An Question Recommendation System for Question Answer Community (Stackoverflow)

Presenter: Haoyu Chen           Haoran Hou
What is Question Answering Community:
Community question answering (cQA) provides a platform for people with diverse background to share information and knowledge.
People need help!

"How do I take care of my pet potato?"

Resolved Question

How do i take care of my pet potato?

i just got a pet potato but i dont know how to take care of it.

2 weeks ago

Julia

Report Abuse
What we decided to work on:

There’s only one style of programming: stackoverflow oriented programming.
Exhibit A:
Result Ranking doesn’t consider about the quality of answers.

<table>
<thead>
<tr>
<th>Title</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>replace() and replaceAll() in Java</td>
<td>2</td>
</tr>
<tr>
<td>Difference between replace and replaceAll [duplicate]</td>
<td>1</td>
</tr>
<tr>
<td>Difference between String replace() and replaceAll()</td>
<td>5</td>
</tr>
<tr>
<td>Replace in C# and replaceAll in Java</td>
<td>2</td>
</tr>
<tr>
<td>Regexp grouping and replaceAll in Java</td>
<td>3</td>
</tr>
<tr>
<td>What is Difference between inheritance and delegation in java [duplicate]</td>
<td>2</td>
</tr>
</tbody>
</table>
Exhibit B:
Result Ranking doesn’t work well in some cases
What we aim to do:

- Find similar questions and list them in more reasonable order.
- Get answers in a faster and more convenient way.
About stackoverflow

- No need for sentiment analysis
- Few duplicated questions
- Provide tags
- Ordered Answer: Voting
- Full data provided

New query

-> Best existing post with most similar query
-> Return best answer
Our thoughts on improvement:

- query-answer matching: After finding similar existing queries, compute the similarity between the new query and the best answer

- Adding tag matching along with query matching

- Find the reasonable ‘return-best-answer’ strategy
query-answer matching
Query: difference replace replaceAll java

Question content

Difference between String replace() and replaceAll()

I am asking this question out of curiosity.

What's the difference between java.lang.String's replace() and replaceAll() methods, other than later uses regex. For simple substitutions like, replace . with / . is there any difference?

59

15

share edit

edited Jul 14 '14 at 7:30

user2864740

27.6k • 4 • 22 • 58

asked May 31 '12 at 5:24

Best answer

In java.lang.String, the replace method either takes a pair of char's or a pair of CharSequence's (of which String is a subclass, so it'll happily take a pair of String's). The replace method will replace all occurrences of a char or CharSetence. On the other hand, both String arguments to replaceFirst and replaceAll are regular expressions (regex). Using the wrong function can lead to subtle bugs.

47

share edit

edited May 31 '12 at 5:33

answered May 31 '12 at 5:28

Only compute new query and existing query
**Adding tag matching**

Compute the similarity between existing queries, as well as their tags.

e.g.
new query: *difference replace replaceall java*

existing query: *difference between string replace() and replaceall()*

tags: java string replace
Find answer:

More votes -> acceptance

Favor vote more than acceptance

Return even if there’s no (good) answer: comments
Let’s start from Solr

Solr is the popular, blazing-fast, open source enterprise search platform built on Apache Lucene

--- The Headline on Solr Official Website
Key Facts on *Stackoverflow data*

**Open** -- Under CC BY-SA 3.0 (ShareAlike and Attribution)
**API** -- E.g. Search Users, Answer, Questions
**Updation** -- every Monday
**Size** -- 8 million questions (28G)

Link: http://data.stackexchange.com/help
Preprocessing Stackoverflow data

Select Useful features -- Tags, QuestionsID, Titles

Convert it into Solr input format

Result: 28G -> 1.6G
Search Flow Chart

Search Java ....

Google app engine

Indexed data

Apache Solr
**Solr similarity algorithm:**

\[
\text{score}(q,d) = \text{coord}(q,d) \cdot \text{queryNorm}(q) \cdot \sum_{t \in q} \left( \frac{\text{tf}(t \text{ in } d) \cdot \text{idf}(t)^2 \cdot t\text{-getBoost()} \cdot \text{norm}(t,d)}{\text{doc.getBoost()} \cdot \text{lengthNorm} \cdot \prod_{\text{field } f \text{ in } d \text{ named as } t} f\text{-getBoost()}} \right)
\]

- Document contains more query's term: the higher
- Make scores between queries comparable
- Normalize document with boost
Let’s Demo Our Tools!

StackOverFlow Questions Search

replace replaceall difference java

- Difference between String replace() and replaceAll()
- String replaceAll() vs. Matcher replaceAll() (Performance differences)
- How resolve a replaceAll of a replaceAll
- Java regex replaceAll multiline
- Java ReplaceAll Problem
Let’s Demo Our Tools!

Features:
- Auto change detection
- Answer overview - (More responsive than StackOverflow version)

Difference:
- Search not just for title, but also tags.
- Show answer with the largest votes

Testing Questions:
- Replace
Demo 1

Title: difference git pull and git fetch

Questions that may already have your answer:

- What's difference between git pull and git fetch + rebase (2)
- What is the difference between 'git remote update', 'git fetch' and 'git pull'? (2)
- 'git pull' and 'git fetch' do not interpret arguments consistently: (design) bug or (design) feature? (2)
- Difference between git pull and git pull --rebase (4)
- What are the differences between 'git pull' and 'git fetch'? (21)
- Difference between git pull and git pull origin master [duplicate] (3)
<table>
<thead>
<tr>
<th>Question</th>
<th>Votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Git fetch and pull with no arguments</td>
<td>10</td>
</tr>
<tr>
<td>What are the differences between 'git pull' and 'git fetch'?</td>
<td>3619</td>
</tr>
<tr>
<td>Git fetch and pull operation on branches</td>
<td>1</td>
</tr>
<tr>
<td>Are git pull and push repository wide operations or branch specific?</td>
<td>2</td>
</tr>
<tr>
<td>git fetch and pull</td>
<td>3</td>
</tr>
</tbody>
</table>

In the simplest terms, **git pull** does a **git fetch** followed by a **git merge**.

You can do a **git fetch** at any time to update your remote-tracking branches under **refs/remotes/**<remote>/**. This operation never changes any of your own local branches under **refs/heads**, and is safe to do without changing your working copy. I have even heard of people running **git fetch** periodically in a separate in the background (although I wouldn't.
Future steps

- Distribute different weight to question title and tags
- Dig more information provided by comments
- Recommend tag using MoreLikeThis feature