



好RRRRRRRR

FINAL PROJECT

華語/西洋音樂歌曲分析

陳怡安/林鼎鈞/黃耀霆/劉少桐/徐雋翔





C



Contents

01

WHY?
發想緣由

02

WHAT?
打算解決什麼問題



03

WHERE?
資料來源

04

HOW?
探索分析





WHY?



一群音樂的重度愛好者，以音樂
為題材無疑是好奇心的驅使。

自身經驗 **1**

2 可取得性

音樂資料橫跨年度大、資料齊全、
易爬取、易整理。

音樂價值 **3**

4 比較性

音樂是大時代的縮影，歲月被譜
寫成曲。每個雋永的片刻都被藝
術家永遠刻劃在黑膠唱片裡。音
樂代表一個時代、一抹思想。

華語/西洋樂皆有完整數據，易於比
較、分析，且網路資源充沛，上手
程度較低。





WHAT?

01

歌手進榜數統計



本組從西洋歌曲出發，因好奇於不同時代的歌手演變，是故以50年間的排行榜和百大金曲下手，統計歌手上榜次數，進而得出最熱門的歌手為何。

02

頻率詞統計



更深入地探討歌詞與歌名，本組意識到，不論是50年內的歌曲、百大金曲或最膾炙人口(hummed)的歌曲，情緒性詞彙(love, like, know)皆是最常出現的。有感於此，我們決定以歌詞的情緒分析為主要命題。

03

歌詞情緒分析



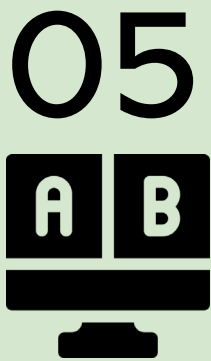
我們好奇於西洋樂壇的情緒用詞習慣，進而將情緒分為十類，並分析每種情緒中最常出現的用字。接著，以15年間的百大金曲為數據依據，統計前30名最正面(Positive)及最負面(Negative)的歌曲。

04

歌曲正負向年度趨勢



本組認為，既然能以情緒劃分歌曲，何不從縱向維度探討西洋樂壇的正負向比例，並從中獲得樂壇的演進趨勢。經過統計及分析，我們發現2019年是西洋樂壇有史以來最負面的一年。



05 樂壇&歌手正負向比較

有感於西洋樂壇的分析結果，本組希望也能自橫向角度，探勘華語樂壇是否也有相同趨勢與同樣的習慣。因此，我們將華語樂與西洋樂進行比較，從而得出兩個結果，分別是華語/西洋歌手和雙方樂壇的正負面比率。

06

歌詞分級分析



以教育部公布的7000單字分級為數據基礎，對Top 100 hummed songs進行分析，探討較傳唱的歌曲是否用詞也較為簡單。結果得出，最為傳唱的歌曲，其用詞頗為極端化，1級和7級用詞占最大宗。



WHERE?

我們的資料哪裡來？

歌手、歌名排行榜

BILLBOARD
KKBOX
其他網站統計

歌詞

MUSIXMATCH
GENIUS
KKBOX

樣本數

5000+
(1970-2020)

合法性

使用網路爬蟲、API
爬取公開網站數據，
研究用途合法

資料來源

Rpubs
Github
相關網站



HOW?

分析歌詞流程

收集

Tidyr套件
Tidyttext套件
TextClean套件
Textdata套件

分析

Ggplot2套件
Wordcloud2套件
Tidyttext套件

報告

網路爬蟲
Python
HTTR套件
RVEST套件
KKBOX API

清洗

Jiebar套件
Dplyr套件
Tibble套件
Tm套件
Qdap套件

視覺化

靜態網站



分工名單

黃耀霆

負責內容：

投影片製作、紀錄討論、網路爬蟲、歌手進榜數統計、頻率詞統計、歌詞情緒分析

投入程度：**6**

陳怡安

負責內容：

書面報告製作、網路爬蟲、歌手進榜數統計、頻率詞統計、歌曲正負向年度趨勢

投入程度：**6**

林鼎鈞

負責內容：

書面報告製作、網路爬蟲、頻率詞統計、歌詞分級分析

投入程度：**5**

劉少桐

負責內容：

網站架設、主題發想、網路爬蟲、歌詞彙整(50個年度)

投入程度：**5**

徐雋翔

負責內容：

網站架設、網路爬蟲、華語/西洋樂壇正負向比較、歌手正負向用詞統計

投入程度：**6**



※投入程度以組員互評之分數取平均並四捨五入至整數位。



靜態網頁 Demo



THANKS!

