

UNIVERSITAT DE LLEIDA
Escola Politècnica Superior
Grau en Enginyeria Informàtica
Treball de fi de grau

Estudi de discussions en xarxes socials: Youtube

Joaquim Picó Mora
49383707Q

Director : R.Bejar
Data : XXXX-XX-XX

Contents

1	Introducció	1
2	Descripció del Problema	2
3	Fase 1: Recollida de Dades	3
3.1	Descripció	3
3.2	Tecnologies	3
3.2.1	Google Cloud Plataform	3
3.2.2	Youtube API	3
3.2.3	Youtube Transcript API	3
3.3	Anàlisis	3
4	Fase 2: Transformació de les Dades	3
4.1	Descripció	3
4.2	Tecnologies	3
4.2.1	HuggingFace Project	3
4.2.2	Text Sumarization	4
4.2.3	ElementTree XML API	4
4.2.4	Sentiment Analysis	4
4.3	Anàlisis	4
5	Implementació Fase 1 i 2	4
5.1	Diagrama de Classes	4
5.2	Recollida de Dades	4
5.3	Construcció de l'arbre	4
5.3.1	Solució Interactiva	4
5.3.2	Solució Automàtica	5
5.4	Construcció XML	5
6	Fase 3: Integració	5
6.1	Descripció	5

6.2	Eina Web Interactiva	5
6.3	Eina d'anàlisi de Discussions	5
7	Conclusions	5
8	Bibliography	5

1 Introducció

Aquest treball pren com a objectiu la transformació i posteriorment l'anàlisi de discussions d'una de les xarxes socials més grans impulsada per una de les empreses tecnològiques més rellevants del món.

Youtube és una plataforma web on els usuaris poden afegir contingut en forma de vídeos, els quals són accessibles per altres usuaris. A la vegada, permet que aquests interactuin de diverses maneres amb els vídeos publicats a la plataforma, per exemple, amb un sistema de votació de likes i dislikes i donant l'opció d'afegir comentaris. Aquesta última característica, fa de Youtube una xarxa plena de debat, discussió i diferents opinions sobre infinitat de temes.

De fet, si parlem de números Youtube compta amb la increïble xifra de més de dos bilions d'usuaris repartits en 100 països que poden gaudir de contingut en 80 llengües i aglomerant diàriament un total de més d'un bilió d'hores de visualització de contingut de la plataforma.

En aquest document es recull la informació de tot el procés de desenvolupament del projecte, començant per una descripció més detallada del problema i seguint amb l'explicació de cada una de les fases que l'han format, així com les decisions d'implementació i problemes que s'han anat trobant durant el camí.

2 Descripció del Problema

Des del departament de GREIA de l'Universitat de Lleida, s'està treballant en una eina d'anàlisi de discussions que es donen en xarxes socials com Twitter, Reddit... Aquest treball consisteix en desenvolupar un modul més d'aquest projecte, una eina capaç de construi un arbre conversacional a partir de la discussió que es forma en els comentaris d'un vídeo de Youtube.

Per tal de realitzar-ho, en primera instància, s'haurà d'aconseguir informació i context sobre el contingut del vídeo per a formar el node arrel de l'arbre i seguidament dels comentaris per tal d'anar formant les branques de conversa. Posteriorment, un cop construït l'arbre, aquest s'haurà de serialitzar en un fitxer XML amb un format determinat per a que l'eina d'anàlisi de discussions ho pugui entendre. Aquest format consisteix en una component on es defineixen per una banda, els nodes del graf arbre i per altra, les arestes que els uneixen.

Finalment, s'integrarà aquesta eina a una aplicació web desenvolupada per a fer l'interacció entre aquesta i l'analitzador de discussions de forma més interactiva i còmoda.

S'ha decidit estructurat el projecte en tres fases:

- Una primera fase de recollida de dades: context del video, likes, comentaris...
- Una segona de transformació de les dades: construcció de l'arbre i parsejat a XML.
- I per últim, una tercerpart d'integració amb l'aplicació web i l'analitzador de discussions.

3 Fase 1: Recollida de Dades

3.1 Descripció

Explicació sobre que consisteix aquesta fase

3.2 Tecnologies

3.2.1 Google Cloud Plataform

Una mica d'explicació de la GoogleCloud Platform

3.2.2 Youtube API

Una mica d'explicació de la api de youtube

3.2.3 Youtube Transcript API

Una mica d'explicació sobre la llibreria de transcripció que s'utilitza

3.3 Anàlisis

Anàlisis realitzat en aquesta fase previ a la implementació. (Analisi dels camps de la api, de com funcionen els comentaris a youtube, de la transcripció de videos...)

4 Fase 2: Transformació de les Dades

4.1 Descripció

Descriuren que consisteix la fase

4.2 Tecnologies

4.2.1 HuggingFace Project

Explicar una mica sobre the HuggingFace Project

4.2.2 Text Sumarization

Explicar una mica sobre el model de sumarització de text

4.2.3 ElementTree XML API

Explicar una mica sobre la llibreria que s'emptra per a construir l'XML

4.2.4 Sentiment Analysis

Explicar una mica sobre el model d'anàlisi de sentiment

4.3 Anàlisis

Anàlisis previ, busqueda e llibreries XML, explicar perquè es fa la sumarització de la transcripció del video, Explicar perquè no es fa servir el CoreNLP d'stanford i s'han buscat altres opcions...

5 Implementació Fase 1 i 2

5.1 Diagrama de Classes

Diagrama de classes de la solució (He de fer el refactor)

5.2 Recollida de Dades

Part de recollida de dades i d'interacció amb les API'S

5.3 Construcció de l'arbre

Part de la construcció de l'arbre, explicació dels conflictes

5.3.1 Solució Interactiva

Explicació de la solució Interactiva

5.3.2 Solució Automàtica

Explicació de la solució Automàtica

5.4 Construcció XML

Explicació implementació construcció XML + Sentiment Analysis

6 Tests

7 Fase 3: Integració

7.1 Descripció

Descripció de la fase

7.2 Eina Web Interactiva

Descripció eina web + Integració

7.3 Eina d'anàlisi de Discussions

Descripció Eina Anàlisi de discussions + Integració

8 Conclusions

9 Bibliography

1- <https://www.youtube.com/intl/en-GB/about/press/>