

УДК 00462

А. М. Мельник, М. П. Дивак, Р. М. Пасічник

МЕТОД ВИЯВЛЕННЯ НЕАКТУАЛЬНОЇ ІНФОРМАЦІЇ В СЕРВІСНО-ОРІЄНТОВАНИХ КОРПОРАТИВНИХ СИСТЕМАХ НА ПРИКЛАДІ СИСТЕМ ОЦІНЮВАННЯ ЯКОСТІ ҐРУНТІВ

Західноукраїнський національний університет, Тернопіль

Анотація. У статті розглянуто важливу науково-прикладну задачу розробки методу виявлення неактуальної інформації, яка є актуальним напрямком розвитку та реалізації веб-орієнтованих інформаційних систем. Проведено аналіз сучасних методів та засобів оцінки неактуальної та недостовірної інформації в сервісно-орієнтованих корпоративних системах та виділено основні проблемні напрямки, які виникають в процесі їх функціонування. Розроблено метод фільтрування даних на основі метрики для оцінки актуальності інформації. Наведено приклад застосування метрики для оцінювання результатів використання різних сервісів аналізу якості ґрунтів та ґрунтових вод. Основними результатами досліджень, наведеними в статті є: метрика оцінки актуальності інформації, яка отримана з використанням сервісів в корпоративних інформаційних системах; метод фільтрування даних на основі метрики оцінки актуальності інформації в рамках досліджуваної предметної області. Особливістю розробленого методу є те, що він може бути реалізований як програмна надбудова до сервісно-орієнтованих інформаційних систем. Використання запропонованих інтелектуальних методів обробки даних, які отримують з використанням сервісів, дозволить підвищити ефективність аналізу неактуальної інформації та скоротити час визначення нерелевантних джерел її надання.

Ключові слова: веб-сервіси, API інтерфейси, неактуальна інформація, сервісно-орієнтована архітектура.

Аннотация. В статье рассмотрены важную научно-прикладную задачу разработки метода выявления неактуальной информации, является актуальным направлением развития и реализации веб-ориентированных информационных систем. Проведен анализ современных методов и средств оценки неактуальной и недостоверной информации в сервисно-ориентированных корпоративных системах и выделены основные проблемные направления, которые возникают в процессе их функционирования. Разработан метод фильтрации данных на основе метрики для оценки актуальности информации. Приведен пример применения метрики для оценки результатов использования различных сервисов анализа качества почв и грунтовых вод. Основными результатами исследований, приведенным в статье являются: метрика оценки актуальности информации, полученной с использованием сервисов в корпоративных информационных системах; метод фильтрации данных на основе метрики оценки актуальности информации в рамках исследуемой предметной области. Особенностью разработанного метода является то, что он может быть реализован как программная надстройка к сервисно-ориентированным информационным системам. Использование предложенных интеллектуальных методов обработки данных, получают с использованием сервисов, достаточно повысит эффективность анализа неактуальной информации и сократит время определения нежелательных источников ее предоставления.

Ключевые слова: веб-сервисы, API интерфейсы, неактуальная информация, сервисно-ориентированная архитектура.

Abstract. The article considers an important scientific and applied task of developing a method for detecting irrelevant information, which is an important area of development and implementation of web-based information systems. The analysis of modern methods and means of evaluation of irrelevant and unreliable information in service-oriented corporate systems is carried out and the main problem areas that arise in the process of their functioning are identified. A method of filtering data based on metrics to assess the relevance of information has been developed. An example of the application of metrics for evaluating the results of using various services for the analysis of soil and groundwater quality. The main results of research presented in the article are: metrics for assessing the relevance of information obtained using services in corporate information systems; method of data filtering based on the metrics of assessing the relevance of information within the studied subject area. The peculiarity of the developed method is that it can be implemented as a software add-on to service-oriented information systems. The use of the proposed intelligent methods of data processing, which are obtained with the use of services, will increase the efficiency of analysis of irrelevant information and reduce the time to identify irrelevant sources of its provision.

Key words: web services, API interfaces, nonrelevant information, outdated information, service-oriented architecture.

DOI: <https://doi.org/10.31649/1999-9941-2021-50-1-45-54>.

Вступ

В сучасному інформаційно-комунікаційному світі дуже важливим питанням є виявлення неактуальної інформації [1-4]. Цьому питанню присвячено багаточисленні науково-прикладні дослідження, а реалізація засобів виявлення такого типу інформації – актуальний напрямок розвитку веб-орієнтованих інформаційних систем [3,5-7].

Зростання доступності інформації, зокрема через використання різноманітних сервісів, викликає необхідність агрегування інформації шляхом фільтрації та ранжування джерел з яких вона отримується, виходячи із цілей кінцевого користувача [8,9].

В останні роки багато досліджень зосереджено на оцінці якості та зручності сервісів, а саме, як вони описуються, узгоджуються та формуються, однак недостатньо уваги приділяється контекстному аналізу самих сервісів, фільтруванню джерел, які можуть надавати неактуальну або недостовірну інформацію, що дуже важливо для підвищення ефективності та коректності структури сервісів та особливостей їх використання [10-14].

Велику частину інформаційних технологій займають корпоративні інформаційні системи [12,15-18]. Розвиток таких систем пов'язаний з автоматизацією практично усіх напрямків людської діяльності. Усе це сприяло розвитку інструментальних та концептуальних засобів для реалізації та підтримки бізнес-систем [18-22].

Одним із таких концептуальних засобів є сервісно-орієнтована архітектура (COA) [13,23-25]. В корпоративних системах сервіси - це прикладне програмне забезпечення, яке забезпечує комунікаційний зв'язок між клієнтами та надавачами послуг. Бізнес процеси в такій програмній архітектурі можна представити сервісами в сервісно-орієнтованій архітектурі [26-31].

Актуальність

Видобування інформації через використання сервісів є дуже важливим завданням, яке дозволяє серед інших можливостей – агрегувати інформацію, фільтрувати та ранжування відповідно до цілей користувача засобами веб-сервісів (на основі SOAP, або REST) або з використанням API методів [4,7,21].

Велика кількість доступних послуг та простих у використанні композиційних інструментів сприяють появі неактуальної інформації на етапі формування послуги та через недосконалу організацію комунікації організованих послуг [4,5,7,32].

Проблема появи неактуальної інформації на етапі формування послуги в основному пов'язані з розробниками та користувачами, які створюють композитні послуги [4-7,23,24]. Проблема комунікації складених послуг є критичною для споживачів послуг, які страждають від небажаних результатів, які надають системи в процесі врахування часових характеристик [5,9,11].

Ці питання стають більш актуальними, коли існує багато типів та обсягів даних, що надаються службами, а в Інтернеті та хмарному середовищі такі дані та послуги мають різні обмеження контексту та якості [7-10]. Невірне опрацювання таких даних може спричинити зниження якості надання таких послуг [4-6]. Розробка методів та програмних засобів виявлення неактуальної інформації в сервісно-орієнтованих системах є актуальним напрямком наукових досліджень.

Мета

Метою статті є підвищення ефективності оцінки неактуальної інформації, отриманої на основі сервісів в сервісно-орієнтованих корпоративних системах шляхом розробки метрики оцінки актуальності інформації та відповідного методу фільтрування даних.

Для досягнення поставленої у роботі мети необхідно розв'язати такі завдання:

- провести аналіз сучасних методів та засобів оцінки неактуальної та недостовірної інформації в сервісно-орієнтованих корпоративних системах та виділити основні проблемні напрямки, які виникають в процесі їх діяльності;
- запропонувати метрику оцінки актуальності інформації, яка отримана з використанням сервісів в корпоративних інформаційних системах;
- запропонувати метод фільтрування даних на основі метрики оцінки актуальності інформації в рамках досліджуваної предметної області;
- здійснити експериментальні дослідження запропонованих в роботі методів та засобів, оцінити їх ефективність.
- сформулювати напрямок подальших наукових досліджень, виходячи із отриманих у роботі напрацювань та результатів.

Задачі

1. Аналіз проблем опрацювання неактуальної та недостовірної інформації в сервісно-орієнтованих корпоративних системах.
2. Розробка метрики оцінки актуальності інформації в сервісно-орієнтованих системах.
3. Реалізація методу фільтрування даних на основі метрики оцінки актуальності інформації.
4. Проведення експериментальних досліджень для підтвердження ефективності запропонованих методів та засобів на прикладі систем оцінювання якості ґрунтів.

Розв'язання задач

В сучасних умовах при використанні зовнішніх сервісів важливим питанням є оцінка актуальності інформації, її контекстної складової, яка пов'язана з використанням сервісів (наприклад, володіння даними), що заважає фільтрувати неактуальну інформацію, з якою розробники та прості користувачі стикаються при виборі необхідних джерел даних [22,23].

На рис. 1 представлено структуру інформаційних потоків при формуванні сервісних послуг на етапі проектування, реалізації та використання.

У всіх зазначених потоках може існувати нерелевантна інформація. З неактуальною інформацією під час складання служби стикається здебільшого розробник, який спирається на величезні джерела інформації для побудови складених сервісів. Розробник сприймає цю проблему, коли інформація про сервіси та дані, що повертаються інструментом композиції сервісів, є непорівнянними або неповними. Під час виконання сервісних запитів простий користувач стикається з нерелевантними інформаційними проблемами, коли дані повертаються інтенсивно, а сервіси не можна порівняти або адекватно оцінити [24, 25].

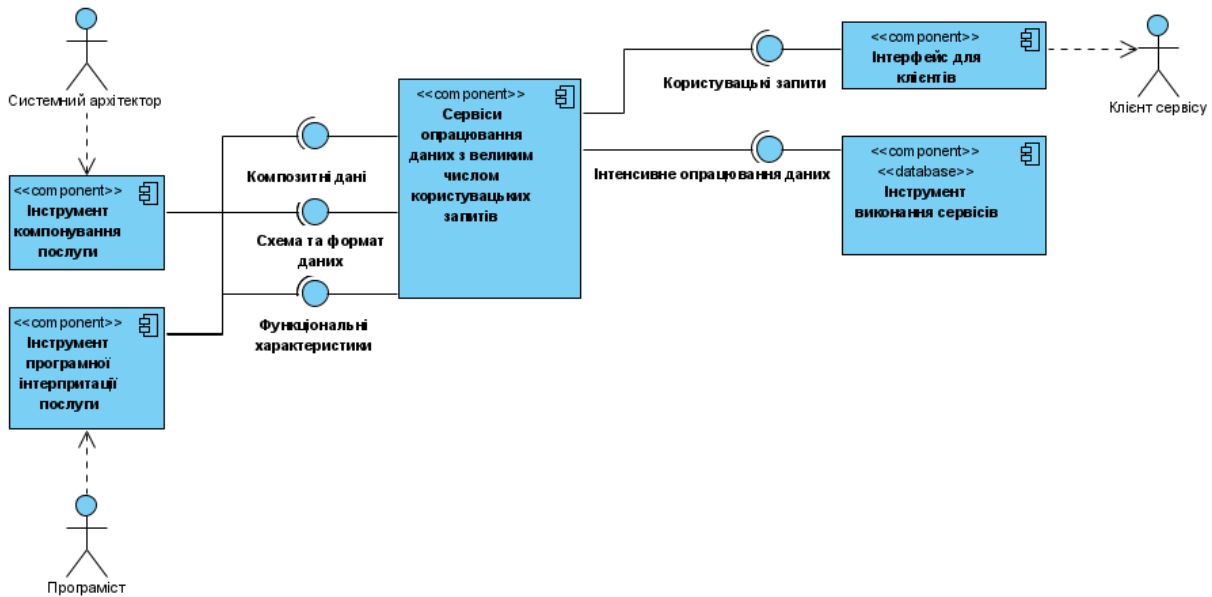


Рисунок 1 – Структура інформаційних потоків при формуванні сервісних послуг

У таблиці 1 наведено деякі приклади неактуальної інформації. Коли ми розглядаємо сценарії складання та виконання послуг у масштабі Інтернету та хмарних служб, в яких послуги надаються різними постачальниками, нерелевантна інформація, пов'язана з цими потоками зростає в багатьох аспектах завдяки різноманітності та складності сервісів та їх описів.

Таблиця 1 – Випадки неактуальної інформації

Причина	Наслідок
Деякі системи не надають опису структур даних	Неструктурований опис даних
Різні програмно-технічні специфікації	Опис різних версій специфікацій, неоднакова семантика при інтерпретації запитів
Неможливість отримати контекст безпосередньо із сервісу	Немає API для оцінки якості
Нерелевантне доменне ім'я, відсутність верифікації довіреної DNS зони	Відсутність можливості визначення достовірності джерела

Наявні нерелевантні інформаційні проблеми при складанні та виконанні сервісів можна вирішити, використовуючи декілька різних методів, наприклад, семантичний аналіз, аналіз даних та аналіз подібності [22,25]. У даній роботі основна увага приділяється питанням оцінки та обміну актуальною інформацією та контекстом інформації про сервіси. Контекстна інформація визначає ситуацію, за якої сервіси формуються та можуть використовуватися, а дані із запитів можуть бути опрацьовані, тоді як буде доведена ефективність джерела даних, яке використовується при формуванні сервісів. Розробка уніфікованого підходу, який поєднає контекстну та якісну складову інформації, отриманої із використанням сервісів дозволить підвищити як якість надання сервісних послуг, так і знизить ймовірність виникнення помилок в процесі розробки інших систем [25-28].

Метрика оцінки актуальності інформації в сервісно-орієнтованих системах

Досліджуючи якість та достовірність інформації, виділено основні характеристики, які можна застосувати для оцінки актуальності інформації про сервіси. Пропонується базову множину показників, які будуть використані для демонстрації ефективності методу оцінки неактуальної інформації в корпоративних інформаційних системах. Розглянемо детальніше ці показники.

Інтерпретованість (I_p) – визначає наявність метаданих та документації для правильної інтерпретації (функціональних та нефункціональних) властивостей сервісів. Ця метрика дозволяє побудувати систему зважених коефіцієнтів для класифікації службових документів, які використовуються при реалізації

сервісів. У таблиці 2 представлено опис властивостей, які використовуються для оцінки інтерпретованості.

Таблиця 2 – Типи показників та інформації, що використовується для оцінки інтерпретованості

Тип (Тип)	Інформація про сервіс	Приклади
P (Патерн)	Сервіси та схеми їх формування та використання	Web Services Description Language, WSDL, умови запиту та відповіді, моделі даних
D (Документація)	Документи	Опис методів API
Nh	Нефункціональні властивості	QoS, локація, локалізація, доступність, надійність, відповідь, час відгуку, актуальність відповіді ресурсу
Fd	Першоджерело	Версії патернів, нефункціональні характеристики
Rs	Достовірність джерела	Довірена доменна зона, зареєстрований програмний продукт в сертифікованому центрі

Виходячи із представлених типів інформації (таблиця 2), показник інтерпретованості I_p оцінюємо таким співвідношенням:

$$I_p = \frac{\sum estimation(type_i) \times v_i}{\sum v_i} \quad (1)$$

де $v(type_i) \in [0,1]$, $\sum v(type_i) = 1$ та $estimation(type_i) \in [0,1]$ – вагові коефіцієнти, які визначають значимість показників для оцінки інтерпретованості та коефіцієнти, які визначають ступінь доступності інформації, відповідно для кожного типу із набору $type_i \in \{P, D, Nh, Fd, Rs\}$.

Оцінка $estimation(type_i)$ може бути отримана з використанням спеціальних програмних засобів (парсерів, аналізаторів), які збирають та класифікують інформацію такого аналізованого типу.

Іншими важливими показниками, від яких залежить актуальність інформації, яку отримуємо на основі сервісів, є повнота F та своєчасність T . Показник повноти (2) визначає відношення між отриманими нефункціональними показниками Nh_p та їх мінімальним визначеним набором Nh_{min} .

Показник Nh_{min} включає всі показники Nh_p , які вважаються релевантними для вибору сервісу, який надається компанією чи організацією, яка, своєю чергою, є власником інформаційної системи. Наприклад, $Nh_{min} \{A, R, An, Tm\}$, де A – доступність, R – надійність, An – відповідь, Tm – тривалість отримання відповіді.

$$F = 1 - \frac{\|Nh_p \cap Nh_{min}\|}{Nh_{min}} \quad (2)$$

Показник своєчасності T визначає наскільки актуальним є опис нефункціональних характеристик Nh . Показник своєчасності (3) визначається як відношення перевірених часових характеристик T_{exp} та їх імплементованих значень T_{norm} :

$$T = 1 - \min\left(\frac{T_{exp}}{T_{norm}}, 1\right) \quad (3)$$

Метод фільтрування даних на основі метрики оцінки актуальності інформації

В сучасних корпоративних системах, які мають власні API для надання сервісів зовнішнім клієнтам не підтримують фільтрацію інформації про сервіси на основі показників якості, тому методи вибору надавачів сервісів повинні містити повну та відфільтровану службу інформацію.

На основі метрики оцінки актуальності інформації можна сформувавши метод фільтрування даних, який базується на двох основних типах описів: на основі всіх службових документів або на основі описів нефункціональних характеристик Nh .

Для реалізації методу фільтрування даних, за допомогою метрики необхідно встановити зважені коефіцієнти та визначити коефіцієнти для різних типів службових документів. Встановлення цих коефіцієнтів можна здійснити, наприклад, на основі службової інформації, яка надається власником сервісної системи, але зважені коефіцієнти для нефункціональних властивостей будуть залежати від сформованого запиту [15,16]. На рис. 2 представлено процедуру визначення зважених коефіцієнтів для нефункціональних властивостей.

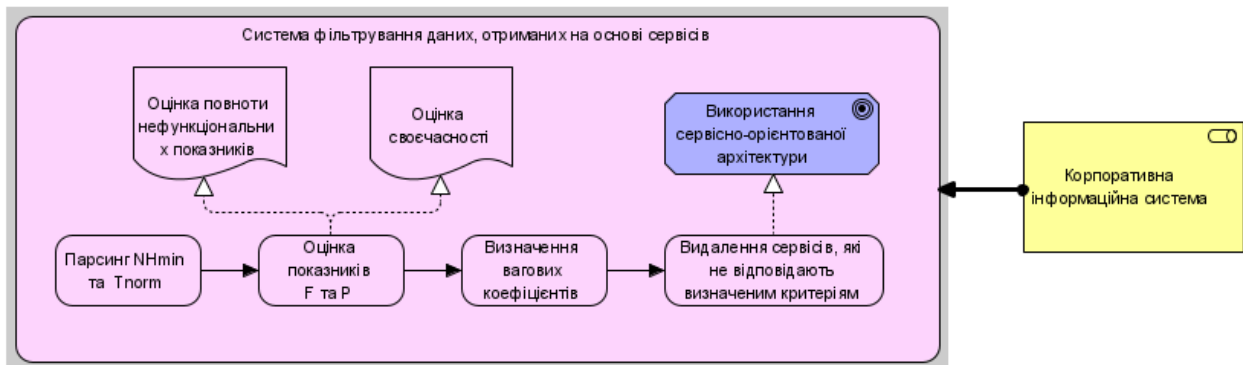


Рисунок 2 – Схема методу фільтрування даних, на основі сервісів

Як бачимо з наведеної схеми, метод фільтрування даних передбачає такі етапи:

- на основі показників оцінки актуальності інформації, які сформовані в метриці інтерпретованості I_p необхідно встановити вагові коефіцієнти та оцінити кожен тип (P, D, Nh, Fd, Rs) показників для кожного сервісу;
- встановлення залежності між інформацією, яка надана корпоративною системою, та показниками, які отримані на основі конкретних запитів;
- на наступному етапі, виходячи із сформованого набору запитів необхідно визначити показники $Nh_{min} \{A, R, An, Tm\}$ та T_{norm} ;
- проведення оцінки на основі метрики оцінки актуальності інформації (показники повноти та своєчасності);
- встановлення порогів для фільтрації нерелевантних джерел надання інформації через сервіси;
- фільтрація сервісів, інформація з яких знаходиться нижче встановленої межі.

Метод фільтрування даних доцільно використовувати перед вибором власника сервісів (корпоративної інформаційної системи) з метою якісного формування комунікаційної складової для отримання лише актуальної та достовірної інформації.

Експериментальні дослідження

Для підтвердження ефективності запропонованих методів було проведено ряд експериментальних досліджень. Суть експериментів полягала у визначенні неактуальних сервісів на основі інформації, яка може бути отримана через зовнішні API інтерфейси, які надають великі відкриті системи на ринку оцінки якості ґрунтів.

Вибір даної предметної області в якості експериментальних досліджень обумовлений використанням запропонованих напрацювань в рамках науково-дослідного проекту «Математичне та програмне забезпечення для ідентифікації та моніторингу особливо небезпечних джерел забруднення ґрунту та ґрунтових вод», № держреєстрації 0120U102040) [31].

Було проаналізовано 10 API-інтерфейсів для оцінки якості ґрунтів, які були повернуті пошуковиком Google при формуванні запиту – «API soil assessment». Для аналізу показників інтерпретованості, повноти та своєчасності із використанням програмних засобів для аналізу різного роду контенту, включаючи сервісний інтерфейс (наприклад, файл WSDL), інформацію про документацію, рейтинг або відгуки користувачів.

Для кожного сервісу покладаємо $estimation(P) = 1$, оскільки їх патерни (схеми) в основному є файлами WSDL, інформація з яких дозволяє класифікувати службову документацію на категорії - {неструктурована, частково структурована, структурована}, що еквівалентно такій оцінці:

$estimation(D) = \{0, 0.5, 1\}$. Мінімальний набір нефункціональних параметрів, які аналізувалися $Nh_{min} = \{доступність, час\ відгуку, відповідь\}$.

На рис. 3 представлено результати оцінки показника інтерпретованості I_p використовуючи різні набори вагових коефіцієнтів для показників:

1) $v(P) = 0.2, v(D) = 0.2, v(Nh) = 0.2, v(Fd) = 0.2, v(Rs) = 0.2$ – варіант, коли усі типи мають однакову вагу.

2) $v(P) = 0.5, v(D) = 0.25, v(Nh) = 0.25, v(Fd) = 0, v(Rs) = 0$ – варіант, коли P – найважливіший тип, D (документація) та Nh (нефункціональні характеристики) мають однакову вагу, а показники Fd та Rs не враховуються, або ступінь їх впливу оцінити складно, враховуючи специфіку предметної області.

3) $v(P) = 0, v(D) = 0.5, v(Nh) = 0.5, v(Fd) = 0, v(Rs) = 0$ варіант, коли аналізується тільки тип D (документація) та Nh (нефункціональні характеристики), які мають однакову вагу при обчисленні інтерпретованості I_p .

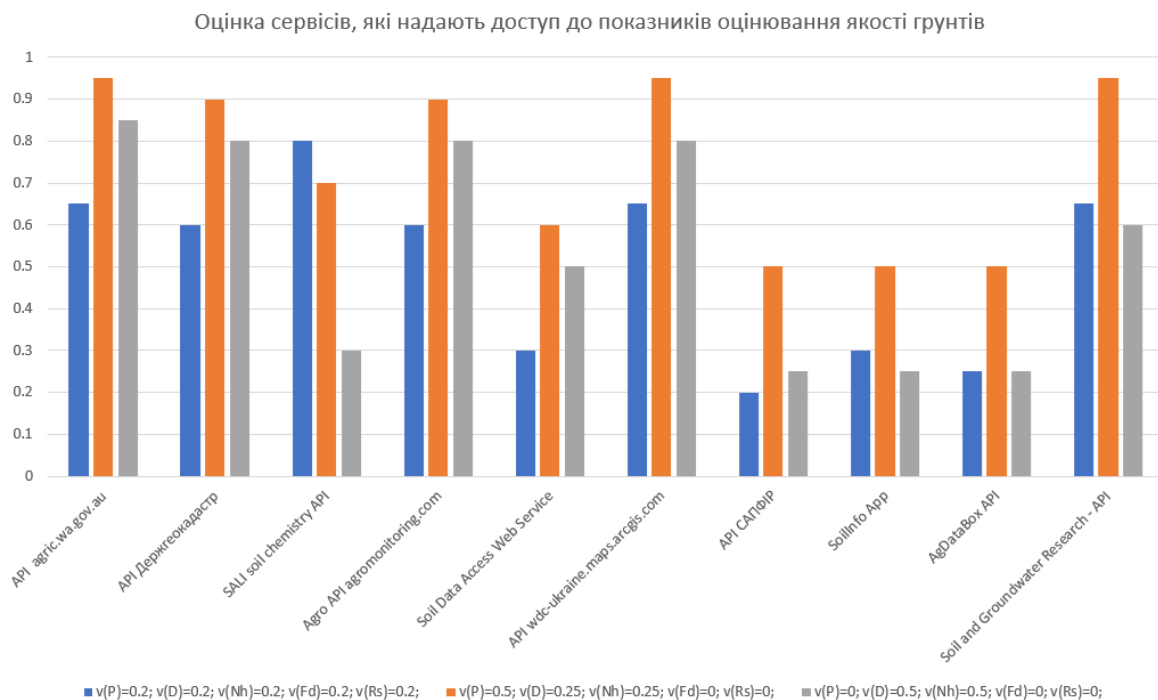


Рисунок 3 – Результати оцінки сервісів, які надають API-інтерфейси для оцінки якості та стану ґрунтів, їх можливого забруднення

Як бачимо з рисунка 3, значення показника інтерпретованості змінюються залежно від коефіцієнтів зважування факторів, які відображають різні вимоги, що можуть ставитися клієнтами в процесі отримання та використання вказаних сервісів. Проте, у всіх трьох випадках наборів вагових коефіцієнтів, які зазначено вище, більшість API-інтерфейсів мають високі значення показника інтерпретованості I_p (зокрема, $I_p \geq 0.5$). Отже, іншу частину сервісів (з меншою інтерпретованістю) можна включити із подальшого аналізу. Використання зазначеного підходу дозволить пришвидшити процес обрання комунікаційних сервісів в залежності від конкретних вимог.

Результати, які отримані під час проведених досліджень використано з метою отримання якісного допоміжного інструментарію для ідентифікації надійних (в інформаційному розумінні) сервісів, які надають інформацію щодо забруднення ґрунту та ґрунтових вод. Така інформація необхідна для розв'язування задач ідентифікації, моделювання та моніторингу особливо небезпечних джерел забруднення ґрунту та ґрунтових вод.

Майбутні дослідження

Подальші дослідження в рамках запропонованих підходів будуть зосереджуватися на розширенні набору показників якості, актуальності та достовірності інформації. Важливим напрямком також є апробація метрики оцінки актуальності та достовірності інформації на інших джерелах її надання та зберігання, наприклад, аналіз інструментів, які базуються на використанні сучасних NoSQL систем управління базами даних. Це доволить значно розширити варіативність використання, а також забезпечить повноту аналізованих інструментів аналізу інформації та засобів її зберігання.

Висновки

Актуальність інформації про сервіси в сервісно-орієнтованих архітектурах відіграє важливу роль для виявлення та зменшення недостовірної інформації та підтримки ефективних методів вибору інформаційного джерела та комунікаційних засобів формування самих сервісів. Відсутність належного способу управління якістю та контекстом у веб-сервісах спонукало до розробки метрики оцінки актуальності інформації в сервісно-орієнтованих системах, а також методу фільтрування даних на основі даної метрики. Використання запропонованих в рамках даної роботи методів дозволило зменшити ступінь отримання неактуальної інформації в процесі формування та виконання відповідних сервісів.

Створені методи програмно реалізовані як надбудова до сучасних рішень і можуть ефективно використовуватися в сервісно-орієнтованих інформаційних системах. Достовірність запропонованих метрики та методів фільтрування даних підтверджена експериментально на прикладі оцінки сервісів, які надають доступ до оцінювання якості ґрунтів та ґрунтових вод.

Список літератури

- [1] Zhou, Xinyi & Zafarani, Reza, «A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities», 2 Dec 2018.
- [2] Ahmad, Iftikhar & Yousaf, Muhammad & Yousaf, Suhail & Ahmad, Muhammad Ovais, «Fake News Detection Using Machine Learning Ensemble Methods», *Complexity in Deep Neural Networks*, 2020.
- [3] S. A. García, G. G. García, M. S. Prieto, A. J. M. Guerrero, and C. R. Jiménez, «The impact of term fake news on the scientific community scientific performance and mapping in web of science», *Social Sciences*, vol. 9, no. 5, 2020.
- [4] S. Akhtar, F. Hussain, F. R. Raja et al., «Improving mispronunciation detection of arabic words for non-native learners using deep convolutional neural network features», *Electronics*, vol. 9, no. 6, 2020.
- [5] Ahmed, S., Hinkelmann, K., Corradini, F., «Combining machine learning with knowledge engineering to detect fake news in social networks – a survey», In: *Proceedings of the AAAI 2019 Spring Symposium*, vol. 12 (2019).
- [6] B. Marr, Coronavirus fake news: how Facebook, Twitter, and Instagram are tackling the problem. *Forbes* (2020). [Online]. Available: <https://www.forbes.com/sites/bernardmarr/2020/03/27/finding-the-truth-about-covid-19-how-facebook-twitter-and-instagram-are-tackling-fake-news>.
- [7] H. Sparks, H. Frishberg, Facebook gives step-by-step instructions on how to spot fake news (2020). [Online]. Available: <https://nypost.com/2020/03/26/facebook-gives-step-by-step-instructions-on-how-to-spot-fake-news/>
- [8] Zhou, X., Zafarani, R.: Fake news: a survey of research, detection methods, and opportunities (2018). arXiv preprint arXiv:1812.00315.
- [9] Ian Chadd, Emel Filiz-Ozbay, Erkut Y. Ozbay. The relevance of irrelevant information. *Experimental Economics*, 2020; DOI: 10.1007/s10683-020-09687-3.
- [10] N. A. D. L. Perera, C. Priyankara and D. W. R. S. Jayasekara, «Identifying Irrelevant Answers in Web Based Question Answering Systems», *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*, Colombo, Sri Lanka, 2020, pp. 11-16, doi: 10.1109/ICTer51097.2020.9325449.
- [11] Alkhodair SA, Ding SH, Fung BC, Liu J (2020) Detecting breaking news rumors of emerging topics in social media. *Inf Process Manag* 57:102018.
- [12] Braşoveanu AM, Andonie R (2019) Semantic fake news detection: a machine learning perspective. In: *International work-conference on artificial neural networks*, Springer, pp 656–667.
- [13] А. М. Мельник, О. З. Миць, «Підвищення ефективності пошуку документів в інтернет із врахуванням подібності веб-сторінок», *Сучасні комп'ютерні інформаційні технології : матеріали III Всеукр. шк.-семінару молодих вчен. і студ. АСІТ'2013 [м. Тернопіль, 17-18 трав. 2013 р.]*, Тернопіль: THEU, 2013, с. 212.
- [14] Li Y, Nie X, Huang R (2018b) Web spam classification method based on deep belief networks. *Expert Syst Appl* 96:261–270.
- [15] Liu Y, Xu S (2016) Detecting rumors through modeling information propagation networks in a social media environment. *IEEE Trans Comput Soc Syst* 3:46–62.

- [16] Naseem U, Razzak I, Musial K, Imran M (2020) Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Gener Comput Syst* 6:91.
- [17] Shu K, Wang S, Lee D, Liu H (2020) Mining disinformation and fake news: concepts, methods, and recent advancements. *arXiv preprint arXiv:2001.00623*.
- [18] Tang D, Qin B, Liu T (2015) Document modeling with gated recurrent neural network for sentiment classification. In: *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp 1422–1432.
- [19] Allen, Jennifer & Howland, Baird & Mobius, Marine & Rothschild, David & Watts, Duncan. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*. 6. eaay3539. 10.1126/sciadv.aay3539.
- [20] М. П. Дивак, А. М. Мельник, О. А. Папа, «Математичне та програмне забезпечення інтелектуального модуля прикладних програмних систем для надання адміністративних послуг щодо проведення екологічної експертизи», *Інформаційні технології та комп'ютерна інженерія*, 49(3), с. 66–76, 2020.
- [21] H. Ahmed, I. Traore, and S. Saad, «Detecting opinion spams and fake news using text classification», *Security and Privacy*, vol. 1, no. 1, 2018.
- [22] Batini, C., Scannapieco, M.: *Data Quality: Concepts, Methodologies and Techniques (Data-Centric Systems and Applications)*. Springer-Verlag New York, Inc., Secaucus (2006).
- [23] Dyvak, M., Papa, O., Melnyk, A., Pukas, A., Porplytsya, N., Rot, A. «Interval Model of the Efficiency of the Functioning of Information Web Resources for Services on Ecological Expertise», *Mathematics*, 8(12), 2116, 2020.
- [24] О. О. Лисенко, Р. І. Кокітко, «Модель аналізу неструктурованої інформації для побудови бази знань корпоративної інформаційної системи», *Сучасні комп'ютерні інформаційні технології: матеріали III Всеукр. шк.-семінару молодих вчен. і студ. АСІТ'2013 [м. Тернопіль, 17-18 трав. 2013 р.]*, Тернопіль: ТНЕУ, 2013, с. 211.
- [25] Truong HL., Comerio M., Maurino A., Dustdar S., De Paoli F., Panziera L. (2010) On Identifying and Reducing Irrelevant Information in Service Composition and Execution. In: Chen L., Triantafillou P., Suel T. (eds) *Web Information Systems Engineering – WISE 2010*. WISE 2010. Lecture Notes in Computer Science, vol 6488. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17616-6_7.
- [26] M. Susla, R. Pasichnyk, A. Melnyk, N. Pasichnyk, O. Vasykiv and O. Androshchuk, «Formalization of Scientific Researches Results in Corporate Knowledge Bases As a Tool of Their Accumulation», *2020 10th International Conference on Advanced Computer Information Technologies (ACIT)*, Deggendorf, Germany, 2020, pp. 488-491, doi: 10.1109/ACIT49673.2020.9208863.
- [27] A. Kovbasistyi, A. Melnyk, M. Dyvak, V. Brych and I. Spivak, «Method for detection of non-relevant and wrong information based on content analysis of web resources», *2017 XIIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, Lviv, 2017, pp. 154–156, doi: 10.1109/MEMSTECH.2017.7937555.
- [28] A. Pukas, A. Simak, O. Syrnyk, L. Horal, V. Shyjko and O. Papa, «Software Module for Data Correctness and Completeness Control in the Academic Staff Performance Appraisal System», *2019 9th International Conference on Advanced Computer Information Technologies (ACIT)*, Ceske Budejovice, Czech Republic, 2019, pp. 277-280, doi: 10.1109/ACITT.2019.8779999.
- [29] M. P. Dyvak, A. V. Kovbasistyi, A. M. Melnyk, L. Y. Turchyn, Y. O. Martsenyuk, «System for web resources content structuring and recognizing with the machine learning elements», *Радіоелектроніка, інформатика, управління*, 3 (Груд 2018). DOI:<https://doi.org/10.15588/1607-3274-2018-3-14>.
- [30] A. Kovbasistyi, A. Melnyk, M. Dyvak, V. Brych et al., «Method for detection of non-relevant and wrong information based on content analysis of web resources», *XIIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, Lviv, 2017, pp. 154–156. DOI: 1109/MEMSTECH.2017.7937555.
- [31] М. П. Дивак, А. М. Мельник, А. В. Ковбасистий, О. А. Папа, «Підхід до математичного моделювання ефективності web-ресурсів», *Оптико-електронні інформаційно-енергетичні технології*, 38, 2 (Бер 2020), 29–37. DOI:<https://doi.org/10.31649/1681-7893-2019-38-2-29-37>.

Стаття надійшла: 22.02.2021

References

- [1] Zhou, Xinyi & Zafarani, Reza, «A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities», 2 Dec 2018.
- [2] Ahmad, Iftikhar & Yousaf, Muhammad & Yousaf, Suhail & Ahmad, Muhammad Ovais, «Fake News Detection Using Machine Learning Ensemble Methods», *Complexity in Deep Neural Networks*, 2020.

- [3] S. A. García, G. G. García, M. S. Prieto, A. J. M. Guerrero, and C. R. Jiménez, «The impact of term fake news on the scientific community scientific performance and mapping in web of science», *Social Sciences*, vol. 9, no. 5, 2020.
- [4] S. Akhtar, F. Hussain, F. R. Raja et al., «Improving mispronunciation detection of arabic words for non-native learners using deep convolutional neural network features», *Electronics*, vol. 9, no. 6, 2020.
- [5] Ahmed, S., Hinkelmann, K., Corradini, F., «Combining machine learning with knowledge engineering to detect fake news in social networks – a survey», In: *Proceedings of the AAAI 2019 Spring Symposium*, vol. 12 (2019).
- [6] B. Marr, Coronavirus fake news: how Facebook, Twitter, and Instagram are tackling the problem. Forbes (2020). [Online]. Available: <https://www.forbes.com/sites/bernardmarr/2020/03/27/finding-the-truth-about-covid-19-how-facebook-twitter-and-instagram-are-tackling-fake-news>.
- [7] H. Sparks, H. Frishberg, Facebook gives step-by-step instructions on how to spot fake news (2020). [Online]. Available: <https://nypost.com/2020/03/26/facebook-gives-step-by-step-instructions-on-how-to-spot-fake-news/>
- [8] Zhou, X., Zafarani, R.: Fake news: a survey of research, detection methods, and opportunities (2018). arXiv preprint arXiv:1812.00315.
- [9] Ian Chadd, Emel Filiz-Ozbay, Erkut Y. Ozbay. The relevance of irrelevant information. *Experimental Economics*, 2020; DOI: 10.1007/s10683-020-09687-3.
- [10] N. A. D. L. Perera, C. Priyankara and D. W. R. S. Jayasekara, «Identifying Irrelevant Answers in Web Based Question Answering Systems», *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*, Colombo, Sri Lanka, 2020, pp. 11-16, doi: 10.1109/ICTer51097.2020.9325449.
- [11] Alkhodair SA, Ding SH, Fung BC, Liu J (2020) Detecting breaking news rumors of emerging topics in social media. *Inf Process Manag* 57:102018.
- [12] Braşoveanu AM, Andonie R (2019) Semantic fake news detection: a machine learning perspective. In: *International work-conference on artificial neural networks*, Springer, pp 656–667.
- [13] A. M. Melnyk, O. Z., «Pidvyshchennia efektyvnosti poshuku dokumentiv v internet iz vrakhuvanniam podobnosti veb-storinok», *Suchasni kompiuterni informatsiini tekhnologii : materialy III Vseukr. shk.-seminaru molodykh vchen. i stud. ASIT2013 [m. Ternopil, 17-18 trav. 2013 r.]*, Ternopil: TNEU, 2013, s. 212.
- [14] Li Y, Nie X, Huang R (2018b) Web spam classification method based on deep belief networks. *Expert Syst Appl* 96:261–270.
- [15] Liu Y, Xu S (2016) Detecting rumors through modeling information propagation networks in a social media environment. *IEEE Trans Comput Soc Syst* 3:46–62.
- [16] Naseem U, Razzak I, Musial K, Imran M (2020) Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Gener Comput Syst* 6:91.
- [17] Shu K, Wang S, Lee D, Liu H (2020) Mining disinformation and fake news: concepts, methods, and recent advancements. arXiv preprint arXiv:2001.00623.
- [18] Tang D, Qin B, Liu T (2015) Document modeling with gated recurrent neural network for sentiment classification. In: *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp 1422–1432.
- [19] Allen, Jennifer & Howland, Baird & Mobius, Marine & Rothschild, David & Watts, Duncan. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*. 6. eaay3539. 10.1126/sciadv.aay3539.
- [20] M. P. Dyvak, A. M. Melnyk, O. A. Papa, «Matematychna ta prohramne zabezpechennia intelektualnoho modulua prykladnykh prohramnykh system dlia nadannia administratyvnykh posluh shchodo provedennia ekolohichnoi ekspertyzy», *Informatsiini tekhnologii ta kompiuterna inzheneriia*, 49(3), s. 66–76. 2020
- [21] H. Ahmed, I. Traore, and S. Saad, «Detecting opinion spams and fake news using text classification», *Security and Privacy*, vol. 1, no. 1, 2018.
- [22] Batini, C., Scannapieco, M.: *Data Quality: Concepts, Methodologies and Techniques (Data-Centric Systems and Applications)*. Springer-Verlag New York, Inc., Secaucus (2006).
- [23] Dyvak, M., Papa, O., Melnyk, A., Pukas, A., Porplytsya, N., Rot, A. «Interval Model of the Efficiency of the Functioning of Information Web Resources for Services on Ecological Expertise», *Mathematics*, 8(12), 2116, 2020.
- [24] O. O. Lysenko, R. I. Kokitko, «Model analizu nestrukurovanoj informatsii dlia pobudovy bazy znan korporatyvnoji informatsiinoji systemy», *Suchasni kompiuterni informatsiini tekhnologii : materialy III Vseukr. shk.-seminaru molodykh vchen. i stud. ASIT2013 [m. Ternopil, 17-18 trav. 2013 r.]*, Ternopil: TNEU, 2013, s. 211.

- [25] Truong HL., Comerio M., Maurino A., Dustdar S., De Paoli F., Panziera L. (2010) On Identifying and Reducing Irrelevant Information in Service Composition and Execution. In: Chen L., Triantafillou P., Suel T. (eds) Web Information Systems Engineering – WISE 2010. WISE 2010. Lecture Notes in Computer Science, vol 6488. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17616-6_7.
- [26] M. Susla, R. Pasichnyk, A. Melnyk, N. Pasichnyk, O. Vasykiv and O. Androshchuk, «Formalization of Scientific Researches Results in Corporate Knowledge Bases As a Tool of Their Accumulation», 2020 10th International Conference on Advanced Computer Information Technologies (ACIT), Deggendorf, Germany, 2020, pp. 488-491, doi: 10.1109/ACIT49673.2020.9208863.
- [27] A. Kovbasistyi, A. Melnyk, M. Dyvak, V. Brych and I. Spivak, «Method for detection of non-relevant and wrong information based on content analysis of web resources», 2017 XIIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), Lviv, 2017, pp. 154–156, doi: 10.1109/MEMSTECH.2017.7937555.
- [28] A. Pukas, A. Simak, O. Syrnyk, L. Horal, V. Shyjko and O. Papa, «Software Module for Data Correctness and Completeness Control in the Academic Staff Performance Appraisal System», 2019 9th International Conference on Advanced Computer Information Technologies (ACIT), Ceske Budejovice, Czech Republic, 2019, pp. 277-280, doi: 10.1109/ACITT.2019.8779999.
- [29] M. P. Dyvak, A. V. Kovbasistyi, A. M. Melnyk, L. Y. Turchyn, Y. O. Martsenyuk, «System for web resources content structuring and recognizing with the machine learning elements», *Radioelektronika, informatyka, upravlinnia*, 3 (Hrud 2018). DOI:<https://doi.org/10.15588/1607-3274-2018-3-14>.
- [30] A. Kovbasistyi, A. Melnyk, M. Dyvak, V. Brych et al., «Method for detection of non-relevant and wrong information based on content analysis of web resources», XIIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), Lviv, 2017, pp. 154–156. DOI: 1109/MEMSTECH.2017.7937555.
- [31] M. P. Dyvak, A. M. Melnyk, A. V. Kovbasistyi, O. A. Papa, «Pidkhid do matematychnoho modeliuвання efektyvnosti web-resursiv», *Optyko-elektronni informatsiino-enerhetychni tekhnolohii*, 38, 2 (Ber 2020), 29–37. DOI:<https://doi.org/10.31649/1681-7893-2019-38-2-29-37>.

Відомості про авторів

Мельник Андрій Миколайович – кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук.

Дивак Микола Петрович – доктор технічних наук, професор, Заслужений діяч науки і техніки, декан факультету комп'ютерних інформаційних технологій.

Пасічник Роман Мирославович – доктор технічних наук, доцент, професор кафедри економічної кібернетики та інформатики.

А. Н. Мельник, Н. П. Дивак, Р. М. Пасечник

МЕТОДЫ ВЫЯВЛЕНИЯ НЕАКТУАЛЬНОЙ ИНФОРМАЦИИ В СЕРВИСНО-ОРИЕНТИРОВАННОЙ КОРПОРАТИВНОЙ СИСТЕМЫ НА ПРИМЕРЕ СИСТЕМ ОЦЕНКИ КАЧЕСТВА ГРУНТОВ

Западноукраинский национальный университет, Тернополь

A. M. Melnyk, M. P. Dyvak, R. M. Pasichnyk

METHOD OF DETECTION OF UPDATED INFORMATION IN SERVICE-ORIENTED CORPORATE SYSTEMS ON THE EXAMPLE OF SOIL QUALITY ASSESSMENT SYSTEMS

West Ukrainian National University, Ternopil