



Graduado en Ingeniería de la Salud

Resistencia a antibióticos en Europa

Antimicrobial Resistance in Europe

Realizado por

Alejandro Pascual Mellado David Cubillos del Toro David Rámirez Arco Nerea Martín Serrano

Tutorizado por

Laura Panizo Jaime





ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA GRADUADO EN INGENIERÍA DE LA SALUD

Resistencia a antibióticos en Europa Antimicrobial Resistance in Europe

Realizado por
Alejandro Pascual Mellado
David Cubillos del Toro
David Rámirez Arco
Nerea Martín Serrano

Tutorizado por **Laura Panizo Jaime**

UNIVERSIDAD DE MÁLAGA

Resumen

En este proyecto se va a desarrollar una página web que incluya información sobre bacterias con resistencia a los antibióticos en Europa.

Se va a comenzar por extraer información de un dataset (en formato XML).

A este archivo se le aplicarán transformaciones XSLT para poder integrar la

información en una página web. Esta página web será diseñada para que los

distintos usuarios puedan visualizar de forma sencillos los datos recogidos.

Para ello se van a crear dos interfaces, una dirigida a investigadores y otra

dirigida a usuarios que simplemente quieran informarse sobre el tema.

Además, se realizarán posibles consultas que puedan realizar los usuarios

sobre el dataset en lenguajes como XQuery y SPARQL.

El proyecto se aloja en el siguiente repositorio de gitHub:

github.com/Antimicrobial-Resistance-in-Europe.

Y la implementacion de la páguina web final está hecha en GitHub Pages,

y se puede acceder con el siguiente enlace:

Archerd6.github.io/Proyecto Estandares de Datos Web.

[a desarrollar conforme se vaya realizando el proyecto]

Palabras clave: Bacteria, Antibiótico, Base de datos, Página web,

HTML, XSLT, XML

2

Índice

1.	Introducción		
	1.1.	Motivación	4
	1.2.	Objetivos	4
	1.3.	Tecnologías usadas	4
2.	Desc	cripción del Dataset	5
3.	Diseño XQuery		
	3.1.	Query 1	7
	3.2.	Query 2	8
4.	Página Web		
	4.1.	Interfaz inicio	9
	4.2.	Interfaz investigadores	10
	4.3.	Interfaz usuarios	11
5.	Integración de datos		
	5.1.	Transformación tabla bacterias	12
	5.2.	Transformación tabla grupos de estudio	13
	5.3.	Transformaciones de las tablas del mapa	14
6.	Grafo RDF		
	6.1.	Clases RDF	15
	6.2.	Propiedades RDF	16
	6.3.	Queries SPARQL	17
		6.3.1. Query 1	17
		6.3.2. Query 2	17
7.	. Conclusiones		18

Introducción

1.1. Motivación

Un problema emergente es la resistencia de los antibióticos que han desarrollado algunas bacterias. Esta resistencia ocurre cuando las bacterias ya no se eliminan con los antibióticos creados con ese objetivo. Esto es un grave problema pues las bacterias pueden causar infecciones con graves consecuencias, y la manera de curar estas infecciones es eliminar la bacteria que lo causa.

Se quiere crear un página web para que tanto el personal médico, como investigadores y otros usuarios puedan tener un listado de las bacterias que hayan desarrollado esta resistencia a antibióticos. Así los usuarios podrán intentar evitar el contacto con estas bacterias, el personal médico podrá saber si existe tratamiento para la infección bacteriana que estén tratando...

1.2. Objetivos

El objetivo es tener una página web donde los usuarios puedan mantenerse informados acerca de este tipo de bacterias. Por un lado, informando a los usuarios sobre en que zonas de Europa se encuentran este tipo de bacterias, y por otro lado, ampliando la información de los investigadores que estén llevando a cabo un estudio relacionado con el tema.

1.3. Tecnologías usadas

Las tecnologías que se van a usar a lo largo de este proyecto son: Visutal Studio, Ł̃TŁX, Protégé, eXist-db

Descripción del Dataset

El dataset elegido guarda información recogida de la base de datos European Antimicrobial Resistance Surveillance (EARS), en concreto la información relativa a aquellas bacterias que han desarrollado resistencia a los antiobióticos. El dataset se encuentra en el siguiente enlace: kaggle.com/datasets/euro-resistance.

Cada una de las filas del dataset representa un estudio realizado sobre un grupo de personas. En las columnas se nos presenta la siguiente información sobre cada estudio:

- Distribución: se refiere a si los datos se extrajeron de un género o grupo de edad en particular.
- RegionName: país de la institución informante.
- Tiempo: el año.
- Categoría: grupo de edad o el género (dependiendo de la distribución).
- Valor: porcentaje de bacterias que fueron resistentes al grupo de antibióticos.
- Bacterias: se refiere a las bacterias en las que se estudió la resistencia.
- Antibiótico: grupo de antibióticos que se usó para matar la bacteria.

En la imagen 1 se encuentra el dataset en formato .XMLSpy (tal cual se descarga de Kaggle).

Se necesita que los datos estén en formato XML para poder realizar las distintas consultas y transformaciones. XML es un metalenguaje de marcado. Un lenguaje de marcado o lenguaje de marcas es una forma de codificar un documento que, junto con el texto,

```
, Distribution, Unit, Time, RegionCode, RegionName, Category, Value, Bacteria, Antibiotic
0, "R - resistant isolates proportion, by age", \%, 2012, BG, Bulgaria, 0-4, 0.0, Acine to bacter spp., Aminogly cosides and the substantial of t
1, "R - resistant isolates proportion, by age", %, 2012, BG, Bulgaria, 5-18, 100.0, Acinetobacter spp., Aminoglycosides
2, "R - resistant isolates proportion, by age",%,2012,BG,Bulgaria,19-64,62.857142857,Acinetobacter spp.,Aminoglycosides
3,"R - resistant isolates proportion, by age",%,2012,BG,Bulgaria,65+,56.52173913,Acinetobacter spp.,Aminoglycosides
4, "R - resistant isolates proportion, by age", %, 2012, CY, Cyprus, 0-4, 33.33333333, Acinetobacter spp., Aminoglycosides
5, "R - resistant isolates proportion, by age",%,2012,CY,Cyprus,5-18,0.0,Acinetobacter spp.,Aminoglycosides
6, "R - resistant isolates proportion, by age", %, 2012, CY, Cyprus, 19-64, 57.142857143, Acinetobacter spp., Aminoglycosides
7, "R - resistant isolates proportion, by age", %, 2012, CY, Cyprus, 65+, 62.5, Acinetobacter spp., Aminoglycosides
8, "R - resistant isolates proportion, by age",%,2012,DE,Germany,0-4,0.0,Acinetobacter spp.,Aminoglycosides
9, "R - resistant isolates proportion, by age",%,2012,DE,Germany,5-18,0.0,Acinetobacter spp.,Aminoglycosides
10, "R - resistant isolates proportion, by age", %, 2012, DE, Germany, 19-64, 9.090909091, Acinetobacter spp., Aminoglycosides
11, "R - resistant isolates proportion, by age", %, 2012, DE, Germany, 65+, 5.479452055, Acinetobacter spp., Aminoglycosides
12, "R - resistant isolates proportion, by age", %, 2012, DK, Denmark, 0-4, 0.0, Acinetobacter spp., Aminoglycosides
13, "R - resistant isolates proportion, by age", %, 2012, DK, Denmark, 5-18, 16.666666667, Acinetobacter spp., Aminoglycosides
14, "R - resistant isolates proportion, by age", %, 2012, DK, Denmark, 19-64, 15.384615385, Acinetobacter spp., Aminoglycosides
15, "R - resistant isolates proportion, by age",%,2012,DK,Denmark,65+,7.142857143,Acinetobacter spp.,Aminoglycosides
16, "R - resistant isolates proportion, by age", %, 2012, FR, France, 0-4, 9.375, Acinetobacter spp., Aminoglycosides
```

Figura 1: Dataset formato .XMLSpy

incorpora etiquetas o marcas que contienen información adicional acerca de la estructura del texto o su presentación Para ello se ha usado la siguiente herramienta online: www.convertcsv.com. Esta herramienta da como salida el archivo .XML de la imagen 2.

Como se puede observar en la imagen, se ha generado un nodo raíz, al que se ha llamado *root*, y por cada fila del dataset se ha generado un nodo *row*, el cual tiene como nodos hijos las columnas del dataset.

```
?xml version="1.0" encoding="UTF-8"?>
<root>
   <Distribution>R - resistant isolates proportion, by age</Distribution>
   <Time>2012</Time>
   <RegionName>Bulgaria</RegionName>
   <Category>0-4</Category>
   <Value>0.0</Value>
   <Bacteria>Acinetobacter spp.
   <Antibiotic>Aminoglycosides
   <Distribution>R - resistant isolates proportion, by age/Distribution>
   <Time>2012</Time>
   <RegionCode>BG</RegionCode>
   <RegionName>Bulgaria</RegionName>
   <Value>100.0</Value>
   <Bacteria>Acinetobacter spp.
   <Antibiotic>Aminoglycosides</Antibiotic>
```

Figura 2: Dataset formato .XML

Diseño XQuery

En esta sección se van a hacer dos posibles consultas que harían distintos usuarios a la base de datos, para lo que se usan las queries o consultas. Una consulta sirve encontrar y extraer elementos y atributos de una base de datos.

Para realizar estas consultas se va a usar XQuery, un lenguaje para realizar consultas en documentos XML. Se basa en funciones, expresiones XPath y predicados.

Las operaciones más comunes para hacer queries son SELECT, que selecciona los datos que serán visibles en la tabla resultado; FROM, para señalar la o las tablas de las cuales quiero obtener la información, WHERE, utilizada para delimitar la información que queremos obtener.

3.1. Query 1

Se procede a implementar una xquery que devuelve el número de estudios de 2012 sobre un antibiótico dado, para ello se declara una variable con el número de estudios que tengan estas características.



Figura 3: XQuery 1

3.2. Query 2

Se procede a implementar una xquery que devuelva el nombre de los antibióticos de estudios de 2014 que tengan un valor superior al 50 por ciento

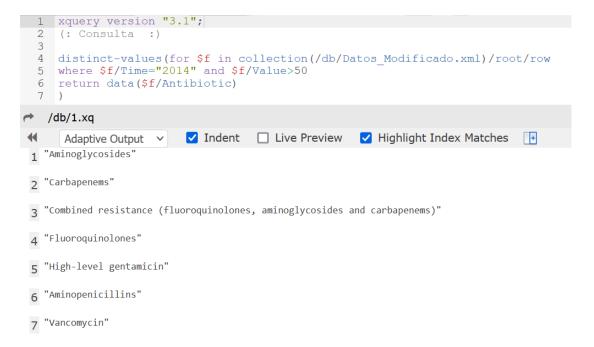


Figura 4: XQuery 2

Página Web

La página Web va a permitir a los usuarios acceder de forma concisa a la información que se quiere presentar. Va a constar dos interfaces, una para investigadores y otra para usuarios que simplemente quieran informarse. Se va añadir una interfaz que se muestre al inicio de la aplicación donde los usuarios elijan que interfaz desean ver.

En la siguiente sección se explica como se van a integrar los datos a la página web, es decir las distintas transformaciones realizadas para poder presentar los datos en las distintas interfaces.

4.1. Interfaz inicio

Esta interfaz va a constar de dos botones, uno que redirija a la interfaz de usuarios y otro a la de investigadores, como se puede comprobar en la imagen 5.



Figura 5: Interfaz Inicio

4.2. Interfaz investigadores

Esta interfaz esta enfocada para personas dedicadas a la investigación. A la izquierda de la interfaz van a encontrarse dos botones y a la derecha una tabla. Al seleccionar uno de los botones de la izquierda (el de la imagen de una bacteria), en la tabla de la derecha se presentará información relacionada con las bacterias, los antibióticos a los que son resistentes y el porcentaje de resistencia de la bacteria en ese estudio (es decir, se filtran de las columnas del dataset aquellas columnas de interés). Esto se refleja en la imagen 6.



Figura 6: Tabla bacterias

En el caso de seleccionar el otro botón en la tabla de la derecha se presentará la información relacionada con los grupos a los que se le realizó el estudio (imagen 7).



Figura 7: Tabla estudios

4.3. Interfaz usuarios

En este interfaz se presenta un mapa de Europa. Delante de este mapa se van a poner una serie de botones por zonas (que incluyan más de un país) para que el usuario pueda hacer click y obtener información sobre aquellas bacterias con resistencia a antibióticos en esa zona de Europa.

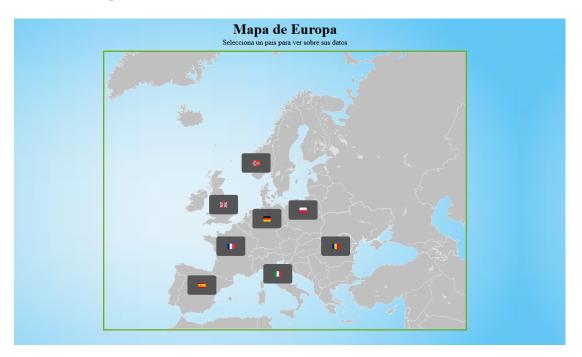


Figura 8: Interfaz usuarios

Integración de datos

En esta sección se va a describir como se van a integrar los datos en la aplicación. Se van a usar transformaciones XSLT, las cuales permiten definir plantillas para transformar un documento XML en otro tipo de documento (HTML en este caso) utilizando sintaxis XML.

5.1. Transformación tabla bacterias

En la interfaz de investigadores, para generar la tabla que aparece al pulsar en la bacteria se ha creado una transformación XSLT.

Al principio de la transformación se ha creado una etiqueta <style> para poder dar formato a la tabla. A continuación, se ha creado la tabla y se han definido el nombre de las columnas. Para poder rellenar la tabla, dentro de cada fila (row) del archivo XSL y se han seleccionado los nodos hijos mencionados anteriormente.

En la imagen 9 se puede ver como se ha usado un for-each para ir recorriendo las filas cuyas bacterias tenga más de un 20 % de resistencia a los antibióticos. Después, se han ordenado las bacterias por mayor resistencia a antibióticos. Por último, se usa un xsl:if para comprobar que los datos no se repiten, pues hay varios estudios donde coinciden los resultados. Si la condición se cumple, se crea una nueva fila en la tabla con os datos seleccionados en los xsl:value-of.

Figura 9: Transformación tabla bacterias

5.2. Transformación tabla grupos de estudio

Se ha realizado otra transformación para el caso en el que el investigador pulse el botón de grupos de estudio. El principio es igual que en el caso anterior, se empieza dándole formato a la tabla y definiendo las columnas. En la imagen 10 se ve como se crean las columnas. Se vuelven a recorrer las distintas filas, esta vez sin condición, y se seleccionan los valores que necesitamos para rellenar la tabla.

Figura 10: Transformación tabla grupos

5.3. Transformaciones de las tablas del mapa

Se ha realizado transformaciones para obtener información sobre aquellas bacterias con resistencia a antibióticos en ese país de Europa

```
h2>Tabla de bacterias</h2></center>
                 Field
                 Distribution
                 Unit
                 Time
                 RegionName
                 Category
                 Value
                 Nombre Bacteria
                 Antibiotico testado
              <xsl:for-each select="//row">
                 <xsl:if test="contains(RegionCode, 'ES')">
                          <xsl:value-of select="FIELD1" />
                           <xsl:value-of select="Distribution" />
                          <xsl:value-of select="Unit" />
                           <xsl:value-of select="Time" />
                           <xsl:value-of select="RegionName" />
                           <xsl:value-of select="Category" />
                           <xsl:value-of select="Value" />
                          <xsl:value-of select="Bacteria" />
                           <xsl:value-of select="Antibiotic" />
```

Figura 11: Transformación tabla grupos

6 Grafo RDF

RDF es un lenguaje de modelado para describir recursos en la web. RDF permite modelar recursos (datos) y sus relaciones mediante un grafo dirigido.

6.1. Clases RDF

Se ha tenido que desarrollar un mapa de clases que están relacionados mediante propiedades de datos. Definiendo las clases, operaciones:



Figura 12: Clases



Figura 13: Relaciones

El mapa se ha jerarquizado en torno al estudio que se esta realizando, de el parten las relaciones.

6.2. Propiedades RDF

Para expresar atributos de las clases se han creado propiedades:



Figura 14: Clases

Estas propiedades complementan a ciertas clases, aportando información sobre ellas. La mayoría se trata de propiedades de una clase a un tipo de dato int o string peor rango de edad se ha declarado como un enumerado.

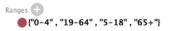


Figura 15: Propiedad enumerado

La figura final, resultado del confeccionar un modelo con relaciones entre las clases (object properties), es el grafo RDF.

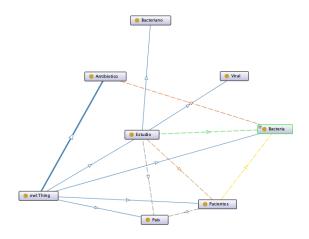


Figura 16: Grafo RDF

6.3. Queries SPARQL

Para poder acceder al modelo .OWL, se ha creado un prefijo llamado sch que almacena el uri del proyecto. A través de el podremos consultar las propiedades de objetos, de clases y las propias clases.

6.3.1. Query 1

Muestra de el id de estudio, el nombre del país donde fue realizado, el nombre de la bacteria con el antibiótico que se testo y el porcentaje de éxito de los estudios por país dado:

```
PREFIX rdf. <a href="http://www.w3.org/1099/102/22-rdf-syntax-n.s#">
PREFIX rdf. <a href="http://www.w3.org/2002/07/owl#">
PREFIX rdf. <a href="http://www.w3.org/2002/07/owl#">
PREFIX xdf. <a href="http://www.w3.org/2001/MLSchema#">
PREFIX xdf. <a href="http://www.semanticweb.org/davidcubillosdeltoro/ontologies/2022/10/untitled-ontology-12">http://www.semanticweb.org/davidcubillosdeltoro/ontologies/2022/10/untitled-ontology-12</a>

PREFIX sch. <a href="http://www.semanticweb.org/davidcubillosdeltoro/ontologies/2022/10/untitled-ontology-12">http://www.semanticweb.org/davidcubillosdeltoro/ontologies/2022/10/untitled-ontology-12</a>

SELECT ?Nomest ?Nomest ?Nomest ?valor.

Past sch.de _ Studio ?Nomeals.

Past sch.sch.ombrePais ?Nompais.

Past sch.sch.ombrePais ?Nompais.

Past sch.sch.nombreBasteria ?Nombact.

Pbact sch.rombreBasteria ?Nombact.

Pant sch.nombreBasteria ?Nomantb.

Pbact sch.valorResistencia ?valor.

FILTER ( CONTAINS(?Nompais, "Bulgaria"))

}
```

Figura 17: Muestra de los estudios de Bulgaria

6.3.2. Query 2

La segunda query, muy similar a la primera, muestro los datos de todos los estudios ordenándolos ascendentemente por el valor de la resistencia que la bacteria tiene al antibiótico. Muestra los campos: nombre de la bacteria, nombre del antibiótico y el valor propio de la resistencia.

```
PREFIX rdf. <a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1909/02/22-rdf-syntax-ns#">http://www.w3.org/2000/17/dwf-schem#</a>
PREFIX rdf. <a href="http://www.w3.org/2000/17/MIK-schem#">http://www.w3.org/2001/17/MIK-schem#">http://www.w3.org/2001/17/MIK-schem#</a>
PREFIX sdf. <a href="http://www.semanticweb.org/davidcubillosdeltoro/ontologies/2022/10/untitled-ontology-12-SELECT">http://www.semanticweb.org/davidcubillosdeltoro/ontologies/2022/10/untitled-ontology-12-SELECT "Nombacd 'Nomantic 'Youndard 'Nomantic 'Post sch.nombreBacteria 'Nombact 'Post sch.nombreBacteria 'Nombact 'Post sch.nombreBacteria 'Youndard 'P
```

Figura 18: Resistencia de una bacteria a su antibiótico

Conclusiones

La página web creada facilita la búsqueda de información sobre bacterias con resistencia a antibióticos, para los dos grupos potenciales de usuarios (personas no especializadas e investigadores).

Las tecnologías de estandarización de datos han sido de gran utilidad para el desarrollo de este trabajo. Es gracias a la familia de lenguajes XML que se ha podido procesar fácilmente el dataset en formato tabular que se obtuvo del repositorio público.

Se han obtenido transformaciones que muestran los datos necesarios para cada una de las vistas gracias a los scripts xslt, que en formato de salida html han sido añadidas al código de cada una de las interfaces de la web.

Además, gracias a esta familia de lenguajes se han podido realizar distintas consultas xquery haciendo uso del lenguaje xpath.

Finalmente, los estandares de RDF y OWL han sido de utilidad para el planteamiento de un modelo de ralaciones semanticas, sobre el que se han realizado consultas SparQL



E.T.S de Ingeniería Informática Bulevar Louis Pasteur, 35 Campus de Teatinos 29071 Málaga