

Análisis de las Palabras más usadas en el Cuarteto Cordobés

Patricio Del Boca - Open Data Córdoba

23 de julio de 2017

Introducción

El objetivo del siguiente análisis es identificar cual es la palabra más usada en las Letras del Cuarteto Cordobés.

Metodología

Para las canciones se scrapeó de las páginas web <http://www.coveralia.com> y <http://www.sonicomusica.com> las letras de los artistas más reconocidos del Cuarteto Cordobés definido en base a la cantidad de canciones existentes para cada uno de los artistas. (lista de artistas más abajo)

Una vez extraídas las letras de las canciones se prosigió unir las fuentes y remover las palabras más comunes del lenguaje español denominadas comunmente en la jerga técnica como: **stop words** (el, la, vos, yo, tu, el, ella. . .). La lista completa de palabras eliminadas del análisis puede verse en el este link.

Luego de extraídas las “stop words” se prosigió a contabilizar la ocurrencia de cada palabra. Se contabilizó la suma de ocurrencias de la palabra en el conjunto total de canciones y luego para cada artista.

Consideraciones Metodológicas

Se analiza y contabiliza la ocurrencia de la palabra mas no el contexto o significado de la misma. Ejemplos:

- La palabra *quiero* aparece segunda en el ranking mas no se detalla el uso (por ejemplo **te quiero** vs **yo quiero**).
- La palabra *vida* no diferencia si se usa como **te quiero mi vida** o **mi vida no es nada sin ti**.
- Tampoco se ha aplicado la técnica conocida como **steeming** es decir, reducir palabras a su raíz. Es por ello que palabras como bailar y bailando aparecen en autores como *la fiesta*.
- La lista de palabras a remover/incluir como *stop_words* es editable y configurable. Por ejemplo, si se desea incorporar y hacer una comparación específica de *el* vs *ella*

Analisis

Cantidad de Canciones por Autor

Se analizaron 10 autores y un total de 3138 canciones.

nombreArtista	canciones
la mona jimenez	721
la barra	421
banda xxi	417
sabroso	348
tru la la	290
la fiesta	269
damian cordoba	209
ulises bueno	166
banda registrada	149
jean carlos	148

15 Palabras más usadas

A continuacion se detallan las 15 palabras más utilizadas en el conjunto de canciones analizadas:

ranking	palabra	n
1	amor	7275
2	quiero	4059
3	vida	2826
4	corazón	2465
5	mujer	1346
6	noche	1330
7	alma	1184
8	cuerpo	1118
9	siento	1013
10	bailar	873
11	piel	847
12	vivir	834
13	loco	787
14	mundo	727
15	besos	719

15 Palabras más utilizadas por cada Autor

A continuación se desglosan las palabras más utilizadas por cada autor

ranking	banda registrada	banda xxi	damian cordoba	jean carlos	la barra
1	amor	amor	amor	amor	amor
2	vida	quiero	quiero	quiero	quiero
3	quiero	corazón	vida	vida	vida
4	corazón	vida	corazón	corazón	corazón
5	piel	calor	noche	amo	noche
6	alma	mujer	duele	cuerpo	mujer
7	amar	cuerpo	quieres	dolor	fiesta
8	queda	bonito	alma	fuego	siento
9	mujer	alma	besos	mujer	vivir
10	siento	linda	volver	dios	bailar
11	volver	siento	cuerpo	estás	loco
12	mundo	acelerao	pasa	dame	alma
13	vivir	ritmo	mujer	alma	cuerpo
14	cielo	besos	miedo	siento	beso
15	noche	llego	corazon	detective	bebo

ranking	la fiesta	la mona jimenez	sabroso	tru la la	ulises bueno
1	amor	amor	amor	amor	amor
2	quiero	quiero	quiero	quiero	quiero
3	vida	corazón	vida	corazón	vida
4	fiesta	vida	corazón	vida	corazón
5	corazón	noche	alma	noche	piel
6	alma	mujer	sabroso	mujer	noche
7	bailar	beso	cuerpo	alma	cuerpo
8	bailando	cuerpo	piel	baila	alma
9	mujer	vivir	amo	bailar	mujer
10	mira	bailar	noche	mundo	siento
11	noche	loco	boca	cuerpo	dime
12	siento	mata	vivir	loco	loco
13	amo	piel	siento	mueve	culpa
14	baila	siento	mujer	dame	vivir
15	piel	huella	loco	dime	quieres