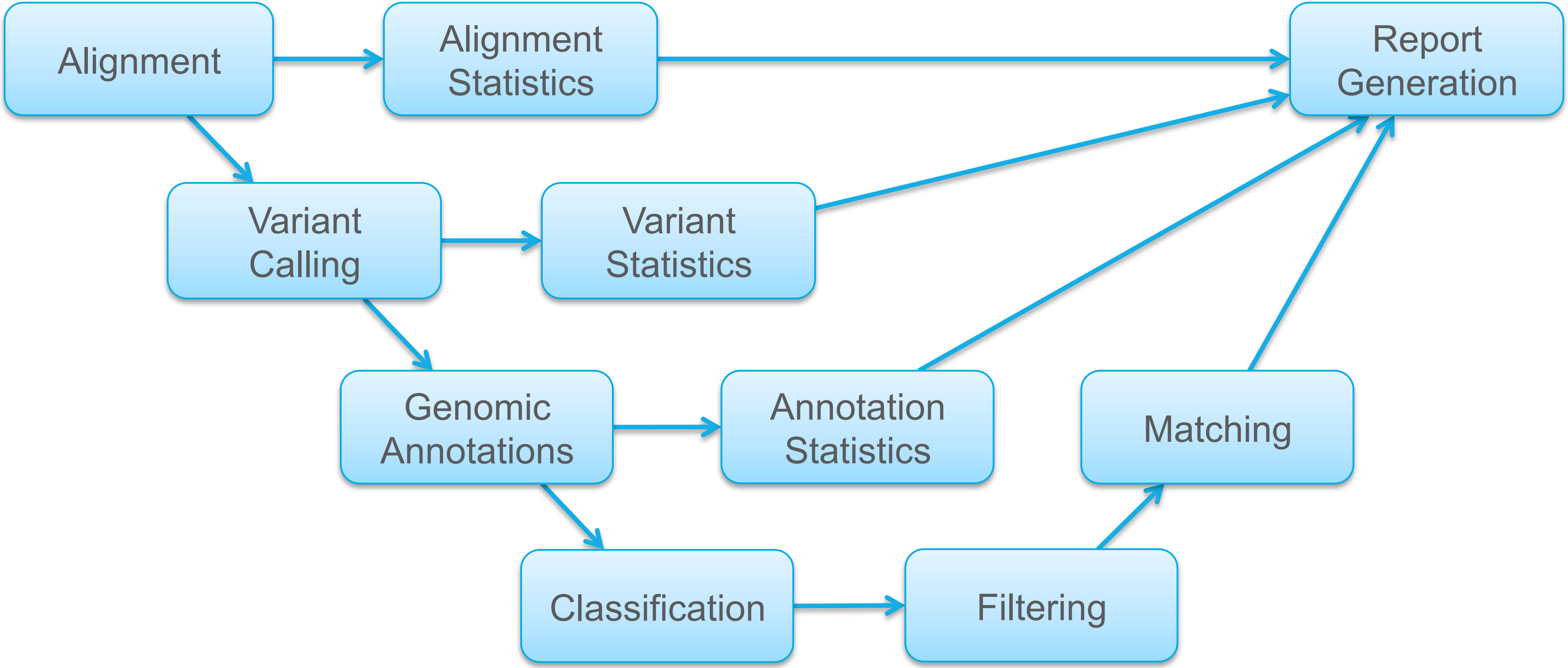
Data-Flow Programming

Motivation

Related Work

CloudKeeper

Example: Genome-Analysis Workflows



Related Work

Academic Workflow Systems

Taverna

UNICORE

Pegasus

Appistry

nextflow

Galaxy wg

Business Process Execution Language

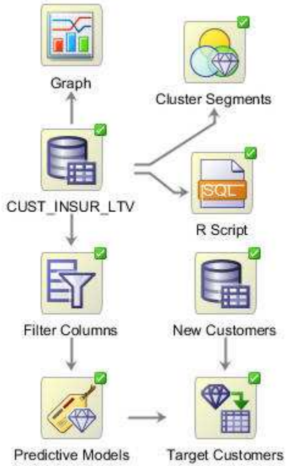
ORACLE
BPEL PROCESS MANAGER

Batch Processing

- Java EE 7 / JSR 352
- spring Batch, XD
- Apache Spark

Domain-Specific Workflow Tools

- Oracle Data Miner



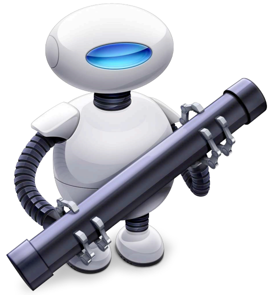
Three Select Aspects for Taxonomy

Embedded as library

Stand-alone application

Textual programming language

Graphical user interface



Binary artifact repository

Social-network-like web site

Software Engineers ↔ Users: Geneticists, etc.

CloudKeeper

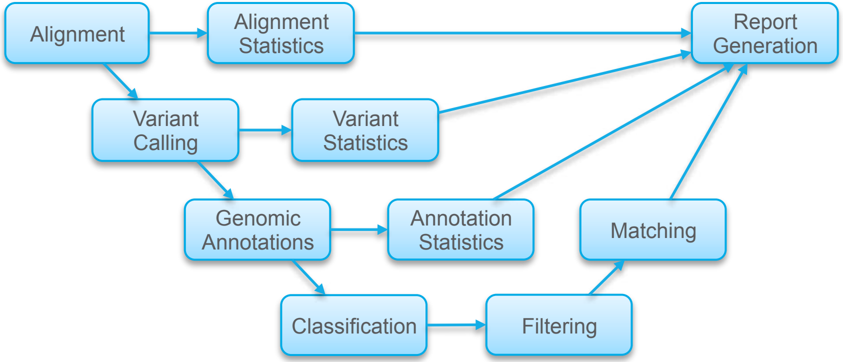
Lifecode's Data-Flow Language and Runtime System

1 Modular data-flow programming "in the large"

2 Scalable: write once, run everywhere

3 Satisfy needs of software engineers and expert users

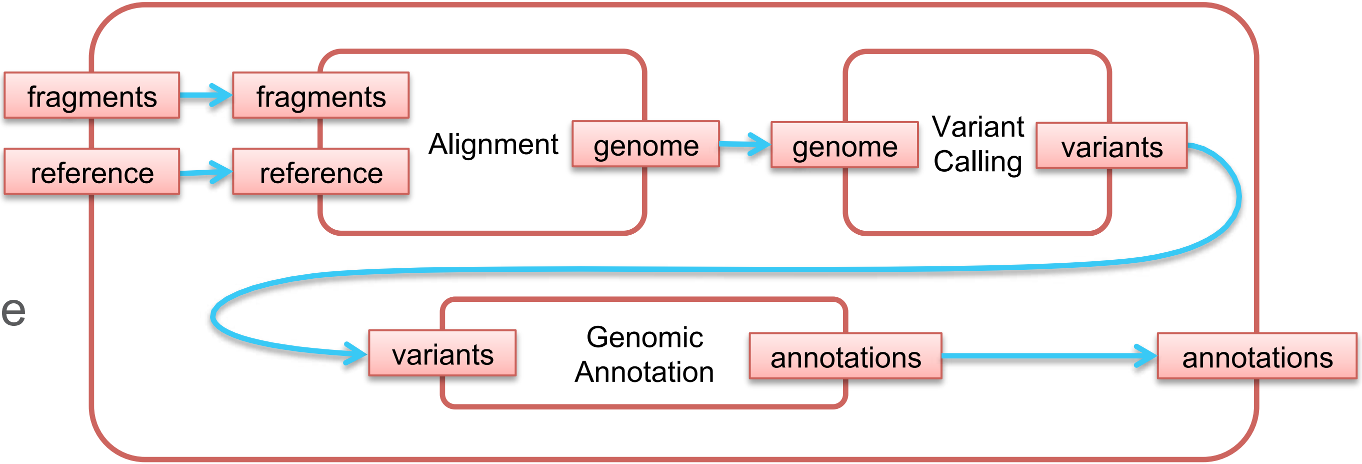
4 Lightweight and easy to deploy



1 Modular Data-Flow Programming “in the Large”

Modules

- Functional units
- In- and out-ports
- Simple or composite



Programming “in the Large”

- Instantiate modules and define connections between ports
- Control flow via special composite modules (e.g., loop modules)

2 Scalable: Write Once, Run Everywhere

Scale Horizontally

- Runs within a single Java Virtual Machine on a laptop
- Same code also runs on a cluster or in the cloud

DRMAA
Distributed Resource Management Application API — www.drmaa.org

GRID ENGINE

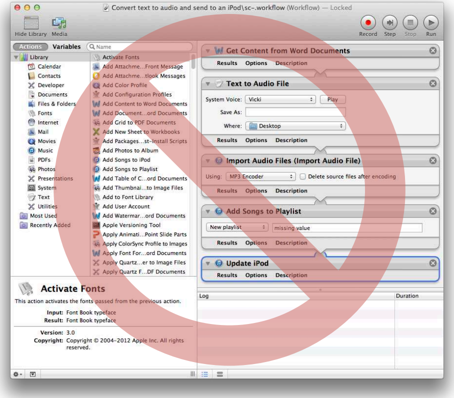
HTCondor
High Throughput Computing

- Transparent data transfer using main memory, file system, S3, etc.
- Checkpointing

3 Satisfy Needs of Software Engineers and Expert Users

Workflow System with Developers in Mind

- Internal DSL on top of Java
  - Works with all IDEs
  - Easy refactoring
- Statically typed and verified


Convert text to audio and send to an iPod\sc-.workflow (Workflow) — Locked
Hide Library
Media
Record
Step
Stop
Run
Actions
Variables
Name
Library
Calendar
Contacts
Developer
Documents
Files & Folders
Fonts
Internet
Mail
Movies
Music
PDFs
Photos
Presentations
System
Text
Utilities
Most Used
Recently Added
Activate Fonts
Add Attachme…Front Message
Add Attachme…tlook Messages
Add Color Profile
Add Configuration Profiles
Add Content to Word Documents
Add Document…ord Documents
Add Grid to PDF Documents
Add New Sheet to Workbooks
Add Packages…st-Install Scripts
Add Photos to Album
Add Songs to iPod
Add Songs to Playlist
Add Table of C…ord Documents
Add Thumbnai…to Image Files
Add to Font Library
Add User Account
Add Watermar…ord Documents
Apple Versioning Tool
Apply Animati…Point Slide Parts
Apply ColorSync Profile to Images
Apply Font For…ord Documents
Apply Quartz…er to Image Files
Apply Quartz F…DF Documents
Get Content from Word Documents
Results Options Description
Text to Audio File
Play
Results Options Description
Import Audio Files (Import Audio File)
Results Options Description
Add Songs to Playlist
Results Options Description
Update iPod
Results Options Description
Log
Duration

System Voice: Vicki
Save As:
Where: Desktop

Using: MP3 Encoder ☐ Delete source files after encoding

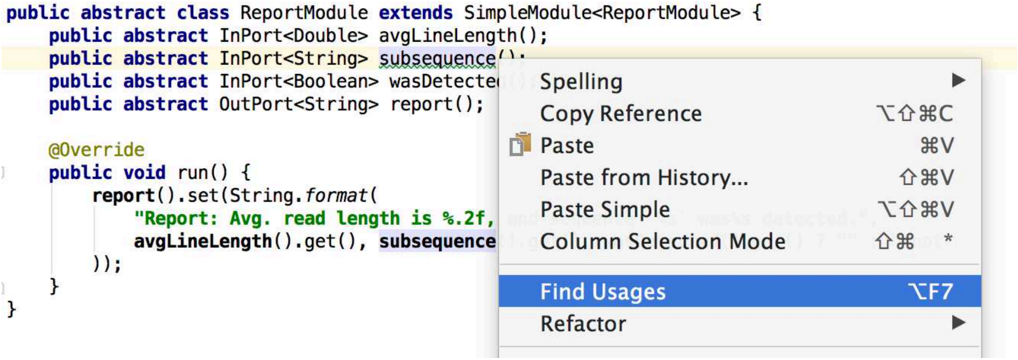
New playlist missing value

Activate Fonts

This action activates the fonts passed from the previous action.

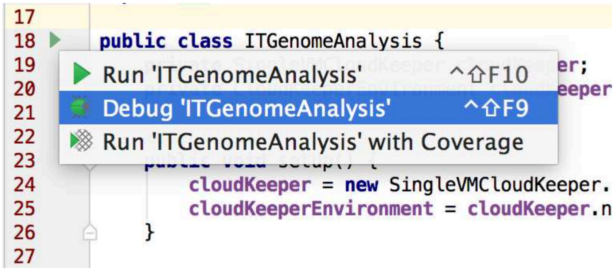
Input: Font Book typeface
Result: Font Book typeface

Version: 3.0
Copyright: Copyright © 2004-2012 Apple Inc. All rights reserved.



```
public abstract class ReportModule extends SimpleModule<ReportModule> {
    public abstract InPort<Double> avgLineLength();
    public abstract InPort<String> subsequence();
    public abstract InPort<Boolean> wasDetected();
    public abstract OutPort<String> report();

    @Override
    public void run() {
        report().set(String.format(
            "Report: Avg. read length is %.2f,
            avgLineLength().get(), subsequence
        ));
    }
}
```



```
public class ITGenomeAnalysis {
                                                         er;
                                                       eeper



        cloudKeeper = new SingleVMCloudKeeper.
        cloudKeeperEnvironment = cloudKeeper.n
    }
```

Run 'ITGenomeAnalysis' ^⇧F10
Debug 'ITGenomeAnalysis' ^⇧F9
Run 'ITGenomeAnalysis' with Coverage

4 Lightweight and Easy to Deploy

CloudKeeper Embedded

- Library, not framework
- Debug in single Java Virtual Machine
- High-level alternative to `ExecutorService`, actors, futures, etc.

JAR

Package Management

- Execute directly from artifact repository
  - Similar to Groovy's Grape

eclipse Aether

Nexus

artifactory

Example

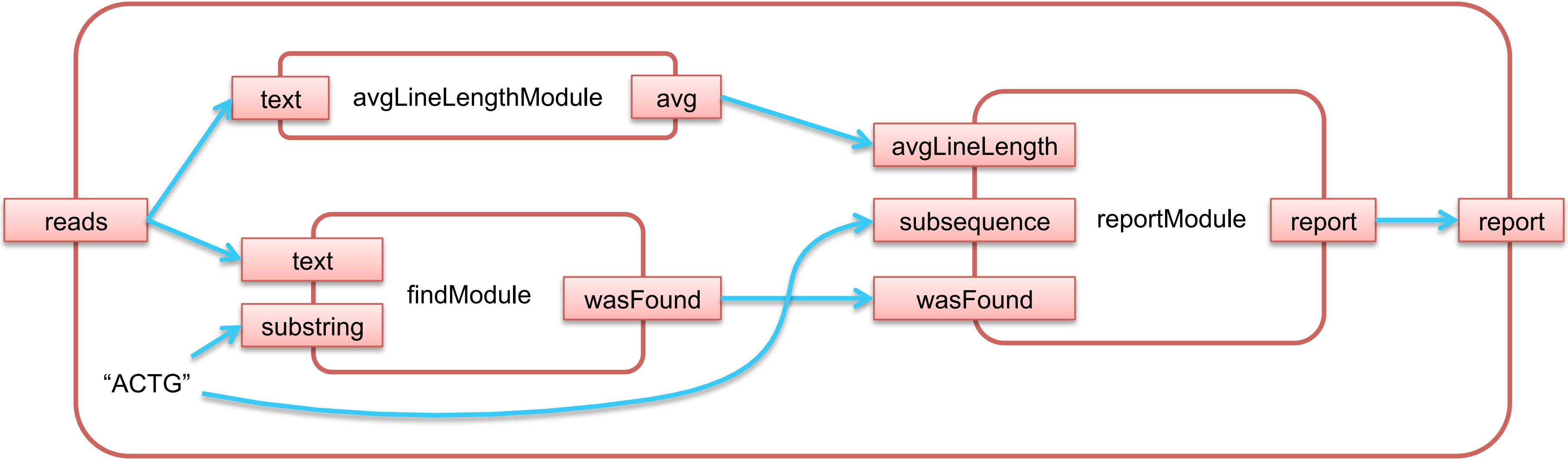
Simple Modules

Composite Modules

Example

```
ACTGTA
CGATA
CGGA
```

Avg. read length is 5.00, and sequence 'ACTG' was detected.



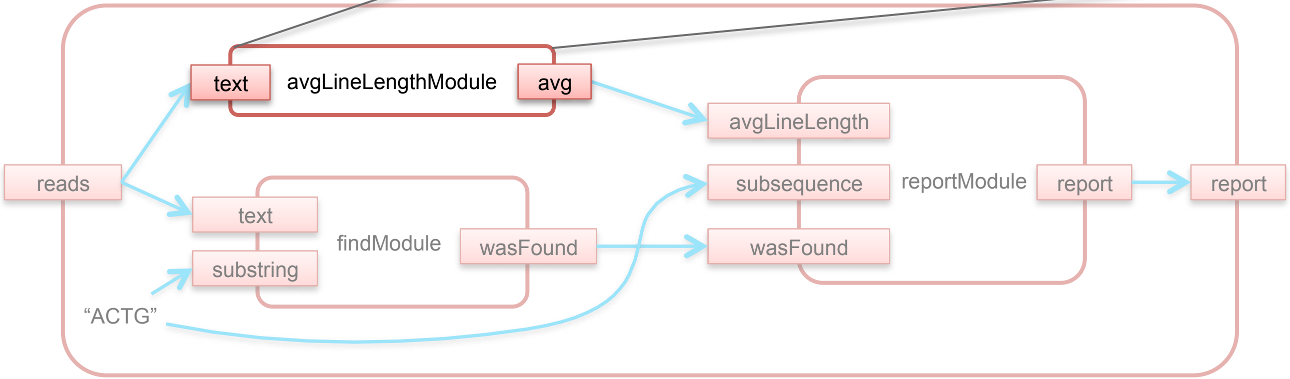
Simple Modules

Internal DSL

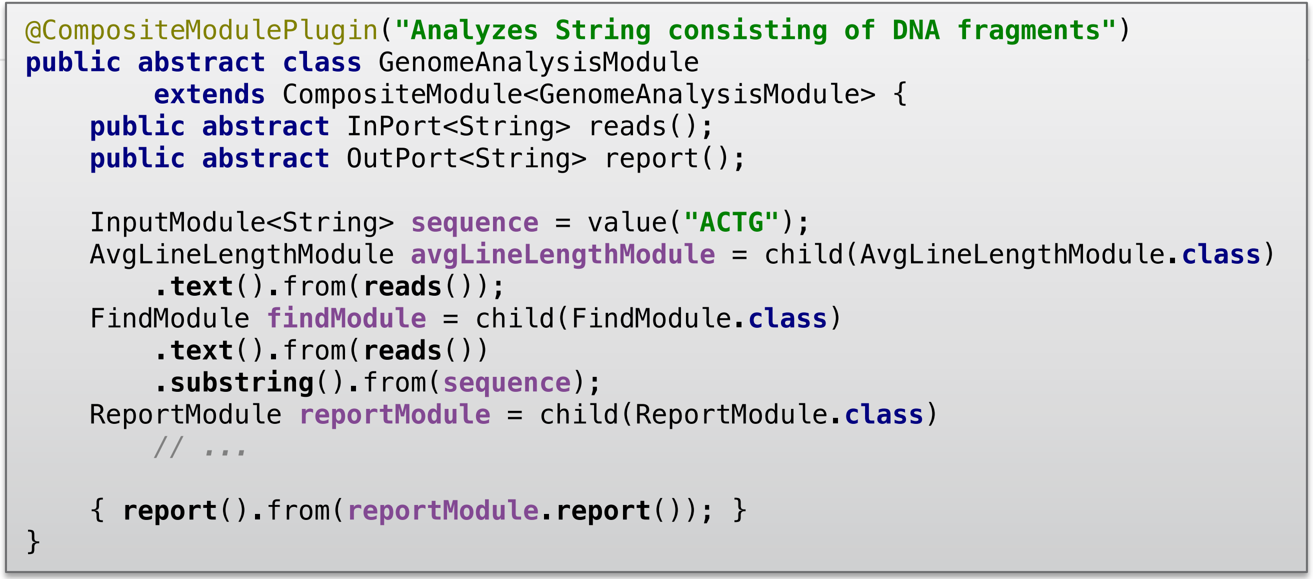
- Module ~ Java class
- Port ~ method
- Metadata using annotations

```
@SimpleModulePlugin("Computes the average line length in a text")
public abstract class AvgLineLengthModule
        extends SimpleModule<AvgLineLengthModule> {
    public abstract InPort<String> text();
    public abstract OutPort<Double> avg();

    @Override
    public void run() throws IOException {
        String text = text().get();
        double avg;
        // ...
        avg().set(avg);
    }
}
```



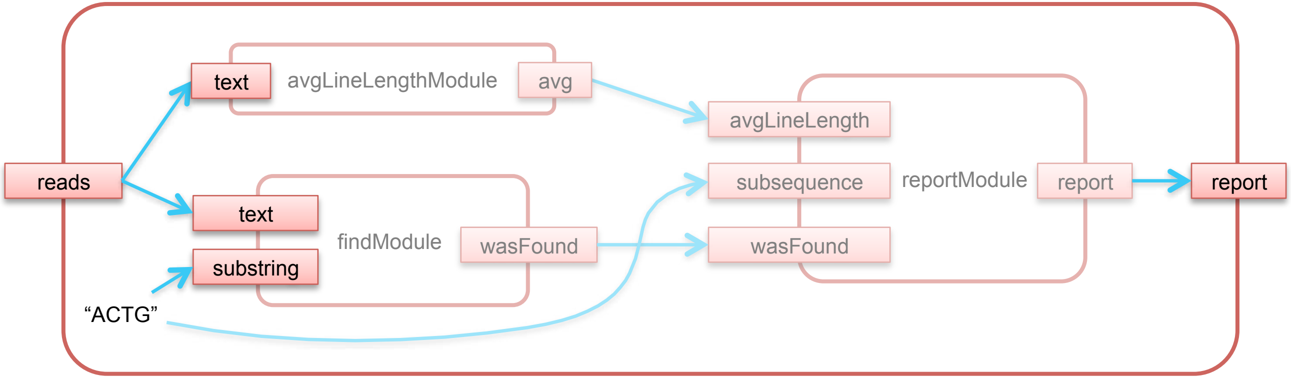
Composite Modules



```
@CompositeModulePlugin("Analyzes String consisting of DNA fragments")
public abstract class GenomeAnalysisModule
        extends CompositeModule<GenomeAnalysisModule> {
    public abstract InPort<String> reads();
    public abstract OutPort<String> report();

    InputModule<String> sequence = value("ACTG");
    AvgLineLengthModule avgLineLengthModule = child(AvgLineLengthModule.class)
        .text().from(reads());
    FindModule findModule = child(FindModule.class)
        .text().from(reads())
        .substring().from(sequence);
    ReportModule reportModule = child(ReportModule.class)
        // ...

    { report().from(reportModule.report()); }
}
```


text
avgLineLengthModule
avg
avgLineLength
reads
subsequence
reportModule
report
report
text
findModule
wasFound
wasFound
substring
“ACTG”