**Cấu trúc bộ code:**

├── src/models

│ ├── dictionary

│ ├── filtered

│ ├── json\_cv

│ ├── object\_service

│ ├── rule

├── src/controller

│ ├── data

│ ├── Filter\_spam\_webservice

│ ├── mongodb

│ ├── mysqldb

│ ├── util

├── src/suppport

│ ├── maxent

│ ├── naivebayes\_main

│ ├── remove\_vietnamese

│ ├── tokenizer

├── src/application

│ ├── Application

Chi tiết:

* 1. src/ models/ dictionary:

Getlib\_ob\_sv.java *– Load từ điển đối tượng, dịch vụ*

Getlib\_sentiment.java *– Load từ điển sentiment*

* 1. src/ models/ filtered:

Filtered.java – Hàm gọi lấy bài viết, từ điển, gán nhãn và tạo multithread

Logic\_tag.java – Rule xác định nhãn cuối cùng sau khi chạy qua các bộ lọc rule

Task\_multithread.java – Xử lí chung toàn bộ thông tin của 1 mention

Task\_test\_sent.jar – Test sentiment

* 1. src/ models/ json\_cv:

object.java - Tạo json object

service.java – Tạo json service

* 1. src/ models/ rule:

Check\_order.java – Check tag mention có phải nhãn order

Check\_question.java – Check tag mention có phải nhãn question

Check\_recom.java – Check tag mention có phải nhãn recomend

Check\_thetindung.java – Check mention có liên quan tới thẻ tín dụng ( cũ )

Rule\_congrat.java – Check mention có phải nhãn congrat

Rule\_spam.java – Check mention có phải nhãn spam

2.1 src/ controller/ data:

FilterClient.java – Liên quan tới service update đối tượng, dịch vụ

FilterMonitazService.java – Liên quan tới service update đối tượng, dịch vụ

Get\_Record.java – Lấy ra mention theo điều kiện

MessageInfo.java – Liên quan tới service update đối tượng, dịch vụ

2.1 src/ controller/ Filter\_spam\_webservice:

HttpClientNew.java – Check mention có phải spam theo model học máy

2.2 src/ controller/ mongodb:

Update.java – Không còn sử dụng

2.3 src/ controller/ mysqldb:

Mysql\_model.java – Connect, query tới Mysql

2.4 src/ controller/ util:

Utils.java – Định nghĩa cấu trúc đối tượng, dịch vụ, mention

3.1 src/ suppport / maxent: - Các hàm liên quan tới học máy trên java

Classifier.java

Filter\_question.java

Filter\_relevance.java

Learning.java

Model\_question.java

Model\_relevance.java

Model\_sub.java

PlainTextEventStream\_sub.java

PlainTextEventStream.java

SentDectec.java

3.2 src/ suppport / naivebayes\_main: - Các hàm liên quan tới học máy trên java

NaiveBayes\_Test.java

3.3 src/ suppport / remove\_vietnamese:

AccentRemover.java – Remove toàn bộ dấu

3.4 src/ suppport / tokenizer:

Accent.java – Remove toàn bộ dấu

PennTokenizer.java – Regex xử lí chuỗi

TextTokenizer.java – Regex xử lí chuỗi

Tokenizer.java – Regex xử lí chuỗi

4.1 src/ application/ Application

Application.java – Run project

Mô tả luồng:

1. Lấy ra mention, load từ điển đối tượng dịch vụ, sentiment
2. Check tag của mention và truyền các info vào task xử lí multithread
3. Task xử lí multithread:
   1. Đưa ra Tag của mention
   2. Check object của mention
   3. Check service của mention
   4. Check object, service của các thông tin liên quan tới mention nếu có (post, page name)
   5. Check spam của mention
   6. Check sentiment của mention theo 2 phương thức: Theo học máy và theo từ điển sentiment
   7. Các rule kết luận đưa ra object và service final của Mention
   8. Update service truyền lên object, service, tag, sentiment của Mention
   9. Update DB đã quét xong mention