

Viveza criolla en la musica uruguaya

Martín Berguer *Instituto de Computación*
Facultad de Ingeniería, Universidad de la República
 Montevideo, Uruguay
 martin.berguer@fing.edu.uy

Pablo Olivera *Instituto de Computación*
Facultad de Ingeniería, Universidad de la República
 Montevideo, Uruguay
 pablo.olivera@fing.edu.uy

Resumen—Este documento se realiza en el marco del proyecto final de la asignatura Bases de Datos No Relacionales, y tiene como objetivo estudiar el abordaje de la creación de un dataset utilizando datos extraídos de la pagina web whosampled.com y almacenados en el motor de base de datos de grafos Neo4j.

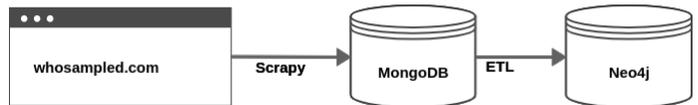


Figura 1: Diagrama de la solución.

I. INTRODUCCIÓN AL PROBLEMA

El proyecto consiste en crear una base de datos de grafos con Artistas relacionados al mundo de la música en Uruguay, tanto individuales como bandas o grupos musicales uruguayos (los cuales a partir de ahora, y por el resto del documento nos referiremos como ‘Artista’ sin pérdida de generalidad y para mantener consistencia) de forma que se puedan obtener métricas en base a las relaciones entre ellos.

El dataset permitirá extraer diversas métricas y buscar relaciones que a priori no sean fáciles de encontrar. Los atributos a almacenar de cada Artista dependerán de lo que esté al alcance de los usuarios en el sitio web whosampled.com. Es en este sentido que se optó por representar la realidad en un modelo de grafos, que sumado a las operaciones orientadas a grafos que nos provee Neo4j, se espera obtener métricas y datos de color sobre los Artistas referentes de Uruguay que a lo mejor de nuestro conocimiento en base a lo investigado hasta el momento son nunca antes vistos en un proyecto académico.

II. TRABAJOS RELACIONADOS

La literatura obre el tema y en particualr sobre la web whosampled.com no es muy extensa, sin embargo, existen dos artículos escritos en Brasil sobre el perfil de las personas que han contribuido datos [de BRITO(2018)], así como una monografía de la comunicación entre las personas del sitio [BRITO(2019)]. También se encontraron artículos relacionados a los géneros de música que más son *sampleados*¹ [Tran(2019)], [Bryan and Wang(2011)]. Es posible que no haya mucha literatura sobre esto, ya que el sitio web no discrimina según el país de origen del Artista, con lo cual fue necesario realizar algunas modificaciones al modelo de datos, detalle que se abordará más adelante en la sección V.

¹terminología para referirse a cuando una canción es utilizada en la creación de otra

III. DESCRIPCIÓN DE LA SOLUCIÓN

La solución, cuyo diagrama se observa en la figura 1, consiste en el desarrollo de un scraper para obtener la información de los Artistas desde whosampled.com, una base de datos documental para guardar los datos en crudo, y una base de grafos para hacer disponible los datos para la etapa experimental. El código fuente y el conjunto de datos se encuentran en https://gitlab.fing.edu.uy/bdnr_2022_grupo03/bdnr.

III-A. Módulo de Scraping

El scraper fue implementado utilizando el framework Scrapy, con el cual, mediante la extensión de clases e implementación de algunas funciones, podemos especificar qué información obtener y cómo guardarla en una base de datos. Mediante parámetros permite configurar cabecales, cantidad de pedidos por segundo, tiempo máximo de ejecución, entre otros. Posteriormente, para almacenar los datos en crudo obtenidos por Scrapy utilizamos MongoDB. En el listado 1 se incluye parse, una de las principales funciones utilizada por Scrapy, que es ejecutada cada vez que se hace un request a una URL, y es donde se puede programar para obtener toda la información que se necesite. Un ejemplo de esto es la función `_get_track_items` que se encarga de obtener la lista de ítems del Artista mediante los cuales se puede conseguir la información de cuáles son los demás Artistas con los cuales se relaciona.

Listado 1 Principales funciones del scraper

```

def parse(self, response):

    # Artist basic info
    whosampled_item = _get_artist_info(response)
    whosampled_item['created_at'] = datetime.now()
    whosampled_item['track_items'] = _get_track_items(
        response.xpath('//article[@class="trackItem"]')
    )
    whosampled_item['parsed_page'] = 1

    # Resolve pagination
  
```

```

if PAGINATION_SUBSTR not in response.url:
    for next_page_url in \
        _get_items_from_pagination(
            response.url, response.xpath(
                '//span[@class="page"]'
            )
        ):
            yield response.follow(
                next_page_url,
                callback=self.parse
            )
else:
    whosampled_item['parsed_page'] = \
        int(response.url.split(PAGINATION_SUBSTR)[1])

    # Linked artists
    whosampled_item['child_links'] = \
        _get_urls_from_track_items(
            whosampled_item['track_items']
        )
    for u in run_for_childs(whosampled_item['child_links']):
        yield response.follow(u, callback=self.parse)

yield whosampled_item

def _get_track_items(track_items_html):
    track_items = []
    for item in track_items_html:
        track_items.append({
            'track_name': item.xpath(
                './header/h3/a/span/text()'
            ).extract_first(),
            'track_url': item.xpath(
                './header/h3/a/@href'
            ).extract_first(),
            'track_connections':
                _get_track_connections(item.xpath(
                    './div/div[@class="track-connection"]'
                ))
        })
    return track_items

```

III-B. Módulo de Grafos

Implementamos una ETL en Python mediante la cual se leen los datos de MongoDB, se transforman para construir la información de los nodos y las relaciones, luego inserta en Neo4j. Para los nodos se utiliza la URI como etiqueta única de identificación, un atributo booleano que indica si es de Uruguay (en la sección V se detallará el porqué de esta variable, así como su funcionamiento) y uno con el nombre del Artista. Las relaciones tienen atributos con el nombre y la URI de la obra (o canción) que generó el vínculo. En 2 se incluye la consulta utilizada para insertar los datos en Neo4j la cual utiliza MERGE para evitar el duplicado de información.

Listado 2 Neo4j query para crear nodos y relaciones

```

MERGE (a1:Artist {{ uri: \${artist1_uri} }})
MERGE (a2:Artist {{ uri: \${artist2_uri} }})
MERGE (a1)-[r:{relationship['type']}]->(a2)
ON CREATE SET a1.from_uruguay =
    '{relationship['band_from_slug']}_in_UY_BANDS_NAMES}'
ON CREATE SET a2.from_uruguay =
    '{relationship['band_to_slug']}_in_UY_BANDS_NAMES}'
ON CREATE SET a1.name =
    '{_sanitize_name(relationship['band_from_name'])}'
ON CREATE SET a2.name =
    '{_sanitize_name(relationship['band_to_name'])}'
ON CREATE SET r.track_name =
    '{_sanitize_name(relationship['track_name'])}'
ON CREATE SET r.track_url =
    '{_sanitize_name(relationship['track_url'])}'
RETURN a1, a2

```

IV. DISEÑO Y CARGA DE LOS DATOS

El grafo se diseñó con los Artistas como nodos y seis tipos de relación entre ellos {WAS_COVERED_IN, WAS_SAMPLED_IN, WAS_REMIXED_IN, IS_A_COVER_OF, SAMPLED, IS_A_REMIX_OF}, que se pueden separar en tres y sus respectivos recíprocos. Por restricciones del dominio del sitio web, solo podemos acceder a un lado de la relación a la vez, en base a esto, exceptuando aquellos Artistas que se logró parsear toda la información, puede suceder que haya relación en una sola dirección. En la figura 2 se observa un ejemplo del grafo para un subconjunto de Artistas uruguayos.

Para la carga de datos se comenzó con el desarrollo del scraper que es capaz de descargar los datos en crudo a una base de datos MongoDB, para agilizar así el proceso de extracción. De forma paralela, se continuó con el desarrollo de la herramienta capaz de leer la información de la base de datos MongoDB creada en el paso anterior y construir el dataset de grafos en Neo4j. Las pruebas de carga y estabilidad del sistema se realizaron en entornos locales de ejecución, y para el almacenamiento en principio se utilizó el servicio AuraDB² que permite hasta 50000 nodos en su plan gratis. Sin embargo, esta restricción hizo muy difícil la tarea de simplificar el pipeline con los Artistas a utilizar, por lo que una vez confirmado el correcto funcionamiento de la ETL, se procedió a crear el entorno de ejecución que consta de una instancia en AWS EC2 donde se instaló Neo4j edición comunitaria, capaz de almacenar toda la información.

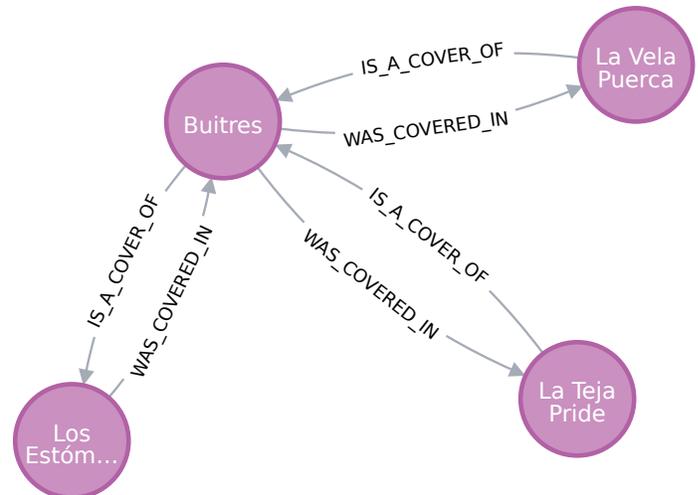


Figura 2: Ejemplo de relaciones entre Artistas uruguayos.

Para la transformación de los documentos a relaciones de grafo en Neo4j, era necesario optimizar el proceso ya que la carga se estaba haciendo muy lenta, con lo cual se optó por crear un índice en la propiedad URI de los nodos con etiqueta *Artist*. Esta acción mejoró considerablemente el proceso de carga de datos, ya que en cada nodo se utilizó el operador MERGE, el cual era encargado de crear ambos nodos en caso de que alguno no existiera, por lo que en cada iteración existe una búsqueda intrínseca. Las consultas a los datos

²[urlhttps://neo4j.com/cloud/platform/aura-graph-database/](https://neo4j.com/cloud/platform/aura-graph-database/)

también vieron mejorada su performance, lo que permitió extraer que los índices son absolutamente necesarios y mejoran considerablemente la misma, algo que es comprobable incluso en bases de datos no relacionales.

IV-A. Colaboración con la web whosampled

Whosampled es un sitio web formado pura y exclusivamente por aportes de la comunidad, lo que lo convierte en un dataset activo continuamente, con métricas realmente impresionantes. A la fecha de este trabajo, más de 27 mil colaboradores, 860 mil canciones y 270 mil Artistas. En relación a esto, y dada la investigación paralela que se realizó para recabar más información sobre las relaciones de las los Artistas uruguayos, se procedió a colaborar con el dataset, introduciendo un total de:

- 11 relaciones nuevas (9 covers, 2 samples).
- 7 artistas nuevos.

Además en el proceso de búsqueda el equipo de investigación fue capaz de identificar dos situaciones claves que potencialmente podían afectar el dataset extraído, esto es:

- La canción Bola de nieve, de la banda La Furia en un comienzo estaba asociada a la banda de rock La Furia y fue necesario reportar el error, el cual fue solucionado en prácticamente que en el acto.
- La canción Mayonesa, de la banda Chocolate en un comienzo estaba asociada a la banda brasileña Chocolate. De forma similar, fue solucionado en un par de horas.

El gran cuello de botella identificado luego de estas colaboraciones es la poca capacidad de los moderadores - 70 personas - para aprobar nuevas solicitudes, luego de contactar a uno de ellos directamente logramos comprender la magnitud de desfase y el elevado tiempo de espera que tiene el sitio en poder procesar cada formulario, en algunos casos más de 10 meses. La calidad de los datos es la prioridad máxima³ en este caso, ya que es la fuente de datos por defecto de muchos de los sellos discográficos más grandes del mundo como Universal Music Group, Sony Music, Warner Music Group, así como de diversos trabajos académicos, algo que es posible evidenciar en Google Scholar⁴.

V. ANÁLISIS DE DATOS

Antes de detallar las métricas y el análisis de los Artistas uruguayos, es necesario retomar un punto que se adelantó en la sección III-B y comentar que el sitio no presenta información relacionada al país del Artista, por lo que fue necesario realizar un extenso trabajo sobre los datos antes de llegar a una cantidad de nodos que permitieran hacer un análisis representativo de la realidad.

Para esto, se procedió a identificar fuentes de datos confiables, las cuales pudieran ser utilizadas como seed del parser. En este sentido, utilizamos tres frentes principalmente: el sitio web oficial de Agadu⁵, un catálogo disponible en wikipedia⁶,

³<https://www.whosampled.com/about/>

⁴[https://scholar.google.com/scholar?hl=en&q="whosampled"](https://scholar.google.com/scholar?hl=en&q=)

⁵https://www.agadu.org/yo_creo_cultura_bandas_grupos_websocios.php

⁶https://es.wikipedia.org/wiki/Categor%C3%ADa:Grupos_de_rock_de_Uruguay_por_g%C3%A9nero

y el sitio web sietenotas⁷ que tiene una interesante base de datos con casi 700 Artistas.

Los resultados macro del scraper en 6 días fueron una base de datos con dos iteraciones a MongoDB, con un total de 230471 (135392+95079) registros documentales, y en Neo4j 129591 nodos y 1164572 Artistas. De estas cifras, se pudo identificar que un total de 57 Artistas eran uruguayos, y un total de 448 relaciones involucran algún nodo uruguayo en cualquiera de los sentidos de la relación.

Particularmente, de las preguntas que se buscaba responder al comienzo de esta investigación, con todos los datos cargados en Neo4j se puede apreciar lo siguiente:

1. El top 10 de artistas Uruguayo con más covers realizados son:

Artista	Links
Carlos Gardel ⁸	40
Los Iracundos	11
Gerardo Matos Rodríguez	10
El Kinto	6
Leo Maslíah	4
Alfredo Zitarrosa	4
Ruben Rada	3
Bajofondo	3
Jaime Roos	3
Natalia Oreiro	3

Listado 3 Query para relaciones en las que le hicieron cover

```
MATCH (n1:Artist {from_uruguay: "True"})
WITH n1, SIZE()-[:IS_A_COVER_OF]->(n1) as links
ORDER BY links DESC LIMIT 10
MATCH (n2)-[]->(n1)
RETURN distinct n1.name as Artista, links
```

Como se puede apreciar en la tabla, el Artista uruguayo que más influenció a otros artistas es Carlos Gardel, seguido de Gerardo Matos Rodriguez, quien además tiene la canción escrita por un uruguayo que más ha sido utilizada por otros artistas, con un total de 5 samples y 19 covers.

2. Los artistas que más influenciaron⁹ a los artistas uruguayos son:

Artista	Links
The Beatles	6
Patricio Rey Y Sus Redonditos De Ricota	3
Jerry Rivera	2
Fito Páez	2
Virgilio y Homero Espósito	2
Los Fabulosos Cádillacs	2
Gilda	2
Blue Swede	1
Turbonegro	1
Miguel Alejandro	1

Listado 4 Query para relaciones en las que le hicieron cover

⁷<https://sietenotas.com/7notas/Interpretes>

⁸Si, era uruguayo

⁹En el sentido de ser el artista con más relaciones con artistas uruguayos

```

MATCH (n1:Artist {from_uruguay: "False"})
WITH n1, SIZE((:Artist {from_uruguay: "True"})-
[:IS_A_COVER_OF]-(n1)) as links
ORDER BY links DESC LIMIT 10
MATCH (n2)-[]-(n1)
RETURN distinct n1.name as Artista, links

```

3. El top 10 de artistas uruguayos que más covers tienen en su repertorio es:

Artista	Links
Azafatas Del Amor	13
Los Iracundos	12
Bajofondo	8
No Te Va Gustar	6
Los TNT	6
Trotsky Vengaran	5
Rubén Rada	5
Los Shakers	5
Los Olimareños	4
Jorge Drexler	4

Listado 5 Query para relaciones en las que le hicieron cover

```

MATCH (n1:Artist {from_uruguay: "True"})
WITH n1, SIZE((:Artist {from_uruguay: "True"})-
[:IS_A_COVER_OF]-(n1)) as links
ORDER BY links DESC LIMIT 10
MATCH (n2)-[]-(n1)
RETURN distinct n1.name as Artista, links

```

/

4. Artistas extranjero más influenciados por artistas uruguayos:

Artista	Links
Andrés Calamaro	3
James Last	2
Mercedes Sosa	2
Julieta Venegas	2
Los Auténticos Decadentes	2
German Wiedemer	2
Miranda!	1
Eydie Gormé	1
Julio Iglesias	1
Luis Miguel	1

Listado 6 Query para relaciones en las que le hicieron cover

```

MATCH (n1:Artist {from_uruguay: "False"})
WITH n1, SIZE((n1)-[:IS_A_COVER_OF]
->(:Artist {from_uruguay: "True"})) as links
ORDER BY links DESC LIMIT 10
MATCH (n2)-[]-(n1)
RETURN distinct n1.name as Artista, links

```

Lo destacable en este caso es que la gran mayoría de los Artistas en este listado son de habla hispana, y los restantes son Directores de orquesta.

5.Cuál es la distancia máxima entre dos artistas uruguayos: 6, efectivamente pudimos comprobar que incluso en Uruguay, la mayor distancia entre dos Artistas se da con la banda Motosierra y alguna de los Artistas de la componente fuertemente conexas uruguayas que se logró encontrar, por ejemplo Buitres. Es necesario realizar hasta 6 saltos entre Artistas para encontrar una relación entre estas, un dato

curioso considerando que ambos fueron contemporáneos, pero una relación que como vimos en otros no es muy significativa.

V-A. Otros datos curiosos

V-A1. Himno uruguayo: Como se anticipaba en la sección anterior, la iniciativa estaba en colaborar con el dataset con todo lo que sea posible. Es en ese sentido, que en la investigación inicial se encontró en numerosos sitios que el himno uruguayo estaba basado en una ópera de Donizetti llamada "Lucrezia Borgia" y en la obertura "Una italiana en Argel" de Rossini. Al escuchar la obra, es notoria la similitud del himno nacional, tanto en melodía, como instrumentos, tiempo, etc., en reiteradas ocasiones, principalmente al finalizar el prólogo, con el fragmento "Tiranos temblad" de la melodía patria. Sin embargo, luego de intentar hacer una solicitud para agregar la relación *sample* entre estas dos canciones, la solicitud fue rechazada arbitrariamente y sin lugar a réplica. Luego de seguir investigando sobre el tema, el asunto se trató académicamente en un artículo¹⁰ de la Facultad de Humanidades de la Universidad de la República, cuyos autores no están identificados en el artículo, pero se menciona que no hay evidencia suficiente para asumir que haya una relación estrecha entre ambas piezas musicales.

V-A2. Remixes: Curiosamente, y como se adelantó al comienzo de este informe, los remixes son un tipo de relación entre Artistas distinta a los covers, y es un tipo especial de relación que consiste en producir un tema usando partes del tema original para hacer una versión nueva. Sin embargo, en todo el subset de datos que corresponden a los artistas uruguayos, no existe ninguna relación de este tipo.

VI. CONCLUSIONES Y TRABAJO FUTURO

VI-A. Conclusiones

Lo primero a destacar es lo acertado de la decisión de utilizar Python y en particular Scrapy como el framework para descargar el contenido del sitio, que no requiere de una compleja configuración, salvo declarar un virtualenv para independizar dependencias con el resto de la configuración ya existente en los entornos de ejecución mencionados en el capítulo dos. Como se puede ver en el código adjunto a este informe, es realmente muy sencillo utilizar la biblioteca Scrapy, ya que provee un SDK que configura lo esencial para poder correr el pipeline de extracción de los datos, y la dificultad de paralelizar o programar la próxima ejecución es prácticamente trivial, así como la configuración para setear cookies, pasos realmente sencillos y con una curva de aprendizaje muy baja. Por otra parte, haber realizado un estudio exhaustivo de las alternativas que habían para generar el dataset primario, y lo aprendido en el curso, hizo muy sencilla la decisión de utilizar MongoDB como destino primario de la información extraída. Nuevamente, al utilizar Python + PyMongo, la biblioteca por defecto para interactuar con MongoDB permitió que los datos fueran cargados rápidamente al servicio MongoDB Atlas, con muy poca latencia y buen desempeño permitiendo descargar aproximadamente un 15 % de la información disponible en el

¹⁰<https://www.fhuce.edu.uy/images/SADIL/Coriun%20Aharonian.pdf>

sitio¹¹ (particularmente todos los registros que se necesitaban para el alcance fijado en esta investigación) en base a los datos que se mencionaron en el capítulo diseño y carga de datos. Así mismo, mencionar que la herramienta Neo4j funciona excelentemente, siendo capaz de almacenar todo el dataset sin mayores inconvenientes, e incluso mejorar la performance considerablemente al agregar índices. Neo4j hace el proceso de realizar consultas muy sencillo y permite operaciones que sería muy difícil realizar en una base de datos SQL. Un problema que habíamos parcialmente identificado en la etapa de análisis antes de comenzar el proyecto, es que el límite de los planes gratis tanto de Mongo Atlas como de Neo4j Aura fue alcanzado en el segundo día de trabajar con el parser y el pipeline posterior de carga de datos, por lo cual se debió recurrir a utilizar algo que no presentara esta restricción, dado que tener los datos por separado dificultaba considerablemente el análisis de los mismos y la ejecución de queries. La decisión de utilizar AWS como servidor centralizado fue muy acertada sumado a que el proyecto fue dockerizado en su mayoría facilitó la tarea de migración entre entornos.

Un aspecto fundamental a destacar en este proyecto de investigación es el sitio web utilizado. WhoSampled presenta información muy útil, de calidad y que los grandes sellos de música y estudios cinematográficos utilicen su información es realmente sorprendente. En ningún momento se identificaron fallas o bajas de servicio lo cual ayudó en el timing de la investigación. Además, el proceso de agregar una nueva entrada con información es realmente simple, tarea que el realizamos un total de 12 veces, contribuyendo y de algún modo pagando la deuda por la extracción de información que se realizó en la primer etapa. Un punto a mejorar es el uso de la API que tienen disponible. Sin embargo no hubo respuesta a los reiterados mensajes que se enviaron con el objetivo de utilizar este servicio para fines académicos, incluso cuando en la página oficial de ellos destacan la cantidad de artículos académicos que se escriben con la información disponible en su web.

Respecto a las métricas y las hipótesis que se habían considerado al comienzo de la investigación, una vez disponible la base de datos Neo4j, el proceso de crear las consultas y extraer los resultados fue trivial, siendo capaces de responder las preguntas planteadas.

VI-B. Trabajo futuro

Sin duda que uno de los problemas más grandes que se enfrentaron en esta investigación fue poder identificar en el sitio el subconjunto de todas las bandas uruguayas, para lo cual fue necesario bastante investigación previa, para alimentar el parser y tener datos con los cuales comenzar a trabajar. Si el sitio guardara la información del país del Artista, no habría tanto lugar para que sucedan situaciones como las detalladas anteriormente de las bandas Chocolate y La Furia.

Además, tener la posibilidad de filtrar Artistas por su país de origen permitiría eventualmente extender el análisis realizado

en este informe para otros países, sin mayor dificultad, y podría ser fácilmente parametrizable y un poco más dinámico.

Como se comentaba en la sección anterior, la base de datos extraída fue capaz de responder todas las preguntas establecidas al comienzo. Sin embargo, el hecho de que los artistas uruguayos no presenten un elevado número de enlaces con otros artistas facilitó la tarea. Si extendemos el alcance y toca estudiar Artistas con canciones masivamente sampleadas o hechas cover, whosampled no tiene una forma muy amigable de mostrar el listado completo de bandas, esto se traduce en procesar tres pasos desde la vista de el Artista, hasta llegar al listado de canciones relacionadas. Nuevamente recalcar que esto no fue necesario, incluso en el caso de la canción uruguaya con más enlaces, pero es algo a considerar si se quiere estudiar más países.

Como último punto a destacar respecto a proveer funcionalidad sobre países es la posibilidad de identificar qué Artistas de un determinado país tienen más influencia en los Artistas de otro país, y luego hacer un segundo grado de influencia, que podría ser tanto el género del grupo así como el género de los covers de otras canciones. Respecto a esto último, comentar que no aplica a nuestro país, ya que los géneros de las bandas uruguayas que se lograron identificar son en su mayoría son encasillados en el mismo segmento de música "popular/rock". Sería interesante poder clasificar la música de Uruguay en los géneros más comunes que existen en el país, como cumbia, murga o tango, por mencionar algunos.

Algunas de las bandas más famosas del sitio cuentan con links a sitios externos, como Wikipedia, o redes sociales. Los integrantes del equipo que llevaron adelante esta investigación no tuvieron acceso en ningún momento a agregar dicha información, por lo que puede ser una característica *premium*¹² para determinados usuarios. Es así que un buen punto a tratar en futuras iteraciones sobre este proyecto sería tener la posibilidad de agregar esta metadata de las bandas y poder sacar métricas respecto a esto, como extender el parser a redes sociales y estudiar sus seguidores.

Relacionado con extender el parser, con funcionalidad actualmente provista por el sitio es la posibilidad de extraer el año de las canciones tanto la original como su sample, cover o remix, esta información es accesible únicamente en la página web del link entre ambas versiones, por lo que hace inviable en el tiempo que se disponía para realizar esta investigación, la tarea de parsear este dato extra y luego procesarlo, siendo así algo que también se puede tener en cuenta para incorporar a futuro.

En la etapa de análisis, se logró identificar un tipo de relación que en principio no era trivial como las que se mencionaron al comienzo del documento. El sitio web tiene la posibilidad de agregar quiénes fueron los productores de una determinada obra, en este caso, al ingresar al detalle de la relación, los productores no aparecerán como los artistas al comienzo del listado sino que tienen su propia sección. Un posible trabajo futuro puede ser estudiar los diversos productores y la relación que mantienen con los Artistas, a modo de ejemplo podemos mencionar a Juan Campodónico

¹¹Los detalles discriminados por artista, canciones, etc. Se encuentran en: <https://www.whosampled.com/about/>

¹²<https://www.whosampled.com/subscription/premium/>

en la versión "Livin' la vida loca" que salió en la película Shrek ¹³.

Por último, remarcar que toda la información que el equipo de investigación utilizó para estudiar y obtener métricas, fue extraída en un total de 6 días de ejecución del scraper, y eso representó aproximadamente un 15 % de la información total del sitio. En caso de tener más días para realizar la investigación sería interesante dejar ejecutando los pipeline de extracción un tiempo extra (extrapolando con la información que tenemos ahora, sería algo así como 2 meses ejecutando 24/7) eso puede dar una idea de el volumen de datos que maneja la web, así como métricas interesantes que no fuimos capaces de visualizar con los datos que se han obtenido dentro del alcance que se había establecido.

REFERENCIAS

- [BRITO(2019)] LUCAS ALVES DE BRITO. SAMPLE HUNTERS: Uma netnografia sobre as práticas comunicacionais da subcultura dos fãs da música sampleada nas comunidades virtuais WhoSampled e Sample Spotters., 2019. URL <http://repositorio.bc.ufg.br/tede/handle/tede/9416>.
- [Bryan and Wang(2011)] Nicholas J Bryan and Gen Wang. Musical influence network analysis and rank of sample-based music. In *ISMIR*, pages 329–334, 2011.
- [de BRITO(2018)] Lucas Alves de BRITO. A subcultura digital dos sample hunters: uma análise dos perfis no site whosampled, 2018.
- [Tran(2019)] Justin Tran. Measuring musical sampling impact through network analysis. 2019.

¹³<https://www.whosampled.com/Eddie-Murphy/Livin%27-La-Vida-Loca/>