УДК 681.3

ПРОГРАМНИЙ МОДУЛЬ ДЛЯ АНАЛІЗУ ЛОГІВ ПОДІЙ НА ОСНОВІ ГРАФІВ

Гіщинський Б.О.[1)], Манжула В.І.[2)]
Тернопільський національний економічний університет
[1)] магістрант; [2)] к.т.н., доцент

I. Постановка проблеми

Аналіз логів подій – це процес перетворення вихідних даних логу в інформацію для вирішення завдань порівняння, аналізу та оптимізації. Слід відзначити, що на сьогоднішній день існує безліч програмних продуктів для аналізу логів. [1,2,3] Крім того обсяг логів подій, які генеруються щодня набувають все більшого значення для організацій, які повинні реєструвати, опрацьовувати інформацію. Вкрай важливо виконувати в режимі реального часу моніторинг, аналіз та звітність логів подій для вирішення будь-яких інцидентів або проблем в області безпеки і боротьби з загрозами для безперервного функціонування бізнесу.

II. Мета роботи

В рамках даної роботи розглядається задача знаходження відмінностей в логах подій. Для того, щоб експерт, який працює над аналізом процесів, міг легко сприймати отриману інформацію про знайдені відмінності в логах подій, результати роботи повинні бути представлені у зручному користувачам вигляді. Метою даної роботи є візуалізація знайдених відмінностей в логах подій на основі графів. На рис 1 показано структуроване дерево логу подій. З допомогою цього дерева, можна зробити наступні висновки про логи подій:

- Процес складається з різних випадків
- Кожен випадок містить в собі різні події
- Події всередині кожного випадку відсортовані
- У подій можуть бути атрибути. Прикладами атрибутів можуть бути дії, час, вартість, ресурс.

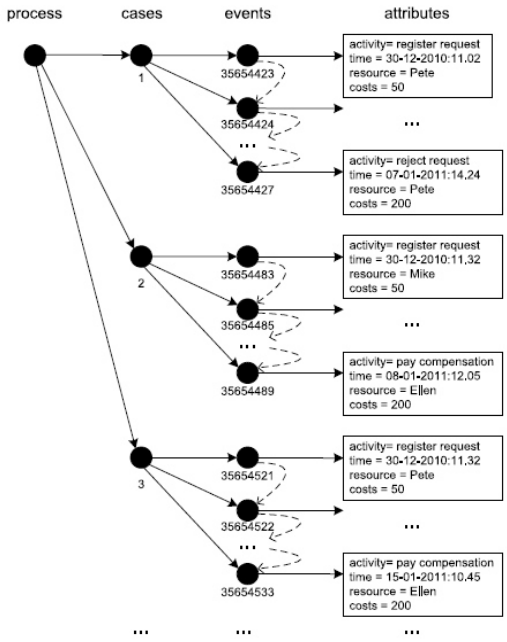

Рисунок 1 - Дерево логу

III. Особливості програмної реалізації

Дана програма призначена для порівняння логів подій за розробленим в рамках роботи алгоритмом. Вхідні дані використовують логи подій у форматі XES. XES є стандартом XML, як основа для логів подій.[4] Його мета полягає в тому, щоб забезпечити загальновизнаний формат для обміну даними логів подій між інструментами і доменами додатків. Відповідно до опису стандарту XES [7], об'єкт даних log на самому верхньому рівні формату містить всю інформацію про події, яка відноситься до одного певного процесу (наприклад, процес обробки страхових запитів). для опису даного об'єкта існує тег <log>. Лог може містити довільну кількість об'єктів послідовностей подій, в тому числі нульову. Кожна послідовність описує виконання одного конкретного випадку процесу з логу події – наприклад, обробка конкретного страхового випадку, одне відвідування веб-сайту конкретним користувачем. Ім'я тегу для зберігання цих даних – <trace>. Кожна послідовність в свою чергу містить довільне число об'єктів-подій, так само може не містити жодного. Події є неподільною частинкою активності, яка сталася в перебігу виконання даного процесу. Прикладами подій є – запис персональної інформації про клієнта в базу, імпорт або експорт зображення з певного сайту. Ім'я тегу для даного об'єкта – <event>.

Описані об'єкти (лог, послідовність, подія) не містять ніякої інформації, а тільки задають структуру документа. Вся змістовна інформація в логу подій зберігається в так званих атрибутах логу. У кожного з атрибутів існує рядок-ключ. Кожен з логів, послідовностей, подій містять довільне число атрибутів, які поділяються на шість типів: рядок, дата, ціле число, число з плаваючою крапкою,

логічний вираз, ідентифікатор. Як додаткова інформації існують вкладені атрибути, створені для того, щоб формат логу був більш гнучким. Вкладеними називають такі атрибути, у яких можуть бути атрибути-нащадки.

На рис. 2 зображена UML діаграма головних компонентів формату XES. У лівій частині зображення все те, що відноситься до базового стандарту XES. Це обов'язково має бути присутнім в будь-якому відповідному XES стандарту логу подій. Базовий стандарт визначає основну структуру логу подій, в той час як всі дані, що стосуються процесу і що зберігаються в логах подій зберігаються в атрибутах і визначені за допомогою розширення базового стандарту, що зображено у правій частині діаграми. На рис. 3 наведено приклад документу формату XES, який зберігає інформацію про процес з двох подій [5]. На основі логів подій будуються системи переходів, на основі яких відбувається пошук відмінностей. Розроблений алгоритм пошуку відмінностей заснований на методі пошуку в ширину на графі з деякими модифікаціями щодо поставленого завдання. За допомогою алгоритму можна знаходити відмінності в логах подій відносно однієї події: видалена (пропущена), додана (зайва) та змінена подія. Також є можливість визначати повністю ідентичні послідовності подій.

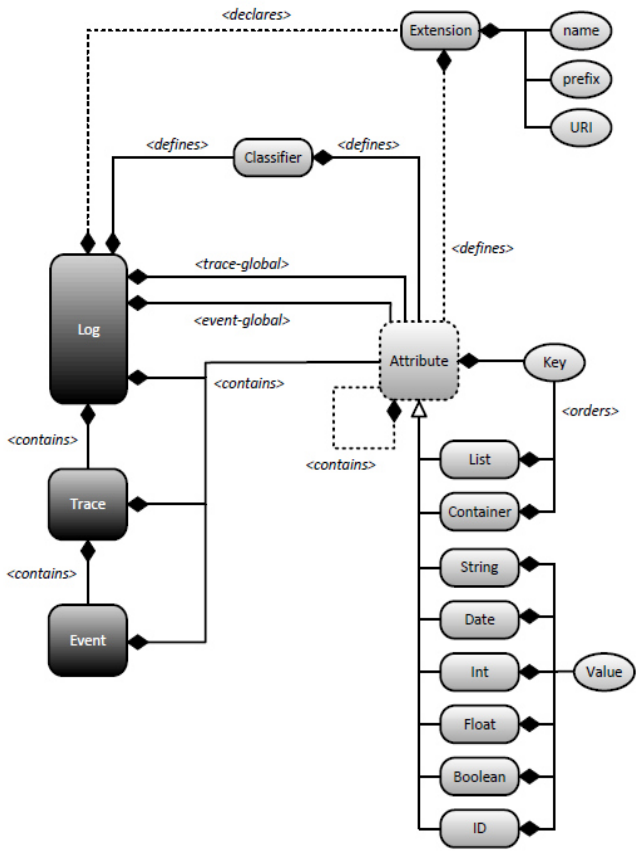

Рисунок 2 - UML-діаграма головних компонентів формату XES

Висновки

У ході роботи було виконано дослідження існуючих підходів для аналізу логів подій. На основі даного аналізу встановлено, що для їх порівняння, аналізу або оптимізації необхідно використовувати підхід на основі теорії графів, з деякими його модифікаціями. Розроблений алгоритм і його реалізація в рамках роботи може бути використаний для порівняння двох логів подій експертами в області роботи з бізнес-процесами.

```
<?xml version="1.0" encoding="UTF-8" ?>
<log xes.version="2.0" xes.features="arbitrary-depth" xmlns="http://www.xes-standard.org
    /">
    <extension name="Concept" prefix="concept" uri="http://www.xes-standard.org/concept.
        xesext"/>
    <extension name="Time" prefix="time" uri="http://www.xes-standard.org/time.xesext"/>
    <global scope="trace">
        <string key="concept:name" value=""/>
    </global>
    <global scope="event">
        <string key="concept:name" value=""/>
        <date key="time:timestamp" value="1970-01-01T00:00:00.000+00:00"/>
        <string key="system" value=""/>
    </global>
    <classifier name="Activity" keys="concept:name"/>
    <classifier name="Another" keys="concept:name system"/>
    <float key="log attribute" value="2335.23"/>
    <trace>
        <string key="concept:name" value="Trace number one"/>
        <event>
            <string key="concept:name" value="Register client"/>
            <string key="system" value="alpha"/>
            <date key="time:timestamp" value="2009-11-25T14:12:45:000+02:00"/>
            <int key="attempt" value="23">
                <boolean key="tried hard" value="false"/>
            </int>
        </event>
        <event>
            <string key="concept:name" value="Mail rejection"/>
            <string key="system" value="beta"/>
            <date key="time:timestamp" value="2009-11-28T11:18:45:000+02:00"/>
        </event>
    </trace>
</log>
```

Рисунок 3 - Приклад документу формату

Список використаних джерел

1. Analyzing Log Analysis: An Empirical Study of User Log Mining by Jessica Lin, –May 20, 2013
2. J. H. Andrews: “Theory and practice of log file analysis.” Technical Report 524, Department of Computer Science, University of Western Ontario, – May 1998.
3. Trace Diff Analysis [Электроний ресурс], Режим доступу: http://www.process mining.org /online/tracediff.
4. Christian W. Gunther, Eric Verbeek. XES Standard Definition. – 2014. – 24 p.
5. Christian W. Gunther. OpenXES Developer Guide. - 2009. - 37 p.