної змінної.

Для візуального розуміння роботи алгоритмів машинного навчання була побудована ROC-крива – графік, що дозволяє оцінити якість класифікації за співвідношенням між часткою об'єктів від загальної кількості носіїв ознаки, вірно класифікованих до загальної кількості об'єктів, що не несуть ознаки, помилково класифікованих, як такі, що мають ознаку.

На рисунку 1 можна побачити графік роботи алгоритму логістична регресія.

Використання моделі машинного навчання

В результаті ансамблювання алгоритмів машинного навчання була побудована прогнозуюча модель, збережена її архітектура та ваги оцінок. Роботу готової моделі було перевірено на тестовому датасеті, який показав точність оцінки негативних коментарів для обраних ознак вище 97%. Для моделі було побудовано REST API, що дало змогу використовувати її як окрему одиницю для інтеграції в різні системи.

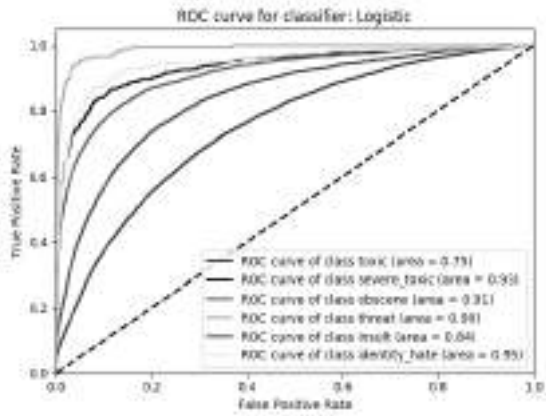

Рис. 1. ROC-крива класифікації ознак алгоритмом логістичної регресії

Висновки

В результаті виконання даної роботи були розглянуті обрані алгоритми машинного навчання для аналізу текстових коментарів. А саме алгоритми Random Forest, Gradient Boosting та логістична регресія. Дані алгоритми були об’єднані в ансамбль для отримання більш точного результату оцінки коментарів, ніж той, який можна отримати використавши певний алгоритм окремо. Приклад роботи кожного з алгоритмів був продемонстрований на графіках ROC кривих для кожної з ознак, за якими класифікується коментар.

Ансамбль алгоритмів дав змогу отримати дуже точну прогнозуючу модель, що на тестових даних показала точність роботи вище 97%.

Література

1. Mitchell T.M. Machine Learning / T.M. Mitchell. – New York: McGraw-Hill Education, 1997. – С. 175.

2. https://towardsdatascience.com/machine-learning-part-18-boosting-algorithms-gradient-boosting-in-python-ef5ae6965be4 (дата звернення 17.10.2019)

3. Guido S. Introduction to Machine Learning with Python / Sarah Guido, Andreas Müller. – New York: O'Reilly Media, 2016. – С. 50.

4. stackabuse.com/random-forest-algorithm-with-python-and-scikit-learn/ (дата звернення 22.10.2019)

Analysis of the calculus basis boundary for redundant codes

Poltorak Vadym
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
v.poltorak@kpi.ua

Abstract. **The paper describes an analysis of the calculus basis boundary for redundant codes, needed to protect data from distortions in communication channel. Here the criteria of minimum code distance limit for the codec quality rating was used, excluding the impact of the data transmission medium and modem properties to the data transmission efficiency estimation. It is implied here boundaries of Gilbert-Varshamov, Hamming, Plotkin and Elias.**

Keywords: calculus basis; redundant codes; communication channel; boundaries for minimum code distance; data transmission.

INTRODUCTION

We know that the initial information coding in general, then error control coding of messages (by redundant codes) and data presentation by different channel signals are different aspects of a single information process - different forms of information presenting. There are many data transmission tasks where we should to correct errors in real-time data flow [1, 2, 3]. The Generic structure of one-directional Data Transmission System (DTS) includes optimal codec for the source and channel codec. The channel codec provides many useful features. The deep meaning of its using is not only to identify and correct certain errors of data received. There are many parameters of DTS that we can vary in order to achieve the better data performance according to certain criteria. For example, we can reduce specific energy consumption or transmission bandwidth to each data unit received with preset speed and reliability. We can increase the speed or distance of data transfer between Source and Acceptor of Messages. Or, we can decrease the device energy consumption in general. It will have to pay by increasing of hardware complexity and algorithms and procedures complexity of its operation to achieve of these improvements. We use different criteria for assessing the quality of the information process in these different aspects. Let we use (for the first step) the criteria of limits for minimum Hamming code distance d for the codec quality rating, excluding the impact of the data transmission medium and modem properties to the data transmission efficiency estimation. For example, we can use the criteria of "limits for minimum code distance d " to justify the use of error control coding for messages in the data transmission channels. It is implied here boundaries of Gilbert-Varshamov, Hamming, Plotkin, Elias [1, 2]. In some cases we can more or less accurately determine the probability of correct data error decoding depending on probability of channel error per one symbol. But it is impossible to completely define the corrective ability of redundant code with only the minimum of d .

PERFORMANCE CRITERIA

Since ancient times, it was adopted a model of high efficiency of "long" redundant codes with high values n of code block lengths [1, 2]. In this case, a more detailed description of properties of these types of codes is often impossible or difficult. It is possible to describe the properties of redundant codes in more detail with small values of n and k , where k is the number of user data in a block of length n that need protection from distortions and errors. That is why in the case of long codes their correcting ability is characterized by the size d of the minimum Hamming distance of the code. In view of the above, the following performance criteria of redundant codes where set out [2]. 1. Among codes with the same n and k (the same relative code rate $R = k / n$), the best is the code that has more d . This code can correct more errors in the block of length n . 2. Among codes with the same n and d (the same $x = d / 2n \approx t / n$, where x is the relative code correction potential and t is a number of errors the code can to correct), the best is the code that has more user data part k of code block. This code has more $R = k / n$ with the same t . 3. Among codes with the same k and d , the best is the one that has less length n . This code has less absolute redundancy $r = (n - k)$ and almost the same relative code redundancy $D = r / n$. From other hand, the best code has more rate $R = k / n$ with more $x = d / 2n \approx t / n$ (or the more t).

CODING EFFECTIVENESS

These characteristics do not consider the complexity of encoding and decoding, error probability, properties of the medium for signal propagation. That is why they say the code has a better performance in terms of pure coding theory without take into account codec procedures complexity and type of modulation. A more detailed analysis of the effectiveness of redundant codes by means of performance criteria mentioned above shows a great impact of the code basis q . Redundant coding efficiency is typically represented as an area that lies between the lower and upper boundaries for minimal code distance d on the plane in Cartesian coordinates of R and x . There is lower boundary for d - the Gilbert- Varshamov one [1, 2]. There are several upper boundaries for d , they are Hamming, Plotkin and Elias ones, and they limit the impact of coding from the upper side [1, 2]. These mathematical boundaries give an idea about the functional relationship between R and x values with the influence of n and q , $R = f(x; n; q)$. These objects are usually considered in terms of the asymptotic approximation of

$n \to \infty$, and typically $q = 2$ [1, 2, 3]. Detailed examination of the boundaries for d shows a significant effect of the code basis q to redundant coding efficiency without asymptotic approximation of $n \to \infty$ [4, 5]. This effect is significantly noticeable even when n takes values of the dozens and hundreds (not of $n \to \infty$): $n = (10...100) \cdot z$. For example, the upper Plotkin boundary gives similar graphs at $n = 20$ and $n = 100$ respectively [4]. It can be seen that impact of n not very differentiates of these graphs. But impact of q has more influence on coding efficiency. Such effects are inherent to other boundaries too.

It was investigated in the work the code rate R cumulative growth influenced by base q increasing, as is presented on Fig.1 for different values of $x = const$. The diagram on Fig.1 shows the conditional limit for q as effect of saturation of value R after q reaches at about 32 values ($q \geq 32$).

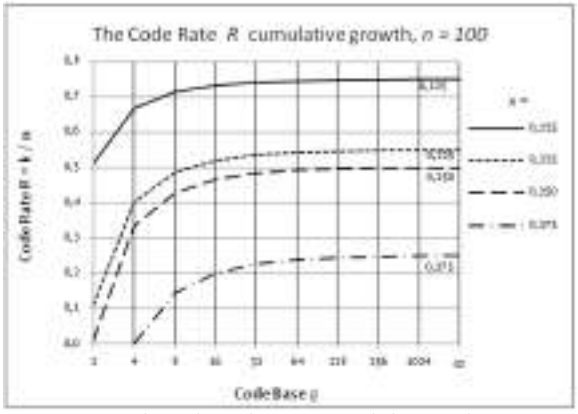

Fig.1. The Code Rate *R* cumulative growth

The code rate R differential comparative growth under base q differential growth reveals a striking effect presented on Fig.2, based on Fig.1 data. The Fig.2 diagram shows a saturation effect for R growth rate under incremental growth of the q.

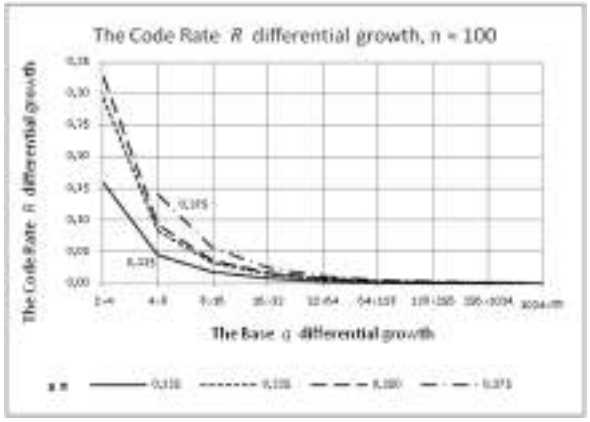

Fig.2. Code rate R differential growth

We can see a negative value R growth rate with value q incremental growth from $q = 2$ up to $q = 32$, for different values of x. It is clear that the R growth rate is becoming smaller and smaller with the q incrementing up to $q = 32$ (of approximately), and it almost does not improve after $q \geq 32$. Observations obtained in the work are typical for other boundaries for q too.

The code efficiency quickly reduces from $q = 2$ to $q >> 2$ and become undifferentiated when $32 \leq q < \infty$ [4]. It is no sense increase q to ∞ to increase the efficiency of coding, because of maximum code efficiency is achievable practically at $q \geq 32$ [4]. A limit found there for q can serve as some "beacon" in designing of channel codec, as motion of $q \to \infty$ means an unlimited increase in complexity of hardware and software implementation, and cannot be considered as desirable. Conversely, a reasonable restriction of code basis $q \leq 32$ gives an almost maximum code efficiency and may provide less expensive performance of codec for DTS then in case of q unfounded growth. Analysis of the q impact on the characteristics of redundant codes reveals another unobvious property. This is the code length n. The famous "stamp" there is still more in thinking that the most effective codes are "long" codes with large values of n (by the criterion of minimum redundancy D), where r is a number of redundant code components (code elements) in the code block [1, 2, 3].

However, our research shows that this is true only if a fixed value $q = const$. That is why the idea of "long" code, as the most effective, was formed that way. For until this time the vast majority of codes used in practice are a purely binary with $q = 2$. The study shows, that high-performance redundant coding can be achieved with relatively small length n, by increasing the code basis $q > 2$, but subject to the limitations that are mentioned above $q \leq 32$. This opens up prospects for exploration and use relatively "short" codes, for which n are compared to q.

References

[1] Peterson W. Error-Correcting Codes / W. [1] Peterson, E. Weldon. – Cambridge: MIT Press, 1972.

[2] Coding Theory / T.[2] Kasami, N. Tokura, Y. Iwadare, Y. Inagaki. – Tokyo: Iwanami, 1975. – 588 p.

[3] Blahut R. E. Theory and Practice of Error Control Codes / Richard [3] Blahut. – Addison: Wesley, 1983. – 293 c.

[4] Полторак В. П. Вплив основи коду на ефективність надлишкового кодування / В. П. [4] Полторак, Н.Вітіщенко. // . Інформатика: Зб. наук. пр. – Київ: Век+. – 2011. – №54. – С. 95–100.

[5] Жураковський Ю.П.Теорія інформації та кодування:Підручн. / Ю. П. [5] Жураковський, В.П.Полторак.–Київ: Вища шк., 2001. – 255с.

БЕЗПЕКА ТА ЗАХИСТ ІНФОРМАЦІЇ

INFORMATION SECURITY

Аналіз особливостей державних стандартів ЕЦП на властивостях еліптичних кривих

Романчук Сергій Анатолійович
КПІ ім. Ігоря Сікорського
Київ, Україна

Анотація. **Наведено особливості побудови криптографічних алгоритмів орієнтоваих на забезпечення цифрового підпису повідомлення, що опираються на властивості еліптичних кривих. Визначено основні відмінності стандарту ДСТУ 4145-2002 і за рахунок чого досягнуто більшої швидкодії. Наведено схеми побудови та верифікації цифрового підпису, що є справедливими незалежно від обраного алгоритму або стандарту.**

Ключові слова: еліптичні криві, цифровий підпис, державний стандарт, криптографія.

Вступ

Нові терміни та поняття приходять у наше життя завдяки інформаційним технологіям, одним з ключових у сучасному світі став термін «електронний цифровий підпис» (ЕЦП). Використовується в якості гарантії ідентифікації та підтвердження юридичної значущості документів. Електронний цифровий підпис - це реквізит електронного документа, призначений для захисту даного електронного документа від підробки чи внесення змін, який формується в результаті криптографічного перетворення інформації через використання секретного ключа підпису і дозволяє однозначно ідентифікувати власника ключа чи спотворення інформації в електронному документі. Користуватися підписом дуже просто. Ніяких спеціальних знань, навичок і умінь для цього не буде потрібно. Кожному користувачеві ЕЦП, який бере участь в обміні електронними документами, генеруються унікальні відкритий і закритий (секретний) криптографічні ключі. Підпис документа відбувається на основі закритого ключа, в підпис включаються наступні дані:

- ім'я файлу відкритого ключа підпису;
- інформація про особу, яка підписує документ;
- дата формування підпису.

Особа, що отримала підписаний документ за допомогою відкритого ключа виконує зворотнє криптографічне перетворення, яке забезпечує валідацію цифрового підпису відправника. Завдяки цьому ЕЦП тісно інтегрувалася в сфери, що пов'язані з фінансовими операціями, документообігом, звітуванням до державних органів, торгівлею, трейдингом, тощо.

Алгоритми ЕЦП

Незважаючи на велику кількість алгоритмів ЕЦП, всі вони є асиметричними і базуються на застосуванні відкритого ключа. Криптографічна стійкість цих алгоритмів забезпечується «односторонньою функцією» та на її основі поділяються на системи, що опираються на:

- розрахунок дискретного логарифму в скінченому полі;
- факторизації добутку двох великих простих чисел;
- задачі дискретного логарифмування на еліптичних кривих у скінченому полі. [1]

Для аналізу проблем реалізації та порівняння вибрано стандарти США, РФ та України. В стандартах ECDSA та ГОСТ Р 34.10-2012 на еліптичних кривих в якості математичної основи використовуються еліптичні криві над простим полем Галуа. Чинний ДСТУ 4145-2002 опирається на властивісті груп точок еліптичних кривих над полями $GF(2^m)$ [2]. Незважаючи на відмінності в деталях реалізації алгоритми можуть бути описані трьома ключовими кроками:

- генерація ключів;
- формування підпису;
- перевірка підпису.

Загальна схема етапу формування ЕЦП наведена на рис. 1.

Оскільки ЕЦП гарантує незмінність відкритого

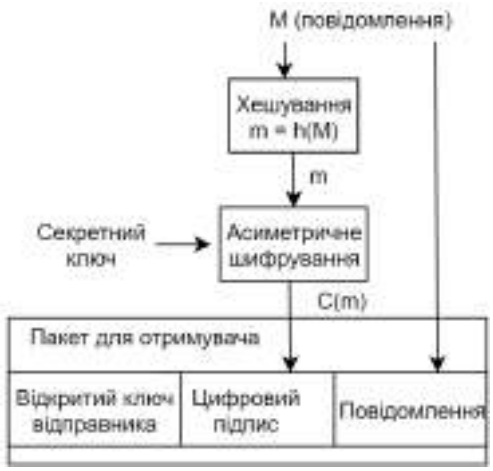

Рис. 1. Схема етапу формування підпису

повідомлення етап верифікації цифрового підпису відрізняється від звичайного декодування зашифрованого повідомлення. Схема етапу верифікаці наведана на рис. 2.

Відповідно до наведених схем для програмної реалізації потрібно виконати моделювання та розробку наступних компонентів системи:

- функція гешування;
- метод генерації відкритого ключа;
- метод генерації секретного ключа;

- блок формування ЕЦП;
- блок формування пакету для відправки;
- блок верифікації ЕЦП.

Вибрані алгоритми наводять рекомендовані функції гешування та довжини ключів, на основі наведених даних може бути проведений короткий аналіз.

Огляд загальних властивостей вибраних алгоритмів наведено у таблиці 1.

Таблиця 1

Загальний аналіз алгоритмів

Алгоритм	**Характеристики (рекомендовано)**		
	Хеш-функція	***Розмір відкритого ключа (біт)***	***Розмір секретного ключа (біт)***
ECDSA	SHA-1 або SHA-2	112-320	80-521
ГОСТ Р34.10-2012	ГОСТ Р34.112012	80-320	256-512
ДСТУ 4145-2002	ГОСТ 34.311	162-768	256-1024

ECDSA – алгоритм з відкритим ключем для створення цифрового підпису, аналогічний до DSA проте, навідміну від нього визначений не над олем

Рис. 2. Схема етапу верифікації

цілих чисел a в групі точок еліптичної кривої. Суттєвою перевагою алгоритму є те, що система базована на еліптичих кривих забезпечує таку ж криптостійкість як та що заснована на дискретному логарифмуванні проте зі значно меншою довжиною ключа.

Алгоритм ГОСТ Р34.10-2012 – російський стандарт, що описує алгоритми формування та перевірки електронного цифрового підпису. Прийнятий і затверджений в 2012 р.

ДСТУ 4145-2002 – український алгоритм який має як переваги так і недоліки. Криптографічний алгоритм який використовується – Нюберга-Рюппеля, який при рівних з DSA якостях відрізняється більшою швидкістю обрахування та перевірки підпису. Очевидно, що алгоритм Нюберга-Рюппеля виграє у DSA за рахунок того, що в останньому використовується ресурсовитратна операція обернення цілих чисел за модулем простого числа. В теоретичому випадку це не так помітно як у випадку реалізації, тому й при переході на еліптичні криві необхідно зводити цілочислову арифметику до необідного мінімуму. В стандарті наведено перелік рекомендованих еліптичних кривих. Незважаючи на переваги стандарт є, здебільшого, декларативним, оскільки відсутня практична можливість розпаралелювання обрахунків при реалізації. Текст стандарту викладено на 39 сторінках складного викладення. В результаті цього для розуміння стандарту недостатньо знань із програмування, потрібно мати хорошу математичну підготовку.

Висновки

В роботі проведено аналіз існуючих алгоритмів формування ЕЦП. Визначено за рахунок чого стандарт ДСТУ 4145-2002 відрізняється високою швидкістю формування та перевірки електронного цифрового підпису.

Література

1. Кочубинский А. Сравнительный анализ характеристик и принципов построения стандартов ЭЦП на свойствах эллиптических кривых [Електронний ресурс] / А. Кочубинский, А. Шаталов // 6. – 2003. – Режим доступу до ресурсу: http://pnzzi.kpi.ua/6/06_p49.pdf.

2. ДСТУ 4145-2002. Інформаційні технології. Криптографічний захист інформації. Цифровий підпис, що ґрунтується на еліптичних кривих. Формування та перевірка. – К.: Держстандарт України, 2003. –39 с.

Рецензент: к.т.н., доц. каф. АУТС, КПІ ім. Ігоря Сікорського В.П. Полторак

Автентифікація зображень на основі методів цифрового підпису

Назарій Калитюк
КПІ ім. Ігоря Сікорського
Київ, Україна

***Анотація*. У статті проведений аналіз підходів до автентифікації зображень. Досліджені можливості застосування алгоритмів цифрового підпису повідомлень для автентифікації зображень.**

Ключові слова: автентифікація, автетифікація зображень, цифровий підпис, точна автентифікація, елективна автентифікація.

Вступ

В наш час методи автентифікації зображень є дуже важливими для значної кількості мультимедійних додатків. Перехоплена та змінена інформація може мати небажані наслідки. Зображення воєнних об'єктів, цифрові нотаріальні документи, дослідницькі документи і тд. Всі ці зображення мають бути захищені.

У зв'язку з доступністю та широким розповсюдженням засобів цифрової оброки зображень, люди можуть виготовляти підробки і різними способами впливати на зображення, що може привезти до матеріальних та навіть людських втрат. Наприклад, якщо замінити діагностичне зображення УЗД чи МРТ пацієнта, то йому може бути поставлений невірний діагноз та невірне лікування, що може призвезти до серйозних наслідків.

Проте існують випадки, коли потрібно визначати зміни в зображені, яке може бути змінене стисненням, відновленням, підвищенням якості та іншими методами обробки. Таким чином, процедури перевірки достовірності зображень можна розділити на 2 групи: точна та селективна автентифікації.

Точна автентифікація

Точна автентифікація застосовується у випадках, коли будь–яка зміна захищеного зображення заборонена. Селективна використовується ж тоді, коли необхідно дозволити алгоритми фільтрації та стиснення зображення [1].

Під час точної автентифікації зображення визнається підробкою у випадку, коли один піксель чи навіть біт був змінений [2]. Проте для більшості випадків такий варіант не підходить.

Для задач точної автентифікації вже знайдені рішення за допомогою традиційної криптографії і технології водяних знаків. Данні рішення забезпечують гарні результати, які задовольняють користувачів, проте деякі дослідження ще варто провести для знаходження більш ефективних можливостей по відновленню підроблених зображень.

Селективна автентифікація

Селективна автентифікація використовує методи на технології напівкрихких водяних знаків і цифрового електронного підпису зображень для забезпечення стійкості до змін. Результати досліджень задовільні, проте проблема ще далека до свого вирішення.

Хотілось би мати можливість стискати зображення в цілях економії пам'яті, відновлювати зображення та підвищувати його якість або навіть змінювати формат. В цьому випадку нам необхідний такий метод автентифікації, котрий допускає конкретні операції обробки зображень. Данні операції змінюють значення пікселів без зміни складу зображення. Таким чином, реальне завдання селективної автентифікації зображень пов'язане із завдання знаходження змістового складу зображення. Іншими словами, нам необхідно визначити тільки ті зміни, які призводять до модифікації візуального образу або до помилкових інтерпретацій зображення, таким як зникнення об'єкта чи поява нового об'єкта. Отже, аби розробити потрібні підходи селективної автентифікації зображень, необхідно розрізняти ті перетворення, які змінюють склад зображення і ті, які його зберігають.

Нажаль, такі обмеження важко реалізувати технічно. Більше того, ці обмеження можуть змінюватись для різних зображені і завдань.

Наступні операції по обробці зображення зберігають склад зображення в більшості випадків, отже, є допустимими для використання методів селективної автентифікації: помилка передачі; шум при передачі; помилка зберігання; квантування та стиснення; геометричні перетворення (поворот, масштабування); методи збільшення якості (фільтрація, обробка рівня сірого); відновлення (зменшення шумів, деконволюція); перетворення форматів.

Перетворення, які змінюють зміст зображення, отже, є недопустимі для методів селективної автентифікації: видалення об'єктів: додавання об'єктів: зміна розташування об'єктів; зміна фону зображення і тд[3].

Селективна автентифікація зображень на основі технології цифрового підпису

Традиційний підпис на паперових документах зазвичай підтверджує надійність. Цифровий підпис необхідний для підпису документу в електронному вигляді і може бути переданий разом із підписаним документом[4].

Принципова різниця між традиційним і цифровим підписом в тому, що будь–яка копія електронного документу ідентична своєму оригіналу, в той час, як копія підписаного паперового документу, як правило, можна відрізнити від оригіналу. Ця відмінність призводить до нової фундаментальної проблеми, пов'язаної з поняттям оригіналу документу, підписаного в електронному вигляді, і з методиками, які забороняють його повторне використання.

В літературі представленні такі алгоритми як схема цифрового підпису Ель–Гамаля, стандард цифрового підпису Digital Signature Standart (DSS), схема Van Heyst–Pedersen. Також цифрові підписи, основані на криптосистемах, таких як RSA та DSA.

Алгоритми формування цифрового підпису застосовуються або безпосередньо до повідомлення або значення хеш–функції від повідомлення. В першому випадку перевіряється дійсність самого повідомлення, а в іншому генерується додаток, який застосовується під час процедури верифікації.

Проте існує проблема, яка полягає в тому, що нам необхідно зрозуміти, яку інформацію підписувати: дані зображення чи його зміст. Насправді, застосування алгоритмів формування цифрового підпису безпосередньо до зображень може призвести до ситуацій, коли зміст зображення не було змінено, а в результаті роботи алгоритму зображення визнано підробкою.

Тому з'являється ситуація, коли зображення можна підписати як звичайний файл, не зважаючи на його специфікації. Проте такий підхід забороняє будь–які зміни зображення, які не впливають на саме зображення.

Таким чином, з'являється необхідність спочатку визначити вагомі елементи зображення та який зміст вони несуть. Це можу бути, наприклад, пухлина на знімку МРТ. Далі цю ділянку зображення необхідно помітити як важливу для цифрового підпису або розробити алгоритми фільтрації ділянки, які не змінять зміст ділянки, або заборонити фільтрацію ділянки. З цього випливає, що ми фокусуємо цифровий підпис тільки на важливій ділянці, зміст якої нас цікавить, і дозволяємо зміну інших частин зображення. Якщо нас цікавить кілька ділянок, то підписуємо зображення з увагою на всі ділянки. Якщо таких ділянок забагато, то залишається лише варіант із повним підписом зображення, яке виключає будь–які зміни. Наразі це оптимальний спосіб, проте вже існують алгоритми машинного навчання, які зможуть контролювати зміни зображення без змін змісту, та вони ще далекі від ідеалу.

Проте для векторних зображень, де дані передаються набором фігур, можна підписати зображення. Більше того, є можливість опису правил зміни фігур, для того аби зображення можна було змінити і не втратити підпис. Такі зображення не потребують стиснення та фільтрації, отже змін, які можуть порушити підпис. Нажаль, для більшості цілей, де потрібен електронний цифровий підпис зображень, векторний формат не годиться. Адже неможна передати векторним зображенням фотографію із супутника, знімок МРТ пацієнта, знімки документів та ін.

Висновки

Таким чином, під час модифікації існуючих схем цифрового підпису необхідно визначити, яка інформація має бути підписана. Тобто, при використанні традиційних алгоритмів цифрового підпису випливає, що потрібно підписувати зміст зображення, а не дані самого зображення. Нажаль, формати растрових зображень були розроблені без можливості цифрового підпису, а розміщення цифрового підпису в метаданих зображення не є достатньо стійким. Тому залишаються лише варіанти розміщення електронного цифрового підпису окремо від зображення, або створення контейнера, який буде містити підпис і зображення. Проте досі не існує стандарту щодо цифрового підпису зображень, тому немає гарантії щодо можливості перевірки підпису різними програмами для перегляду та обробки зображень.

Література

1. Kutter M. Digital watermaking of color images using amplitude modulation / M. Kutter, F. Jordan, F. Bossen // J Electron Imaging / M. Kutter, F. Jordan, F. Bossen., 1998. – С. 326–332.

2. N. Memon. Proceedings of the SPIE international conference on security and watermarking of multimedia contents II / V. Poorvi, Y. Boon-Lock, M. Yeung // Distortion bounded authentication techniques / N.Memon, V. Poorvi, Y. Boon-Lock, M. Yeung., 2000. – (Vol. 3971). – С. 164–174.

3. Dittmann J. Content-based digital signature for motion pictures authentication and content- fragile watermarking / J. Dittmann, A. Steinmetz // Proceedings of the IEEE international conference on multimedia computing and systems / J. Dittmann, A. Steinmetz., 1999. – С. 209–213.

Брассар Д. Современная криптология / Дж Брассар., 1999. – 107 с.

ABSTRACTS

INFORMATION SYSTEMS AND TECHNOLOGIES

Page 11

An object recognition subsystem for automotive autonomous control systems

Mykhailo Khlivnenko
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
xlivnenko.michael@gmail.com

Andrii Pysarenko
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine

Abstract. **An analysis of existing hardware object recognition methods for autonomous control systems. The research of quality indicators of machine learning recognition algorithms was described. Based on the analysis, the structure of the object recognition system model for auto-mobile autonomous control systems was proposed.**

Keywords**:** ***autonomous control systems, object detection algorithms, ACF, YOLO, LIDAR.***

Page 13

Design of architecture for text research system

Andrii Zubrytskyi
Igor Sikorsky Kyiv Polytechnic institute
Kyiv, Ukraine
andreizubritskiy@gmail.com

Abstract. **This article describes the designed system architecture for text analysis of ukrainian language by linguistic researchers. In order to design the functionality of the system better the existing solutions were analyzed and comparative analysis of them was made. During design modern development tools were used and focus was on designing system modules in such way that they could be easily added to the system.**

Keywords**:** ***system architecture, text analysis system, web application design.***

Page 15

On reability simulating and evaluating in cloud services system

Eugene Pokrovskyi
Oleg Morgal
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
ea.pokrovski@gmail.com
m_olegm@ukr.net

Olena Savchuk
Olexandr Pokhylenko
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
savchuk_ll@ukr.net
pokhilenko.alex@gmail.com

Abstract. . **Some problems of simulating and estimating reliability in the cloud services system are considered. An example of a typical structure analyzes the cloud service errors at the request and service stages. Markov model is explored for the waiting queue regime. An approach to the cloud data security is proposed which, unlike traditional approaches, guarantees confidentiality and security of information.**

Keywords: cloud services; reliability simulating; data security in cloud environments**.**

Page 17

The impact and unforeseen challenges of E-procurement systems in Canada

Bodak Bohdan
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
bohdan.bodak@outlook.com

Doroshenko Anatoliy
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
a-y-doroshenko@urk.net

Abstract. **Automation in public procurement sector had become widespread across Europe and North America in mid-1990-s. Canada was among pioneers to implement the competitive process, which aims to create the best opportunities for Canadians, while enhancing transparency, competition, and fairness. Procurement data is available to view and download allowing any supplier to look for opportunities, access all listed tenders, bookmark and share searches. According to studies, the government was able to save up to 15% of budgetary funds on procurement services whilst reducing additional costs and time with the introduction of online system. Therefore, it is crucial to analyze the success and issues in development of an online procurement system in Canada in order to take over the experience and key concepts.**

Keywords: E-government, procurement systems, tenders, goods and services, E-procurement, fair competition, buy and sell.

Page 19

The use of radio-frequency identification in information systems

Kharabet Rodion
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine

Pysarenko Andrii
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
andrew.pisarenko@gmail.com

Abstract. **The article describes the use of radio-frequency identification in information systems. It covers the application of RFID in the housekeeping process, in particular, goods monitoring. The existing solutions and their pros and cons were reviewed. To increase the efficiency and convenience of shopping the software and hardware system based on radio-frequency identification was proposed. The article contains an explanation of its workflow and interaction with users and outlines the advantages of this system compared to existing systems..**

Keywords**:** ***radio-frequency identification, RFID, information systems, housekeeping, barcode, esp8266, automation.***

Page 21

An integrated information system of monitoring and control for smart farm based on IoT

Karine Diachenko
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
diachenkokarine@gmail.com

Pysarenko Andrii
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
andrew.pisarenko@gmail.com

Abstract. **An integrated information system of monitoring and control for smart farm based on Internet of Things and algorithms that are used to analyse big data and predict animal health are offered.**

Keywords**:** ***smart farm, sensors, Internet of Things, monitoring, control, algorithms.***

Page 23

Remote automatic software deployment system

Andrew Telenyk
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
andrzej.telenik@gmail.com

***Abstract*: The purpose of the article is to review available means of solving the problem, to clarify the aspects of their deployment, and to describe author's vision for solving the problem. The best solution to the problem will be the implementation of remote deployment system coded in C # language, based on the principles of object-oriented programming and client-server architecture, which will allow easy installation of software within the computer network.**

***Keywords*: *software deployment; IT-infrastructure; local network; software upgrade; interface; quiet installation*.**

Page 25

Developing of the E-government System based on Java for Online Voting

Alhawawsha Mohammad
Taras Shevchnko National University of Kyiv
Kyiv, Ukraine
mhawawsha@gmail.com

Anatoly V. Anisimov
Taras Shevchnko National University of Kyiv
Kyiv, Ukraine

***Abstract.* The current article is focused on proposing a new electronic voting system that will assist the government in conducting the online voting for the elections. The system will be secure with the end to end encryption to ensure that no data theft occurs during the transfer. The key demerit of the system is secure from the end of the government. It has to be ensured that all the external threats are kept at bay when using the online voting system. The secondary research approach is considered to mine the information required to conduct the study and develop the system. Overall, it can be stated that the new system will allow the government and the public to have improved voting experience if implemented successfully.**

***Keywords: Java-based E-government system, online voting, privacy theft, technological adoption, internet, encryption, hackers*.**

INFORMATION PROCESSING IN COMPLEX SYSTEMS

Page 31

Automatization of search for errors in raw data and creation of SDTM datasets for medical research

Kateryna Khololovych
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
k.9609.khololovich@gmail.com

Maksym Bukasov
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
bukasov@gmail.com

Abstract. **This article addresses the problem of automatization of search for errors in raw data and creation of SDTM (Standard Data Tabulation Model) datasets for medical research, and proposes a solution using the SAS system version 9.4, in the integrated development environment SAS Enterprise Guide 7.1. This solution will greatly improve performance and quality of data processing.**

Keywords: data processing for medical research, CDISC, Standard Data Tabulation Model (SDTM), SAS, SAS Enterprise Guide 7.1.

Page 33

Comments analyzing using machine learning

Oleh Pysarenko
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
oa.pisarenko@gmail.com

Anatoliy Doroshenko
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
a-y-doroshenko@ukr.net

***Abstract*. This paper demonstrates the result of implementing a service for analyzing text comments using machine learning. To do this, the approach of assembling machine learning algorithms was chosen, which included Gradient Boosting, Random Forest and logistic regression algorithms. The predictive model was trained and its accuracy was evaluated based on test data.**

Keywords*: *Machine learning, Tensorflow, Gradient Boosting, Random Forest, Logistic regression.

Page 35

Analysis of the calculus basis boundary for redundant codes

Poltorak Vadym
Igor Sikorskii Kyiv Polytechnic Institute
Kyiv, Ukraine
v.poltorak@kpi.ua

Abstract. **In this paper, algorithms of factorization of natural numbers are investigated and compared. The classic The paper describes an analysis of the calculus basis boundary for redundant codes, needed to protect data from distortions in communication channel. Here the criteria of minimum code distance limit for the codec quality rating was used, excluding the impact of the data transmission medium and modem properties to the data transmission efficiency estimation. It is implied here boundaries of Gilbert-Varshamov, Hamming, Plotkin and Elias.**

Keywords: boundaries for minimum code distance; data transmission.

INFORMATION SECURITY

Page 39

National EDS standards properties analysis

Serhii Romanchuk
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine

Abstract. **Existed digital signature systems were researched. The digital signature algorithms were analyzed with opinion of software implementation. It's done analysis that help understand strong and weak sides of existed algorithms. It shows basic schemes of digital signature creation.**
Keywords: Eliptic curves, digital signature, standards, cryptographic algorithm.

Page 41

Image authentification based on digital signatures

Nazarii Kalytiuk
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine

Abstract. P**aper analyzes different approaches of image authentication. The author investigates the possibility of using digital signature algorithms that ensure message authenticity for image authentication.**

Keywords: authentication, image authentication, digital signature, strict authentication, selective authentication.