

BÁO CÁO ĐỒ ÁN TỐT NGHIỆP

NGHIÊN CỨU HỆ THỐNG ĐỀ XUẤT NỘI DUNG ÁP DỤNG KỸ THUẬT
CONTENT-BASED FILTERING TRONG ỨNG DỤNG NGHE NHẠC
TRỰC TUYẾN INCOGNITO MUSIC

Sinh viên thực hiện : HOÀNG MINH TÀI
Mã số sinh viên : 6051071102
Giảng viên hướng dẫn: ThS. TRẦN PHONG NHÃ

CẤU TRÚC BÀI THUYẾT TRÌNH

1. Lý do chọn đề tài
2. Giới thiệu về ứng dụng
3. Giới thiệu về hệ thống đề xuất nội dung
4. Ứng dụng Incognito Music
5. Kết quả đạt được và kiến nghị



Lý do chọn đề tài

Những ứng dụng nghe nhạc miễn phí đang gây khó chịu cho người dùng:

1. Quảng cáo
2. Giới hạn số lần chuyển bài
3. Không cho phép tải bài hát
4. Tự động thay đổi danh sách phát
5. Chất lượng âm thanh không cao



Giới thiệu về ứng dụng

Các công nghệ

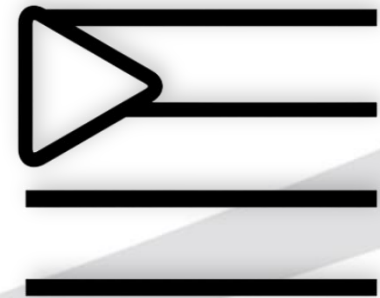
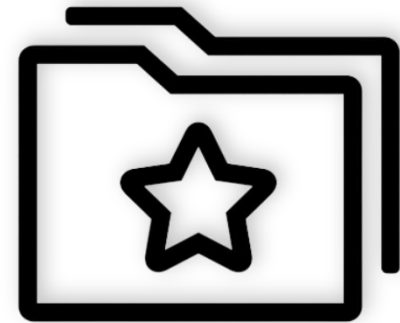
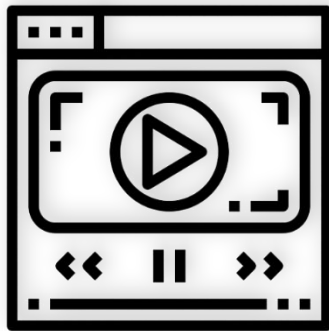
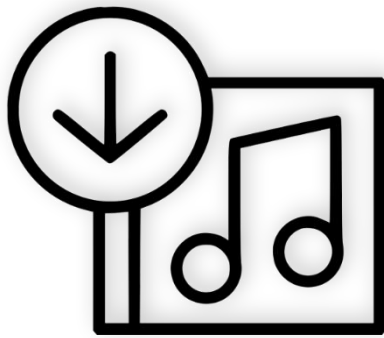


Firebase Realtime Database



Giới thiệu về ứng dụng

Các tính năng nổi bật



Giới thiệu về hệ thống đề xuất nội dung

Các công nghệ



Flask



Giới thiệu về hệ thống đề xuất nội dung

Các thư viện



NumPy



Giới thiệu về hệ thống đề xuất nội dung

Cách thức hoạt động

```
1 songs_df = pd.DataFrame(songs).dropna().drop_duplicates(['id'])
2
3 song_stat_df = songs_df.sample(5,ignore_index=True)
4 song_stat_df.drop(columns=['320kbps','album_id','duration','has_lyrics','image','perma_url','url','release_date','genre','album','language',
5 # to convert string into list of strings
6 def splitting(text):
7     text=str(text).split()
8     return text
9 song_stat_df['title'] = song_stat_df['title'].apply(splitting)
10 song_stat_df['artist'] = song_stat_df['artist'].apply(splitting)
11 song_stat_df['all_tags'] = song_stat_df['title'] + song_stat_df['artist']
12 song_stat_df.drop(columns=['title','artist'],inplace=True)
13 song_stat_df
```

✓ 0.1s

	id	all_tags
0	iy9qZR_OGa0	[Agust, D, '해금', Official, MV, HYBE, LABELS]
1	Ayteq9RzXvs	[Đế, Vương, -, Đình, Dũng, Dinh, Dung]
2	u14rrcxENDw	[MC, 張天賦, -, 老派約會之必要, A, Gentleman's, Guide, t...
3	F5tS5m86bOI	[LA, LỪNG, /, Vũ., (Original), Vũ, Official]
4	0ovYubZL2A4	[Trăm, Năm, Không, Quên, -, Quang, Hà, Live, C...

Giới thiệu về hệ thống đề xuất nội dung

Cách thức hoạt động

```

1 def convert_lower(text):
2     l=[]
3     for item in text:
4         l.append(item.lower())
5     return l
6 song_stat_df['all_tags']=song_stat_df['all_tags'].apply(convert_lower)
7 song_stat_df

```

✓ 0.0s

	id	all_tags
0	iy9qZR_OGa0	[agust, d, '해금', official, mv, hybe, labels]
1	Ayteq9RzXvs	[để, vương, -, đình, dừng, dinh, dung]
2	u14rrcxENDw	[mc, 張天賦, -, 老派約會之必要, a, gentleman's, guide, t...
3	F5tS5m86bOI	[lạ, lòng, /, vũ., (original), vũ, official]
4	0ovYubZL2A4	[trăm, năm, không, quên, -, quang, hà, live, c...

```

1 from nltk.stem.porter import PorterStemmer
2 ps=PorterStemmer()
3 def stemming(text):
4     l=[]
5     for i in text:
6         l.append(ps.stem(i))
7     return l
8 song_stat_df['all_tags'] = song_stat_df['all_tags'].apply(stemming)
9 song_stat_df['all_tags'] = song_stat_df['all_tags'].apply(lambda x: " ".join(x))
10 song_stat_df

```

✓ 0.0s

	id	all_tags
0	iy9qZR_OGa0	agust d '해금' offici mv hybe label
1	Ayteq9RzXvs	để vương - đình dừng dinh dung
2	u14rrcxENDw	mc 張天賦 - 老派約會之必要 a gentleman' guid to old-fash...
3	F5tS5m86bOI	lạ lòng / vũ. (original) vũ offici
4	0ovYubZL2A4	trăm năm không quên - quang hà live concert tu...

Giới thiệu về hệ thống đề xuất nội dung

Cách thức hoạt động

$$TF(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad \{1\}$$

$$\{2\}$$

$$IDF(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|} \quad \{3\}$$

$$\{4\}$$

$$Tfidf_{(t,d,D)} = Tf_{(t,d)} * Idf_{(t,D)}$$

Trong đó:

{1} là số lần xuất hiện của thuật ngữ t trong tài liệu d.

{2} là tổng số từ trong tài liệu d.

{3} là tổng số tài liệu trong tập văn bản D.

{4} là số văn bản chứa từ nhất định, với điều kiện t xuất hiện trong văn bản ($TF_{(t,d)} \neq 0$)

```
1 from sklearn.feature_extraction.text import TfidfVectorizer
2 vectorizer = TfidfVectorizer()
3 all_tags = song_stat_df['all_tags']
4 data = vectorizer.fit_transform(all_tags)
5 print("Unique vocabulary: ", vectorizer.vocabulary_)
6 data.toarray()

✓ 0.0s

Unique vocabulary: {'agust': 0, '해금': 41, 'offici': 24, 'mv': 21, 'hybe': 11,

array([[0.42841136, 0., 0., 0., 0.,
        0., 0., 0., 0., 0.,
        0., 0.42841136, 0., 0., 0.,
        0.42841136, 0., 0., 0., 0.,
        0., 0.42841136, 0., 0., 0.28691208,
        0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0.,
        0., 0.42841136],
```

Giới thiệu về hệ thống đề xuất nội dung

Cách thức hoạt động

```
1 from sklearn.metrics.pairwise import cosine_similarity
2 similarity = cosine_similarity(data)
3 similarity
```

✓ 0.0s

```
array([[1.          , 0.          , 0.04599996, 0.07040474, 0.          ],
       [0.          , 1.          , 0.          , 0.          , 0.          ],
       [0.04599996, 0.          , 1.          , 0.03934248, 0.          ],
       [0.07040474, 0.          , 0.03934248, 1.          , 0.          ],
       [0.          , 0.          , 0.          , 0.          , 1.          ]])
```

Ứng dụng Incognito Music

Quét mã QR để cài đặt ứng dụng (Android)



Kết quả đạt được và kiến nghị

Kết quả

Về chuyên môn, em đã làm quen với việc sử dụng Flutter để thiết kế và phát triển được một ứng dụng đáp ứng ổn các chức năng yêu cầu đề ra. Biết cách sử dụng Python xây dựng hệ thống đề xuất nội dung, tích hợp các thuật toán để xử lý dữ liệu và tạo API. Học và vận dụng thành công NoSQL.

Về thái độ, đồ án đã giúp em nâng cao thêm khả năng xử lý logic và giải quyết vấn đề trong lập trình. Cải thiện khả năng đọc tài liệu và học những công nghệ mới trong thời gian ngắn. Nâng cao tính tập trung trong công việc.

Kết quả đạt được và kiến nghị

Kiến nghị

Vận dụng thêm kĩ thuật collaborative filtering vào hệ thống đề xuất nội dung để tăng thêm trải nghiệm cho người dùng.

Phát triển chức năng cho người dùng đăng tải bài hát của chính mình và hướng đến một hệ thống giải trí đa nền tảng.



CẢM ƠN MỌI NGƯỜI ĐÃ LẮNG NGHE