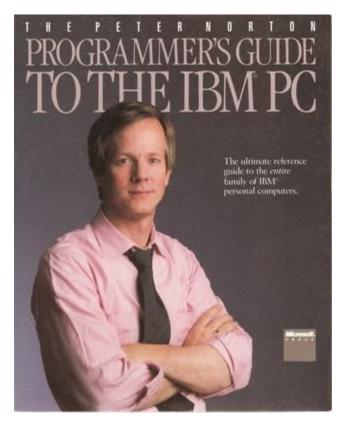


Christoph Treude

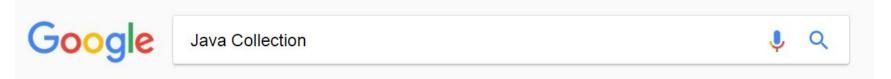
Bimodal Software Documentation

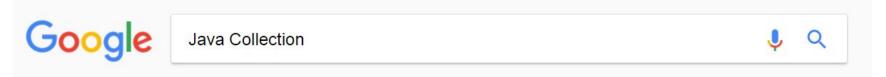
adelaide.edu.au seek LIGHT

Software Documentation



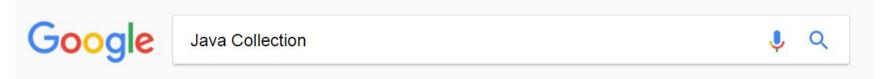
[1985]



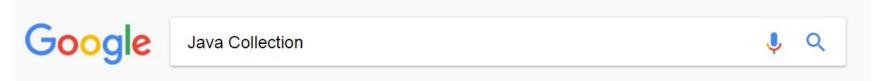


100%





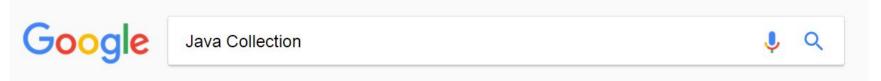


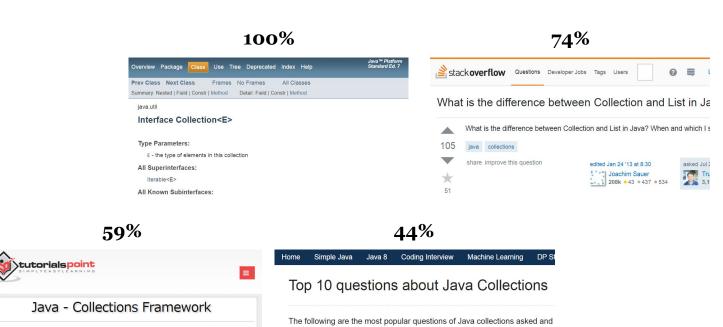




59%





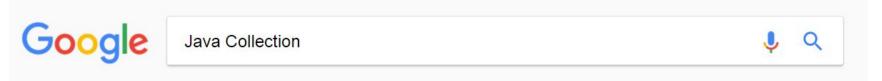


○ Previous Page Next Page ○
Prior to Java 2, Java provided ad hoc classes such as Dictionary, Vector, Stack, and Properties to store and manipulate groups of objects. Although these classes were quite useful, they lacked a central, unifying theme. Thus, the way that you used Vector was different from

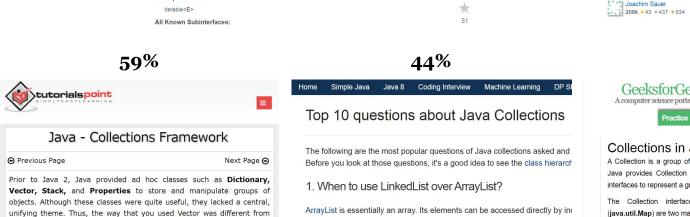
ArrayList is essentially an array. Its elements can be accessed directly by inc

1. When to use LinkedList over ArrayList?

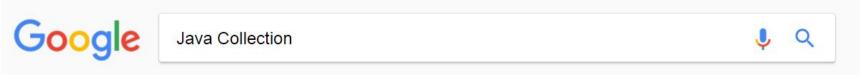
Before you look at those questions, it's a good idea to see the class hierarch

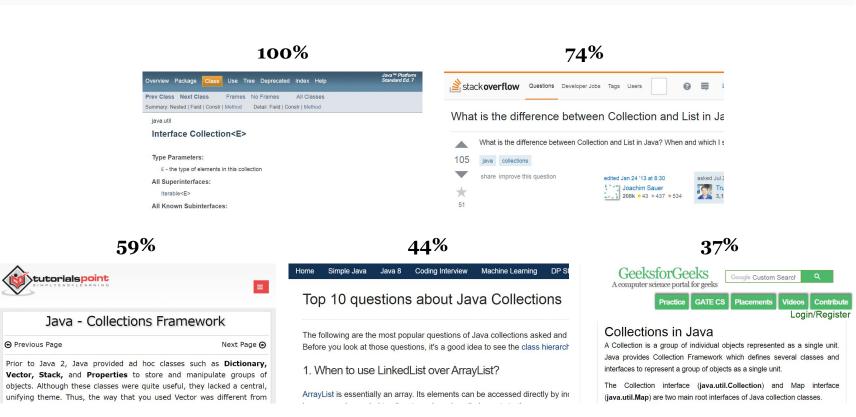












162 different domains in the top 10 for 99 queries

Tensorflow Python API: 309 different domains in the top 10 for 2,192 queries



Tensorflow Python API: 309 different domains in the top 10 for 2,192 queries



jQuery Event API: 75 different domains in the top 10 for 57 queries



Navigating documentation is not trivial



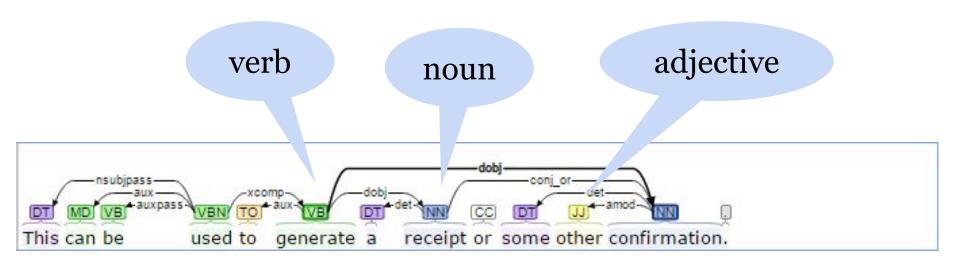
Navigating documentation is not trivial



Common Tasks

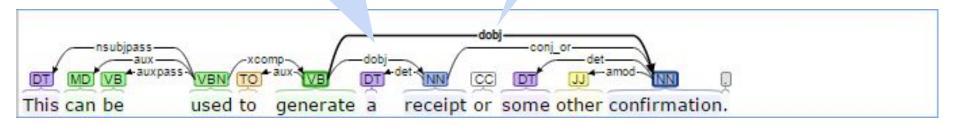
<u>Link</u> **Link** <u>Link</u> **Link** <u>Link</u> **Link** <u>Link</u> **Link**

Extracting tasks from documentation

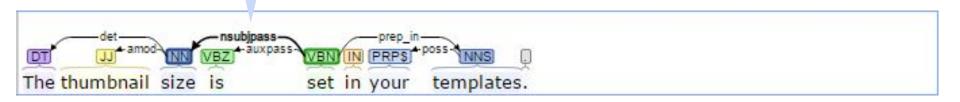


direct object: generate confirmation

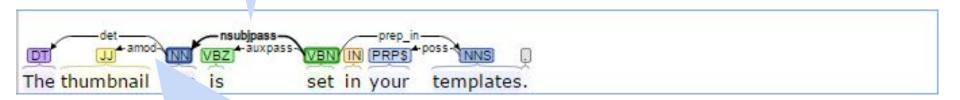
direct object: generate receipt



passive nominal subject: set size



passive nominal subject: set size



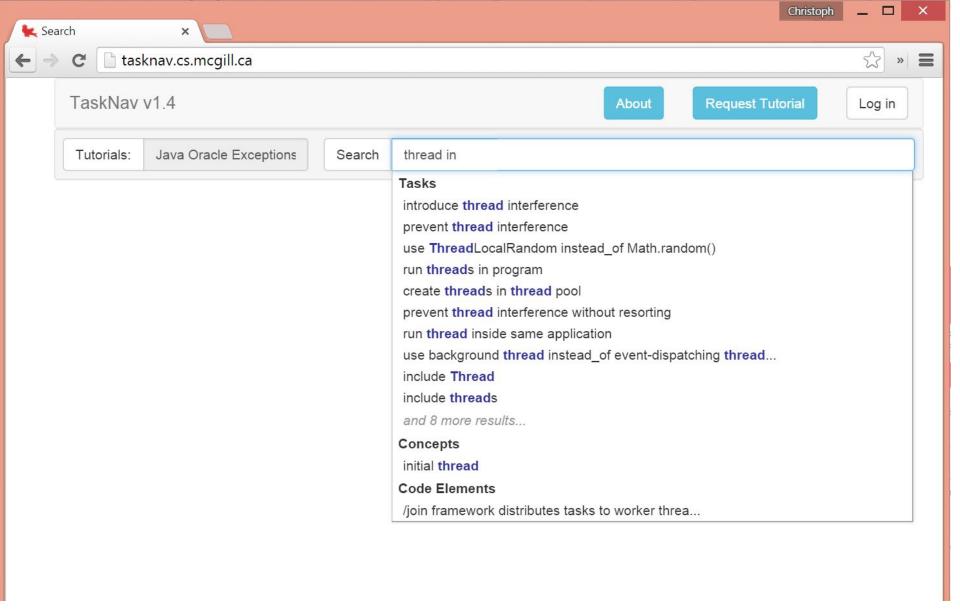
adjective modifier: set thumbnail size

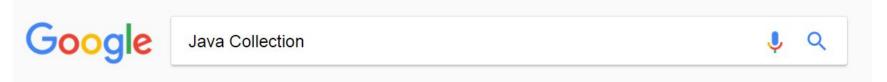
passive nominal subject: set size

preposition:set thumbnail sizein templates



adjective modifier: set thumbnail size







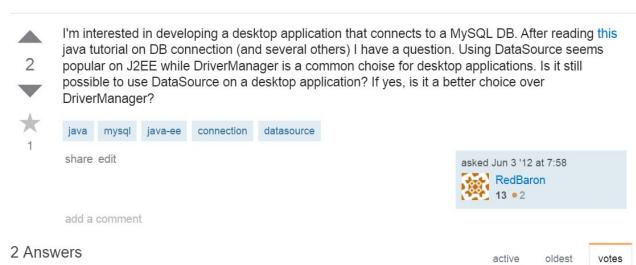








DataSource and DriverManager on J2SE





I'll quote the javadoc:

2



An alternative to the DriverManager facility, a DataSource object is the preferred means of getting a connection.

Normally you use <code>DriverManager</code> when you just want a connection for one time while with <code>DataSource</code> you get other features such as connection pooling and distributed transactions.

Hope this helps.

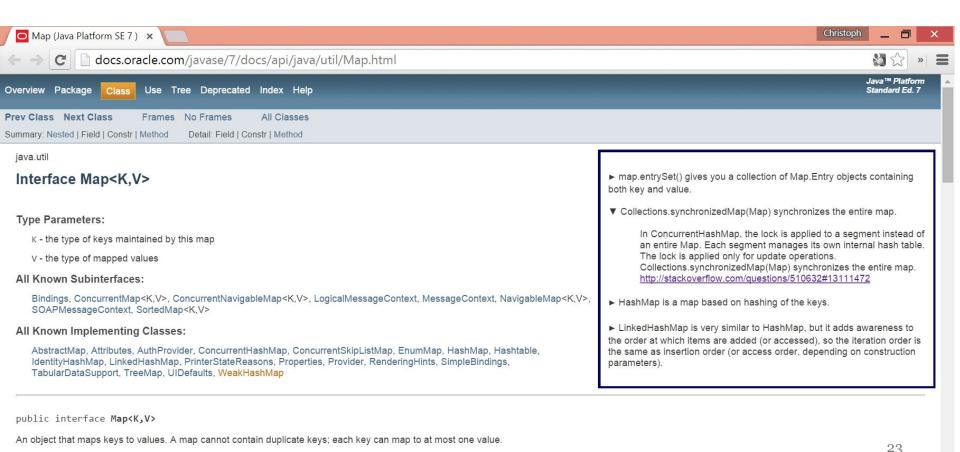
insight sentence

a sentence from Stack
Overflow that is related to a
particular API type and that
provides insight not
contained in the API
documentation of that type



Supervised Insight Sentence Extractor

Augment API documentation with insights from Stack Overflow



This interface takes the place of the Dictionary class, which was a totally abstract class rather than an interface.

Bimodal software documentation



Challenges in Analyzing Documentation

- Software documentation is technical and often contains references to code elements
- Natural language text written by software developers may not obey all grammatical rules, e.g.,
 - sentences that are grammatically incomplete
 - content that has not authored by a native speaker







Returns the C++ variable.



Returns the C++ variable.



Returns the C++ variable.



Returns the C++ variable.



Returns the C + + variable .



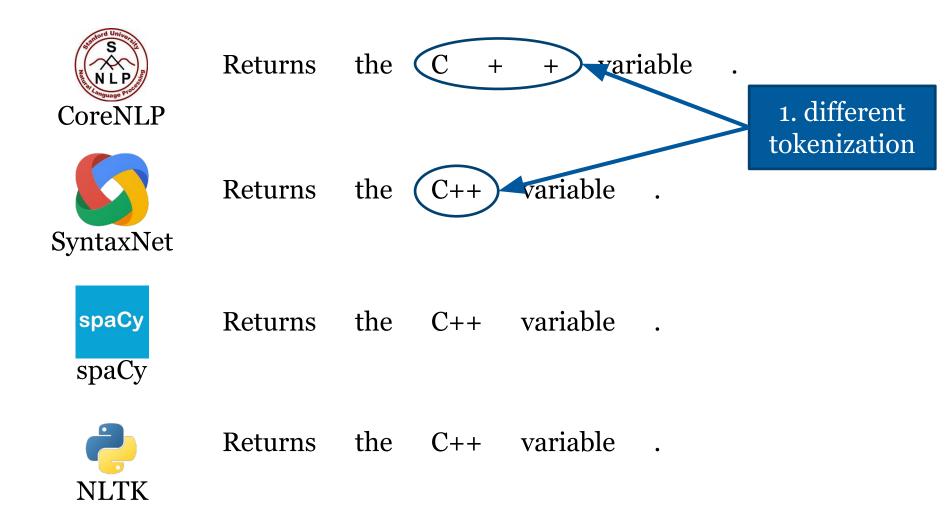
Returns the C++ variable .



Returns the C++ variable



Returns the C++ variable





1. different tokenization





Returns the
$$C++$$
 variable VBZ DT NNP NN



Returns the C++ variable .
$$NNS$$
 DT NN JJ .



Returns the C + + variable

NNS DT NN JJ CC JJ

1. different tokenization



Returns the C++ variable . \overline{VBZ} DT NNP NN .

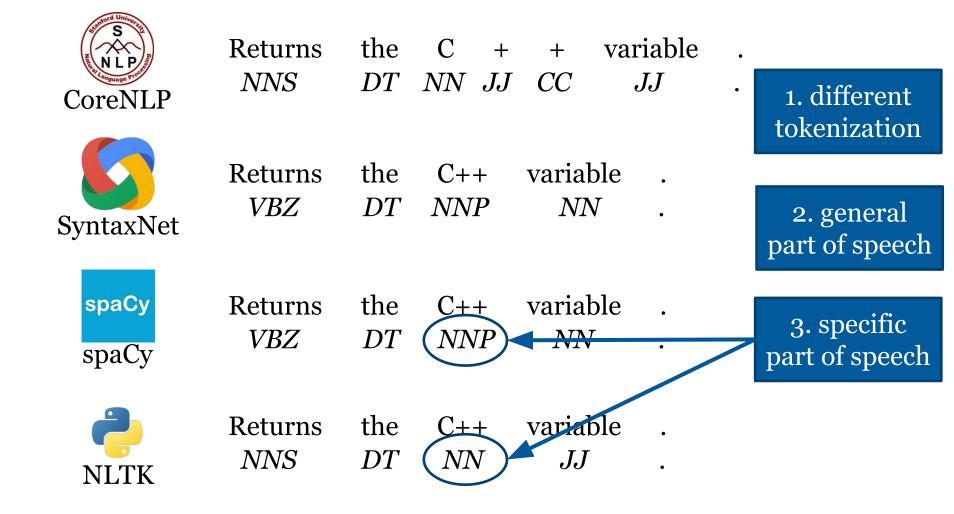
2. general part of speech



Returns the C++ variable NND = NDD = N



Returns the C++ variable NNS DT NN JJ .











Only between 60% and 71% of tokens from Stack Overflow, GitHub, and the Java API Documentation were assigned the same part-of-speech tag by all four libraries.

1. different tokenization

2. general part of speech

3. specific part of speech

Bimodal software documentation



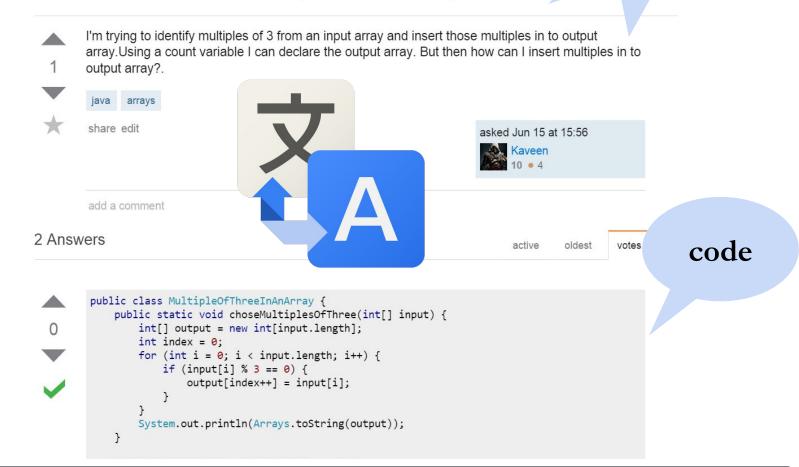
Bimodal software documentation

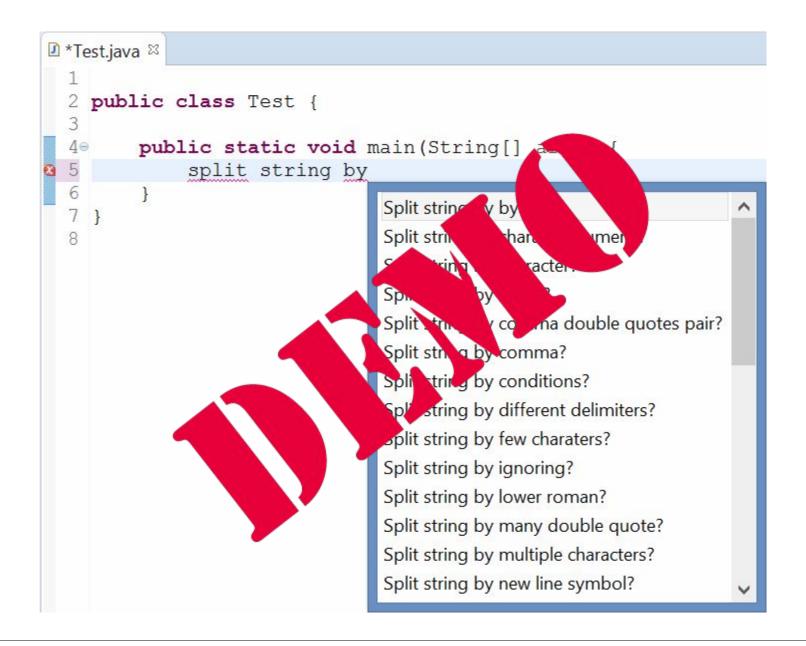
tasks



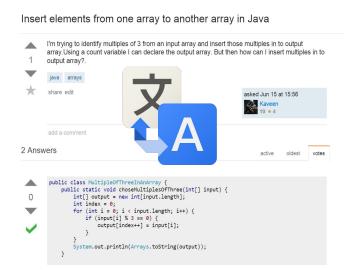
Code Snippet Content Assist

tasks



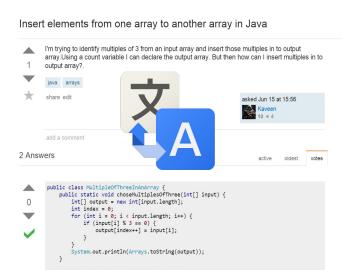






The integration of natural language and code in documentation





The integration of natural language and code in documentation

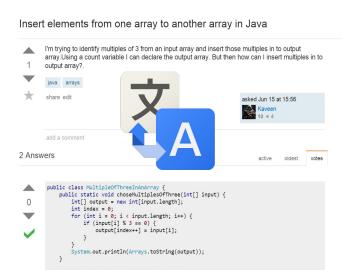


```
■ *Test.java 

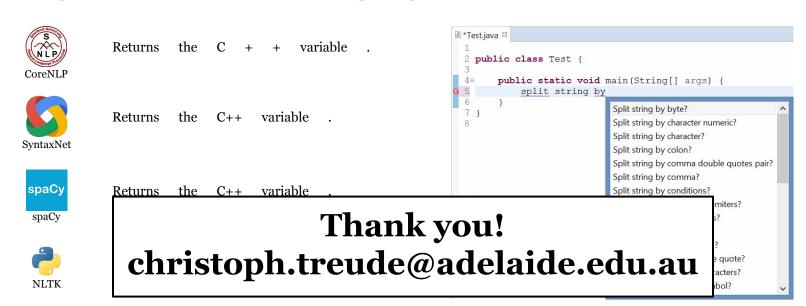
□
    public class Test {
           public static void main(String[] args) {
                 split string by
                                        Split string by byte?
  7 }
                                        Split string by character numeric?
                                        Split string by character?
                                        Split string by colon?
                                        Split string by comma double guotes pair?
                                        Split string by comma?
                                        Split string by conditions?
                                        Split string by different delimiters?
                                         Split string by few charaters?
                                        Split string by ignoring?
                                         Split string by lower roman?
                                         Split string by many double quote?
                                        Split string by multiple characters?
                                         Split string by new line symbol?
```

creates challenges & opportunities for software engineering tools.





The integration of natural language and code in documentation



creates challenges & opportunities for software engineering tools.