



# DH@Madrid Summer School 2016

## Tecnologías de mercado específicas para poesía: TEI-XML

Helena Bermúdez Sabel

[helena.bermudez@usc.es](mailto:helena.bermudez@usc.es)



European  
Research  
Council

Madrid, 27 junio al 1 de julio de 2016



# Introducción

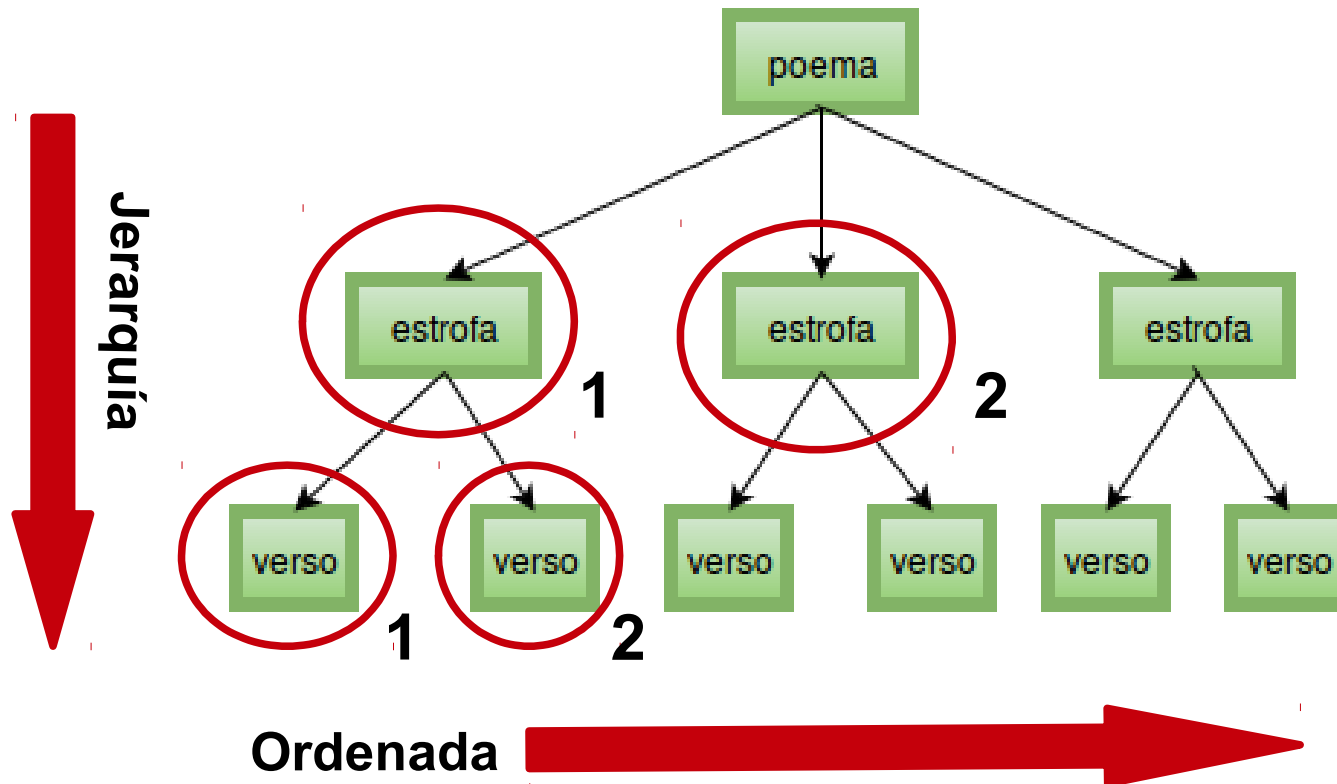
- Marcado de textos
  - ¿Qué es XML?
- Text Encoding Initiative
  - **Módulo específico para poesía**



# Marcar un texto es...

- ... modelar la estructura inherente y las propiedades semánticas de los documentos culturales a través de:
  - jerarquías
  - estructuras ordenadas





# Marcar un texto es...

... modelar la estructura inherente y las propiedades semánticas de los documentos culturales a través de:

- jerarquías
- estructuras ordenadas
- lenguaje humano + lenguaje informático

→ **XML**



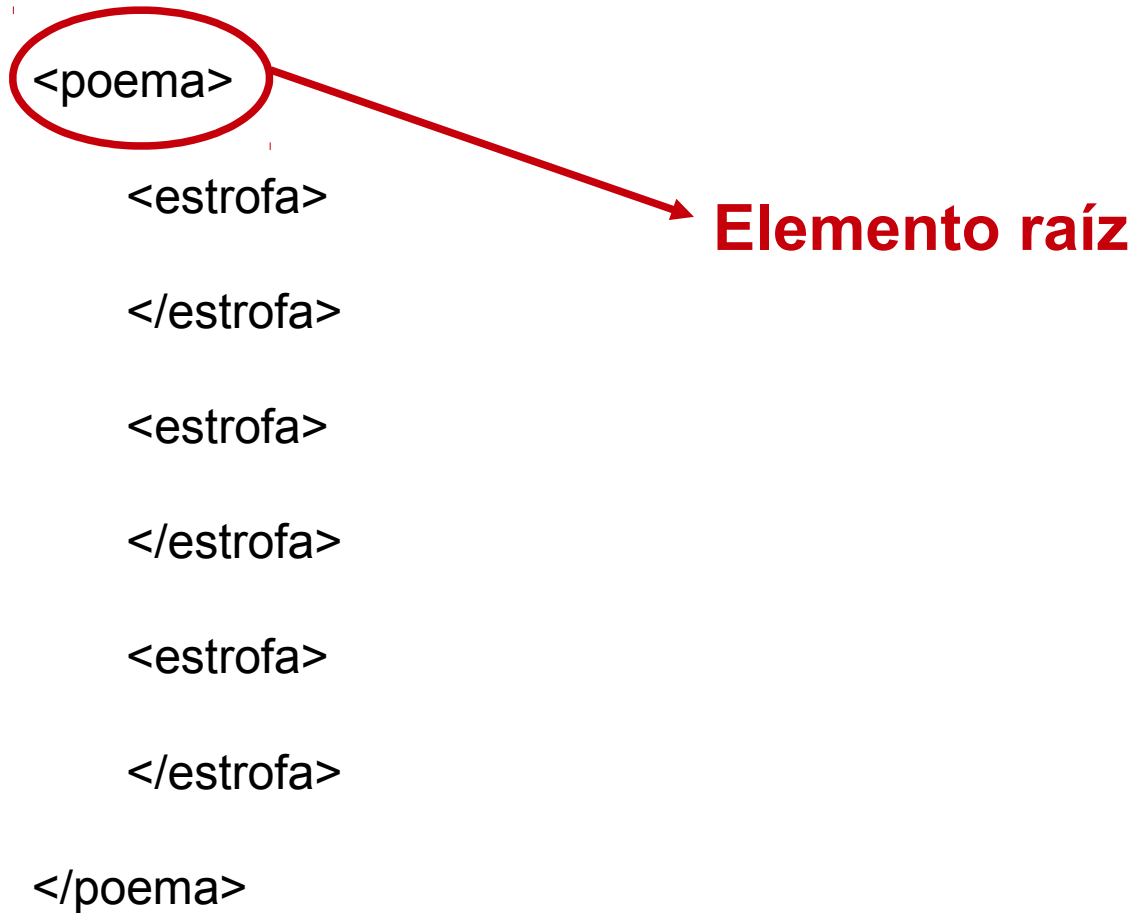
# ¿Qué es XML?

- Formato de almacenamiento y transmisión de datos
- Definido por el World Wide Web Consortium (W3C)
- Uso muy extendido

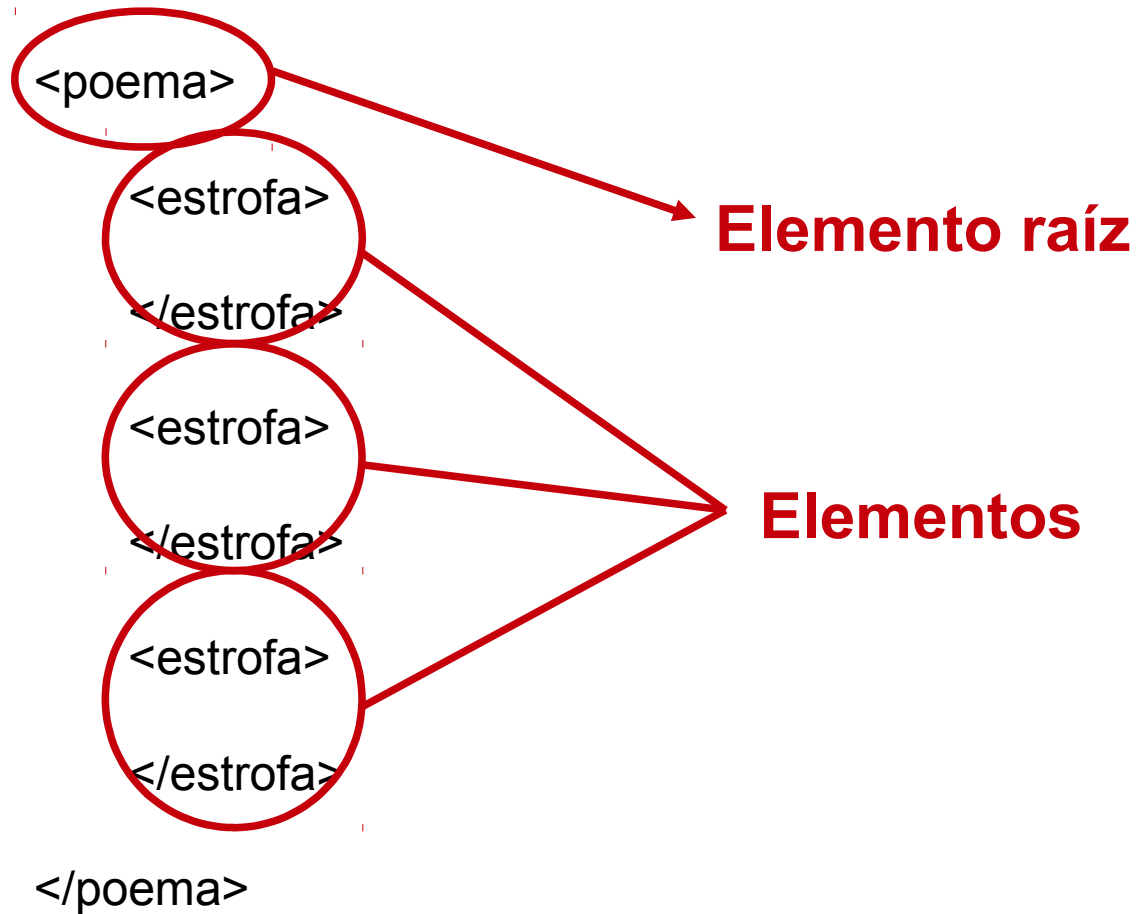


<poema>  
    *contenido*  
</poema>





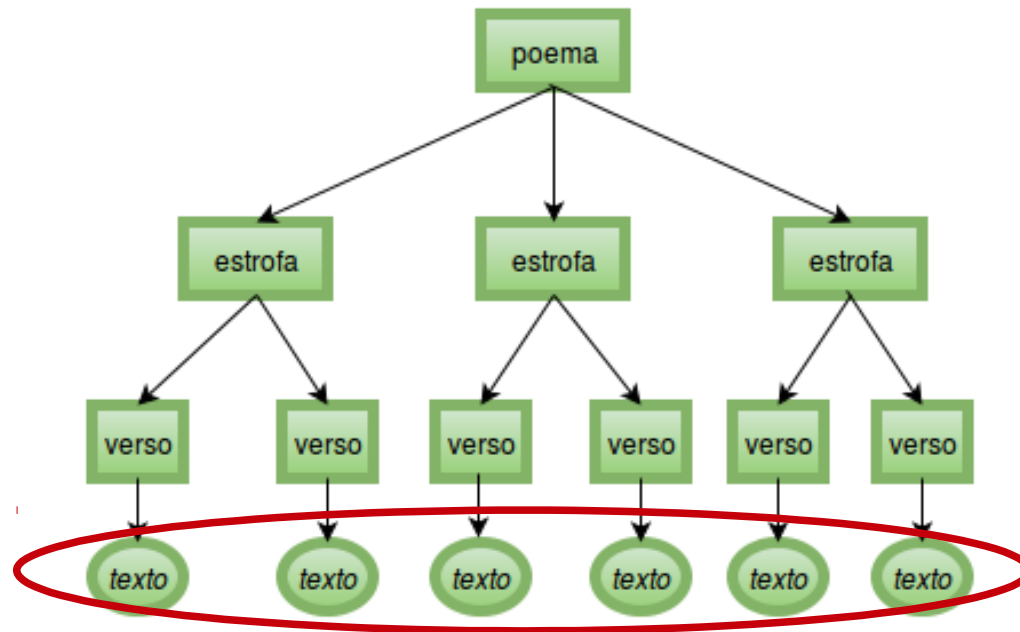




```
<poema>
  <estrofa>
    <verso>Al alba venid, buen amigo,</verso>
    <verso>al alba venid.</verso>
  </estrofa>
  <estrofa>
    <verso>Amigo el que yo más quería,</verso>
    <verso>venid al alba del día.</verso>
  </estrofa>
  <estrofa>
    <verso>Amigo el que yo más amaba,</verso>
    <verso>venid a la luz del alba.</verso>
  </estrofa>
</poema>
```

**Nodo textual**





```
<poema>
  <estrofa>
    <verso>Al alba venid, buen amigo,</verso>
    <verso>al alba venid.</verso>
  </estrofa>
  <estrofa tipo="pareado">
    <verso>Amigo el que yo más quería,</verso>
    <verso>venid al alba del día.</verso>
  </estrofa>
  <estrofa>
    <verso>Amigo el que yo más amaba,</verso>
    <verso>venid a la luz del alba.</verso>
  </estrofa>
</poema>
```

**Atributo**



# Sintaxis

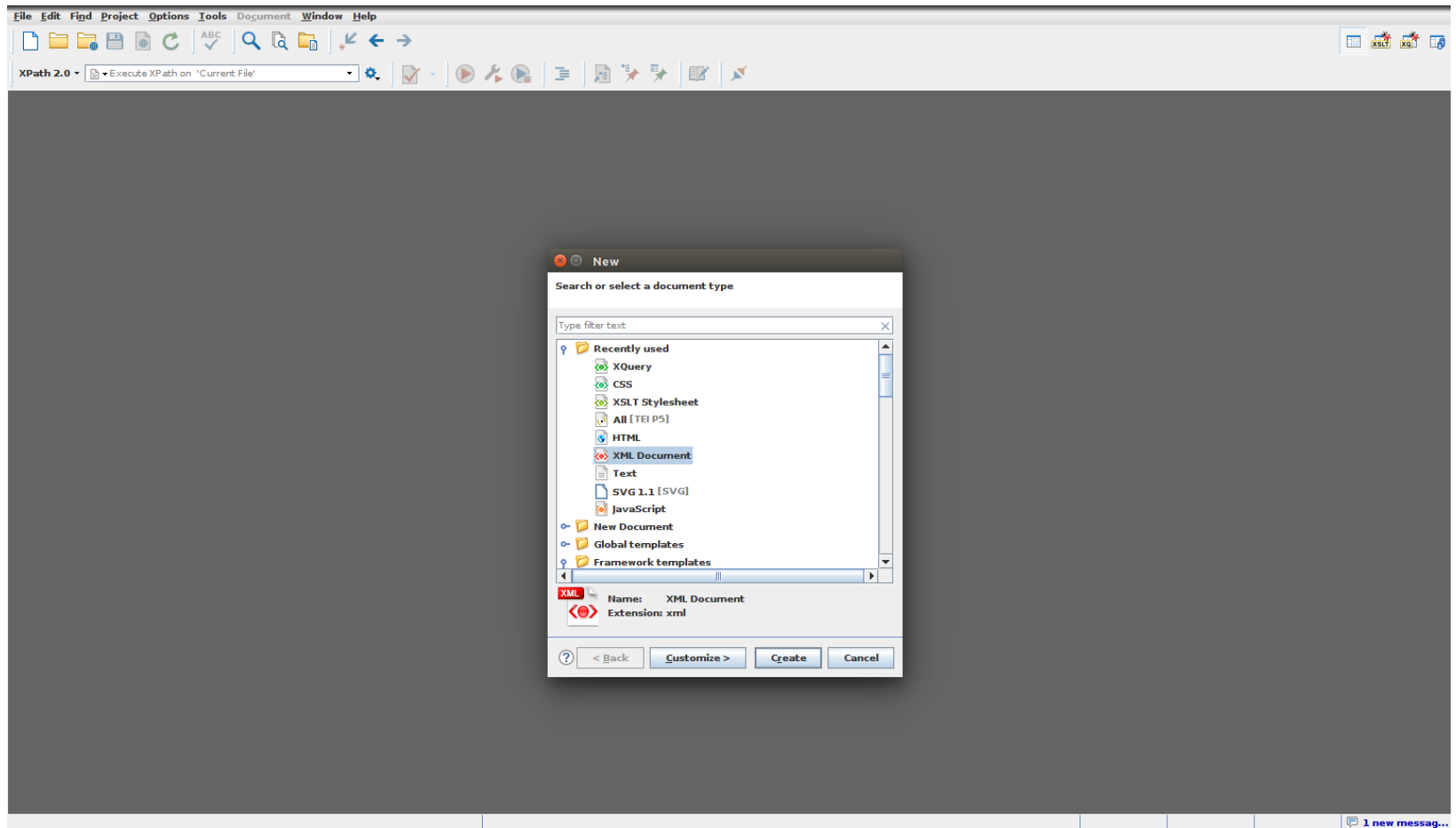
- Todo se encuentra perfectamente delimitado
- Un único elemento raíz
- Ausencia de los caracteres: **&** , **<** → **&amp;** ;  
**&lt;** ;
- Elementos anidados (**sin  
sobrelapamientos**)



# Ejercicio 1

- Inicia <oXygen/> XML Editor
- **File > New... > XML Document**





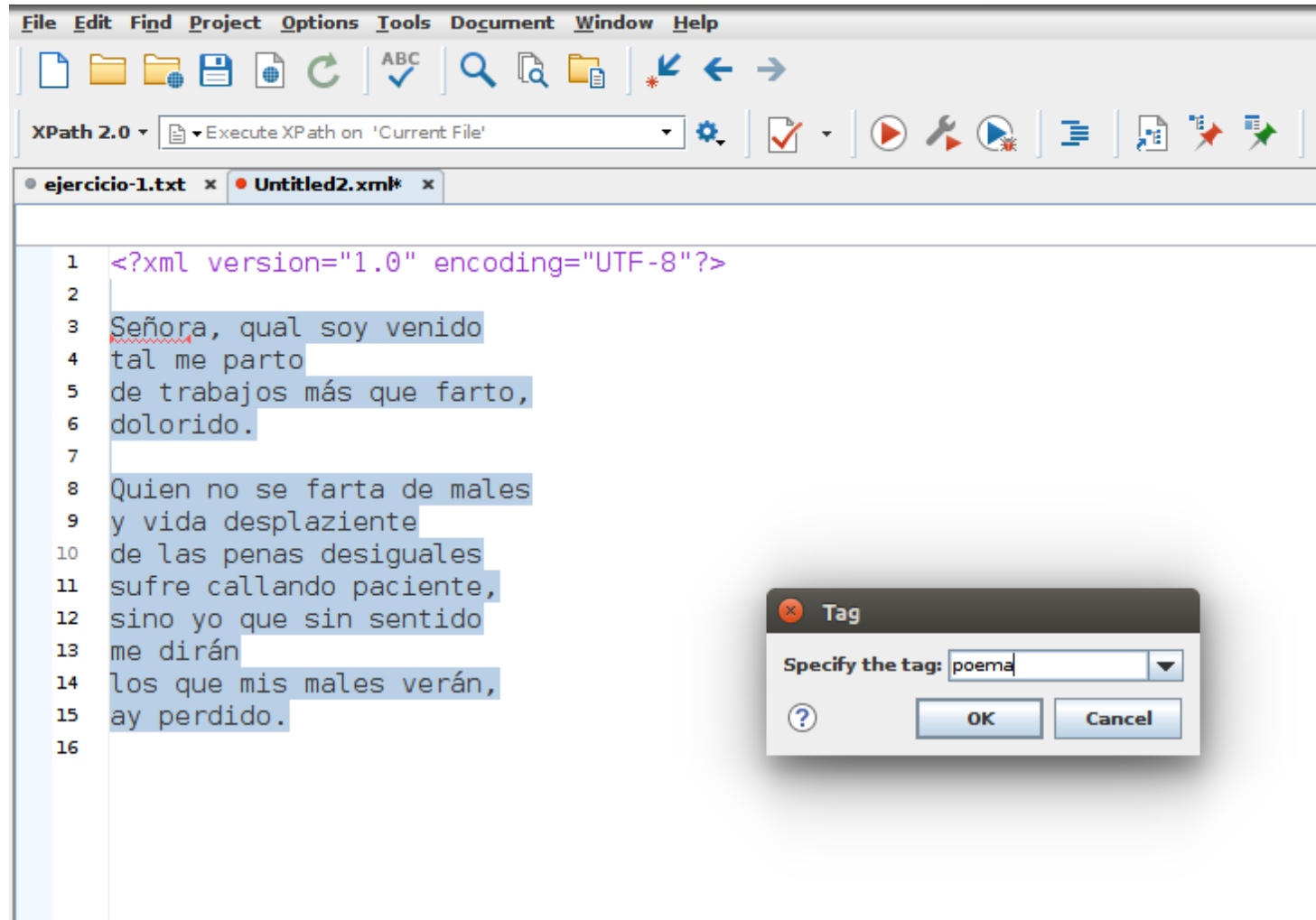
# Ejercicio 1

- Abre el documento “Texto Ejercicio 1”
- **Clic derecho > Guardar página como... > nombre-archivo.txt** (para evitar problemas de codificación de caracteres)
- Copia el contenido y pégalo en el documento XML creado en <oxygen/>
- Selecciona el texto y pulsa **Ctrl + e**
- Escribe poema en el diálogo emergente





# Ejercicio 1



The screenshot shows an XML editor window with two tabs: 'ejercicio-1.txt' and 'Untitled2.xml'. The 'Untitled2.xml' tab is active, displaying an XML document. The document starts with a declaration: `<?xml version="1.0" encoding="UTF-8"?>`. Below this, a poem is being wrapped in XML tags. The text of the poem is:   
Señora, qual soy venido  
tal me parto  
de trabajos más que farto,  
dolorido.  
  
Quien no se farta de males  
y vida desplaziente  
de las penas desiguales  
sufre callando paciente,  
sino yo que sin sentido  
me dirán  
los que mis males verán,  
ay perdido.

A 'Tag' dialog box is open, prompting the user to 'Specify the tag:'. The text 'poema' is entered in the input field. The dialog box has 'OK' and 'Cancel' buttons.



# Ejercicio 1

- Añade más elementos y algún atributo al texto
- Utiliza **Ctrl + e** para crear un elemento que “encerrará” el fragmento de texto que selecciones
- Otros atajos de `<oXygen/>`



# Ejercicio 1

¿Está tu  
documento **bien**  
**formado**?



- Comienza siempre por Buscar/Reemplazar los caracteres “prohibidos”
- Utiliza siempre un editor especializado para evitar problemas de validación
- Aprende a utilizar **expresiones regulares** para marcar automáticamente tus textos



# Otros conceptos: validez

- Uso correcto del vocabulario: los elementos y atributos usados existen en ese lenguaje
- Uso correcto de la gramática: los elementos se usan en el lugar adecuado y siguiendo el orden establecido

 **Schema**

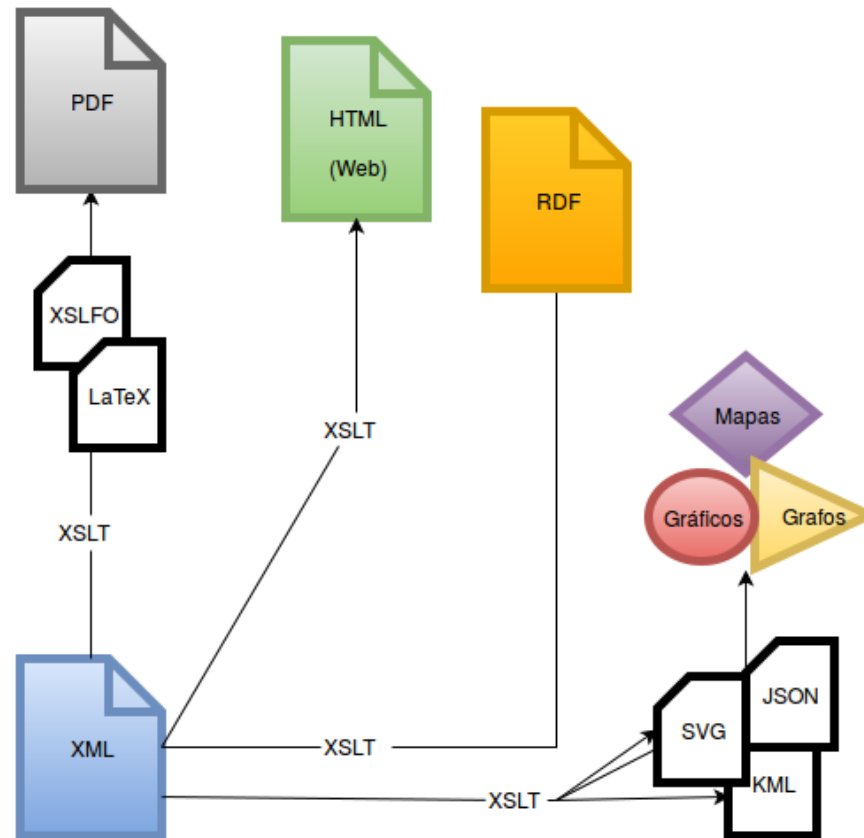


# ¿Por qué XML?

- Sintáxis accesible
- Legible
- Máximo control del input y output
- Independiente de software o hardware
- Uso extenso
- Variedad de software (código abierto + privativo) que lo soportan



# ¿Por qué XML?



# ¿Por qué XML?

```
<front>
<div>
<listWit>
<witness corresp="#A">Cancioneiro da Ajuda</witness>
<witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
</listWit>
</div>
</front>
<body>
<div type="poem" corresp="#A82B186b">
<head>
<title>
<app>
<rdg wit="#A" hand="#q">
<idno>A82</idno>
<locus from="21r" to="21r"/></rdg>
<rdg wit="#B" hand="#d">
<idno>B186</idno>
<locus from="48r" to="48v"/></rdg>
</app>
</title>
<name role="author" ref="#PGarBu"/>
</head>
<lg type="cobra">
<l n="1">
<app>
<rdg wit="#A #B">De</rdg>
</app>
<app>
<rdg wit="#A">quantos</rdg>
<rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></seg></rdg>
</app>
<app>
<rdg wit="#A">mui</rdg>
<rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
</app>
<app>
<rdg wit="#A">coitados</rdg>
<rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
</app>
<app>
<rdg wit="#A #B">son</rdg>
</app>
</l>
<l n="2">
<app>
<rdg wit="#A #B">a que</rdg>
```



**Testemunho: A**  
**Texto: A2**  
**Localização: 1r**  
**Copista: α**

1. Sennor fremosa grand enueia ei
2. eu a tod ome que ueio morrer
3. e **segud** ora o meu coñ<sup>o</sup>çer
4. enquant este faço mui gran razon
5. ca ei por uos e no meu coraçon
6. tan gran **cuita** que mil uezes meten
7. sennor sen **falla et** sen tod o sen
8. e **nõ** uus queredes de **min** doer
9. Pero **sē**nor ùa ren u9 direi
10. con tod est ora nõ ei eu poder

**Testemunho: B**  
**Texto: B92**  
**Localização: 24v**  
**Copista: c**

1. Senhor fremosa **muy** grãd enueia ey
2. eu a tod ome que ueio morrer
3. e segund ora o meu conhocer
4. enquant **est** faço **muy** gram razon
5. ca **ey** por uos e no meu coraçon
6. tan **gram** coyta que mil uezes etem
7. senhor sen fala essem tod o **sem**
8. e non **u9** queredes **d** mj doer
9. Pero senhor **hũa rē** u9 direy
10. **cõ** tod est ora nõ ey eu poder





# ¿Por qué XML?

```
<front>
<div>
  <listWit>
    <witness corresp="#A">Cancioneiro da Ajuda</witness>
    <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
  </listWit>
</div>
</front>
<body>
  <div type="poem" corresp="#A82B186b">
    <head>
      <title>
        <app>
          <rdg wit="#A" hand="#q">
            <idno>A82</idno>
            <locus from="21r" to="21r"/></rdg>
          <rdg wit="#B" hand="#d">
            <idno>B186</idno>
            <locus from="48r" to="48v"/></rdg>
        </app>
      </title>
      <name role="author" ref="#PGarBu"/>
    </head>
    <lg type="cobra">
      <cl re="1">
        <app>
          <rdg wit="#A #B">De</rdg>
        </app>
        <app>
          <rdg wit="#A">quantos</rdg>
          <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></seg></rdg>
        </app>
        <app>
          <rdg wit="#A">mui</rdg>
          <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
        </app>
        <app>
          <rdg wit="#A">coitados</rdg>
          <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
        </app>
        <app>
          <rdg wit="#A #B">son</rdg>
        </app>
      </cl>
      <cl re="2">
        <app>
          <rdg wit="#A #B">a que</rdg>
        </app>
      </cl>
    </lg>
  </div>
</body>
```



Cantiga: A224, B39S, VS

Autor: Afonso Lopez de Baian

Período: 1246-1280 (período 3)

- |   |        |               |    |   |
|---|--------|---------------|----|---|
| A | Sennor | que grau oi a | mi | e |
| B | Senhor | que grau oi a | mj | e |
| V | Senhor | que grau oi a | mi | e |
- |   |    |    |         |     |          |
|---|----|----|---------|-----|----------|
| A | de | me | auer de | uos | a partir |
| B | de | m  | auer de | nos | a partir |
| V | de | m  | auer de | uos | a partir |
- |   |    |     |    |      |      |      |        |
|---|----|-----|----|------|------|------|--------|
| A | ca | sei | de | pran | pois | m eu | partir |
| B | ca | sey | de | pram | poys | m eu | partir |
| V | ca | sey | de | pram | poys | m eu | partir |
- |   |     |    |   |       |     |     |    |
|---|-----|----|---|-------|-----|-----|----|
| A | que | mi | a | uerra | per | bõa | fe |
| B | que | mh | a | ueira | per | boa | fe |
| V | que | mh | a | uerra | per | boa | fe |



# ¿Por qué XML?

```
<front>
<div>
  <listWit>
    <witness corresp="#A">Cancioneiro da Ajuda</witness>
    <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
  </listWit>
</div>
</front>
<body>
  <div type="poem" corresp="#A82B186b">
    <head>
      <title>
        <app>
          <rdg wit="#A" hand="#q">
            <idno>A82</idno>
            <locus from="21r" to="21r"/></rdg>
          <rdg wit="#B" hand="#d">
            <idno>B186</idno>
            <locus from="48r" to="48v"/></rdg>
        </app>
      </title>
      <name role="author" ref="#PGarBu"/>
    </head>
    <lg type="cobra">
      <l n="1">
        <app>
          <rdg wit="#A #B">De</rdg>
        </app>
        <app>
          <rdg wit="#A">quantos</rdg>
          <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></seg></rdg>
        </app>
        <app>
          <rdg wit="#A">mui</rdg>
          <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
        </app>
        <app>
          <rdg wit="#A">coitados</rdg>
          <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
        </app>
        <app>
          <rdg wit="#A #B">son</rdg>
        </app>
      </l>
      <l n="2">
        <app>
          <rdg wit="#A #B">a que</rdg>
        </app>
      </l>
    </lg>
  </div>
</body>
</div>
</front>
```



## Nasalização progressiva

### Testemunho

Ms. B	c	d	e	b	(49.25%)
Ms. A	α				(47.76%)
Ms. V	v				(10.45%)

### Período

Per. 3					(79.1%)
Per. 2					(20.9%)



# ¿Por qué XML?

```
<front>
  <div>
    <listWit>
      <witness corresp="#A">Cancioneiro da Ajuda</witness>
      <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
    </listWit>
  </div>
</front>
<body>
  <div type="poem" corresp="#A82B186b">
    <head>
      <title>
        <app>
          <rdg wit="#A" hand="#a">
            <idno>A82</idno>
            <locus from="21r" to="21r"/></rdg>
          <rdg wit="#B" hand="#d">
            <idno>B186</idno>
            <locus from="48r" to="48v"/></rdg>
          </app>
        </title>
        <name role="author" ref="#PGarBu"/>
      </head>
      <lg type="cobra">
        <l n="1">
          <app>
            <rdg wit="#A #B">De</rdg>
          </app>
          <app>
            <rdg wit="#A">quantos</rdg>
            <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></seg></rdg>
          </app>
          <app>
            <rdg wit="#A">mui</rdg>
            <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
          </app>
          <app>
            <rdg wit="#A">coitados</rdg>
            <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
          </app>
          <app>
            <rdg wit="#A #B">son</rdg>
          </app>
        </l>
        <l n="2">
          <app>
            <rdg wit="#A #B">a que</rdg>
          </app>
        </l>
      </lg>
    </div>
  </body>
```

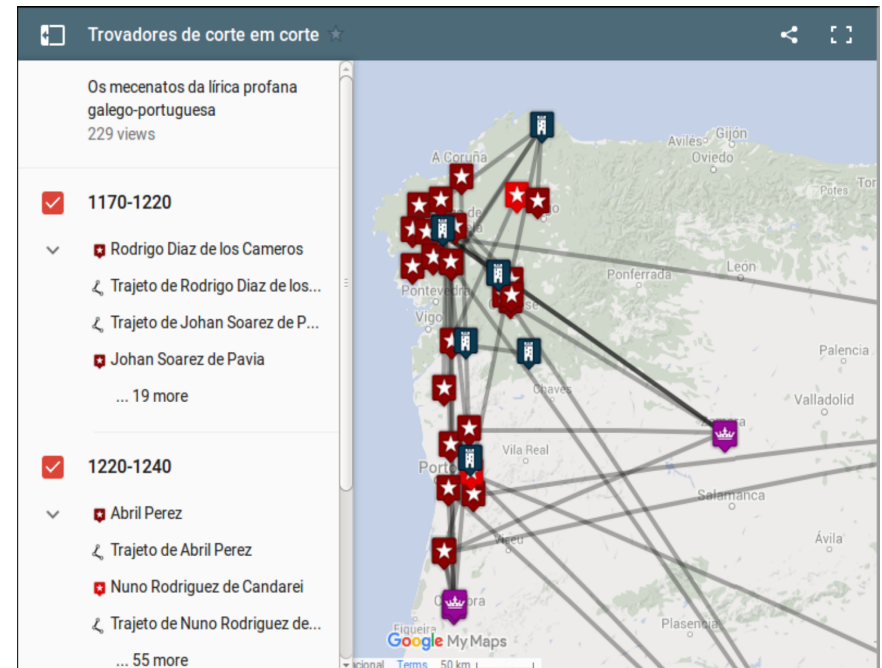


Fenómeno	Testemunho	Localização	Copista	Autor	Período	Cantiga	V.
quix] qs	A	11v-12r	α	Martin Soarez	1230-1270 (3)	A48, B160	18
quige] qs	A	22v-23r	α	Pero Garcia de Buralès	1240-1270 (3)	A89, B193	13



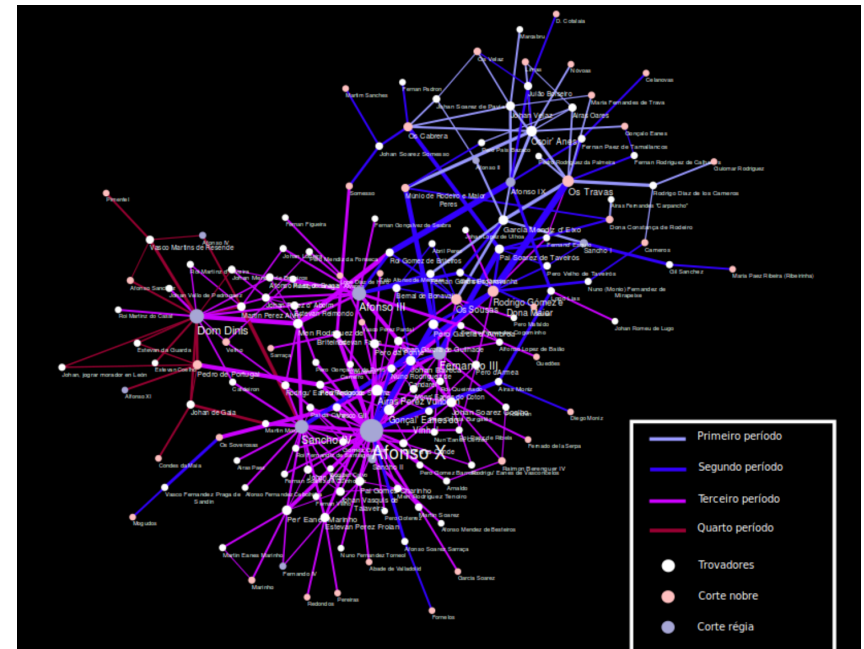
# ¿Por qué XML?

```
<front>
<div>
  <listWit>
    <witness corresp="#A">Cancioneiro da Ajuda</witness>
    <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
  </listWit>
</div>
</front>
<body>
<div type="poem" corresp="#A82B186b">
  <head>
    <title>
      <app>
        <rdg wit="#A" hand="#q">
          <idno>A82</idno>
          <locus from="21r" to="21r"/></rdg>
          <rdg wit="#B" hand="#d">
            <idno>B186</idno>
            <locus from="48r" to="48v"/></rdg>
        </app>
      </title>
      <name role="author" ref="#PGarBu"/>
    </head>
    <lg type="cobra">
      <cl n="1">
        <app>
          <rdg wit="#A #B">De</rdg>
        </app>
        <app>
          <rdg wit="#A">quantos</rdg>
          <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></seg></rdg>
        </app>
        <app>
          <rdg wit="#A">mui</rdg>
          <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
        </app>
        <app>
          <rdg wit="#A">coitados</rdg>
          <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
        </app>
        <app>
          <rdg wit="#A #B">son</rdg>
        </app>
      </cl>
      <cl n="2">
        <app>
          <rdg wit="#A #B">a que</rdg>
        </app>
      </cl>
    </lg>
  </div>
</body>
```



# ¿Por qué XML?

```
<front>
  <div>
    <listWit>
      <witness corresp="#A">Cancioneiro da Ajuda</witness>
      <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
    </listWit>
  </div>
</front>
<body>
  <div type="poem" corresp="#A82B186b">
    <head>
      <app>
        <title>
          <rdg wit="#A" hand="#a">
            <idno>A82</idno>
            <locus from="21r" to="21r"/></rdg>
          <rdg wit="#B" hand="#d">
            <idno>B186</idno>
            <locus from="48r" to="48v"/></rdg>
          </app>
        </title>
        <name role="author" ref="#PGarBu"/>
      </head>
      <lg type="cobra">
        <cl n="1">
          <app>
            <rdg wit="#A #B">De</rdg>
          </app>
          <app>
            <rdg wit="#A">quantos</rdg>
            <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></seg></rdg>
          </app>
          <app>
            <rdg wit="#A">mui</rdg>
            <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
          </app>
          <app>
            <rdg wit="#A">coitados</rdg>
            <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
          </app>
          <app>
            <rdg wit="#A #B">son</rdg>
          </app>
        </cl>
        <cl n="2">
          <app>
            <rdg wit="#A #B">a que</rdg>
          </app>
        </cl>
      </lg>
    </div>
  </body>
</front>
```



- Conjunto de normas para codificar documentos culturales
- Consorcio internacional que mantiene y desarrolla una lengua estandarizada de codificación XML para etiquetar textos en Humanidades
- Comunidad de proyectos de investigación e individuos que usan las directrices TEI



# XML - TEI

Investigador/a

XML

TEI



**Conceptos**

```
<elemento>  
  contenido  
</elemento>
```

```
<elemento atributo="valor">  
  contenido  
</elemento>
```

**Lengua / Sintaxis**

```
<lg>
```

```
<l>
```

```
<l type="menor">
```

**Vocabulario**



# ¿Por qué TEI?

- Eficiencia del modelo
- Mercado descriptivo → Recursividad
- Economía
- Interoperabilidad
- Intercambio
- Tránsferencia tecnológica





# Usar TEI



<div> <lg> <l> <lb/>  
<seg> <caesura/>  
@met @real @rhyme  
@type @subtype  
@enjamb



# Usar TEI

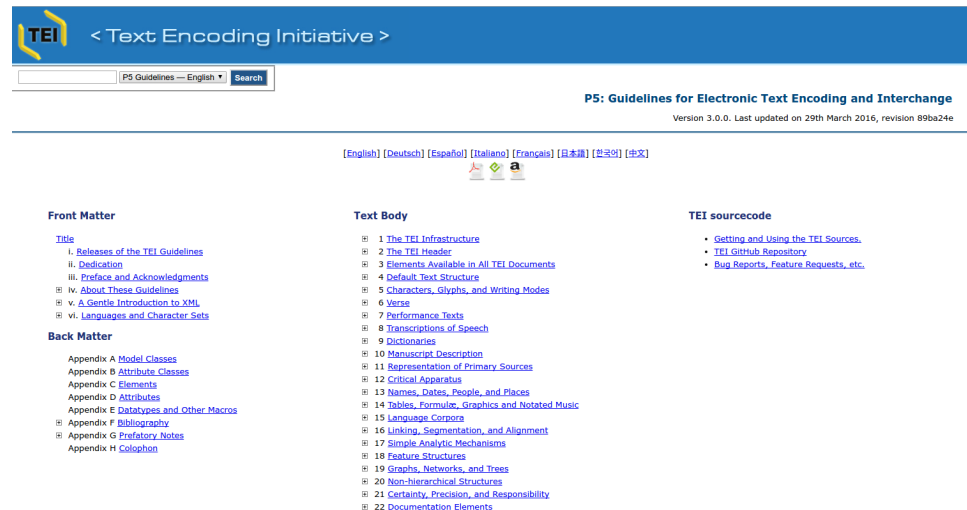


<lg type="chevy-chase-stanza" met="-+-  
+--+/-+--+-" rhyme="abab cdcd">  
<l n="1">Und frische Nahrung, neues  
Blut</l>  
<l n="2" real="+--+--">Saug' ich aus freier  
Welt;</l>  
<l n="3" real="+--+--+--">Wie ist Natur so  
hold und gut,</l>  
<l n="4" real="---+--">Die mich am Busen  
hält!</l>  
<l n="5">Die Welle wieget unsern  
Kahn</l>  
<l n="6">Im Rudertakt hinauf,</l>  
<l n="7">Und Berge, wolkig himmelan,</l>  
<l n="8">Begegnen unserm Lauf.</l>  
</lg>



# Usar TEI

- Text Encoding Initiative
- P5: Guidelines for Electronic Text Encoding and Interchange



The screenshot shows the TEI P5 Guidelines website. The header is blue with the TEI logo and the text "< Text Encoding Initiative >". Below the header is a search bar with the text "P5 Guidelines — English" and a "Search" button. The main content area is white and features the title "P5: Guidelines for Electronic Text Encoding and Interchange" and the version information "Version 3.0.0. Last updated on 29th March 2016, revision 89ba24e". Below this is a navigation bar with links for [English], [Deutsch], [Español], [Italiano], [Français], [日本語], [한국어], and [中文]. The main content is organized into three columns: "Front Matter", "Text Body", and "TEI sourcecode". The "Front Matter" column includes links for Title, Releases of the TEI Guidelines, Dedication, Preface and Acknowledgments, About These Guidelines, A Gentle Introduction to XML, Languages and Character Sets, Back Matter, Appendix A Model Classes, Appendix B Attribute Classes, Appendix C Elements, Appendix D Attributes, Appendix E Datatypes and Other Macros, Appendix F Bibliography, Appendix G Prefatory Notes, and Appendix H Colophon. The "Text Body" column includes links for 1 The TEI Infrastructure, 2 The TEI Header, 3 Elements Available in All TEI Documents, 4 Default Text Structure, 5 Characters, Glyphs, and Writing Modes, 6 Verse, 7 Performance Texts, 8 Transcriptions of Speech, 9 Dictionaries, 10 Manuscript Description, 11 Representation of Primary Sources, 12 Critical Apparatus, 13 Names, Dates, People, and Places, 14 Tables, Formulae, Graphics and Notated Music, 15 Language Corpora, 16 Linking, Segmentation, and Alignment, 17 Simple Analytic Mechanisms, 18 Feature Structures, 19 Graphs, Networks, and Trees, 20 Non-hierarchical Structures, 21 Certainty, Precision, and Responsibility, and 22 Documentation Elements. The "TEI sourcecode" column includes links for Getting and Using the TEI Sources, TEI Github Repository, and Bug Reports, Feature Requests, etc.



# Ejercicio 2

- Descarga y guarda el *schema* de este ejercicio
- Abre <oXygen/>
- **File** > **New**
- En el diálogo emergente, filtra los resultados escribiendo “**TEI**”
- Selecciona **TEI P5** > **All**



# Ejercicio 2

- Haz click en **Document** > **Schema** > **Associate Schema**
- En el diálogo emergente, asocia el *schema* que has descargado dentro de la opción **URL**
- Descarga “Texto Ejercicio 2”



# Ejercicio 2

- TEI y la importancia de los metadatos:  
completa el elemento `<teiHeader>` (+ info)
- Etiqueta la composición (+ info)

