



Programación Orientada a Objetos con Java

Paradigmas de la programación 2025



¿Qué vamos a desarrollar?

- Una aplicación de consola en Java.
- Debe parsear feeds de noticias desde distintos sitios.
- También mostrarlos en pantalla de forma clara y legible.
- Además detectar entidades nombradas con heurísticas.

¿Qué es un feed rss?

- Es un formato XML para distribuir contenido en la web.
- Se utiliza para difundir información actualizada frecuentemente a usuarios que se han suscrito a la fuente de contenidos.

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <rss xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:media="http://search.yahoo.com/mrss/" xmlns:atom="http://www.w3.org/2005/Atom" xmlns:nyt="http://www.nytimes.com/namespaces/rss"
3 <channel>
4   <title>NYT &gt; Business</title>
5   <link>https://www.nytimes.com/section/business</link>
6   <atom:link href="https://rss.nytimes.com/services/xml/rss/nyt/Business.xml" rel="self" type="application/rss+xml"></atom:link>
7   <description></description>
8   <language>en-us</language>
9   <copyright>Copyright 2025 The New York Times Company</copyright>
10  <lastBuildDate>Thu, 10 Apr 2025 01:51:38 +0000</lastBuildDate>
11  <pubDate>Thu, 10 Apr 2025 01:07:27 +0000</pubDate>
12  <image>
13    <title>NYT &gt; Business</title>
14    <url>https://static01.nyt.com/images/misc/nyt_logo_rss_250x40.png</url>
15    <link>https://www.nytimes.com/section/business</link>
16  </image>
17  <item>
18    <title>Trump Reverses Course on Global Tariffs, Announcing 90-Day Pause</title>
19    <link>https://www.nytimes.com/2025/04/09/us/politics/trump-tariffs-stocks-china.html</link>
20    <guid isPermaLink="true">https://www.nytimes.com/2025/04/09/us/politics/trump-tariffs-stocks-china.html</guid>
21    <atom:link href="https://www.nytimes.com/2025/04/09/us/politics/trump-tariffs-stocks-china.html" rel="standout"></atom:link>
22    <description>The president further raised already steep tariffs on China, saying that Beijing should not have retaliated against his earlier trade actions.</description>
23    <dc:creator>Ana Swanson and Tony Romm</dc:creator>
24    <pubDate>Wed, 09 Apr 2025 23:57:46 +0000</pubDate>
25    <category domain="http://www.nytimes.com/namespaces/keywords/des">Customs (Tariff)</category>
26    <category domain="http://www.nytimes.com/namespaces/keywords/des">International Trade and World Market</category>
27    <category domain="http://www.nytimes.com/namespaces/keywords/des">United States Economy</category>
28    <category domain="http://www.nytimes.com/namespaces/keywords/des">United States International Relations</category>
29    <category domain="http://www.nytimes.com/namespaces/keywords/des">United States Politics and Government</category>
30    <category domain="http://www.nytimes.com/namespaces/keywords/nyt_per">Trump, Donald J</category>
31    <category domain="http://www.nytimes.com/namespaces/keywords/nyt_per">Bessent, Scott</category>
32    <category domain="http://www.nytimes.com/namespaces/keywords/nyt_geo">China</category>
33    <media:content height="1801" medium="image" url="https://static01.nyt.com/images/2025/04/09/multimedia/09markets-tariffs-pause-hzwk/09markets-tariffs-pause-hzwk-mediumSquareAt3
34    <media:credit>Eric Lee/The New York Times</media:credit>
35  </item>
```


Orden sugerido de trabajo

1-Leer un archivo de suscripciones JSON

módulo subscription
módulo parser

- Parsear el archivo JSON subscriptions.json ubicado en config.
- Agregamos las URLs a otro objeto de tipo SingleSubscription (en representación de una sola suscripción)
- Agregar cada una de ellas a una lista de subscriptions definida por la misma clase Subscription.

Orden sugerido de trabajo

2-Realizar petición HTTP al servidor del feed

módulo httpRequester

- Tomar la URL del feed que se parseo en el punto 1.
- Establecer la conexión HTTP.
- Obtener el contenido del feed en una cadena de texto.

Orden sugerido de trabajo

3-Parsear el archivo XML

módulo feed
módulo parser

- Obtener los atributos de cada <item>.
- Solo necesitaremos
 - Title
 - Description
 - PubDate
 - link

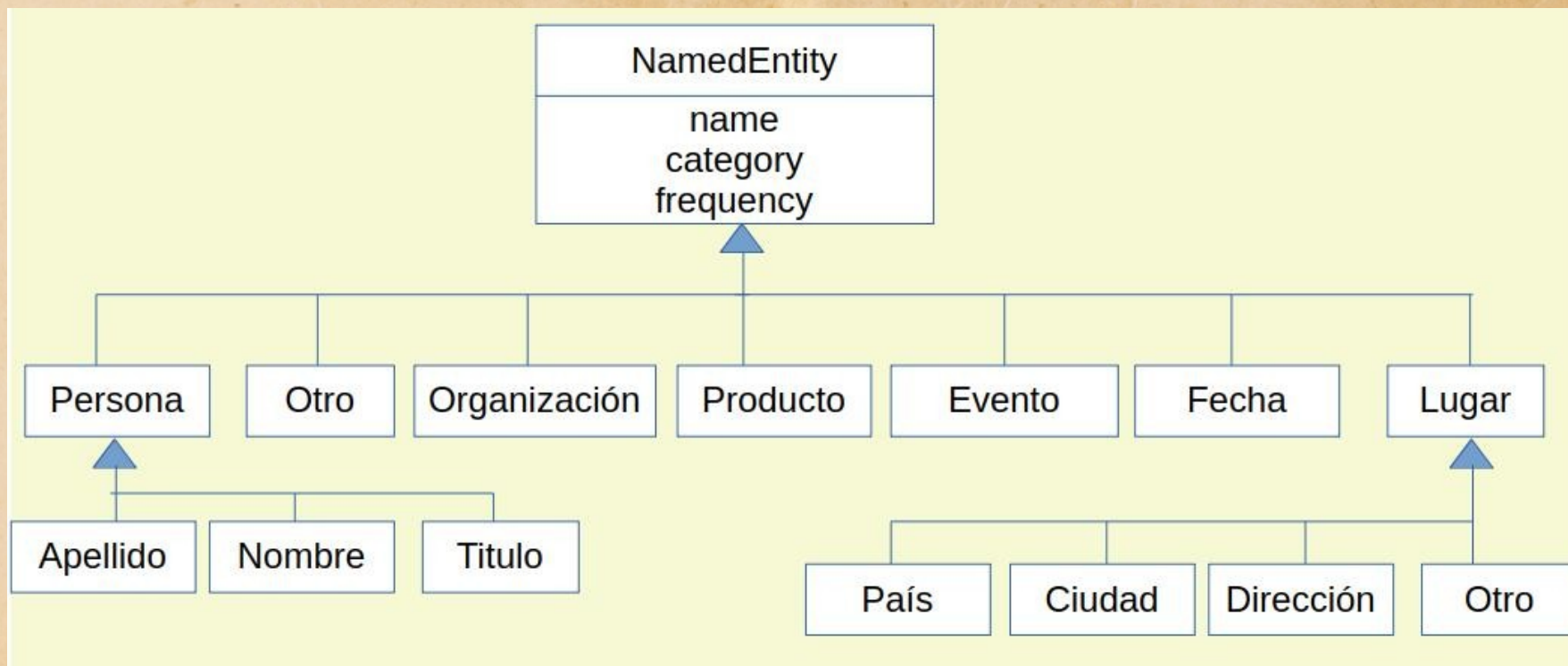
Orden sugerido de trabajo

4-Imprimir por pantalla los artículos del feed

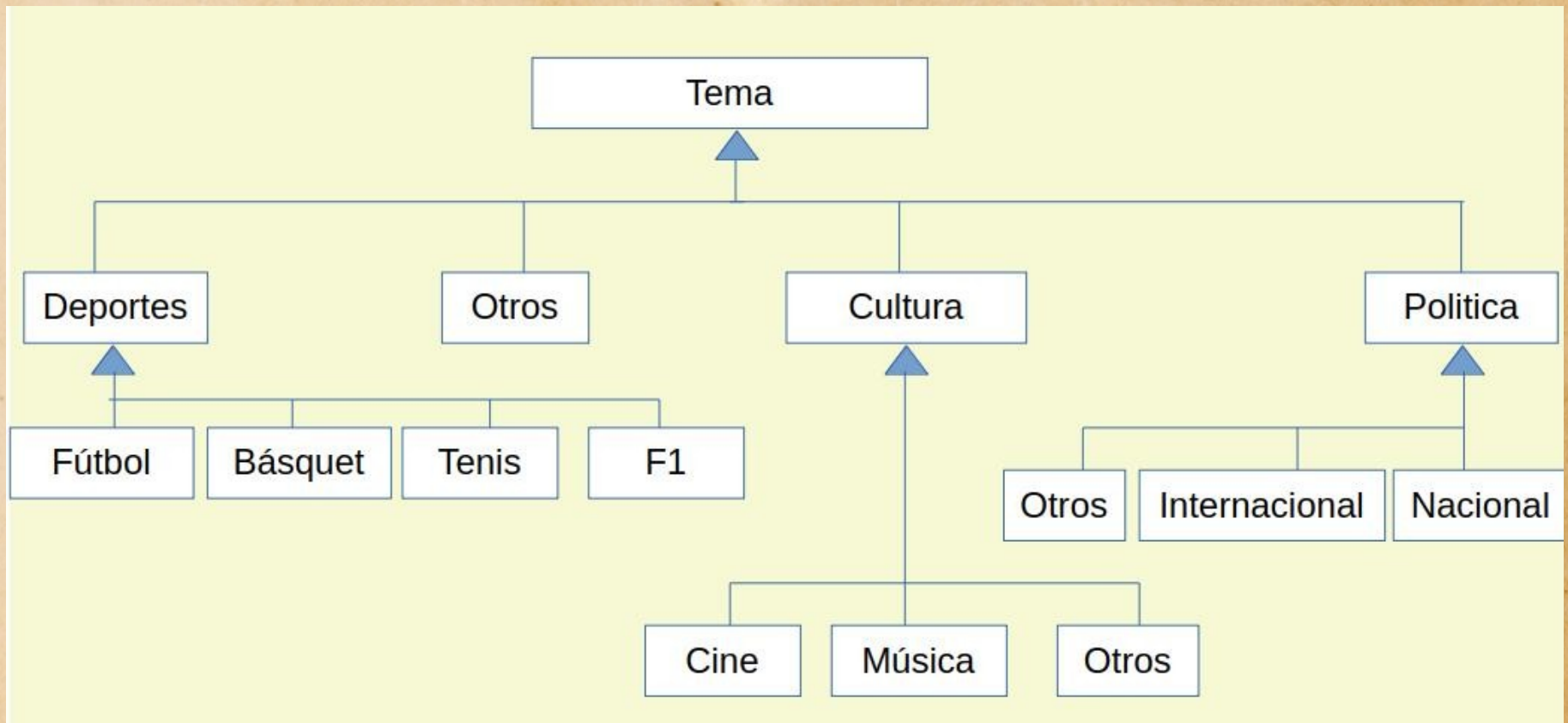
módulo NamedEntity
módulo feed

- Si no se pasan parametros solo se imprime el articulo
- Si se pasa un parametro (podria ser -ne) se debe contar las ocurrencias de las entidades nombradas en los textos de cada artículo.
- Se dan dos heurísticas:
 - RandomHeuristic: Utiliza palabras random para identificar si una palabra es entidad.
 - QuickHeuristic: Tiene un mapa de palabras que no son entidades, las demás sí lo serán.

Jerarquía de NamedEntity



Jerarquía de Tema



Preguntas