Internship Report
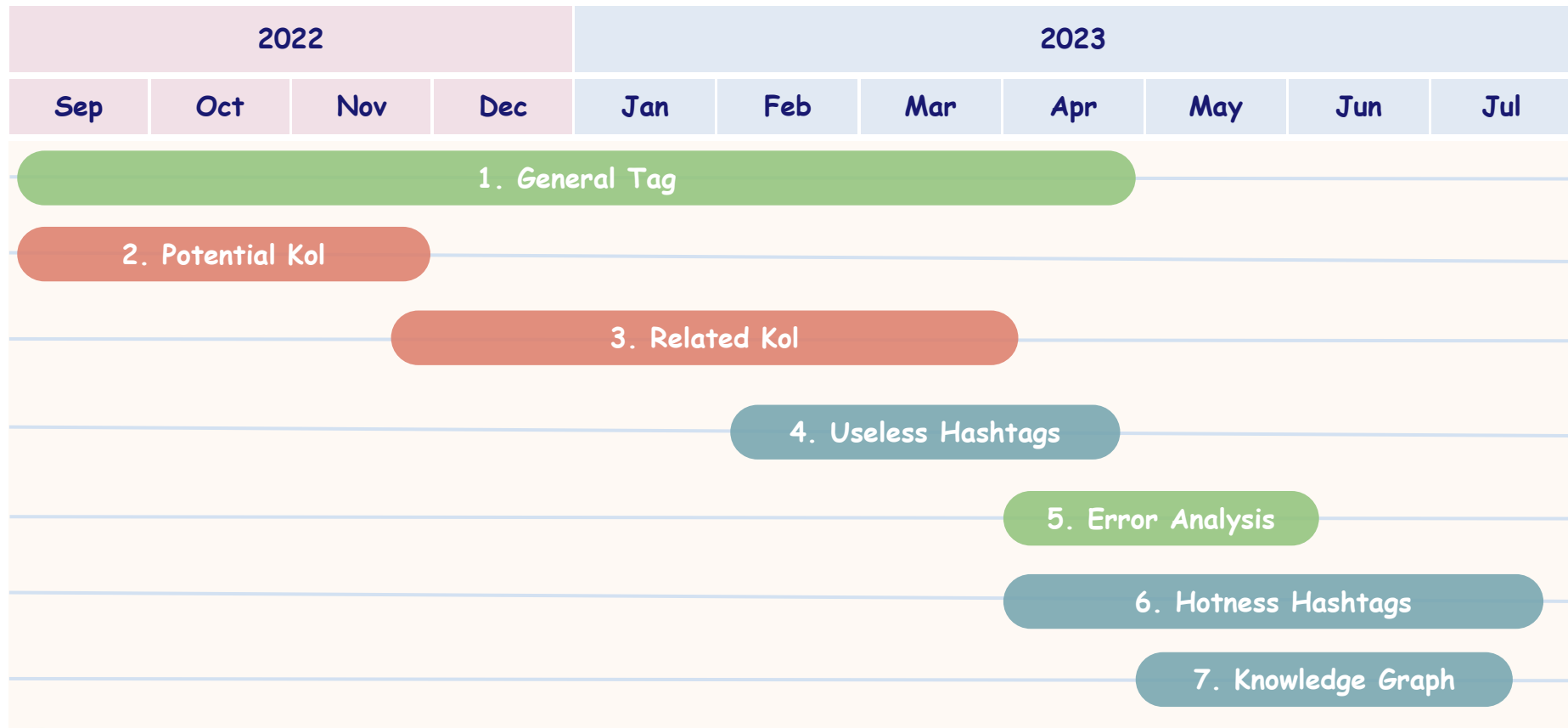
iKala AI Team - DA Intern - Antio Tseng

Mentor - Joy Shih, Mark Yu, Stephen Chen, Wendy Wen

2022.09.19 - 2023.7.31

General Tag

What is General Tag?

It is a text classifier with more training data, more categories and more accuracy than the original Autotag.

What did I do in General Tag?

黃阿瑪的後宮生活
5月24日 · 🌐

EX

阿瑪：「有沒有人願意給害羞膽小的經理一個遮風擋雨的家呀~」
---以下為子民口吻---
《貓咪送養 在台中 》

Auto Tag:
/pet
General Tag:
/pets/pet_rescue_adoption

姓名：經理（原本前主任叫他貓咪）
性別：男生
年齡：1歲左右
個性：害羞、膽小、給他一點時間應該會撒嬌、
結紮：以結紮
健康狀況：一切良好

小故事：原本是移工覺得可愛所以抓回來養（籠養）但最近移工要回去越南不再回來台灣了，想要放生他讓他自生自滅，但我又覺得經理這樣很可憐所以想找認養人，希望能找到牠的家人要永遠照顧牠愛牠。謝謝各位

聯絡方式：https://www.facebook.com/pohsuan.chen.77?mibextid=LQQJ4d

1.Web Crawling

Successfully crawled 100K+ articles in various categories using both dynamic and static web scraping techniques.

2. Labeling

Assist ML engineer in labeling articles across 19 level 1 categories and 263 level 2 categories.

Potential Kol

Potential kol

What is Potential Kol?

Predicting fan growth potential for KOLs across different platforms using machine learning & deep learning methods.

Normal kol

How to Predict Potential Kol?

Utilizing LightGBM model, predicting the potential of KOL in the next eight weeks using data from the initial six weeks. Ultimately, achieving an average precision of 70%.

Decline kol

曾敬驊
平台：FB
基期粉絲數：37k
基期後八週粉絲成長率：117.9%

EX

iKala | A Human-Centered AI Company
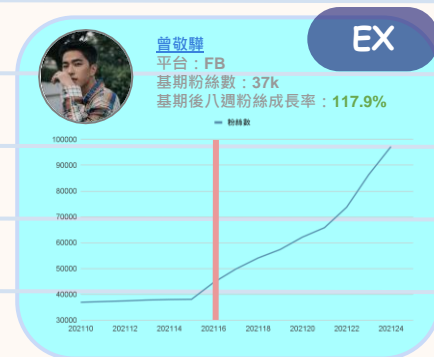
Related Kol

What is Related Kol?

Find out the related KOLs that KOLs interact with in the community to help KOL
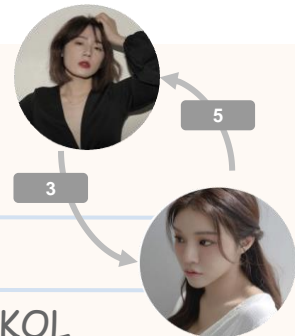
Radar propose feasible and novel cooperation plans for different business matching cases.

What is the definition of Related Kol?

Rule 1

The ranking of two-way related kol is always higher than that of one-way

Rule 2

Sort by number of mentions

EX	Related KOL	KOL platform name	Mention times	關係
1	ube	ube__	3	雙向
2	LIN YEN	linyenchou	2	雙向
3	Smirk	smir.k	1	雙向
4	Eros_ericwoohoo	ericwoohoo	1	雙向
5	魏如昀	qqqqqueen	1	雙向
6	「目靑」	6.21_	1	雙向
7	HSIN HO	hsiniho	1	雙向
8	舒安	an_2095	1	雙向
9	林志泓	jimmy1987909	20	單向
10	吳羿蓉Ara Wu	ara.wu	10	單向

iKala | A Human-Centered AI Company

High Frequency & Useless Hashtags

#台中美甲？   #傳說對決？

What is High Frequency & Useless Hashtags?

We often encounter various unrelated hashtags in posts.

We want to exclude those high-frequency & useless hashtags.

How did we do it?

Process

1. 文章分類 — We classified all posts according to the Autotag.

2. 高頻篩選 — defined a high-frequency threshold

3. 無用篩選 — used TF-IDF to filter out high-frequency irrelevant hashtags in various categories of posts.

Error Analysis

What is Error Analysis?

We cross-referenced the predicted results of 'Autotag' and 'Generaltag,' analyzed the differences, and used the insights to improve the 'Generaltag' predictions.

1 'Generaltag' has predictions while 'Autotagv2' has empty labels.

2 'Generaltag' has empty labels while 'Autotagv2' has predictions.

3 'Generaltag' predictions span multiple labels across level 1 categories.

	GeneralTag 空標	GeneralTag 有結果	GeneralTag 多標籤
AutoTag 空標	-	O **1**	O **3**
AutoTag 有結果	O **2**	-	O **3**

Hotness & Trending Hashtags

Hotness Hashtags

僅以 hashtags 熱度分數進行排名，掌握當週高討論度的 hashtags。 🔥

Trending Hashtags

考量 hashtags 熱度的成長趨勢，掌握近期爆紅或有成長趨勢的 hashtags。 📊

What is Hotness & Trending Hashtags?

The goal of this project is to redesign a formula to measure the popularity of hashtags and use this formula as the basis for developing trending hashtags.

How did we do it?

熱度公式

$$\log_{15}(\text{weekly\_voice} + 1) * \log_{2}(\text{weekly\_kol\_count} + 1)$$

排序方式

基期週熱度分數 — 前四期熱度平均

EX

情人節快樂　元宵節快樂　happyvalentinesday　20220214　valentineday

iKala | A Human-Centered AI Company
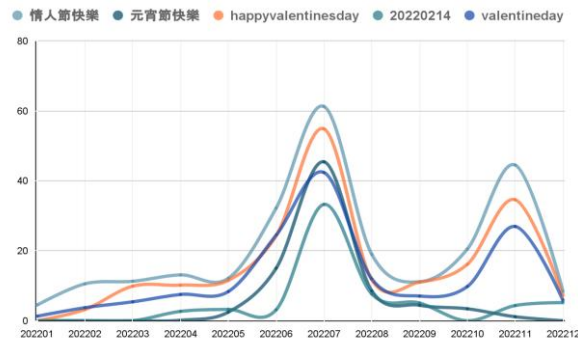
Knowledge Graph

What is Knowledge Graph?

Using LLM, we create knowledge graphs for various categories of social media posts by extracting entities, attributes, and relationships.

What did I do in Knowledge Graph?

1. Defining the entities, attributes, and relationships.

xN

2. Adding and modifying prompts for the knowledge graph.

xN

iKala | A Human-Centered AI Company