

Impacts of Coding Practices on Readability

発表者：ダニエル

文献: Rodrigo Magalhães dos Santos, Marco Aurélio Gerosa

背景

- problems related to **low code readability**, impacting software projects negatively.
- Low readability can lead to issues like declines in contributions received by open source projects and delays in developers' contributions.
- The survey was conducted among two groups of developers:
 - Computer science bachelor students
 - Professional programmers from a large Brazilian software company

目的

- Identify coding conventions with a positive impact on readability.
- Investigate how coding practices influence readability levels.
- Explore the influence of readers' characteristics on their perceptions of readability.
- Provide insights into coding standards and practices to guide the creation of coding standards.

関連研究

| Edward L. Thorndike

- Found out that the more often a word is used, easier it is to understand
- The difficulty of a word is related to how often it is used

| Buse and Weimer,

- Developed one of the first code readability models, which used static code attributes and developer opinions to establish weights.
- they surveyed opinions of 120 software developers about 100 code snippets.

| Scalabrino et al,

- Developed a code readability model focusing on textual features.

調査方法

- | Developed a custom web application tailored to the survey's requirements.
- Designed the application to make all data fields optional to minimize participant dropout.
- Recorded all participant actions for subsequent analysis.
- The application included screens for language selection, instructions, code snippet pairs, and profile data collection.
- | Invited Participants:
 - Invited two groups of developers:
 - 55 students of Software Engineering from an American University
 - 7 professional programmers from a Brazilian software development company.

調査方法

| Code Snippets Composing:

- Used the GitHub database accessible through Google BigQuery to search for code snippets.
 - Selected 11 coding practices based on the following model metrics:
 - Average blank lines, keywords per line, indents per line, etc.
 - Number of identifiers using dictionary words.
- (1) Buse and Weimer [6]:
 - (a) Average blank lines;
 - (b) Average keywords per line;
 - (c) Average indents per line;
 - (d) Maximum number of identifiers in a single line;
 - (e) Average “.” characters;
 - (f) Maximum length of a single line;
 - (g) Average “(” and “{” per line;
 - (h) Average line length;

In your opinion, which code snippet is the most readable? Sample 1/10

```
1 @Override
2 public PhysicalOperation visitMarkDistinct(MarkDistinctNode node, LocalExecutionPlanContext context) {
3     PhysicalOperation source = node.getSource().accept(this, context);
4     List<Integer> channels = getChannelsForSymbols(node.getDistinctSymbols(), source.getLayout());
5     Optional<Integer> hashChannel = node.getHashSymbol().map(channelGetter(source));
6     MarkDistinctOperatorFactory operator = new MarkDistinctOperatorFactory(context.getNextOperatorId(), node.getId(),
7     return new PhysicalOperation(operator, makeLayout(node), source);
8 }
9 }
```

This

```
1 @Override
2 public PhysicalOperation visitMarkDistinct(MarkDistinctNode node,
3     LocalExecutionPlanContext context) {
4     PhysicalOperation source = node.getSource().accept(this, context);
5     List<Integer> channels = getChannelsForSymbols(node.getDistinctSymbols(),
6     source.getLayout());
7     Optional<Integer> hashChannel = node.getHashSymbol().
8     .map(channelGetter(source));
9     MarkDistinctOperatorFactory operator =
10     new MarkDistinctOperatorFactory(context.getNextOperatorId(), node.getId(),
11     source.getTypes(), channels, hashChannel, joinCompiler);
12     return new PhysicalOperation(operator, makeLayout(node), source);
13 }
```

This

Comments about your vote (optional):

Back

Next

リサーチクエスションと調査結果

RQ1: How coding practices influence the readability levels as perceived by readers?

two-tailed test

- Significance level: 0.05
- Null hypothesis: The votes are equally distributed (50/50) between adhering and violating snippets.
- Alternative hypothesis: The votes are not equally distributed, indicating that the coding practice has an effect.

リサーチクエスションと調査結果

RQ1: How coding practices influence the readability levels as perceived by readers?

Table 1: Received Votes per Code Snippet Pair and P-Values of Tests for Proportions

Code Practice	Short Description	Adherent Votes	Violating Votes	% of Violating	Total Votes	Proportion Test P-Value
P.1	Blank following curly braces	45	7	13.46% ▲	52	0.0000
P.2	Opening braces at the same line as clause's	16	37	69.81% ▼	53	0.0060
P.3	Blanks separating related instructions	28	28	50.00% ◆	56	1.0000
P.4	Line lengths not exceeding 80 chars	44	15	25.42% ▲	59	0.0003
P.5	Indents as 4 spaces	21	29	58.00% ▼	50	0.3222
P.6	At most 3 levels of code block nesting	34	19	35.85% ▲	53	0.0545
P.7	Avoid mult. statements on a same line	48	10	17.24% ▲	58	0.0000
P.8	Avoid fully qualified names	52	6	10.34% ▲	58	0.0000
P.9	Avoid child properties; use references	45	9	16.67% ▲	54	0.0000
P.10	Variables to store intermediary logicals	21	34	61.82% ▼	55	0.1056
P.11	Names using dictionary words	42	12	22.22% ▲	54	0.0001

▲ : 過半数の票を獲得しました
▼ : コードスニペットの違反
◆ : 引分

乖離票の割合 < 0.5 は可読性が向上していることを意味します
乖離票の割合 > 0.5 は可読性が低下することを意味します

リサーチクエスションと調査結果

RQ1: How coding practices influence the readability levels as perceived by readers?

Table 1: Received Votes per Code Snippet Pair and P-Values of Tests for Proportions

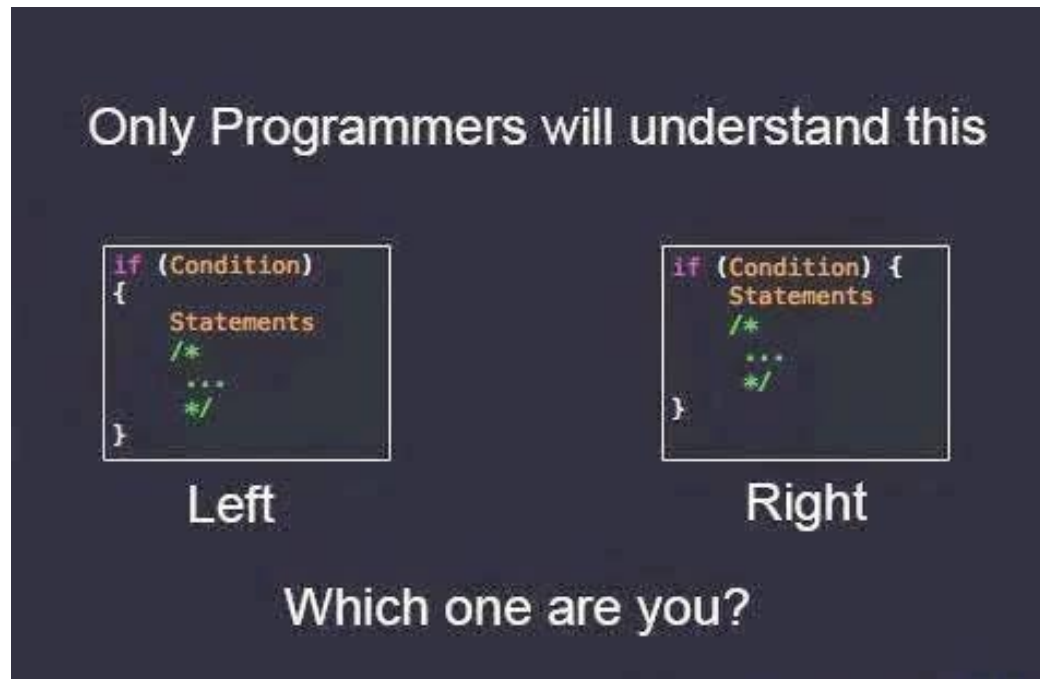
Code Practice	Short Description	Adherent Votes	Violating Votes	% of Violating	Total Votes	Proportion Test P-Value
P.1	Blank following curly braces	45	7	13.46% ▲	52	0.0000
P.2	Opening braces at the same line as clause's	16	37	69.81% ▼	53	0.0060
P.3	Blanks separating related instructions	28	28	50.00% ◆	56	1.0000
P.4	Line lengths not exceeding 80 chars	44	15	25.42% ▲	59	0.0003
P.5	Indents as 4 spaces	21	29	58.00% ▼	50	0.3222
P.6	At most 3 levels of code block nesting	34	19	35.85% ▲	53	0.0545
P.7	Avoid mult. statements on a same line	48	10	17.24% ▲	58	0.0000
P.8	Avoid fully qualified names	52	6	10.34% ▲	58	0.0000
P.9	Avoid child properties; use references	45	9	16.67% ▲	54	0.0000
P.10	Variables to store intermediary logicals	21	34	61.82% ▼	55	0.1056
P.11	Names using dictionary words	42	12	22.22% ▲	54	0.0001

▲ : 過半数の票を獲得しました
▼ : コードスニペットの違反
◆ : 引分

同じ行で中括弧を開くと読みやすさが低下します

リサーチクエスションと調査結果

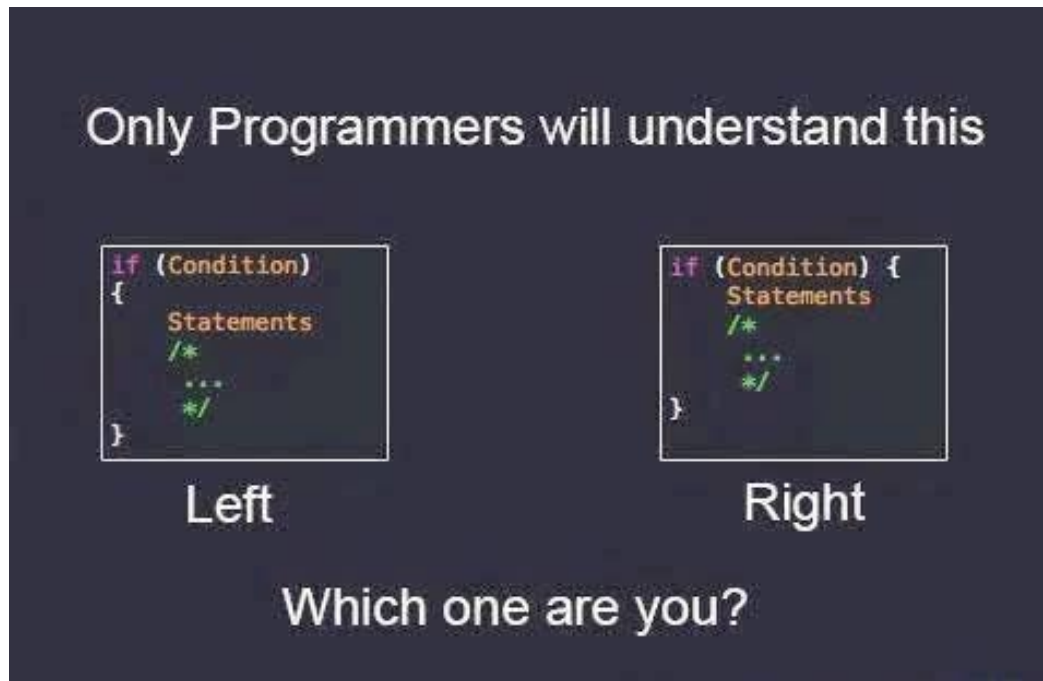
RQ1: How coding practices influence the readability levels as perceived by readers?



- 学生は,行数やページ数を減らすために同じ行に中括弧を入れています。
- 行内に 1 つの括弧文字が単独で表示されるのは煩わしいと思います。

リサーチクエスションと調査結果

RQ1: How coding practices influence the readability levels as perceived by readers?



- プロのプログラマーとしては、大規模なアプリケーションでは、**グループ化**の感覚を与えるため
- 中括弧だけを含む行がいくつかあることを許容できることを理解しています。

リサーチクエスションと調査結果

RQ2: What is the influence of readers' characteristics over their perceptions of readability?

Fisher's Exact tests

- Votes on snippet pairs were categorized as either "adhering" or "violating" based on the type of snippet being voted on.

Table 3: Frequencies of adhering/violating votes per gender on Practice P.7

	Female	Male
Adherent	2	33
Violating	5	4

Table 4: Frequencies of adhering/violating votes per gender on Practice P.10

	Female	Male
Adherent	4	9
Violating	1	27

リサーチクエスションと調査結果

RQ2: What is the influence of readers' characteristics over their perceptions of readability?

Fisher's Exact tests

Table 2: P-Values for the series of Fisher's Exact Tests

Factors							
Practice	Gender	Java Experience	Programming Experience	Age Bracket	Schooling	Learning Profile	Group (Student/ Professional)
P.01	0.2040	1.0000	0.3834	0.7274	1.0000	1.0000	0.1798
P.02	0.1520	0.3019	0.2010	0.3063	0.5117	0.5586	0.1550
P.03	0.0994	0.4909	0.0747	0.2109	0.6304	0.7385	0.1927
P.04	0.3470	1.0000	1.0000	0.8209	0.3643	1.0000	1.0000
P.05	0.6303	0.1714	0.6569	0.8539	0.6840	0.0538	0.6378
P.06	1.0000	1.0000	0.8713	0.7429	0.5042	0.0558	0.6117
P.07	0.0020	1.0000	0.6066	1.0000	0.7479	0.1125	0.5770
P.08	0.2476	1.0000	0.1438	0.3068	0.6730	1.0000	0.4970
P.09	1.0000	0.3082	1.0000	1.0000	0.7381	0.6037	0.5743
P.10	0.0284	0.5192	0.1290	0.1755	0.0513	0.1957	0.3588
P.11	0.1147	1.0000	0.1662	0.5475	0.7514	1.0000	0.0667

- P.7 Avoid mult. statements on a same line
- P.10 Variables to store intermediary logicals

However, further analysis showed that the correlations do not suggest a significant difference in how men and women think about readability

考察

- 実践 P.7 および P.10 と性別との相関関係は別として、プロフィール要素と開発者の意見との間に有意な相関関係は見つかりませんでした。
- 収集されたデータセットをさらに分析して、コードの特徴と可読性の間の新たな相関関係を発見することができます。

評価されたコーディング手法 11 件のうち 8 件で、開発者が認識する読みやすさに影響を与える証拠が示されました。