

Automated Query Reformulation for Efficient Search based on Query Logs From Stack Overflow

Kaibo Cao, Chunyang Chen, Sebastian Baltes
Christoph Treude, Xiang Chen

2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)
pp1273-1285.

情報検索における問題

- Stack Overflow (SO) は開発者にとって貴重な情報源となっている^[4]
- しかし、開発者が必要とする情報を効率的に探し当ててすることは難しい^{[5][6]}

適当な情報取得が困難である要因

検索におけるギャップ

- ユーザの意図とクエリ → 数単語で表現
- クエリと投稿内容 → 類義語, 省略形, 誤字など

[4] R. Abdalkareem, E. Shihab, and J. Rilling, "What do developers use the crowd for? a study using stack overflow," IEEE Software, vol. 34, no. 2, pp. 53–60, 2017.

[5] X.Xia,L.Bao,D.Lo,P.S.Kochhar,A.E.Hassan,andZ.Xing,"What do developers search for on the web?" Empirical Software Engineering, vol. 22, no. 6, pp. 3149–3185, 2017.

[6] C. Chen and Z. Xing, "Towards correlating search on google and asking on stack overflow," in Proceedings of 2016 IEEE 40th Annual Computer Software and Applications Conference. IEEE, 2016, pp. 83–92.

Query Reformulation

Query: do and while in java

Search Results

Advanced Search Tips

Results for **do and while in java**

35,563 results

Relevance Newest More ▾

7178 votes

A: "Thinking in AngularJS" if I have a jQuery background?

do the same thing, but **in** the AngularJS version anyone looking at the template knows what's supposed to happen. Whenever a new member of the development team comes on board, she can look at this **and** ... the reason you don't **do** that is that this is like half-jQuery, half-AngularJS, **and** no good. The problem here is that the developer is trying to "**do** jQuery" **in** the context of AngularJS. That's never ...

answered Feb 21 '13 by Josh David Miller

Reformulation: do and while **loop** in java

Search Results

Advanced Search Tips

Results for **do and while loop in java**

4,419 results

Relevance Newest More ▾

327 votes

Q: How do I make a delay in Java?

I am trying to **do** something **in Java** **and** I need something to wait / delay for an amount of seconds **in a while loop**. **while** (true) { if (i == 3) { i = 0; } ceva[i].setSelected(true ...); // I need to wait here ceva[i].setSelected(false); // I need to wait here i++; } I want to build a step sequencer **and** I'm new to **Java**. Any suggestions? ...

java wait sleep thread-sleep

asked Jun 8 '14 by ardb

Visiting Post: While loop in method Java

0 votes

Q: While loop in method Java

I want to achieve this menu to keep looping to receive input when i enter the wrong input other than 1,2,3. How **and** where to put my **while loop/ do while loop**? I am new **in JAVA**. After the user input ... other than 1,2 or 3 it should prompt the menu again. May i know how? Thanks. How **and** where to put my **while loop/ do while loop**? import java.util.*; public class InputMenu { public void ...

java loops while-loop do-while

asked Jan 6 '18 by Syamil Fuad

スレッド

1. 欲しい情報をクエリで検索



意図した情報が検索結果にない

2. クエリを修正して再度検索



意図した情報が検索結果に含まれる

3. 意図した情報が含まれる投稿
のページに移る

Query Reformulation

Query: do and while in java

Search Results [Advanced Search Tips](#)

Results for **do and while in java**

35,563 results

Relevance Newest More ▾

7178 votes

A: "Thinking in AngularJS" if I have a jQuery background?

do the same thing, but in the AngularJS version anyone looking at the template knows what's supposed to happen. Whenever a new member of the development team comes on board, she can look at this **and** ... the reason you don't **do** that is that this is like half-jQuery, half-AngularJS, **and** no good. The problem here is that the developer is trying to "do jQuery" in the context of AngularJS. That's never ...

answered Feb 21 '13 by Josh David Miller

Reformulation: do and while **loop** in java

Search Results [Advanced Search Tips](#)

Results for **do and while loop in java**

4,419 results

Relevance Newest More ▾

327 votes

Q: How do I make a delay in Java?

I am trying to **do** something in **Java** **and** I need something to wait / delay for an amount of seconds in a **while loop**. **while** (true) { if (i == 3) { i = 0; } ceva[i].setSelected(true ...); // I need to wait here ceva[i].setSelected(false); // I need to wait here i++; } I want to build a step sequencer **and** I'm new to **Java**. Any suggestions? ...

8 answers

スレッド

1. 欲しい情報をクエリで検索



意図した情報が検索結果にない

2. クエリを修正して再度検索



意図した情報が検索結果に含まれる

意図した情報を取得するためにクエリを
修正することは煩わしい

(特に、対象のドメインに精通していないユーザ)

Automated Query Reformulation

- SEQUEARの提案
(Software-specific Query Reformulation)
 - 適切な結果が得られるクエリに自動修正する手法

事前調査

- クエリにはどのような特徴があるか？
- なぜQuery Reformulationが行われるのか？
- Query Reformulationによる変更の規模は？

データセット

- 対象期間：2017年12月～2018年11月
- リクエスト数：747,421,780 件

データ収集・整形

- SOサイト内で検索しているリクエストの抽出
 - Googleなどブラウザから直接投稿に遷移しているものは除外
- リクエストにイベントタグを付与
 - searching, post visiting, question list
- Cookieを用いて検索したユーザの特定
- ユーザ識別子ごとにデータを時系列に整列

クエリにはどのような特徴があるか？

- クエリ内容，長さ，高度検索内容进行分析

調査方法

クエリ内容

- クエリ文字列に対して前処理
- N-gram（～ 4-gram）を適用し，頻出表現を調査

クエリ長

- 空白で分割し単語数をカウント

高度検索内容

- 項目ごとにどのくらい利用されているかを調査

クエリ内容は？

- 頻出する表現の上位10件

Rank	1-gram	2-gram	3-gram	4-gram
1	python	how to	how to use	how to create a
2	java	in python	how to get	how to make a
3	file	what is	how do i	how do i use
4	string	in java	how to create	exception in thread "main"
5	android	failed to	how to make	ImportError: No module named
6	c#	in swift	could not find	how to get the
7	error	in r	how to change	no such file or
8	array	unable to	how can i	how to add a
9	sql	how do	how to find	how to check if
10	list	not found	how to install	such file or directory

1-gramの場合

- プログラミング言語 (python, java, sql, など)
- プラットフォーム (android)
- データ型 (string, array, list, など)

クエリ内容は？

- 頻出する表現の上位10件

Rank	1-gram	2-gram	3-gram	4-gram
1	python	how to	how to use	how to create a
2	java	in python	how to get	how to make a
3	file	what is	how do i	how do i use
4	string	in java	how to create	exception in thread "main"
5	android	failed to	how to make	ImportError: No module named
6	c#	in swift	could not find	how to get the
7	error	in r	how to change	no such file or
8	array	unable to	how can i	how to add a
9	sql	how do	how to find	how to check if
10	list	not found	how to install	such file or directory

2-gram, 3-gram, 4-gramの場合

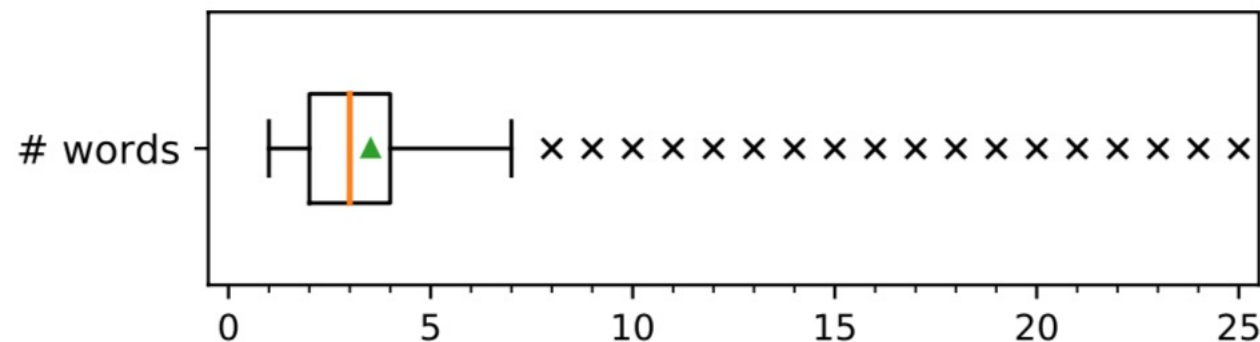
- “how to”というフレーズが頻出
- “exception”や“ImportError”が4-gramで上位
 - エラーコードをコピーしている

クエリの長さは？

伝統的な検索エンジンについて

ユーザは検索に引っかけられない状況を回避するために
少ない単語数でクエリを構成する [21]

(結果)



25% : 2 (単語), 中央値 : 3 (単語), 75% : 4 (単語)

SOにおける検索クエリは従来の検索エンジンと同様に
数単語で構成されていることがわかった

[21] P. Mahdabi, M. Keikha, S. Gerani, M. Landoni, and F. Crestani, "Building queries for prior-art search," in Proceedings of Information Retrieval Facility Conference. Springer, 2011, pp. 3–15.

高度検索内容は？

- 上位10項目

Type	Example	Proportion
tag	[powershell] job output	73.53%
user	user:8945947	10.35%
declare specific phrase	ios "save to files"	10.00%
exclude phrase	accordion -jquery	3.88%
Wildcard	[xamarin*] does not support...	0.74%
question only	is:question powershell version	0.44%
#answers	answers:0 firebase	0.28%
multiple tags	[scipy] or [numpy] array vs matrix	0.26%
score	safari cache score:3	0.19%
creation date	oauth read gmail created:05-04-2015..	0.06%

- クエリの8.74%が高度検索を利用している
- タグやユーザID, キーワードをがよく利用される

なぜQuery Reformulationが行われるのか？

- 384スレッドを手動で分類

調査方法

- スレッドの最初と最後のクエリを抽出
 - 著者2人でクエリの変化を4つのカテゴリに分類
 - Add, Modify, Delete, Other
 - 分類が異なった場合は議論で決定
-
- 分類の結果 Cohen's kappa^[24] = 0.83

値が1に近いほど両者の分類が一致している

[24] A.J.Viera,J.M.Garrettetal.,“Understandinginterobserveragreement: the kappa statistic,” Fam med, vol. 37, no. 5, pp. 360–363, 2005.

なぜQuery Reformulationが行われるのか？

Category	Sub-category	Example	Proportion
Add	Software or platform	why to use sha1 => why to use sha1 in android	21.61%
	Detailed requirement	db file => open db file	18.49%
Category subtotal			40.10%
Modify	Spelling and syntax check	.net string. empty => .net string. empty	19.01%
	Simplify and refine	list inside list for sightly => nested list in sightly	12.50%
	Turn to related information	python program freezes => python program hangs	2.08%
Category subtotal			33.59%
Delete	Detailed or unnecessary words	C# update a keypairvalue in a Dictionary => C# update Dictionary	15.10%
	Specific information in error message	Property 'getData' does not exist on type 'ReactInstance' => does not exist on type 'ReactInstance'	6.51%
	Symbols or web links	:= dbms_datapump.open => dbms_datapump.open	1.56%
Category subtotal			23.18%
Others			3.13%

- Add (40.1%) : ドメイン追加, 詳細化
- Modify (33.6%) : 誤字の修正, 言い換え
- Delete (23.2%) : 抽象化

Query Reformulationによる変更の規模は？

- 変更前と変更後のクエリの類似度を分析

類似度の計算方法

- 変更前後のクエリペアを作成
 - 1,121,185のクエリペア
- 文字レベルの最長共通部分列（LCS）の抽出
 - Longest Common Subsequence
 - [31]のテキストマッチングアルゴリズムを適用
- 類似度を計算

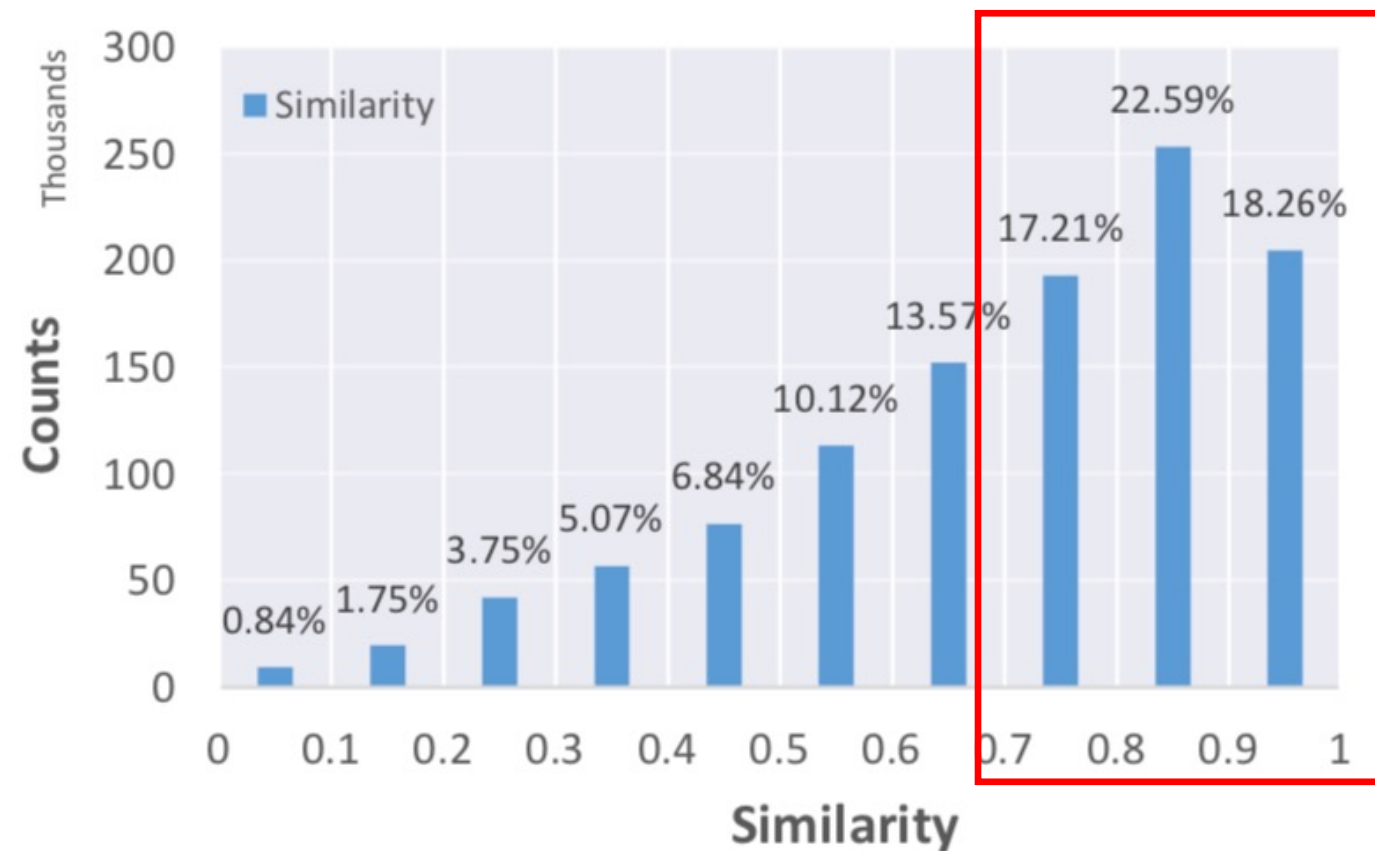
$$\text{similarity}(\text{original}, \text{reformulated}) = \frac{2 \times N_{\text{match}}}{N_{\text{total}}}$$

N_{match} : 最長共通部分列

N_{total} : 変更前後のクエリ文字数の合計

[31] L. Bergroth, H. Hakonen, and T. Raita, "A survey of longest common subsequence algorithms," in Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000. IEEE, 2000, pp. 39–48.

Query Reformulationによる変更の規模は？



過半数のクエリペアは
かなり類似している

Similarity (0.7～) : 約58%

- クエリそのものを書き換えるような大規模な変更はあまり行われなない
- 一部書き換えなどの小規模な変更が主に行われる

Implication

事前調査

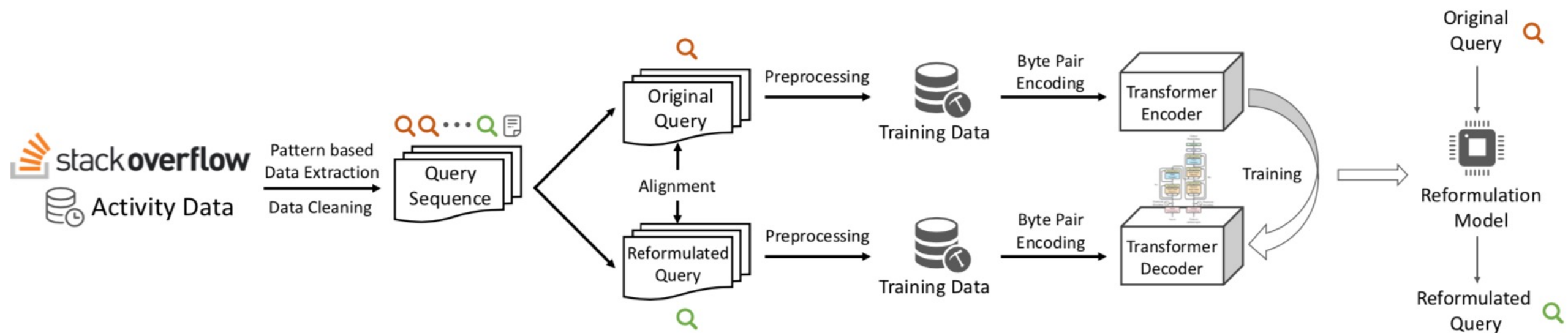
- クエリにはどのような特徴があるか？
 - 基本的には数単語で構成されている
 - エラーログやコードスニペットで検索される場合もある
- なぜQuery Reformulationが行われるのか？
 - Add, Modify, Deleteなど理由はさまざま（一般化は難しい）
- Query Reformulationによる変更の規模は？
 - 多くは小規模な変更にとどまる

モデル構築の検討

- Query Reformulationにはさまざまなタイプが存在するため、ルールベースは難しい
→ ディープラーニングでモデル構築
- 大規模な変更は、ほぼ新しいクエリ生成に近い
→ 小規模な変更を対象（Similarityが高い）
- 入力クエリと対応する修正クエリを出力
→ 機械翻訳問題（辞書引きをイメージ）

SEQUER

- Similarity 0.7以上の小規模な変更（p.15参照）を行っているクエリ対象に自動修正



- 機械翻訳で用いられるTransformer^[33]を利用
- 未知語対策としてBPE（Byte Pair Encoding）^[36]を適用

[33] A.Vaswani,N.Shazeer,N.Parmar,J.Uzkoreit,L.Jones,A.N.Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Proceedings of Advances in neural information processing systems, 2017, pp. 5998– 6008.

[36] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," *arXiv preprint arXiv:1508.07909*, 2015.

評価実験

- ベースライン
 - LanguageTool
 - GooglePS ^[19]
 - Head-qs ^[18]
 - seq2seq ^[17]
 - seq2seq + Attention
- 評価指標
 - GLEU ^{[52][53]}
 - MaxMatch ^[55]
 - ExactMatch

- [17] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Proceedings of Advances in neural information processing systems*, 2014, pp. 3104–3112.
- [18] A. Sordani, Y. Bengio, H. Vahabi, C. Lioma, J. Grue Simonsen, and J.-Y. Nie, “A hierarchical recurrent encoder-decoder for generative context-aware query suggestion,” in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 553–562.
- [19] R. C. Cornea and N. B. Weininger, “Providing autocomplete suggestions,” Feb. 4 2014, uS Patent 8,645,825.
- [52] C. Napoles, K. Sakaguchi, M. Post, and J. Tetreault, “Ground truth for grammatical error correction metrics,” in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2015, pp. 588–593.
- [53] —, “Gleu without tuning,” *arXiv preprint arXiv:1605.02592*, 2016.
- [55] D. Dahlmeier and H. T. Ng, “Better evaluation for grammatical error correction,” in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2012, pp. 568–572.

結果

Approach	EM@1	EM@5	EM@10	GLEU	M ² @P	M ² @R	M ² @F1
LanguageTool	—	—	5.87	53.27	15.31	6.11	8.73
GooglePS	6.20	8.14	8.58	61.07	24.67	25.84	25.24
HRED-qs	10.40	17.63	20.01	56.85	31.80	25.82	28.50
seq2seq	11.02	23.11	28.23	61.30	36.30	25.62	30.04
seq2seq+Attn.	14.53	28.47	33.77	62.93	35.93	21.15	26.62
SEQUER	22.21	33.47	39.37	67.68	39.67	31.97	35.41

- 全てにおいてベースラインを超える性能

ID	Original Query	GooglePS	seq2seq + Attn.	SEQUER
1	pandas delete last characters in strin	python delete last characters in string	pandas delete last characters in string	python delete last characters in string
2	playing sound in swift3	playing sound in swift 3	—	play sound in swift 3
3	swift grab currentdate	—	swift grab current date	swift get current date
4	assign string to number C3	—	assign string to number C#	assign string to number C#
5	do and while in java	do and while loop in java	—	do and while loop in java
6	requests negotiate	requests negotiate python	python requests negotiate	[python] requests negotiate
7	/usr/bin/ld: skipping incompatible libpthread.so	—	libskipping incompatible libpthread.so	/usr/bin/ld: skipping incompatible libpthread.so
8	subtracting the pandas series rf from all columns	—	subtracting the pandas series rf from all columns	subtracting the pandas series rf from all columns
9	<trigger>Missing report definition</trigger>	—	<trigger>Missing report definition</trigger>	<trigger>Missing report definition</trigger>
10	opencv scale image	—	—	opencv resize image
11	a* search	—	—	—
12	df -	—	—	—
13	res	—	—	—
14	vo	—	—	—

(誤字) assign string to number **C3** → assign string to number **C#**

shift押し忘れて”3”となっていたところを”#”に変更

(タグ追加) requests negotiate → **[python]** requests negotiate

pythonのrequestsモジュールと推測し，pythonタグを追加

まとめ

(目的)

- ユーザが意図する情報を取得するためのクエリ修正自動化

(やったこと)

- Query Reformulationの分析
 - 内容, タイプ分け, 変更規模, など
- Query Reformulationの自動化手法を構築
 - Transformerを利用したディープラーニングモデル
 - 全ての既存ツールと比べて良いパフォーマンス

所感

- モデル構築までの流れが整っており理解しやすい
 - モデル構築をする前に調査を行い、モデルが対象とするスコープを明確にしているところは参考になる
- 未知語問題（Out-of-Vocabulary）対策で使われていた Byte Pair Encoding が気になった
 - 低頻出単語への対応は自分の研究の課題でもある
- 評価指標の解釈が全然わからなかった
 - 指標が示す値が相対的にどの位置付けにあるのか？
- NLPするなら Transformer モデルの勉強はしておくべき
 - BERTなどの元になっている最近主流のモデル

参考サイト

- Transformerモデル (Qiita)
 - <https://qiita.com/halhorn/items/c91497522be27bde17ce>
 - <https://qiita.com/omiita/items/07e69aef6c156d23c538>
- Byte Pair Encoding (SlideShare)
 - <https://www.slideshare.net/ssuserd79a5c1/2019bpe>