Paper:

Background on RegEx TIFFANY

Regular expressions are present in a vast amount of scripting languages (Perl, Java, Python, et cetera) and serve as a powerful programming tool using algebraic notation (Medeiros, Mascarenhas, & Ierusalimschy, 2014) (Campeanu & Santean, 2009). The syntax of regular expressions specify a pattern a user is searching for in a string or a full string against a list of strings (Medeiros, Mascarenhas, & Ierusalimschy, 2014). A modification of regular expressions called the extended regular expressions with back-references (regex), creates a means of creating expressions for patterns (repetitions) by performing a usual pattern match, but then “backtracking when a particular path through the expression makes the match fail”. (Schmid, 2013) (Medeiros, Mascarenhas, & Ierusalimschy, 2014). This backtracking or back-referencing refers to the revisiting of the earlier subexpression when a match fail occurs, to provide a more accurate pattern search throughout a string (Schmid, 2013). “For example, r:=( 1 (a|b)\*)1c\1 is a regex, where \1 is a backreference to the referenced subexpression in between the parentheses (1 and)1 (Schmid, 2013). In this example r denotes the set within the string, a|b are alterations of the search pattern, and c denotes the frequency (Schmid, 2013). Ad-hoc optimizations are used to reduce the amount of backtracking of regex implementation, which in turn decreases the running time of the pattern search (Medeiros, Mascarenhas, & Ierusalimschy, 2014).

When implementing regex as a programming tool, there are specifics in syntax which is dependent on the programming language. In regards to our GFinder algorithm, Java programming language was used. When regex is introduced within a class in Java, the pattern is first specified as a string, then compiled into an instance of the class (Oracle, 2014). The pattern is then “used to create a Matcher object that can match arbitrary character sequences against the regular expression” (Oracle, 2014). Figure [#] is an example of a typical regex invocation sequence.

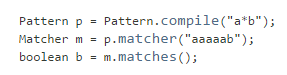


Figure [#]: An example of a typical regex invocation sequence.

Oracle (2014). Oracle: Class pattern. Retrieved from http://docs.oracle.com/javase/7/docs/api/java/util/regex/Pattern.html

Reference:

Campeanu C., & Santean, N. (2009). On the intersection of regex languages with regular expression. *Theoretical Computer Science*, 410, 2336-2344. <http://dx.doi.org/10.1016/j.tcs.2009.02.022>

Schmid, M.L. (2013). Inside the class of regex languages. International Journal of Foundations of Computer Science, 24(7), 1117-1134. Doi: 10.1142/S0129054113400340.

Medeiros, S., Mascarenhas, F., & Ierusalimschy, R. (2014). From regrexes to parsing expression grammars. Science of Computer Programming, 93, 3-18. Doi: 10.1016/j.scico.2012.11.006.

Oracle (2014). Oracle: Class pattern. Retrieved from http://docs.oracle.com/javase/7/docs/api/java/util/regex/Pattern.html

Let's have TIFFANY cover the installation of Java, Eclipse and the BioJava packages (please include relevant screen shots -- if you use any online resources to help with this just toss those on the references thread in discussion

To be able to enhance, troubleshoot, and to modify GFinder in any way, the use of an integrated development environment (IDE) is needed. Eclipse is the IDE our group used in the development of the GFinder algorithm. In order to manipulate and test the algorithm or associated methods of the algorithm, proper installations of this IDE, along with the Java Development Kit (JDK) is central. All instructions provided are based on the Windows operating systems.

Before installing Eclipse, a user must first download a JDK. In Appendix [#] you will find a Java SE download link as well as its commercial installation guide. The user can click on the link or copy/paste the url in any internet browser. When on the Java SE downloads home page click on the download button under JDK in the section labeled Java SE 8u65 / 8u66 (or the most up-to-date version). The next page would take you to a variety of downloads, the user should choose the proper Windows x86 or Windowsx64 download, downloading the specified version of Windows, under the Java SE Development Kit 8u65 section (or the most up-to-date version). A user can find the computer’s Windows operating system version by right-clicking on “This PC” or “My Computer” from your hard drive application and choosing properties. Use the Java platform, standard edition installation guide found in Appendix [#] for a step by step process for the proper download. Also, save the download file in a specific folder to be able to find to for later use.

Java SE download:

<http://www.oracle.com/technetwork/java/javase/downloads/index.html>

Java platform, standard edition installation guide:

<http://docs.oracle.com/javase/8/docs/technotes/guides/install/install_overview.html>

Once the JDK is downloaded and installed, Eclipse can then be downloaded and installed. In Appendix [#], you will find a link to download Eclipse. Click or copy/paste the url in the browser of your choice to start this process. There are many IDEs found on the Eclipse downloads home page, scroll to the Eclipse package labeled, “Eclipse IDE for Java Developers” and click on the appropriate download as mentioned above. Choose the default mirror download found in the gray and orange box on the left side of the page. This download should be saved to a familiar folder for instillation.

Eclipse IDE for Java Developers download:

<http://www.eclipse.org/downloads/>

To install Eclipse, the zip file of the Eclipse download should be unzipped first. To unzip a file, right-click the fine, select the WinZip command, then click “extract to here”. This creates a folder named eclipse. Create a shortcut of the eclipse.exe file onto your desktop for convenience of setup and use.

To start the setup process of Eclipse, double click the shortcut. A workspace launcher window should populate, where the pathway of all Java projects will be saved is designated. Accepting the default location would be beneficial if you are a beginner Java programmer (after setup, the file pathway can be found by selecting properties under the file tab in this application). Make sure the box labeled “use this as the default and do not ask again” box is checked before clicking OK. After installing Eclipse, restart your computer. Once your system is restarted, open Eclipse and exit out of the welcome message.

To run our GFinder algorithm we will provide you with a numerous .java and .jar files via email. To export the algorithm (GFinder.java) for modifications, first a java package would have to be made. To create a java package click on the drop down arrow of the new icon and select java project. The user would then name the project by any name and keep the default settings. After this step is complete, the provided GFinder.java file could be uploaded for modification. To import this file, a user should right-click the package name and click on import. After selecting the import function a window will populate with multiple folders. The user should choose the general folder, then double-click archive file. After the pathway is selected of the GFinger.java file, click the Finish button. After this file is uploaded, under the src file within your created java package, double-click the GFinder file and to the right the java code will populate. For this code to run the biojava.jar files we will provide would have to be imported as well. To import the provided .jar files, first right-click the created package, and select “add external archives” from the build path tab. Click on the browse button and select the .jar files that were provided, then click open. The imported .jar files will then be located under the “Referenced Libraries” folder within the created package.

<http://www3.ntu.edu.sg/home/ehchua/programming/howto/eclipsejava_howto.html>

<http://www.wikihow.com/Download,-Install,-and-Run-JDK-and-Eclipse>

<http://www.eclipse.org/downloads/index.php?show_instructions=TRUE>

### [Eclipse IDE for Eclipse Committers 4.5.1](http://www.eclipse.org/downloads/packages/eclipse-ide-eclipse-committers-451/mars1)

Eclipse Luna

<https://www.cs.umd.edu/eclipse/install_java8_luna.html>

<http://scn.sap.com/community/abap/eclipse/blog/2015/06/03/install-eclipse-luna--all-installation-steps>

What is a compiler?

http://whatis.techtarget.com/definition/compiler

Java and Eclipse: Download/Installation Instructions:

https://www.cs.cmu.edu/~pattis/15-1XX/common/handouts/javaeclipseinstallation.html

Installing Eclipse:

https://www.cs.umd.edu/class/fall2004/cmsc131/EclipseTutorial/install.html

Java SE Downloads:

http://www.oracle.com/technetwork/java/javase/downloads/index.html

Eclipse IDE for Eclipse Committers

4.5.1 https://eclipse.org/downloads/

How to add biojava.jar

https://www.cs.duke.edu/courses/fall07/cps004g/assign/final/shotgun/addlibrary.html

BIOJAVA RESOURCES:

Main BioJava repository:

http://biojava.org/wiki/Main\_Page

NCBI Fetch (being used to retrieve FASTA sequence):

http://biojava.org/wiki/BioJava:CookBook:ExternalSources:NCBIFetch

NCBI QBLAST Service (being used to run BLAST search on candidate sequences):

http://biojava.org/wiki/BioJava:CookBook3:NCBIQBlastService

(the problem, the analysis, the recommendations)

Presentation:

* Discuss NCBIFetch and its final configuration (TIFFANY)
* Final version of regex search (initial to final version to show evolution) (TIFFANY)