Identificación de platillos de comida en base a ingredientes empleando Spark

Angel Vidal Larreategui Castro¹

¹Universidad Nacional de Ingeniería. Facultad de Ciencias. Ciencias de la Computación

Agosto, 2023



Tabla de contenidos

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- Trabajos futuros



Contenido

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- Trabajos futuros



Introducción

En la gastronomía peruana se puede encontrar una diversidad de platillos entre todas las regiones que conforman a nuestro país.



Figura: Platillos peruanos.



Contenido

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- Trabajos futuros



Problema

La elección sobre qué platillo preparar en base a lo que se tenga en la cocina es un problema recurrente en muchos peruanos.



Figura: ¿Qué puedo cocinar hoy?



Contenido

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- Trabajos futuros



Objetivos

Objetivo general

 Realizar la inferencia de platillos de comida en base a ingredientes para poder contextualizarlo y aplicarlo en el campo de la gastronomía peruana.

Objetivos específicos

- Revisar técnicas de Big Data.
- Revisar el funcionamiento y procesamiento de datos de Spark.
- Mejorar el diseño de una ontología relacionada a la gastronomía peruana.



Contenido

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- Trabajos futuros



Comidas de la región costera del Perú

Perú posee una gran diversidad gastronómica.



Figura: Platillos de la costa peruana



Web scraping



Figura: Proceso de extracción de la información de un sitio web.



Apify



Figura: Apify es una forma más inteligente de extraer información de sitios web.



Ontologías

- Se puede definir una ontología como como una definición formal de propiedades, relaciones entre entidades que van a existir para un determinado dominio en particular.
- Comprende un esquema semántico de clases, propiedades y diversos tipos de relaciones que se encuentren, además de una taxonomía de conceptos, clases y también de ciertos individuos específicos.



Contenido

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- 8 Trabajos futuros



Desarrollo

El proceso de llegó a dividir en 3 partes.

- 1. Extracción de la información.
- 2. Procesamiento de la data en Spark.
- 3. Elección de conocimientos y construcción de la ontología.



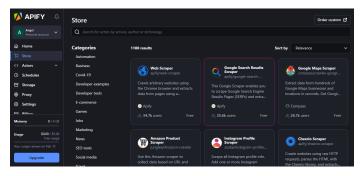


Figura: Interfaz de Apify, donde se muestran los diversos bots que aloja dicha plataforma.



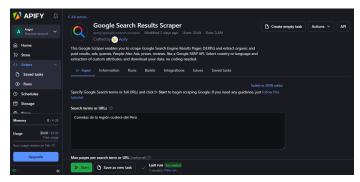


Figura: Interfaz del bot Google Search Results Scraper.



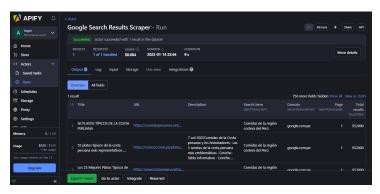


Figura: Resultados que muestra el bot al haberle solicitado una búsqueda.



55 platos típicos de la costa peruana más representativos según la historia (y sus recetas)





Figura: Página elegida para poder extraer la data de las comidas costeras peruanas.



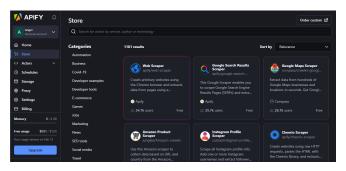


Figura: Bot Web Scraper para poder extraer la data del sitio web elegido.



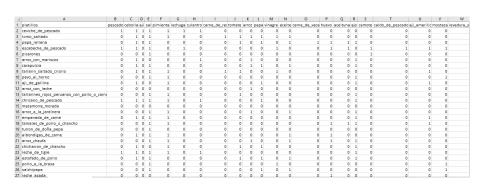


Figura: Visualización de la data obtenida en Excel.



Procesamiento de la data en Spark

```
val ingredientCount = readLine("Ingrese la cantidad de ingredientes: ").toInt

val ingredientNames = (1 to ingredientCount).map { _ => readLine("Ingrese el nombre de un ingrediente: ")}

val columns = ingredientNames.map(column => s""""$column"""").mkString(" AND ")

val query = s"SELECT platillos FROM peruvian_dishes WHERE $columns"

val jdbcDF = spark.read.format("jdbc").
    option("url", "jdbc:postgresql:postgres").
    option("user", "postgres").
    option("password", "021299").
    option("dbtable", s"($query) AS t").
    load()

jdbcDF.select("platillos").show()
```

Figura: Procesamiento de la data obtenida en Spark.



En primer lugar, se hará la elección del conocimiento **Platillo** y realizará su construcción dentro de la presente ontología mostrada en lenguaje OWL.

```
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#Platillos">
    <rdfs:comment xml:lang="es">Clase que representa un platillo.</rdfs:comment>
    <rdfs:label xml:lang="es">Platillo</rdfs:label>
</owl:Class>
```

Figura: Conocimiento Platillo dentro de la ontología en lenguaje OWL.



```
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#Segundo">
   <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Platillos"/>
   <rdfs:comment xml:lang="es">Clase que representa un platillo de segundo.</rdfs:comment>
   <rdfs:label xml:lang="es">Segundo</rdfs:label>
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#SegundoMaritimo">
   <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Segundo"/>
   <rdfs:comment xml:lang="es">Clase que representa un segundo marítimo.</rdfs:comment>
   <rdfs:label xml:lang="es">SegundoMaritimo</rdfs:label>
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#SegundoNoMaritimo">
   <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Segundo"/>
   <rdfs:comment xml:lang="es">Clase que representa un segundo no marítimo./rdfs:comment>
   <rdfs:label xml:lang="es">SegundoNoMaritimo</rdfs:label>
```

Figura: Clase Segundo perteneciente a Platillo

```
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#Sopa">
   <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Platillos"/>
   <rdfs:comment xml:lang="es">Clase que representa un platillo de sopa.</rdfs:comment>
   <rdfs:label xml:lang="es">Sopa</rdfs:label>
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#SopaMaritima">
   <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Sopa"/>
   <rdfs:comment xml:lang="es">Clase que representa una sopa marítima.</rdfs:comment>
   <rdfs:label xml:lang="es">SopaMaritima</rdfs:label>
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#SopaNoMaritima">
   <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Sopa"/>
   <rdfs:comment xml:lang="es">Clase que representa una sopa no marítima.
   <rdfs:label xml:lang="es">SopaNoMaritima</rdfs:label>
```

Figura: Clase **Sopa** perteneciente a **Platillo**.



De manera análoga, también se construyen las clases **Postres** y **Refrescos**. El conocimiento **Ingredientes** también es necesario dentro de la ontología pues a partir de ellos se realizan los platillos.

Figura: Conocimiento Ingredientes dentro de la ontología en lenguaje OWL.



```
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#Carnes">
    <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Ingredientes"/>
    <rdfs;comment xml:lang="es">Clase que representa carnes.</rdfs;comment>
    <rdfs:label xml:lang="es">Carnes</rdfs:label>
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#Granos">
    <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Ingredientes"/>
    <rdfs:comment xml:lang="es">Clase que representa granos.</rdfs:comment>
    <rdfs:label xml:lang="es">Granos</rdfs:label>
<owl:Class rdf:about="http://www.semanticweb.org/usuario/smartcooking#Aderezos">
    <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/usuario/smartcooking#Ingredientes"/>
    <rdfs:comment xml:lang="es">Clase que representa aderezos.</rdfs:comment>
    <rdfs:label xml:lang="es">Aderezos</rdfs:label>
```

Figura: Clases Carnes, Granos, Aderezos pertenecientes al conocimiento Ingredientes.

CIENCIAS

27 / 47

Es importante también definir los individuos pertenecientes a cada clase que se ha definido.

Figura: Ejemplo de como se define un individuo dentro de la ontología, en este caso siendo el individuo **Agua**.



Es necesario también definir una propiedad que pueda relacionar los conocimientos **Platillo** e **Ingredientes**.

Figura: Propiedad **contieneIngredientes** que relaciona los conocimientos **Platillos** e **Ingredientes**.



Para apreciar como funciona dicha propiedad, se define un individuo perteneciente a **Platillos**.

Figura: Individuo arroz_con_mariscos empleando la propiedad contieneIngredientes para relacionarse con sus ingredientes.



Contenido

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- 8 Trabajos futuros



```
scala> 23/07/20 15:55:51 WARN ProcfsMetricsGetter: Exception when trying to compute pagesize, as a result reporting of ProcessTree metrics is stopped val ingredientCount = readLine("Ingrese la cantidad de ingredientes: ").toInt warning: there was one deprecation warning (since 2.11.0); for details, enable `:set ting -deprecation' or `:replay -deprecation'
Ingrese la cantidad de ingredientes: ingredientCount: Int = 4
```

Figura: Obtención de la cantidad de ingredientes deseados por el usuario.



```
scala> val ingredientNames = (1 to ingredientCount).map { _ => readLine("Ingrese el nombre de un ingrediente: ")}
warning: there was one deprecation warning (since 2.11.0); for details, enable `:set ting -deprecation' or `:replay -deprecation'
Ingrese el nombre de un ingrediente: Ingrese el nombre de un ingrediente: Ingrese el nombre de un ingrediente: Ingrese el accollection.immutable.IndexedSeq[String] = Vector(cebolla, sal, pimienta, aji)
```

Figura: Obtención de los nombres de los ingredientes ingresados por el usuario.



```
scala> val columns = ingredientNames.map(column => s"""$column""").mkString(" AND
")
columns: String = "cebolla" AND "sal" AND "pimienta" AND "aji"
```

Figura: Generación de lista de columnas para la consulta SQL.



```
scala> val query = s"SELECT platillos FROM peruvian_dishes WHERE $columns"
query: String = SELECT platillos FROM peruvian_dishes WHERE "cebolla" AND "sal" AND
"pimienta" AND "aji"
```

Figura: Construcción de la consulta SQL.





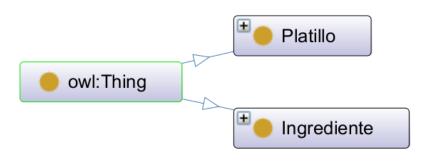


Figura: Clases Platillo e Ingrediente.



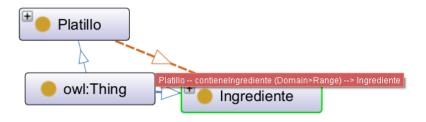


Figura: Propiedad *contieneIngrediente* visible entre las clases Platillo e Ingrediente.



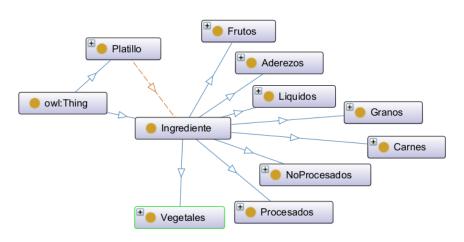


Figura: Clases y subclases de la ontología.





Figura: Inferencia de platillos a partir de ciertos ingredientes.



```
warning: there was one deprecation warning (since 2.11.0); for details, enable `:setting -deprecati
on' or `:replay -deprecation'
Ingrese el nombre de un ingrediente: Ingrese el nombre de un ingrediente: Ingrese el nombre de un i
ngrediente: ingredientNames: scala.collection.immutable.IndexedSeq[String] = Vector(cebolla, papa,
pan)
scala> val columns = ingredientNames.map(column => s"""$column"""").mkString(" AND ")
columns: String = "cebolla" AND "papa" AND "pan"
scala> val query = s"SELECT platillos FROM peruvian dish<u>es WHERE $columns"</u>
query: String = SELECT platillos FROM peruvian dishes WHERE "cebolla" AND "papa" AND "pan"
scala> val jdbcDF = spark.read.format("jdbc").
        option("url", "jdbc:postgresql:postgres").
        option("user", "postgres").
        option("password", "021299").
         option("dbtable", s"($query) AS t").
        load()
dbcDF: org.apache.spark.sql.DataFrame = [platillos: string]
scala> idbcDF.select("platillos").show()
     platillos
aji_de_gallina
```

Figura: Inferencia con 3 ingredientes.





Figura: Consulta dentro de la ontología realizada en Protege.



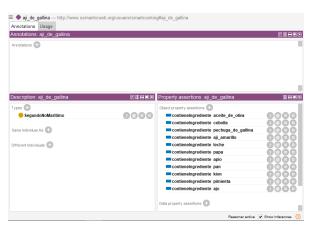


Figura: Individuo ají_de_gallina y la propiedad contieneIngrediente.

Contenido

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- 8 Trabajos futuros



Conclusiones

- Se realizó la inferencia de platillos de comida costeños peruanos utilizando Spark y la reconstrucción de la ontología de un sistema de cocina inteligente, ampliando su aplicabilidad en la gastronomía peruana.
- Se revisó correctamente el funcionamiento y procesamiento de datos en Spark relacionados con la información de platillos costeños peruanos.
- Se comprendió el funcionamiento de Spark en la inferencia y procesamiento de datos para realizar consultas relacionadas con la gastronomía peruana. logró significativamente mejorar la ontología de un sistema de cocina inteligente, permitiendo consultas exitosas sobre platillos costeños peruanos y su integración con una aplicación móvil mediante la API de OWL.



Contenido

- Introducción
- 2 Problema
- Objetivos
- 4 Conocimientos previos
- Metodología y desarrollo
- 6 Resultados
- Conclusiones
- Trabajos futuros



Trabajos futuros

- La ontología mostrada puede también expandirse al campo de la salud, pues se podrían agregar clases relacionadas al valor nutricional de cada alimento para que se pueda tener conocimiento del valor nutricional de cada platillo.
- La revisión de la documentación de la API de owl para que la ontología desarrollada en Protege pueda conectarse de forma correcta con la aplicación Android.

