

# [WeTok] Report

The following is a report template to help your team successfully provide all the details necessary for your report in a structured and organised manner. Please give a straightforward and concise report that best demonstrates your project. Note that a good report will give a better impression of your project to the reviewers.

*Here are some tips to write a good report:*

- *Try to summarise and list the bullet points of your project as much as possible rather than give long, tedious paragraphs that mix up everything together.*
- *Try to create diagrams instead of text descriptions, which are more straightforward and explanatory.*
- *Try to make your report well structured, which is easier for the reviewers to capture the necessary information.*

*We give instructions enclosed in square brackets [...] and examples for each sections to demonstrate what are expected for your project report.*

*Please remove the instructions or examples in italic in your final report.*

## Table of Contents

- [WeTok] Report
  - Table of Contents
  - Team Members and Roles
  - Conflict Resolution Protocol
  - Application Description
    - Application Use Cases and or Examples
  - Application UML
  - Application Design and Decisions
    - Data Structures
      - AVL Tree
    - Design Patterns
      - Singleton
      - Template
      - DAO
    - Grammars

- Surprise Item
  - (i) Ranking algorithm
  - (ii) Simple personalisation
- Summary of Known Errors and Bugs
- Testing Summary
  - AVL Tree Testings
- Implemented Features
- Team Meetings

## Team Members and Roles

UID	Name	Role
[uid]	Xinyu Kang	[role]
[uid]	Xinyue Hu	[role]
u6684233	Yuxin Hong	[role]
[uid]	Zhaoting Jiang	[role]

## Conflict Resolution Protocol

1. Conflictors organize their views, list the pros and cons of their perspectives.
2. Initiate a zoom meeting with all team member.
3. Conflictors presenting the confliction and why disagree with the other.
4. Team member do a vote and share the reason.
5. As a result, merge conflictors's solutions or take more agreed solutions.

## Application Description

*[What is your application, what does it do? Include photos or diagrams if necessary]*

*Here is a pet specific social media application example*

*PetBook is a social media application specifically targetting pet owners... it provides... certified practitioners, such as veterians are indicated by a label next to their profile...*

## Application Use Cases and or Examples

*[Provide use cases and examples of people using your application. Who are the target users of your application? How do the users use your application?]*

*Here is a pet training application example*

*Molly wants to inquiry about her cat, McPurr's recent troublesome behaviour*

- 1. Molly notices that McPurr has been hostile since...*
- 2. She makes a post about... with the tag...*
- 3. Lachlan, a vet, writes a reply to Molly's post...*
- 4. ...*
- 5. Molly gives Lachlan's reply a 'tick' response*

*Here is a map navigation application example*

*Targets Users: Drivers*

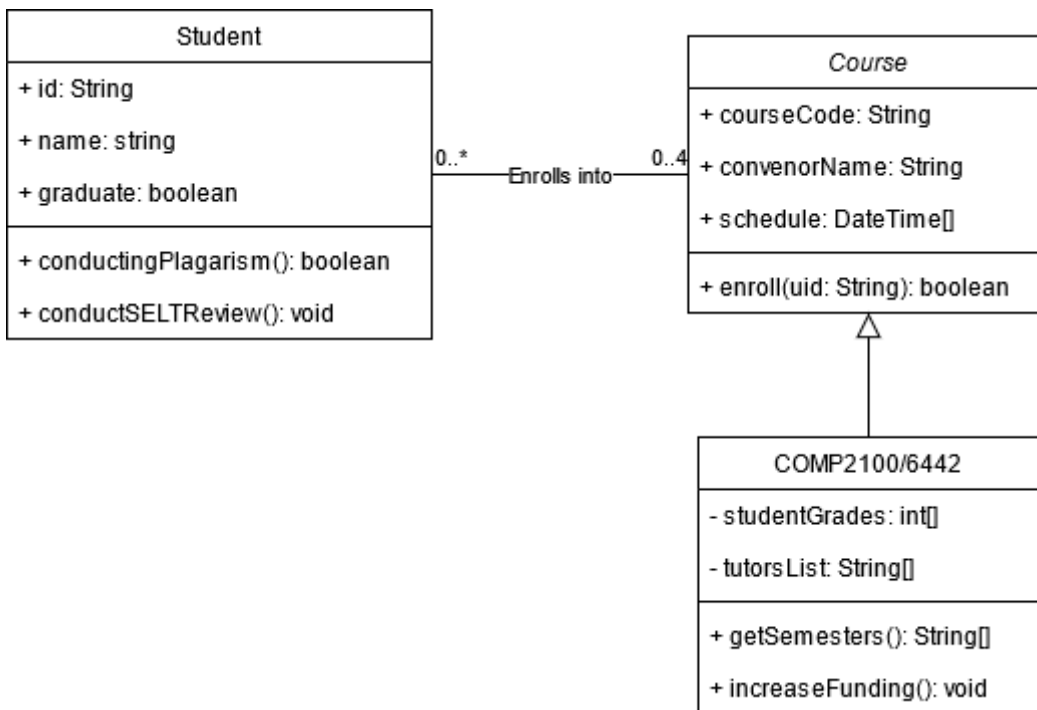
- Users can use it to navigate in order to reach the destinations.*
- Users can learn the traffic conditions*
- ...*

*Target Users: Those who want to find some good restaurants*

- Users can find nearby restaurants and the application can give recommendations*
- ...*

*List all the use cases in text descriptions or create use case diagrams. Please refer to <https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-use-case-diagram/> for use case diagram.*

## **Application UML**



[Replace the above with a class diagram. You can look at how we have linked an image here as an example of how you can do it too.]

## Application Design and Decisions

Please give clear and concise descriptions for each subsections of this part. It would be better to list all the concrete items for each subsection and give no more than 5 concise, crucial reasons of your design. Here is an example for the subsection *Data Structures* :

I used the following data structures in my project:

1. *LinkedList*
  - *Objective:* It is used for storing xxxx for xxx feature.
  - *Locations:* line xxx in XXX.java, ..., etc.
  - *Reasons:*
    - It is more efficient than ArrayList for insertion with a time complexity  $O(1)$
    - We don't need to access the item by index for this feature
2. ...
3. ...

## Data Structures

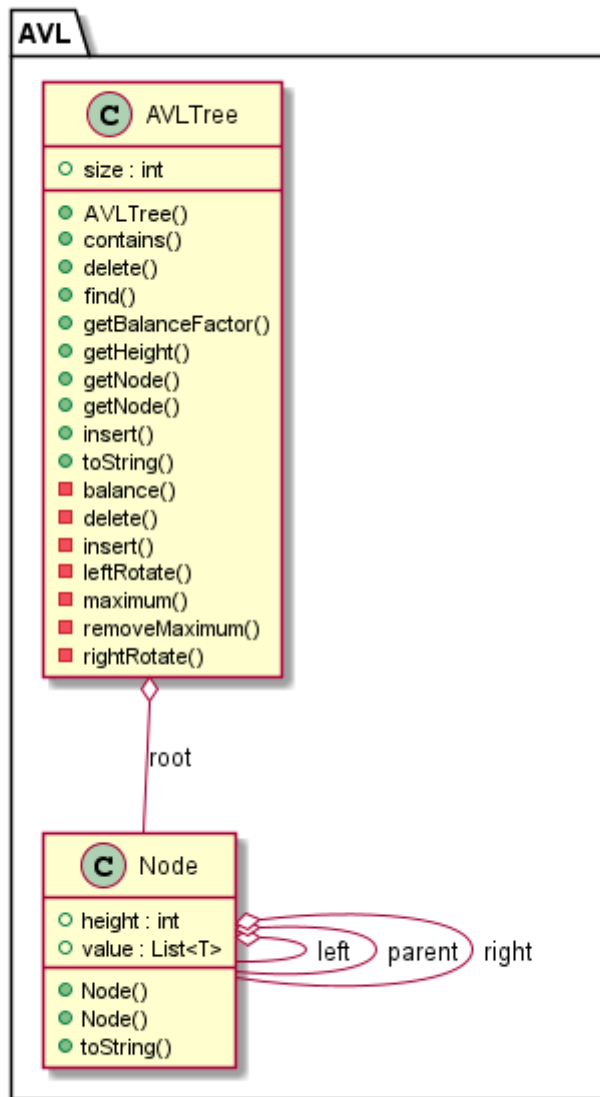
### AVL Tree

We selected AVL tree as the data structure of posts operations. AVL tree is a self-balancing binary search tree. It controls the height of the tree and prevents it from becoming skewed. For operations considered they have time complexity:

- Rotations to achieve balanced tree  $O(1)$
- Insertion  $O(\log(n))$
- Deletion  $O(\log(n))$
- Search  $O(\log(n))$
- Max/Min  $O(\log(n))$

It is an efficient data structure and here is the UML of our implementation:

### AVL's Class Diagram



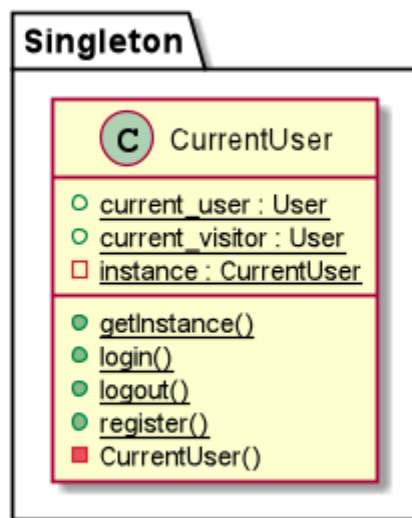
# Design Patterns

We used three design patterns: Singleton, Template, DAO

## Singleton

- Location: *CurrentUser.java*
- We used Singleton design pattern in CurrentUser to make sure there must be exactly one instance of the current user.

### SINGLETON's Class Diagram

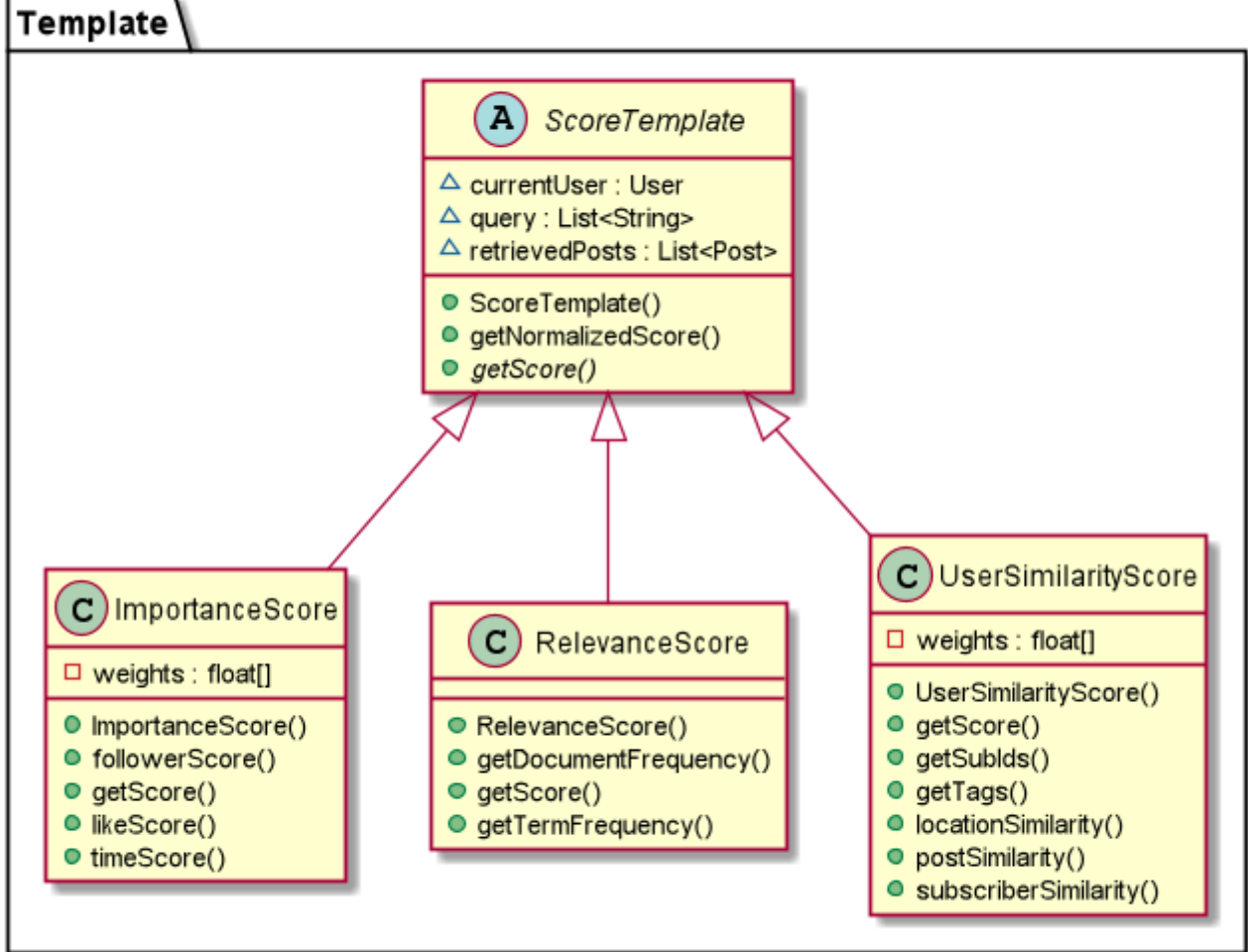


PlantUML diagram generated by SketchIt! (<https://bitbucket.org/pmesmeur/sketch.it>)  
For more information about this tool, please contact [philippe.mesmeur@gmail.com](mailto:philippe.mesmeur@gmail.com)

## Template

- Location: *ScoreTemplate.java* (abstract class), *RelevanceScore.java* (concrete class), *ImportanceScore.java* (concrete class), *UserSimilarity.java* (concrete class)
- We want to sort the posts based on three criteria. There is a common process (i.e., calculate scores, normalize scores) to scoring the posts but each criterion has its own scoring logic.
- So, we used a score template to define the skeleton of the scoring algorithm, and let subclasses redefine certain steps of the algorithm without changing the main structure.

## TEMPLATE's Class Diagram

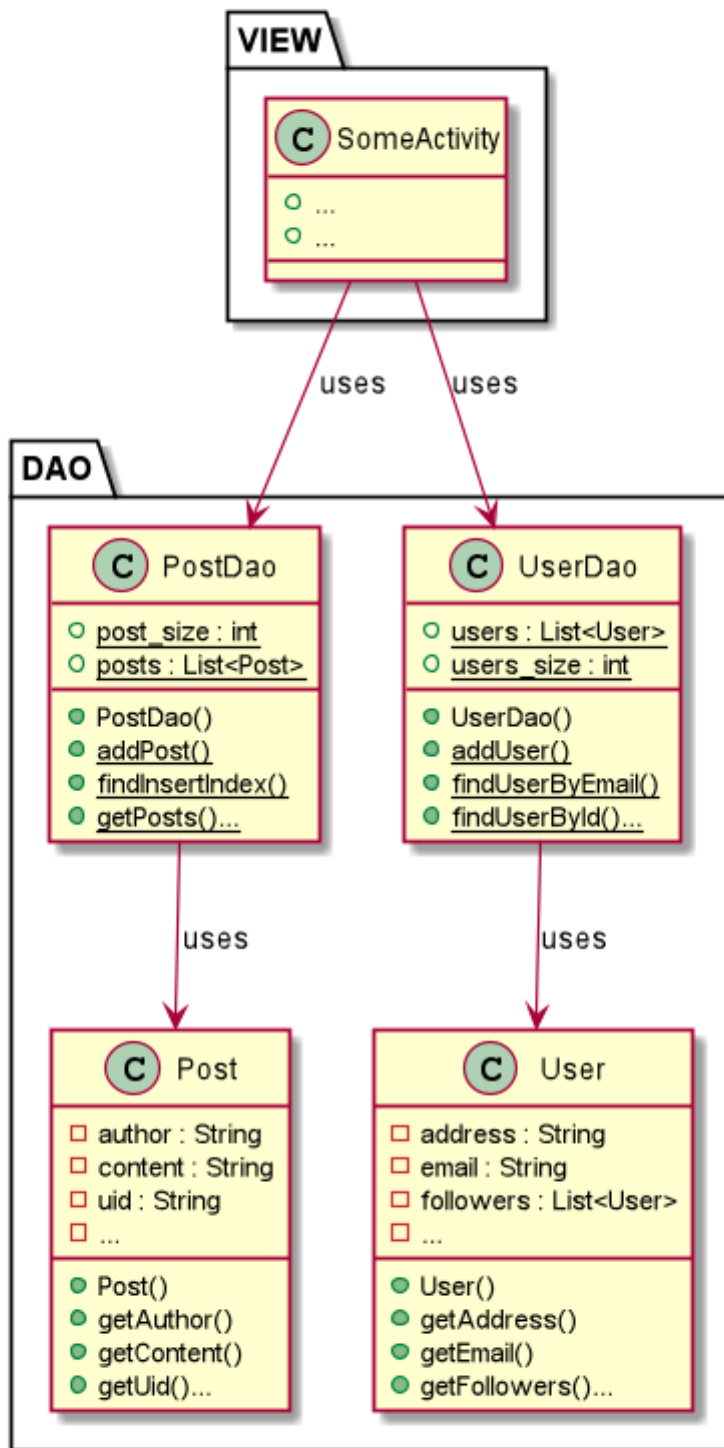


PlantUML diagram generated by SketchIt! (<https://bitbucket.org/pmestmeur/sketch.it>)  
For more information about this tool, please contact [philippe.mesmeur@gmail.com](mailto:philippe.mesmeur@gmail.com)

## DAO

- Location: *UserDao.java*, *PostDao.java*
- We want to decouple domain logic from persistence mechanisms and avoid exposing details of the data storage. The two DAO classes provided convenient access to user data in whole project.

## DAO's Class Diagram



PlantUML diagram generated by SketchUML (<https://bitbucket.org/pmesmeur/sketch.it>)  
For more information about this tool, please contact [philippe.mesmeur@gmail.com](mailto:philippe.mesmeur@gmail.com)

## Grammars



## Production Rules

$\langle \text{exp} \rangle ::= \langle \text{term} \rangle \mid \langle \text{term} \rangle \mid \langle \text{exp} \rangle$

$\langle \text{term} \rangle ::= \langle \text{factor} \rangle \mid \langle \text{factor} \rangle \& \langle \text{term} \rangle$

$\langle \text{factor} \rangle ::= \langle \text{tag} \rangle \mid \langle \text{exp} \rangle$

According to this grammar, we can parse *AND* and *OR* operations. For example, suppose we have the expression **condition1 & condition2 | condition3**, human will process it as **(condition1 & condition2) | condition3**, so does our grammar will do. Another example is **condition1 & (condition2 | condition3 & condition4)**, human will process it as **condition1 & ((condition2 | (condition3 & condition4)))**, so does our grammar will do. In application, the advantages of our grammar is to filter tag of post with *AND* and *OR* operation. It takes tag as its condition and only return post that satisfied given condition in expression.

## Tokenizer and Parsers

Generally speaking, the advantages of our design is we to search multiple tags at once and this is also the design approach of our grammar. Operations *AND* and *OR* can be applied on multiple tags search, that is, the intersection and union of single search result. By search *#tag1&#tag2*, the intersection of individual search result of *#tag1* and *#tag2* will be returned. By search *#tag1|#tag2*, the union of individual search result of *#tag1* and *#tag2* will be returned. And of course the operators can be freely used to create more expressions. By default, the precedence of *AND* operation is greater than *OR* operation. And precedence can be changed parentheses *LBRA* and *RBRA*.

## Surprise Item

### Ranking algorithm

Based on the *Pariser talk*, we decided to sort the posts by three criteria: relevance; importance; user similarity.

- **Relevance**

The class *RelevanceScore* evaluates the relevance of the query and the retrieved posts by calculating tf-idf scores. The score of post *d* given query  $q = (t_1, t_2, \dots, t_m)$  is:

$$score_{tfidf}(d, q) = \sum_{i=1}^m tfidf_{t_i, d}$$

$$tfidf_{t, d} = tf_{t, d} \times idf_t$$

$$idf_t = \log \frac{N}{1 + df_t}$$

$df_t$  = the number of posts in all posts that contain a term  $t$

$tf_{t, d}$  = the number of occurrences of tag  $t$  in post  $d$

$N$  = the total number of posts

We expect the posts containing more "rare words" that appeared in the query to be more relevant. The rare word here refers to the word that does not exist in many posts.

- **Importance**

The class *ImportanceScore* evaluates the importance of the retrieved posts by calculating the post time, post likes, and the user's followers:

$$score_{importance} = w_{time} \times score_{time} + w_{follower} \times score_{follower} + w_{like} \times score_{like}$$

$$score_{time} = \frac{1}{1 + \#days \text{ between query and post}}$$

$$score_{follower} = \min(\frac{1}{4} \log_{10}(\#followers \text{ of post}), 0)$$

$$score_{like} = \min(\frac{1}{4} \log_{10}(\#likes \text{ of post}), 0)$$

$w_{time}$ ,  $w_{follower}$ ,  $w_{like}$  are initially set to be 0.7, 0.2, 0.1

We expect the posts with more likes, the posts sent by users with more followers, and the latest posts to be more important to society, and thus have higher ranking scores.

- **User Similarity**

The class *UserSimilarityScore* evaluates the similarity of the current user and the senders of the posts by evaluating their subscribers, posts, and addresses:

$$score_{similarity} = w_{subscriber} \times score_{subscriber} + w_{post} \times score_{post} + w_{address} \times score_{address}$$

$$score_{subscriber} = \frac{1}{1 + \#common\ subscribers}$$

$$score_{post} = \frac{1}{1 + \#common\ tags\ from\ past\ posts}$$

$$score_{address} = 1 \text{ if current user and post sender live in same area, } = 0 \text{ otherwise}$$

$w_{subscriber}, w_{post}, w_{address}$  are initially set to be 0.5, 0.4, 0.1

The post sent by the user who is considered as more 'similar' with the current user will have a lower ranking score. We expect to use this algorithm to offset some negative effects of the 'Filter Bubbles'.

- **Overall Score**

$$score = w_{tfidf} \times score_{tfidf} + w_{importance} \times score_{importance} + w_{similarity} \times score_{similarity}$$

$w_{tfidf}, w_{importance}, w_{similarity}$  are initially set to be 0.6, 0.2, 0.2

All of the weights introduced in this algorithm were intuitively chose by our group members. In practice we should use machine learning techniques to train the best weights.

We expect that the posts having strong relevance with the query, the posts which are important to society, and the posts sent by the users who are "different" with the user, to have better ranking in the search results presenting.

The more and more intelligent recommendation systems create the 'Filter Bubbles' for users. We intend to "break" the bubble by implemtening the ranking algorithm from users similarity dimension. The score of the posts created by the users who are very "similar" with you will be scaled down. People will have more chances to receive different viewpoints.

## Simple personalisation

### Other

[What other design decisions have you made which you feel are relevant? Feel free to separate these into their own subheadings.]

# Summary of Known Errors and Bugs

*[Where are the known errors and bugs? What consequences might they lead to?]*

*Here is an example:*

1. Bug 1:

- A space bar ( ' ') in the sign in email will crash the application.
- ...

2. Bug 2:

3. ...

*List all the known errors and bugs here. If we find bugs/errors that your team do not know of, it shows that your testing is not through.*




## Testing Summary

### AVL Tree Testings

Number of test cases: 13

✓ AVLTreeTest (SearchTree.AVL)	171 ms
✓ searchByKeyTest	31 ms
✓ rightRotateTest	94 ms
✓ deleteNotExistException	15 ms
✓ deleteTwoChildrenNodeTest	0 ms
✓ deleteLeafTest	0 ms
✓ balanceFactorTest	0 ms
✓ insertInOrderTest	0 ms
✓ sameKeyInsertTest	0 ms
✓ deleteSingleChildNodeTest	0 ms
✓ advancedRotationsTest	15 ms
✓ leftRotateTest	16 ms
✓ searchByNotExistKeyTest	0 ms
✓ deleteRootTest	0 ms

Code coverage:

Element	Class, %	Method, %	Line, %
 AVLTree	100% (1/1)	94% (16/17)	91% (99/108)
 AVLTreeTest	100% (2/2)	100% (15/15)	100% (207/207)
 Node	100% (1/1)	100% (3/3)	100% (25/25)

Types of tests created:

1. *sameKeyInsertTest* : test whether the AVL tree can insert multiple values into one nodes
2. *insertInOrderTest* : test whether the AVL tree can correctly insert nodes in order
3. *leftRotateTest* : test whether the AVL tree can correctly left rotate
4. *rightRotateTest* : test whether the AVL tree can correctly right rotate
5. *balanceFactorTest* : test whether the class can generate correct balance factors
6. *advancedRotationsTest* : test whether the AVL tree can handle complex rotations
7. *deleteNotExistException* : test whether the AVL tree can throw an exception when deleting a node that not exists
8. *deleteLeafTest* : test whether the AVL tree can delete a leaf node
9. *deleteSingleChildNodeTest* : test whether the AVL tree can correctly delete a node with one child
10. *deleteTwoChildrenNodeTest* : test whether the AVL tree can correctly delete a node with two children
11. *deleteRootTest* : test whether the AVL tree can correctly delete the root node
12. *searchByKeyTest* : test whether the AVL tree can correctly find the node based on the key
13. *searchByNotExistKeyTest* : test whether the AVL tree returns null when searching a key that not exists

## Parser and Tokenizer Tests

Number of test cases: 12

✓ com.example.parserAndTokenizerTest.ParserTest	13 ms
✓ testSimpleCase	8 ms
✓ testSimpleAnd	1 ms
✓ testSingleTag	1 ms
✓ testIllegalProductionException	2 ms
✓ testMidCase	1 ms
✓ testSimpleOr	0 ms
✓ com.example.parserAndTokenizerTest.TokenizerTest	6 ms
✓ testOrToken	3 ms
✓ testExceptionToken	2 ms
✓ testMidTokenResult	0 ms
✓ testFirstToken	1 ms
✓ testAndToken	0 ms
✓ testAdvancedTokenResult	0 ms

Code coverage:

100% classes, 93% lines covered in package 'com.example.wetok.parserAndTok.

Element ▲	Class, %	Method, %	Line, %
AndExp	100% (1/1)	100% (4/4)	100% (13/13)
Exp	100% (1/1)	100% (0/0)	100% (1/1)
OrExp	100% (1/1)	100% (4/4)	100% (13/13)
Parser	100% (2/2)	100% (5/5)	93% (69/74)
TagExp	100% (1/1)	100% (3/3)	100% (8/8)
Token	100% (3/3)	85% (6/7)	70% (12/17)
Tokenizer	100% (1/1)	100% (4/4)	100% (30/30)

Tests created for Tokenizer:

1. *testAndToken* : test whether *AND* token can be recognize.
2. *testOrToken* : test whether *OR* token can be recognize.
3. *testFirstToken* : test whether first *LBRA* token can be recognize.
4. *testMidTokenResult* : test whether *TAG* and *OR* token at middle can be recognize.
5. *testAdvancedTokenResult* : test whether tokens from (*#weekend* | *#mood*) & *#time* can be tokenized.
6. *testExceptionToken* : test whether tokenizer throw exceptions as expected.

Tests created for Parser:

1. *testSingleTag* : test whether single-tag-search is correct.
2. *testSimpleAnd* : test whether search tag with *AND* operation is correct.
3. *testSimpleOr* : test whether search tag with *OR* operation is correct.
4. *testSimpleCase* : test whether search tag with multiple *AND* operations is correct.
5. *testMidCase* : test whether search tag with *AND* and *OR* operations is correct.
6. *testIllegalProductionException* : test whether parser throw exceptions as expected.

## Ranking Tests

Number of test cases: 8

✓	✓	com.example.rankingTest.ImportanceScoreTest	12 ms
	✓	followerScoreTest	12 ms
	✓	timeScoreTest	0 ms
	✓	likeScoreTest	0 ms
✓	✓	com.example.rankingTest.RelevanceScoreTest	3 ms
	✓	twoTagRelevantScoreTest	3 ms
	✓	singleTagRelevantScoreTest	0 ms
✓	✓	com.example.rankingTest.UserSimilarityScoreTest	2 ms
	✓	postRelevantScoreTest	1 ms
	✓	subscriberRelevantScoreTest	1 ms
	✓	locationRelevantScoreTest	0 ms

Code coverage:

66% classes, 72% lines covered in package 'com.example.wetok.ranking'			
Element	Class, % ▲	Method, %	Line, %
Rank	0% (0/2)	0% (0/5)	0% (0/35)
ImportanceScore	100% (1/1)	100% (6/6)	94% (32/34)
RelevanceScore	100% (1/1)	100% (4/4)	100% (29/29)
ScoreTemplate	100% (1/1)	50% (1/2)	38% (5/13)
UserSimilarityScore	100% (1/1)	100% (7/7)	90% (70/77)

Tests created for Ranking Tests:

1. ImportanceScoreTest: 3 tests for 3 component of calculating importance score: time, like, follower.
2. RelevanceScoreTest: 2 tests for 1 tag relevent and 2 tags relevant.
3. UserSimilarityScoreTest: 3 tests for 3 component of calculating user similarity score: location, subscriber, post.

## Implemented Features

### Improved Search

1. Search functionality can handle partially valid and invalid search queries. (medium)

### UI Design and Testing

1. UI must have portrait and landscape layout variants as well as support for different screen sizes. Simply using Android studio's automated support for orientation and screen sizes and or creating support without effort to make them look reasonable will net you zero marks. (easy)

### Greater Data Usage, Handling and Sophistication

1. *User profile activity containing a media file (image, animation (e.g. gif), video). (easy)*
2. *Deletion method of either a Red-Black Tree and or AVL tree data structure. The deletion of nodes must serve a purpose within your application (e.g. deleting posts). (hard)*

#### *User Interactivity*

1. *The ability to micro-interact with 'posts' (e.g. like, report, etc.) [stored in-memory]. (easy)*
2. *The ability for users to 'follow' other users. There must be an adjustment to either the user's timeline in relation to their following users or a section specifically dedicated to posts by followed users. [stored in-memory] (medium)*

#### *User Interactivity*

1. *Use Firebase to implement user Authentication/Authorisation. (easy)*

## **Team Meetings**

*Here is an example:*

- [Team Meeting 1](#)
- [Team Meeting 2](#)
- [Team Meeting 3](#)
- ...

*Either write your meeting minutes here or link to documents that contain them. There must be at least 3 team meetings.*