

ДОДАТОК А

Перелік джерел посилання за науковими напрямками керівника та науковців
кафедри програмної інженерії

2. Sergiy Zagorodnyuk, Bohdan Sus, Iona Revenchuk, Oleksandr Bauzha Information Security of Users Rights Assignment via the Software Solutions Based on LDAP // Problem of Infocommunications. Science and Technolpgy (PIC S&T'2020), Kharkiv, Ukraine- 6-9 October 2020.

9. Гриб Р.В., Ревенчук І.А. Дослідження методів та інструментів для групування помилок журналу тестування за допомогою алгоритмів подібності рядків: матер VII Міжнар наук.-практ конф. "Scientific Research: Theoretical Foundations and Practical Applications". Відень 24-26.01.2024. Р.102-105 URL: https://isu-conference.com/wp-content/uploads/2024/01/Scientific_research_theoretical_foundations_and_practical_applications.pdf (дата звернення: 25.05.2024).

ДОДАТОК Б

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



Ім'я користувача:
Нечволод Вадим Юрійович каф. ПІ

ID перевірки:
1016292687

Дата перевірки:
29.05.2024 05:45:59 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
29.05.2024 10:27:16 EEST

ID користувача:
94949

Назва документа: 2024_М_ПІ_ІПЗм-22-3_Гриб_Р_В_скорочений

Кількість сторінок: 40 Кількість слів: 6224 Кількість символів: 48226 Розмір файлу: 1.32 MB ID файлу: 1016086775

0.35% Схожість

Найбільша схожість: 0.21% з Інтернет-джерелом (<http://medbib.in.ua/teorii-psihičeskogo-razvitiya-40237.html>)

0.21% Джерела з Інтернету

18

Сторінка 42

0.14% Джерела з Бібліотеки

1

Сторінка 42

2.38% Цитат

Цитати 1

Сторінка 43

Вилучення списку бібліографічних посилань вимкнене

0% Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи

2

ДОДАТОК В
Слайди презентації

Харківський національний університет радіоелектроніки

Кафедра Програмної інженерії

Кваліфікаційна робота магістра

Дослідження методів та інструментів для
групування помилок журналу тестування
за допомогою алгоритмів подібності
рядків

Виконав: ст. гр. ІЗПм-22-3 Гриб Р.В

Керівник: доцент каф. ПІ Ревенчук І.А

2

Актуальність дослідження

- Підвищення рівня складності програмних систем
- Збільшення кількості помилок
- Зниження якості кінцевого продукту
- Підвищення витрат на розробку
- Велика кількість часу на тестування

Аналіз предметної області

- Ключовий та незамінний етап розробки – тестування програмного забезпечення
- Ефективна класифікація та управління помилками
- Виявлення помилок на ранніх етапах розробки коштує набагато менше ніж їх виправлення після відправки продукту на ринок
- Групування помилок тестування сприяє підвищенню ефективності та швидкості вирішення проблем при розробці

Аналіз предметної області

- Найефективніші методи групування помилок – групування за допомогою алгоритмів подібності рядків

STRING SIMILARITY

Алгоритми подібності рядків

- Косинусна подібність – використовується для оцінки ступеня подібності між векторами у просторі
- Відстань Левенштейна – визначає мінімальну кількість операцій редагування, необхідних для того, щоб перетворити один рядок у інший
- Індекс Жаккара – використовується у багатьох галузях для вимірювання схожості між множинами
- Подібність Джаро-Вінклера – враховує як точні збіги символів, так і відмінності в їх порядку у рядках
- Коефіцієнт Соренсена-Дайса – обчислює подібність як відношення подвоєного перетину до суми розмірів множин

Постановка задачі

Теоретично

- Рішення багатокритеріальної задачі

Експериментально

- Практична реалізація групування алгоритмами

Необхідно:

- 1) Оцінити алгоритми за такими критеріями: точність та швидкість обробки даних, обсяг використаних ресурсів, стійкість до шуму, вартість реалізації.
- 2) Теоретична частина: обираються середні значення показників алгоритмів для рішення багатокритеріальної задачі.
- 3) Практична частина: програмно реалізувати обрані алгоритми та експериментальним шляхом перевірити результат вирішення багатокритеріальної задачі.

Рішення багатокритеріальної задачі

Назва алгоритму	Точність даних (%)	Швидкість обробки	Обсяг використаних ресурсів	Стійкість до шуму	Вартість	Z*
CosSimilarity	0	0,5	1	0,33	1	0,169
Levenshtein	0,26	1	0,66	1	1	0,263
JaccardIndex	0,87	0	0	1	0	0,152
JaroWinkler	0,14	1	0,33	0	0,5	0,163
SorensenDice	1	0,5	0,66	0,33	0,5	0,250
β	0,266	0,333	0,2	0,133	0,066	
α	0,440	0,333	0,377	0,375	0,333	

Опис прийнятих проєктних рішень

string-comparison

JavaScript implementation of tdebatry/java-string-similarity

A library implementing different string similarity, distance and sortMatch measures. A dozen of algorithms (including Levenshtein edit distance and siblings, Longest Common Subsequence, cosine similarity etc.) are currently implemented. Check the summary table below for the complete list...

- string-comparison
 - Download & Usage
 - Overview
 - Normalized, metric, similarity and distance
 - (Normalized) similarity and distance
 - Levenshtein
 - Longest Common Subsequence
 - Metric Longest Common Subsequence
 - Cosine similarity
 - Sorensen-Dice coefficient
 - Jaro-Winkler similarity
 - API
 - Methods

Install

```
> npm i string-comparison
```

Repository

github.com/Rabbitzcc/js-string-compari...

Homepage

github.com/Rabbitzcc/js-string-compari...

Weekly Downloads

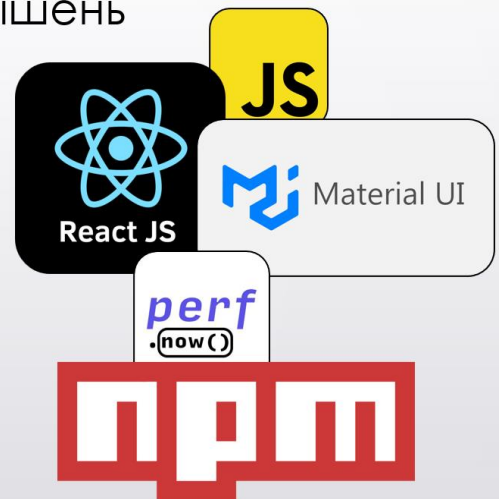
17 660

Version 1.3.0 License MIT

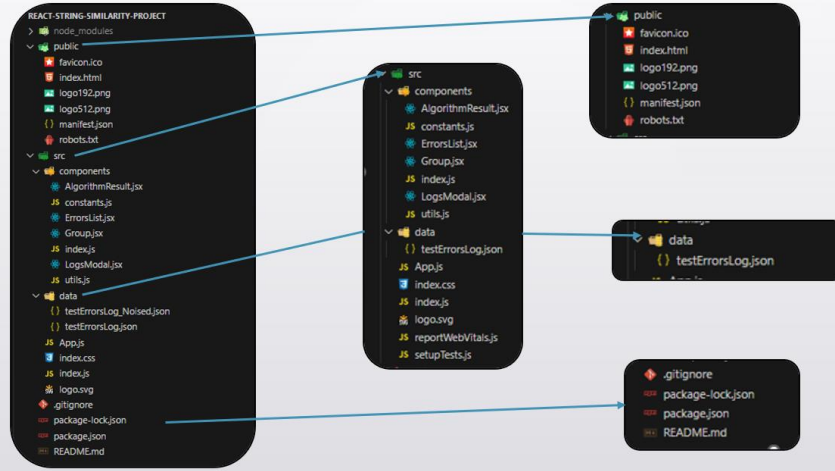
Unpacked Size 33.8 kB Total Files 10

Issues 0 Pull Requests 0

Last publish 6 months ago



Опис програмної реалізації – Архітектура проєкту



Опис програмної реалізації – Модель даних

```

testErrorsLog.json X
src > data > {} testErrorsLog.json > {} errors > {} 35 > message
You, 2 weeks ago | 1 author (You)
1  {
2  "errors": [
3  { "message": "File not found", "errorCode": 1001 },
4  { "message": "Database connection failed", "errorCode": 1002 },
5  { "message": "Invalid input data", "errorCode": 1003 },
6  { "message": "Permission denied", "errorCode": 1004 },
7  { "message": "Authentication failed", "errorCode": 1005 },
8  { "message": "Server timeout", "errorCode": 1006 },
9  { "message": "Unexpected end of file", "errorCode": 1007 },
10 { "message": "Invalid token", "errorCode": 1008 },
11 { "message": "Internal server error", "errorCode": 1009 },
12 { "message": "Resource not available", "errorCode": 1010 },
13 { "message": "Invalid username or password", "errorCode": 1011 },
14 { "message": "Missing required parameters", "errorCode": 1012 },
15 { "message": "Duplicate entry found", "errorCode": 1013 },
16 { "message": "Invalid file format", "errorCode": 1014 },
17 { "message": "Service temporarily unavailable", "errorCode": 1015 },
18 { "message": "Network error", "errorCode": 1016 },
19 { "message": "Invalid API key", "errorCode": 1017 },
20 { "message": "Insufficient permissions", "errorCode": 1018 },
21 { "message": "Resource limit exceeded", "errorCode": 1019 },

```

```

{
  "message": "TypeError: Cannot read property 'pop' of undefined",
  "errorCode": 2034
},

```

```

{
  "message": "API response: Gateway Timeout (Server response delayed)",
  "errorCode": 3029
},

```

```

{
  "message": "Material-UI error: ListItem component requires a 'button' prop.",
  "errorCode": 5014
},

```

Опис програмної реалізації – Реалізація алгоритмів

The diagram illustrates the implementation of an error grouping algorithm. It features a main code block on the left and three callout boxes on the right, each pointing to a specific part of the code.

```

//main code snippet
export const getGroupedLogSimilarity = ({ logs, algorithmKey }) => {
  //measure memory usage
  const startMemory = performance.memory.usedJSHeapSize / 1024 / 1024;
  //measure time
  const start = now();
  //get algorithm by its' name
  const similarityMethod = getAlgorithm(algorithmKey);
  //array with grouping result
  let groupedErrors = [];
  //run through all errors and use similarity algorithm
  logs.forEach(log => {
    let foundGroup = false;
    for (let group of groupedErrors) {
      if (
        similarityMethod.similarity(log.message, group[0].message) >
        THRESHOLD_VALUE
      ) {
        group.push(log);
        foundGroup = true;
        break;
      }
    }
    if (!foundGroup) {
      groupedErrors.push(log);
    }
  });
  //finish measuring the time
  const end = now();
  //finish measuring the memory usage
  const endMemory = performance.memory.usedJSHeapSize / 1024 / 1024;
  return {
    grouped: groupedErrors,
    times: end - start,
    memoryUsage: Math.abs(endMemory - startMemory),
  };
};

```

Callout 1: //measure memory usage
const startMemory = performance.memory.usedJSHeapSize / 1024 / 1024;

Callout 2: //measure time
const start = now();
//get algorithm by its' name
const similarityMethod = getAlgorithm(algorithmKey);

Callout 3: //array with grouping result
let groupedErrors = [];

Callout 4: //run through all errors and use similarity algorithm
logs.forEach(log => {
 let foundGroup = false;
 for (let group of groupedErrors) {
 if (similarityMethod.similarity(log.message, group[0].message) > THRESHOLD_VALUE) {
 group.push(log);
 foundGroup = true;
 break;
 }
 }
 if (!foundGroup) {
 groupedErrors.push(log);
 }
}

Callout 5: //finish measuring the time
const end = now();
//finish measuring the memory usage
const endMemory = performance.memory.usedJSHeapSize / 1024 / 1024;
return {
 grouped: groupedErrors,
 times: end - start,
 memoryUsage: Math.abs(endMemory - startMemory),
};

Опис експериментального дослідження

Вихідні дані

АЛГОРИТМИ ПОДІБНОСТІ РЯДКІВ У ГРУПУВАННІ ПОМИЛОК ЖУРНАЛУ ТЕСТУВАННЯ

The screenshot shows a web interface for viewing test error logs. At the top, there is a search bar with the text "Алгоритми подібності рядків". Below it, the title "Журнал помилок тестування:" is displayed. A list of error codes is shown, ranging from 1001 to 1012. A "SEARCH" button is located at the bottom right of the list.

Алгоритми подібності рядків

Журнал помилок тестування:

- 1001 File not found
- 1002 Database connection failed
- 1003 Invalid input data
- 1004 Permission denied
- 1005 Authentication failed
- 1006 Server timeout
- 1007 Unexpected end of file
- 1008 Invalid token
- 1009 Internal server error
- 1010 Resource not available
- 1011 Invalid username or password
- 1012 Missing required parameters

SEARCH

Опис експериментального дослідження

Виконання алгоритмів

АЛГОРИТМИ ПОДІБНОСТІ РЯДКІВ У ГРУПУВАННІ ПОМИЛОК ЖУРНАЛУ ТЕСТУВАННЯ

ПЕРЕГЛЯНУТИ ВИХОДНІ ДАНІ

Cosine Similarity Algorithm	Levenshtein Similarity Algorithm	Jaccard Index Similarity Algorithm	Jaro Winkler Similarity Algorithm	Sorensen Dice Similarity Algorithm
Groups amount: 6 Time spent: 2.60 ms Memory usage: 0.00 MB	Groups amount: 94 Time spent: 280.60 ms Memory usage: 6.88 MB	Groups amount: 39 Time spent: 201.20 ms Memory usage: 1.95 MB	Groups amount: 20 Time spent: 7.30 ms Memory usage: 0.00 MB	Groups amount: 92 Time spent: 72.20 ms Memory usage: 1.62 MB
1001 File not found 1002 Database connection failed 1003 Invalid input data 1005 Authentication failed 1007 Unexpected end of file 1008 Invalid token 1013 Duplicate entry found 1014 Invalid file format 1022 Action not allowed 1026 Resource query failed 2002 ReferenceError: Undefined variable 2009 RangeError: Maximum call stack size exceeded	1001 File not found 1002 Database connection failed 1003 Invalid input data 1004 Permission denied 1005 Authentication failed 1006 Server timeout 1007 Unexpected end of file 1008 Invalid token 1009 Internal server error	1001 File not found 1002 Database connection failed 1023 Data integrity compromised TypeError: Cannot read property 'setState' of undefined TypeError: Cannot read property 'includes' of undefined 3021 API response: Conflict (Data integrity compromised) 3024 API response: Unprocessable Entity (invalid request) API response: Internal Server Error (Database connection failed)	1001 File not found 3008 API response: Not Found 1002 Database connection failed 1004 Permission denied 1005 Authentication failed 1022 Action not allowed 1023 Data integrity compromised 1026 Resource query failed 2004 TypeError: Assignment to constant variable 2007 TypeError: Invalid operand types 2012 TypeError: Cannot assign to read only property	1001 File not found 1002 Database connection failed 1003 Invalid input data 1004 Permission denied 1005 Authentication failed 1006 Server timeout 1007 Unexpected end of file 1008 Invalid token 1009 Internal server error 3004 API response: Internal Server Error

Опис експериментального дослідження

Результат Групування

АЛГОРИТМИ ПОДІБНОСТІ РЯДКІВ У ГРУПУВАННІ ПОМИЛОК ЖУРНАЛУ ТЕСТУВАННЯ

ПЕРЕГЛЯНУТИ ВИХОДНІ ДАНІ

Cosine Similarity Algorithm	Levenshtein Similarity Algorithm	Jaccard Index Similarity Algorithm	Jaro Winkler Similarity Algorithm	Sorensen Dice Similarity Algorithm
Groups amount: 6 Time spent: 1.70 ms Memory usage: 0.00 MB	Groups amount: 74 Time spent: 220.10 ms Memory usage: 0.23 MB	Groups amount: 17 Time spent: 63.60 ms Memory usage: 0.69 MB	Groups amount: 14 Time spent: 4.30 ms Memory usage: 0.00 MB	Groups amount: 69 Time spent: 42.20 ms Memory usage: 0.00 MB
1001 File not found 1002 Database connection failed 1003 Invalid input data 1005 Authentication failed 1007 Unexpected end of file 1008 Invalid token 1009 Internal server error 1010 Resource not available 1011 Invalid username or password 1013 Duplicate entry found 1014 Invalid file format 1018 Insufficient permissions 1019 Resource limit exceeded 1026 Unexpected response from	1001 File not found 1002 Database connection failed 1026 Database query failed 1003 Invalid input data 1004 Permission denied 1005 Authentication failed 1006 Server timeout 1007 Unexpected end of file 1008 Invalid token 1017 Invalid API key	1001 File not found 1002 Database connection failed 1005 Authentication failed 1007 Unexpected end of file 1013 Duplicate entry found 1014 Invalid file format 4008 ESLint: 'function' is not defined. 1003 Invalid input data 1004 Permission denied 1011 Invalid username or password 1018 Insufficient permissions 1023 Data integrity compromised	1001 File not found 1008 Invalid token 1013 Duplicate entry found 1028 Resource not found 3008 API response: Not Found 4008 ESLint: 'function' is not defined. 1002 Database connection failed 1004 Permission denied 1005 Authentication failed 1007 Unexpected end of file 1010 Resource not available 1022 Action not allowed 1023 Data integrity compromised	1001 File not found 1028 Resource not found 1002 Database connection failed 1003 Invalid input data 1004 Permission denied 1005 Authentication failed 1006 Server timeout 1007 Unexpected end of file 1008 Invalid token 1017 Invalid API key

Groups amount: 14
Time spent: 4.30 ms
Memory usage: 0.00 MB

Опис експериментального дослідження

15

Перевірка стійкості до шуму

```
testErrorLog_Noised.json X
src > data > {} testErrorLog_Noised.json > {} errors > {} 200
%> 2 days ago | 1 minute (now)

{
  "errors": [
    {
      "message": "File not found", "errorCode": 1001 },
    {
      "message": "Databaseconnection failed", "errorCode": 1002 },
    {
      "message": "Invalid input data", "errorCode": 1003 },
    {
      "message": "Permission (error)", "errorCode": 1004 },
    {
      "message": "Authentication failed", "errorCode": 1005 },
    {
      "message": "Server timeout", "errorCode": 1006 },
    {
      "message": "Unexpected error of file", "errorCode": 1007 },
    {
      "message": "Invalid token", "errorCode": 1008 },
    {
      "message": "Internal server error", "errorCode": 1009 },
    {
      "message": "Resource not available", "errorCode": 1010 },
    {
      "message": "Invalid username /> password", "errorCode": 1011 },
    {
      "message": "Missing required parameters", "errorCode": 1012 },
    {
      "message": "Duplicate entry found", "errorCode": 1013 },
    {
      "message": "Maximum call stack size exceeded", "errorCode": 1014 },
    {
      "message": "Type: Cannot read", "errorCode": 1015 }
  ]
}
```

АЛГОРИТМИ ПОДІБНОСТІ РЯДКІВ У ГРУПУВАННІ ПОМИЛОК ЖУРНАЛУ ТЕСТУВАННЯ

Cosine Similarity Algorithm

Groups amount: 8
Time spent: 2.80 ms
Memory usage: 0.00 MB

- 1001 File not found
- 1002 Databaseconnection failed
- 1003 Invalid input data
- 1005 Authentication failed
- 1007 Unexpected error of file
- 1008 Invalid token

Levenshtein Similarity Algorithm

Groups amount: 84
Time spent: 272.70 ms
Memory usage: 7.26 MB

- 1001 File not found
- 1002 Databaseconnection failed
- 1003 Invalid input data
- 1004 Permission (error)
- 1005 Authentication failed
- 1006 Server timeout
- 1007 Unexpected error of file
- 1008 Invalid token
- 1009 Internal server error
- 1010 Resource not available

Jaccard Index Similarity Algorithm

Groups amount: 36
Time spent: 194.20 ms
Memory usage: 1.88 MB

- 1001 File not found
- 1002 Databaseconnection failed
- 1023 Data integrity compromised
- 2047 Type: Cannot read property 'testState' of undefined
- 2043 Type: Cannot read property 'includes' of undefined
- 3021 API response: Conflict (Data integrity compromised)
- 3024 API response: Unprocessable Entity (invalid request)
- 3028 API response: Internal Server Error (Database connection failed)
- ESLint 'strict' is not

Jaro Winkler Similarity Algorithm

Groups amount: 20
Time spent: 7.10 ms
Memory usage: 0.00 MB

- 1001 File not found
- 3008 API response: Not Found
- 1002 Databaseconnection failed
- 1004 Permission (error)
- 1005 Authentication failed
- 1022 Action not allowed
- 1026 Data-base query failed
- 2004 Type: Error: Assignment to constant variable
- 2007 Type: Error: Invalid operand type
- 2011 Type: Error: Cannot assign to read only property
- Type: Cannot read

Sorensen Dice Similarity Algorithm

Groups amount: 62
Time spent: 58.00 ms
Memory usage: 1.53 MB

- 1001 File not found
- 1002 Databaseconnection failed
- 1003 Invalid input data
- 1004 Permission (error)
- 1005 Authentication failed
- 1006 Server timeout
- 1007 Unexpected error of file
- 1008 Invalid token
- 1009 Internal server error
- 3001 API response: Internal Server Error

$$\text{Точність} = \frac{\sum \text{справильних результатів} * 100\%}{\sum \text{помилки}}$$

Опис експериментального дослідження

16

Результати

Назва алгоритму	Точність даних (%)	Швидкість обробки	Обсяг використаних ресурсів	Стійкість до шуму	Вартість
CosSimilarity	62,2	Дуже Висока	Низький	Середня	Низька
Levenshtein	96,17	Низька	Високий	Дуже висока	Висока
JaccardIndex	79,1	Середня	Середній	Низька	Середня
JaroWinkler	95,8	Дуже Висока	Низький	Висока	Низька
SorensenDice	94,1	Висока	Низький	Дуже висока	Низька

■ – середньо ■ – відмінно



ВИСНОВКИ

1. Алгоритм знаходження подібності Джаро-Вінклера найефективніший серед найпопулярніших алгоритмів подібності рядків
2. Алгоритм Левенштейна найточніший серед алгоритмів
3. Результат даного дослідження підвищить точність та ефективність обробки помилок, сприяє скороченню часу на тестування, забезпечує швидший вихід продуктів на ринок
4. Впровадження алгоритму Джаро-Вінклера для групування помилок у журналі тестування стане вагомим кроком до оптимізації процесу розробки програмних продуктів

ДОДАТОК Г

Апробація результатів роботи

Гриб Р.В., Ревенчук І.А. Дослідження методів та інструментів для групування помилок журналу тестування за допомогою алгоритмів подібності рядків: матер VII Міжнар наук.-практ конф. "Scientific Research: Theoretical Foundations and Practical Applications". Відень 24-26.01.2024. Р.102-105
URL: https://isu-conference.com/wp-content/uploads/2024/01/Scientific_research_theoretical_foundations_and_practical_applications.pdf (дата звернення: 25.05.2024).

ДОДАТОК Д

Експертний висновок результатів перевірки кваліфікаційної роботи на
відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)

програмної інженерії
(кафедра)

ІПЗМ-22-3
(група)

Гриб Р.В.

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

Експерт

(підпис)

Нечволод В.Ю.

(прізвище, ініціали)

29.05.2024