Upper Arlington High School

**IB Computer Science HL Internal Assessment**

**For Joe Sheridan**

**Calvin Kinateder**

**IB Computer Science HL**

**Dr. Kahle**

**10 / 06 / 2018**

**Criterion A: Planning**

The problem I am tackling for my IA is for the ComOps department at AEP. ComOps deals with the buying and selling of power. They needed a way to find news about the price of power and really all things that would interest that department on consumer sites. The goal of my program will be to get a list of webpages that could be potentially relevant to a certain keyword. Google does not do this for them because Google a) does not do the best job of sorting by relevance in this instance, and b) does not write to a database. The client’s name is Joe Sheridan and he is in the ComOps department. A family member of mine has worked with him on a few occasions and that’s how I got in contact with him.

My Java program will search a list of sources for a given keyword and write the output to a database sorted by relevance. The sources will include most consumer news sites and will also include a Twitter API as well as the option to add more sources. Pre-installed sources will include

* <http://bbc.com/>
* <http://foxnews.com/>
* <http://www.huffingtonpost.com/>
* <http://www.nbcnews.com/>
* <http://www.nytimes.com/>
* <http://www.wsj.com/>
* <http://www.usatoday.com/>
* <http://news.google.com/>
* <http://www.rollcall.com>
* <http://www.latimes.com>
* <http://www.wired.com>
* <http://www.cnn.com>
* <http://www.npr.org>
* <http://abcnews.go.com/>
* <https://www.usnews.com/news>
* <https://www.yahoo.com/news/>
* <https://www.ap.org/en-us/>

All sources will be searched recursively but will never read duplicate links. Loops and recursion will be an essential part needed for searching. It will need to write to a database, likely Derby. GUI will be a very large part of this as I will not be the one using it. It must be intuitive, easy to follow, and include a help menu. The sorting will be done in two stages: one, a ‘dumb’ classifier (in Java) will read the data and sort it by the number of times the search keyword appears, and two, a ‘smart’ classifier (in Python). The smart classifier will be in Python because of the wealth of libraries available to use. First, the program will start with a list of sources. Then, those sources will be read into a class that will explore each link, read the article it points to, and search it for the given keyword. If they keyword is found, it will write it to a file AND a Vector set. The Vector data structure will be used to hold all links of any relevance to the keyword. Vector will be used because part of the program will involve multithreading, and Vector is thread-safe. Then, the Python classifier will take that file as input and explore each link in the file. A Latent Dirichlet Allocation model will be used to sort the data based suspected relevance to the keyword. This will output to a database, and this database will be able to be accessed by the user but in only read-only mode. The user will then be able use that data however they wish.

**Criterion B: Record of Tasks List**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task  # | Planned action | Planned outcome | Time estimated | Target completion date | Criterion |
| **1** | **Meet with Joe** | **Get project idea.**  **Stimulate thought** | **short** | **Completed may 2018** |  |
| **2** | **Meet with Joe** | **Get plan**  **Get enough to work on**  **for a while** | **1 hour** | **Completed August 2018** |  |
| **3** | **Work on project** | **get work done**  **No goal, just as much**  **as possible** | **15 hrs** | **8/23-8/30** |  |
| **4** | **Make test cases** |  |  |  |  |
| **5** | **Test test cases** |  |  |  |  |
| **6** | **Flowchart** |  |  |  |  |
| **7** | **UML diagram** |  |  |  |  |
| **8** | **Write psuedocode** |  |  |  |  |
| **9** | **Write real code** |  |  |  |  |
| **10** | **Make movie** |  |  |  |  |
| **11** | **Get feedback from client** |  |  |  |  |
| **12** | **Write-up** |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column1** | **Column2** | **Column3** | **Column4** | **Column5** |
| Task | Planned action | Planned outcome | Time estimated | Target completion date |
| 1 | Work on code | hit count works! |  | 2018-10-11 (13 minutes ago) |
| 2 | Work on code | added error color codes |  | 2018-10-10 (6 hours ago) |
| 3 | Work on code | added search settings and error color flashing |  | 2018-10-09 (28 hours ago) |
| 4 | Work on code | CHANGED SOURCES TO VECTOR |  | 2018-10-09 (35 hours ago) |
| 5 | Work on code | Merge origin/master |  | 2018-10-09 (2 days ago) |
| 6 | Work on code | added link and stuff |  | 2018-10-08 (2 days ago) |
| 7 | Work on code | delimit in comma csv |  | 2018-10-08 (2 days ago) |
| 8 | Work on code | Delete Mon Oct 8 13:49:14 EDT 2018.txt |  | 2018-10-08 (2 days ago) |
| 9 | Work on code | added purposes to each file, 10 |  | 2018-10-08 (2 days ago) |
| 10 | Work on code | added log\_MAC.command |  | 2018-10-08 (3 days ago) |
| 11 | Work on code | fixed scroller not updating and added catch to prohibit source modification while running |  | 2018-10-08 (3 days ago) |
| 12 | Work on code | updated working tree |  | 2018-10-08 (3 days ago) |
| 13 | Work on code | updated IA.docx |  | 2018-10-06 (4 days ago) |
| 14 | Work on code | fixed some formatting |  | 2018-10-06 (4 days ago) |
| 15 | Work on code | fixed some formatting |  | 2018-10-06 (4 days ago) |
| 16 | Work on code | about to migrate to full on LINK |  | 2018-10-06 (5 days ago) |
| 17 | Work on code | open out opens links.txt |  | 2018-10-05 (5 days ago) |
| 18 | Work on code | ICONS WORK!!! |  | 2018-10-05 (5 days ago) |
| 19 | Work on code | mainly just updated working tree |  | 2018-10-05 (5 days ago) |
| 20 | Work on code | added tooltips |  | 2018-10-04 (7 days ago) |
| 21 | Work on code | changed screenshot |  | 2018-10-04 (7 days ago) |
| 22 | Work on code | made some things more consistent |  | 2018-10-04 (7 days ago) |
| 23 | Work on code | renamed variables and took out the trash |  | 2018-10-04 (7 days ago) |
| 24 | Work on code | update |  | 2018-10-03 (8 days ago) |
| 25 | Work on code | improved quit time by 99% |  | 2018-10-02 (8 days ago) |
| 26 | Work on code | source editor doesnt work but ok to revert to |  | 2018-10-02 (9 days ago) |
| 27 | Work on code | addasource sucks so were gonna fix that next |  | 2018-10-01 (9 days ago) |
| 28 | Work on code | fixed scrolling stuff |  | 2018-10-01 (9 days ago) |
| 29 | Work on code | more cosmetic |  | 2018-10-01 (10 days ago) |
| 30 | Work on code | added constant for output size and formatted scroller a little |  | 2018-10-01 (10 days ago) |
| 31 | Work on code | Added automatic refresh plus complete notification |  | 2018-09-29 (12 days ago) |
| 32 | Work on code | minor |  | 2018-09-28 (12 days ago) |
| 33 | Work on code | BIG ONE : THREAD STOP WORKS |  | 2018-09-28 (12 days ago) |
| 34 | Work on code | modified some names |  | 2018-09-28 (12 days ago) |
| 35 | Work on code | added uml project |  | 2018-09-28 (13 days ago) |
| 36 | Work on code | updated writeup |  | 2018-09-27 (2 weeks ago) |
| 37 | Work on code | added reuters |  | 2018-09-26 (2 weeks ago) |
| 38 | Work on code | Added apache common codec |  | 2018-09-26 (2 weeks ago) |
| 39 | Work on code | change to vector |  | 2018-09-25 (2 weeks ago) |
| 40 | Work on code | continuing linkset |  | 2018-09-25 (2 weeks ago) |
| 41 | Work on code | started linkset |  | 2018-09-25 (2 weeks ago) |
| 42 | Work on code | Added Link Class |  | 2018-09-25 (2 weeks ago) |
| 43 | Work on code | ADDED ia |  | 2018-09-24 (2 weeks ago) |
| 44 | Work on code | replace datasets |  | 2018-09-19 (3 weeks ago) |
| 45 | Work on code | added files |  | 2018-09-19 (3 weeks ago) |
| 46 | Work on code | fixed source updating problem and added datasets |  | 2018-09-18 (3 weeks ago) |
| 47 | Work on code | update sc |  | 2018-09-17 (3 weeks ago) |
| 48 | Work on code | added verbose option |  | 2018-09-17 (3 weeks ago) |
| 49 | Work on code | updated scroller |  | 2018-09-16 (3 weeks ago) |
| 50 | Work on code | added long out |  | 2018-09-14 (4 weeks ago) |
| 51 | Work on code | update |  | 2018-09-14 (4 weeks ago) |
| 52 | Work on code | moved scroll to scroller |  | 2018-09-14 (4 weeks ago) |
| 53 | Work on code | disregard source changes, but scroll works |  | 2018-09-14 (4 weeks ago) |
| 54 | Work on code | finished message is good |  | 2018-09-14 (4 weeks ago) |
| 55 | Work on code | fixed concurrency error when clicking extract when running |  | 2018-09-13 (4 weeks ago) |
| 56 | Work on code | MAJOR MAJOR MAJOR - STOP FINALLY WORKS OKAY - some sync probs tho |  | 2018-09-13 (4 weeks ago) |
| 57 | Work on code | update |  | 2018-09-13 (4 weeks ago) |
| 58 | Work on code | progress output works well |  | 2018-09-13 (4 weeks ago) |
| 59 | Work on code | Updated formatting |  | 2018-09-13 (4 weeks ago) |
| 60 | Work on code | Extends thread not implements runnable. now can access extractor vars so cool |  | 2018-09-12 (4 weeks ago) |
| 61 | Work on code | OK to revert - progress bar works great |  | 2018-09-11 (4 weeks ago) |
| 62 | Work on code | dont revert just updating progress |  | 2018-09-10 (4 weeks ago) |
| 63 | Work on code | MAJOR. REVERT TO HERE. Progress works! |  | 2018-09-10 (4 weeks ago) |
| 64 | Work on code | background runner works well |  | 2018-09-10 (4 weeks ago) |
| 65 | Work on code | extractorcontainer is replaced by background runner |  | 2018-09-10 (4 weeks ago) |
| 66 | Work on code | just backing up work DO NOT REVERT |  | 2018-09-10 (4 weeks ago) |
| 67 | Work on code | update working tree dont revert to here |  | 2018-09-10 (4 weeks ago) |
| 68 | Work on code | wouldn't necessarily revert to here, just updating tree for desktop |  | 2018-09-08 (5 weeks ago) |
| 69 | Work on code | modified props and little ui change |  | 2018-09-07 (5 weeks ago) |
| 70 | Work on code | added sources and authentication |  | 2018-09-05 (5 weeks ago) |
| 71 | Work on code | added refresh src button |  | 2018-09-03 (5 weeks ago) |
| 72 | Work on code | fixed readme.md |  | 2018-09-03 (5 weeks ago) |
| 73 | Work on code | Delete sgmain.PNG |  | 2018-09-03 (5 weeks ago) |
| 74 | Work on code | Update README.md |  | 2018-09-03 (5 weeks ago) |
| 75 | Work on code | fixed readme.md |  | 2018-09-03 (5 weeks ago) |
| 76 | Work on code | fixed readme.md |  | 2018-09-03 (5 weeks ago) |
| 77 | Work on code | fixed readme.md |  | 2018-09-03 (5 weeks ago) |
| 78 | Work on code | revisited main layout |  | 2018-09-03 (5 weeks ago) |
| 79 | Work on code | works about to change executor to larger scope ok to revert to |  | 2018-09-03 (5 weeks ago) |
| 80 | Work on code | about to change to executor.submit - OK TO REVERT |  | 2018-09-02 (6 weeks ago) |
| 81 | Work on code | updated ok to revert |  | 2018-09-01 (6 weeks ago) |
| 82 | Work on code | source editor works ok to revert still needs work though |  | 2018-09-01 (6 weeks ago) |
| 83 | Work on code | executerservice works just like thread rn. but added sources. multithreaded speed confirmed. |  | 2018-09-01 (6 weeks ago) |
| 84 | Work on code | Added some functions for joining threads, still stuck. beginning port to executor service. ok to revert to |  | 2018-09-01 (6 weeks ago) |
| 85 | Work on code | search works well now extract with threads still a little glitchy ok to revert to i guess |  | 2018-08-30 (6 weeks ago) |
| 86 | Work on code | multithread works but still dont revert yet |  | 2018-08-30 (6 weeks ago) |
| 87 | Work on code | DO NOT REVERT TO THIS COMMIT -- BETA -- ADDED EXTRACTOR CONTAINER |  | 2018-08-30 (6 weeks ago) |
| 88 | Work on code | single threaded works |  | 2018-08-30 (6 weeks ago) |
| 89 | Work on code | added twitter custom checkbox |  | 2018-08-29 (6 weeks ago) |
| 90 | Work on code | better code |  | 2018-08-29 (6 weeks ago) |
| 91 | Work on code | attached remote javadoc to jtwitter: https://www.winterwell.com/software/jtwitter/javadoc/ |  | 2018-08-29 (6 weeks ago) |
| 92 | Work on code | added libs |  | 2018-08-29 (6 weeks ago) |
| 93 | Work on code | writes as it goes |  | 2018-08-28 (6 weeks ago) |
| 94 | Work on code | writes as it goes |  | 2018-08-28 (6 weeks ago) |
| 95 | Work on code | Changed main ui |  | 2018-08-28 (6 weeks ago) |
| 96 | Work on code | Fixed broken project |  | 2018-08-28 (6 weeks ago) |
| 97 | Work on code | Fixed broken project |  | 2018-08-28 (6 weeks ago) |
| 98 | Work on code | Delete project.xml |  | 2018-08-28 (6 weeks ago) |
| 99 | Work on code | Delete project.properties |  | 2018-08-28 (6 weeks ago) |
| 100 | Work on code | Delete genfiles.properties |  | 2018-08-28 (6 weeks ago) |
| 101 | Work on code | Delete build-impl.xml |  | 2018-08-28 (6 weeks ago) |
| 102 | Work on code | Delete manifest.mf |  | 2018-08-28 (6 weeks ago) |
| 103 | Work on code | Delete links.txt |  | 2018-08-28 (6 weeks ago) |
| 104 | Work on code | Delete build.xml |  | 2018-08-28 (6 weeks ago) |
| 105 | Work on code | Delete acountss.txt |  | 2018-08-28 (6 weeks ago) |
| 106 | Work on code | Delete Java 8 Articles |  | 2018-08-28 (6 weeks ago) |
| 107 | Work on code | remove storygrab folder |  | 2018-08-28 (6 weeks ago) |
| 108 | Work on code | tst |  | 2018-08-28 (6 weeks ago) |
| 109 | Work on code | extractor works recursively |  | 2018-08-28 (6 weeks ago) |
| 110 | Work on code | changed extractor |  | 2018-08-28 (6 weeks ago) |
| 111 | Work on code | added links |  | 2018-08-28 (6 weeks ago) |
| 112 | Work on code | added nonfunct xtractor |  | 2018-08-27 (6 weeks ago) |
| 113 | Work on code | added period |  | 2018-08-27 (6 weeks ago) |
| 114 | Work on code | added help menu |  | 2018-08-27 (6 weeks ago) |
| 115 | Work on code | fixed last commit |  | 2018-08-27 (6 weeks ago) |
| 116 | Work on code | Update README.md |  | 2018-08-27 (6 weeks ago) |
| 117 | Work on code | Update README.md |  | 2018-08-27 (6 weeks ago) |
| 118 | Work on code | Update README.md |  | 2018-08-27 (6 weeks ago) |
| 119 | Work on code | Merge origin/master |  | 2018-08-27 (6 weeks ago) |
| 120 | Work on code | BIG ONE: Removed access to create account panel-it was not necessary and it was insecure |  | 2018-08-27 (6 weeks ago) |
| 121 | Work on code | Update README.md |  | 2018-08-27 (6 weeks ago) |
| 122 | Work on code | Update README.md |  | 2018-08-27 (6 weeks ago) |
| 123 | Work on code | Update README.md |  | 2018-08-26 (6 weeks ago) |
| 124 | Work on code | added screenshots |  | 2018-08-26 (6 weeks ago) |
| 125 | Work on code | added two icons |  | 2018-08-26 (6 weeks ago) |
| 126 | Work on code | fixed corners by changing JFrame shape |  | 2018-08-25 (7 weeks ago) |
| 127 | Work on code | fixed mac issue with spawning pwdfield in create account |  | 2018-08-25 (7 weeks ago) |
| 128 | Work on code | add logo |  | 2018-08-25 (7 weeks ago) |
| 129 | Work on code | change author from calvi to my actual name |  | 2018-08-25 (7 weeks ago) |
| 130 | Work on code | Cleaned up functions and improved create accounts screen |  | 2018-08-25 (7 weeks ago) |
| 131 | Work on code | Buttons and labels now work for release too |  | 2018-08-25 (7 weeks ago) |
| 132 | Work on code | Big one: fixed buttons toggling and create account only works if it fields are filled all the way To do: pressed not clicked for buttons |  | 2018-08-25 (7 weeks ago) |
| 133 | Work on code | added writing accounts to file and loading |  | 2018-08-24 (7 weeks ago) |
| 134 | Work on code | added create account |  | 2018-08-24 (7 weeks ago) |
| 135 | Work on code | update readme |  | 2018-08-24 (7 weeks ago) |
| 136 | Work on code | update readme |  | 2018-08-24 (7 weeks ago) |
| 137 | Work on code | Added Accountmanager and user classes. |  | 2018-08-24 (7 weeks ago) |
| 138 | Work on code | added checked and unchecked icons - may not be used |  | 2018-08-24 (7 weeks ago) |
| 139 | Work on code | login creates new screen |  | 2018-08-24 (7 weeks ago) |
| 140 | Work on code | Made it look nicer and added icons |  | 2018-08-24 (7 weeks ago) |
|  | Work on code | add lib |  | 2018-08-24 (7 weeks ago) |
|  | Work on code | update readme |  | 2018-08-24 (7 weeks ago) |
|  | Work on code | update readme |  | 2018-08-24 (7 weeks ago) |
|  | Work on code | Create README.md |  | 2018-08-23 (7 weeks ago) |
|  | Work on code | Fixed opacity problem and added X |  | 2018-08-23 (7 weeks ago) |
|  | Work on code | First commit |  | 2018-08-23 (7 weeks ago) |

**Criterion C: Interview and All Communication**

[Greetings]

Joe: Alright, so what is this school project.

Calvin: Well I’m taking IB computer science and we do a project with a client so I was wondering if you could maybe help me

Joe: Yes I probably could.

Calvin: Great! So do you have any problems you would need me to code a solution for?

Joe: Well we do a lot with news. We have commercial news sites that we use for getting things like the price of coal and stuff like that, but we don’t have anything to do that for consumer news sites. = **Problem**

Calvin: So what would that entail?

Joe: Well you would need to handle a lot of data and parse into links sorted by relevance. It would need to be stored in a database for the output. People on my team would use it so it would need to be easy to use. It would probably be a good idea to multithread because it’s a lot of data.

Calvin: Ok cool I’ll update you with my progress.

Joe: Sounds good.

[Calvin leaves]

Hi Calvin!

I hope you had a good summer! I spoke with my management earlier and we can only use data which is also available to the public, but that doesn’t mean we can’t still do something cool and useful to AEP. I had a couple of ideas, but I’m open to anything.

Com Ops uses a tremendous amount of information (for example news stories) from a large number of disparate sources. There is usually a lot of data presented that is of no use, but we still have to sift through it to find the few useful items. One idea I had was to write a program that could analyze various sites and produce a list of things we might be interested in. The most fun option might be to use the free (standard version) Twitter API (<https://developer.twitter.com/en/docs.html>), but you could always just pull in a couple of news sources instead. You could use a string compare or regex to find specific strings that would be of interest. Regular expressions are a little scary at first, but they are useful for a lot of different situations and worth learning. You mentioned that you were interested in machine learning, if you want to take that on you could use automatic classification tools (that will be a lot of work though). If you wanted to build a GUI, the user could input keywords to look for. You could write the results to an HTML file (maybe just links and a subject line) or even a database if you want to get ambitious. Just one idea, and if you like it we can flesh out the requirements.

Another option I thought of is to use the publicly available Data Miner 2 from the PJM website (<https://www.pjm.com/markets-and-operations/etools/data-miner-2.aspx>). PJM is a bulk electricity market that AEP sells our generation (energy created from our power plants) into and buys our load (the electricity our customers actually use). We can go over what the market is and isn’t if you want to go this route. Data Miner 2 has a web service API (<https://www.pjm.com/-/media/etools/data-miner-2/data-miner-2-api-guide.ashx?la=en>), though I’ve never used it so I’m not sure how complicated it is. Pretty much anything publicly available on Data Miner 2 would be useful. If you wanted you could store it into a database like derby (<https://db.apache.org/derby/>) or something.

Let me know what you think. I’m open to just about anything.

Thanks!

-Joe

Thanks for working to let me help you with some programs. I'm not really sure what the second option's goal would be, do you think you could elaborate a little on it? The first option looks like it would work well for me though, and I've done some regex stuff before in linux although I'll need to brush up a little on using that with Java. I also did some GUI work last year for a project so I could make a front end for it too. I'm interested in the ML automatic classification feature for that as well, and it would be really great to pipe the output to a database. What are the requirements for that project?

Also, when would you need me to come down to AEP next?

Thanks,

Calvin Kinateder