String Processing

- Often need to go through long string and
  - Extract information from it
  - Modify it

- There are tools in the API for doing this
  - However, you need to be able to write your own code because the tools may not meet your needs
  - Will also look at API later

- Sample problems:
  - Extract all of the words from text (strip punctuation and spaces)
  - Find all of the URLs in some file
  - Look for key words: insert material there
  - ???
Simple String Processing

- **input** has words separated by a single space
  - No extraneous text to worry about
  - Want output as an array of Strings

```java
ArrayList<String> words = new ArrayList<String>();
while (input.length() > 0) {
    int spPos = input.indexOf(" "); // find boundary
    if (spPos >= 0) {
        String word = input.substring(0, spPos); // extract
        words.add(word);
        input = input.substring(spPos + 1); // chop off front
    } else {
        words.add(input);
        input = " ";
    }
}
String[] output = words.toArray(new String[0]);
```
Simple String Processing

- Remove all punctuation and spaces from input
  - Want output as String

```java
String output = "";
for (int k = 0; k < input.length(); k++) {
    char c = input.charAt(k); // get char
    if (Character.isLetter(c)) { // decide
        output += c; // built output
    }
}
```
Simple String Processing

- Replace all occurrences of "which" in \textit{input} with "that"
  - Want output as \textit{String}
    \begin{verbatim}
    String output = "";
    while (input.length() > 0) {
        int loc = input.indexOf("which"); // find key
        if (loc >= 0) {
            output += input.substring(0, loc) + "that";
            String word = input.substring(0, spPos); // build
            input = input.substring(loc + 1); // chop off front
        } else {
            output += input;
            input = "";
        }
    }
    \end{verbatim}
  - What are some flaws in this code?
String Processing Using API

- Can use split method of String class to get an array
  - Want output as String array
  - Specify separator string
  - For single space separator use:

    ```java
    String[] output = input.split(" ");
    ```

  - `split` has two forms and many options, so look at the API
String Processing Using API

- Can use Scanner with String as input
  - Want output as String array
  - Specify separator string
  - For single space separator use:

```java
ArrayList<String> words = new ArrayList<String>();
Scanner scan = new Scanner(input);
while (scan.hasNext()) {
    words.add(scan.next());
}
String[] output = words.toArray(new String[0]);
```

- Scanner has many additional methods to process String