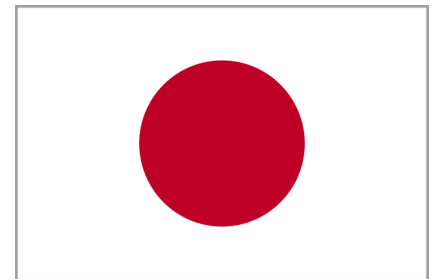
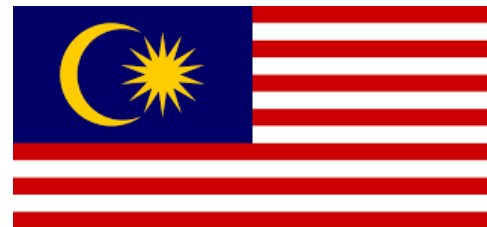


A Comparative Study of Coding Styles and Method Naming Conventions in Open Source Software: Japan vs. Malaysia

発表者：ダニエル



背景

- **Introduction to Open Source Software Development**
- **Importance of Coding Style and Method Naming**
- **Cross-Cultural Influences on Coding Practices**
 - マレーシア: コーディングには英語を使用し、世界標準に準拠しています。
 - 日本: 英語と日本語が混在し、独自のコーディング規則が存在する可能性があります。
- **Domestic and international project percentage**

Approach

- コーディング スタイルの分析:
 - 日本とマレーシア(または世界標準)のオープンソース プロジェクトで使用されている一般的なコーディング スタイルを比較します。
- メソッドの命名規則の調査:
 - 両国の開発者が使用しているメソッドの命名規則を特定し、比較します。
- | 日本: オープンソース プロジェクトは多数ありますが、国際協力は限られています。
- | マレーシア: オープンソース プロジェクトは全体的に少ないですが、国際協力の度合いは高いです。

Aim

- **Improved Code Readability**
- **Enhanced Collaboration**
- **Standardization of Coding Practices**

Research Question

- RQ: GitHub 上の日本とマレーシアのオープンソース プロジェクトでは、コーディング スタイルとメソッドの命名規則はどのように異なりますか?
- RQ: 言語の違いは、日本とマレーシアのオープンソース プロジェクトのコードの読みやすさと保守性にどのような影響を与えますか?
- RQ: マレーシアのオープンソース プロジェクトから、日本のプロジェクトにおける国際協力を改善するためにどのようなベスト プラクティスを導き出せますか? またその逆はどうですか?
- RQ: 国際協力の強化により、日本のオープンソース プロジェクトの品質は向上するでしょうか?

Research method

1. 日本とマレーシアのユーザーからリポジトリを抽出する



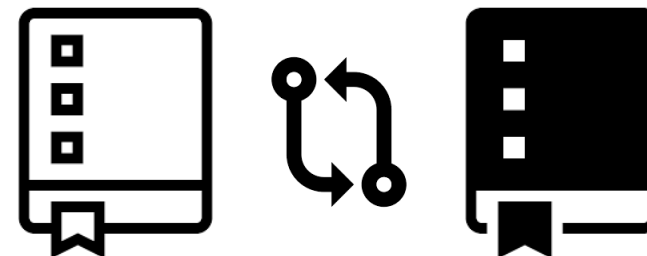
2. 抽出されたリポジトリをフィルターする



3. フィルタリングされたリポジトリが分析する



4. マレーシアと日本の分析結果を比較する



Research method

Expected result:

- Criteria for Analyzing Coding Practices:

Criteria	Description
Indentation	Whether spaces or tabs are used for indentation; consistency across the project
Method Naming	Language used for method names (Japanese, English); consistency in naming style
Variable Naming	Language and style used for variable names (camelCase, PascalCase, underscores)
Commenting	Language of comments; detail and frequency of comments
Documentation	Language of documentation and inline comments
Direct Translation	Presence of direct translations of terms from Japanese to English or vice versa
Coding Style Consistency	Overall consistency in coding style throughout the project
Use of English Keywords	Frequency and context of English keywords in code
Readability	Overall readability of code for an international audience
International Collaboration	Presence and impact of international contributors in the project

Research method

Expected result:

- Example:

Aspect	Japanese Projects	Malaysian Projects
Indentation Style	Mixed (tabs and spaces, influenced by conventions)	More standardized (mostly spaces)
Method Naming Convention	Mixed (Japanese and English names)	Predominantly English names
Use of Underscores (_)	Common in method names and variables	Less common, more camelCase or PascalCase
Direct Translation of Terms	Frequent use of direct translations from Japanese	Consistent use of English terms
Commenting Practices	Detailed comments, often in Japanese	Concise comments, always in English
Variable Naming	Mix of Japanese and English	Predominantly English
Coding Style Consistency	Variable, influenced by individual preferences	More consistent, aligns with global standards
Documentation Language	Japanese	English
Code Readability	Can vary, especially for non-Japanese speakers	Generally high, easy for international contributors

Query

Criteria	Description
Indentation	Whether spaces or tabs are used for indentation; consistency across the project

検索クエリ:

- 場所: {日本/マレーシア}
- followers数: <10
- 異なるプログラミング言語で検索

抽出されたリポジトリ:

- ユーザーごとに 3 つのリポジトリ
- 合計 5 ユーザー

1時間あたり5,000リクエスト

```
Rate Limit: 5000 requests remaining
Rate limit resets at: 2024-07-24 03:22:25 JST
Current time: 2024-07-24 02:22:26 JST
Time left until rate limit reset: 3598.1920986175537 seconds
```

Filter

Criteria	Description
Indentation	Whether spaces or tabs are used for indentation; consistency across the project

フィルター:

- 特定のプログラミング拡張機能のみ抽出を許可
- 各リポジトリのファイル制限は30

```
languages = ['C', 'Java', 'JavaScript', 'PHP', 'Python', 'Ruby']
```

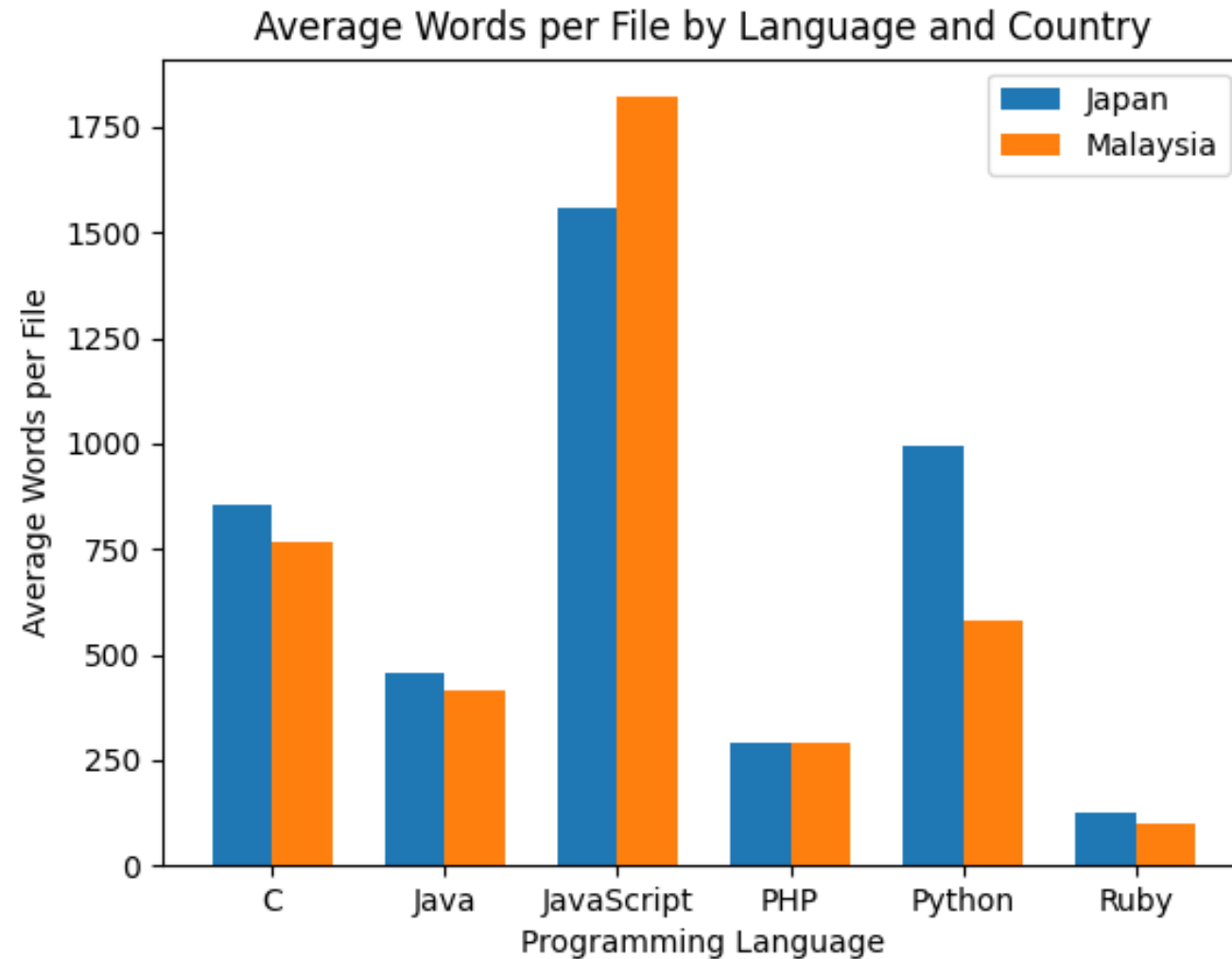
```
HEADERS = {'Authorization': f'token {GITHUB_TOKEN}'}  
ALLOWED_EXTENSIONS = {'.py'} # Only extract Python files
```

```
def collect_repo_files(owner, repo, path="", file_limit=30):  
    contents = fetch_repo_contents(owner, repo, path)  
    if not contents:  
        return []
```

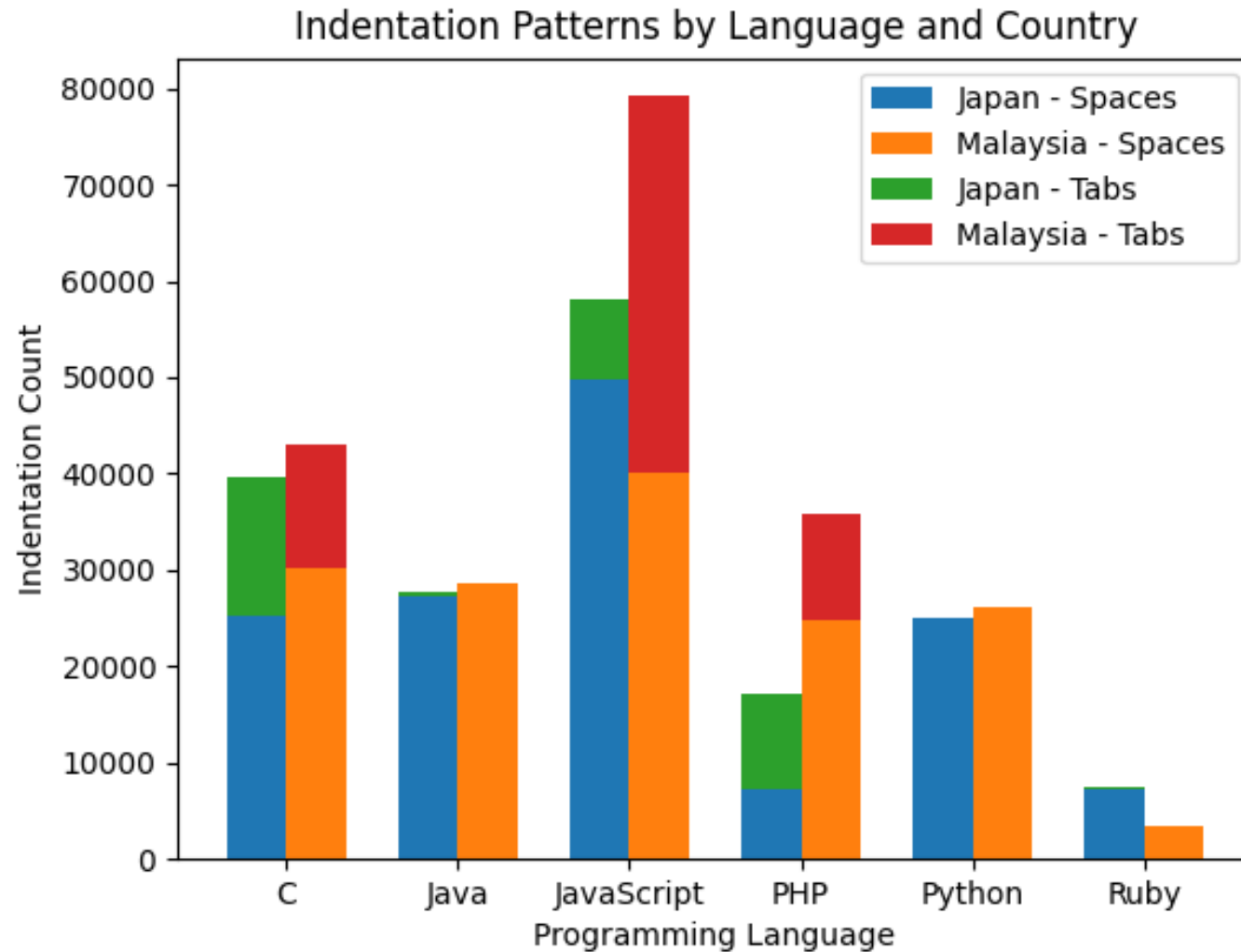
Analysis

Language	Country	Indentation Pattern (Total)	Avg Indentation per File (Spaces)	Avg Indentation per File (Tabs)	Weighted Avg Indentation by Words (Spaces)	Weighted Avg Indentation by Words (Tabs)	Total Words	Avg Words per File
C	Japan	Spaces: 38724, Tabs: 13371	1018.88	83.73	9206.54	187.19	206093	855.16
	Malaysia	Spaces: 34963, Tabs: 10784	576.14	44.25	4563.33	57.95	190300	767.34
Java	Japan	Spaces: 31175, Tabs: 448	936.47	3.88	6419.52	1.76	110219	455.45
	Malaysia	Spaces: 27758, Tabs: 67	667.79	0.27	2089.89	0.36	113413	412.41
JavaScript	Japan	Spaces: 39901, Tabs: 8225	1951.15	167.51	7512.27	4108.11	212175	1560.11
	Malaysia	Spaces: 40689, Tabs: 36815	757.21	544.56	8958.26	13591.44	406131	1821.22
PHP	Japan	Spaces: 8001, Tabs: 9899	145.39	116.26	221.39	307.82	57179	290.25
	Malaysia	Spaces: 12191, Tabs: 11953	268.15	112.72	636.35	1174.92	89599	291.85
Python	Japan	Spaces: 31195, Tabs: 17	2417.37	0.21	13200.22	0.24	109107	991.88
	Malaysia	Spaces: 22863, Tabs: 37	1240.43	0.22	3771.82	0.18	129702	581.62
Ruby	Japan	Spaces: 7771, Tabs: 3	154.17	0.02	526.08	0.00	29015	123.47
	Malaysia	Spaces: 3377, Tabs: 126	35.40	0.96	160.27	1.09	26517	97.13

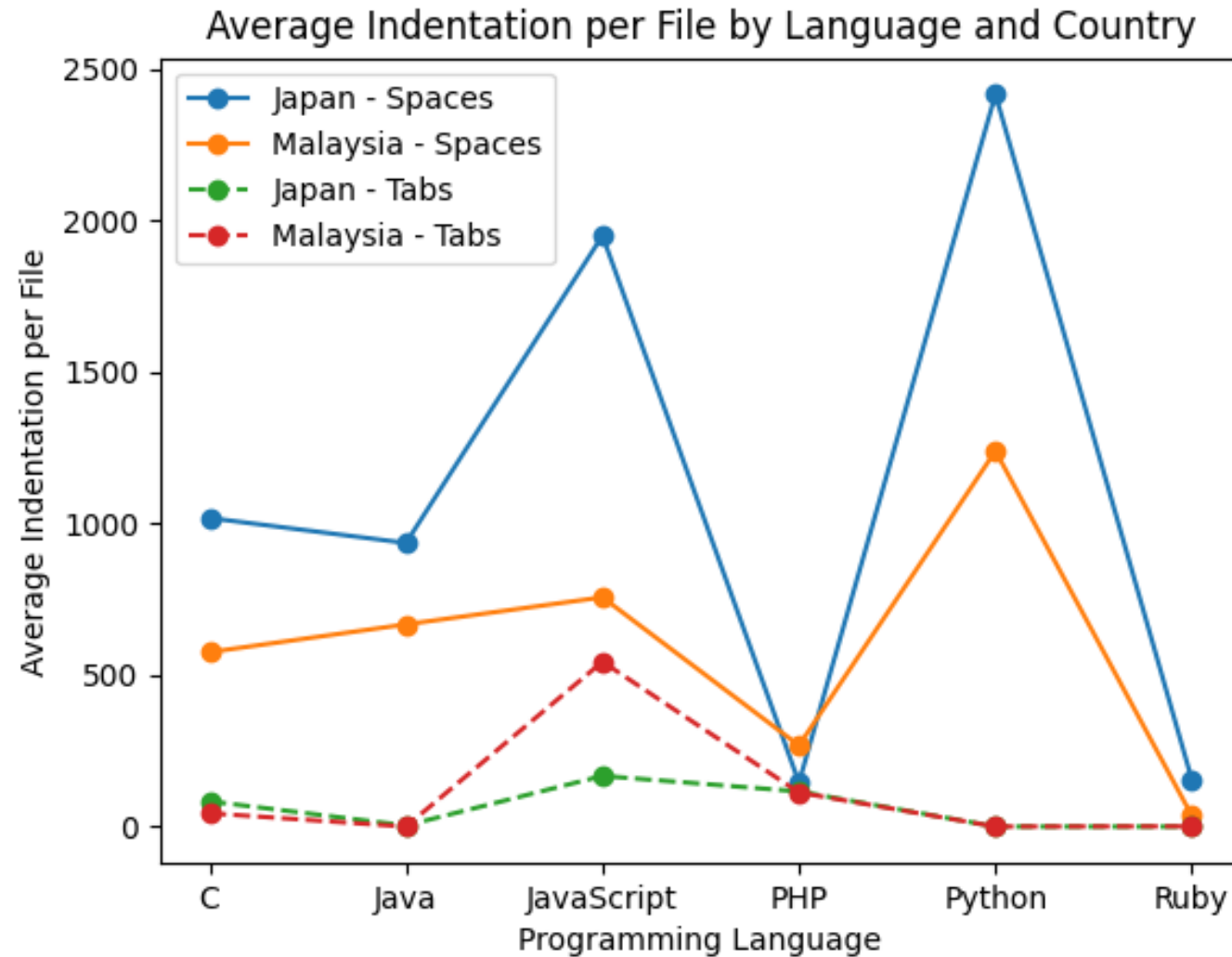
Comparison



Comparison



Comparison



Conclusion

- 日本とマレーシアのインデントパターンの違いは、すべての言語で均一ではない。
- 結果は、タブの使用（またはインデントレベル）がプログラミング言語に応じて、2つの国間で一貫していないことを示している。
- 日本では、ほとんどの言語で一般的に一貫性があり深いインデントが好まれる
- マレーシアでは、JavaScript と PHP のインデントが特に深く、より多様性が見らる

結果は一貫していないが、同じプログラミング言語内で国によってコーディングスタイルの違いが見らる。

今度の課題

Criteria	Description
Indentation	Whether spaces or tabs are used for indentation; consistency across the project
Method Naming	Language used for method names (Japanese, English); consistency in naming style
Variable Naming	Language and style used for variable names (camelCase, PascalCase, underscores)
Commenting	Language of comments; detail and frequency of comments
Documentation	Language of documentation and inline comments
Direct Translation	Presence of direct translations of terms from Japanese to English or vice versa
Coding Style Consistency	Overall consistency in coding style throughout the project
Use of English Keywords	Frequency and context of English keywords in code
Readability	Overall readability of code for an international audience
International Collaboration	Presence and impact of international contributors in the project