

LMSGI03.1 Documentos XML



☰ Definición de esquemas y vocabularios en xml

☰ Sintaxis XML

☰ Espacios de nombres

?

Autoevaluación

CONTENIDO PDF

☰ Contenido PDF

Definición de esquemas y vocabularios en xml



XML permite guardar datos en documentos manteniendo una estructura común y así se pueden compartir los datos de una manera eficaz.

Introducción

XML (eXtensible Markup Language ó Lenguaje de Etiquetas Extendido) es un lenguaje desarrollado por W3C (World Wide Web Consortium) que está basado en SGML (Standard

Generalized Markup Language-Lenguaje de Marcado Generalizado Estándar). Es un lenguaje de etiquetas, que estructuran y guardan de forma ordenada la información proporcionando a webs y a aplicaciones una forma realmente potente de guardar la información.

Al igual que en HTML un documento XML es un documento de texto con extensión ".xml", compuesto de parejas de etiquetas, estructuradas en árbol, que describen una función en la organización del documento. Es decir, un documento XML es texto plano (sin formato) y contiene marcas (etiquetas) definidas por el desarrollador, cosa que no sucede en HTML que, además, fue designado para mostrar datos mientras que XML fue designado para decir qué información muestran los datos.

XML es un metalenguaje, es decir, puede ser empleado para definir otros lenguajes, llamados dialectos XML. Por ejemplo, algunos lenguajes basados en XML son:

- GML (Geography Markup Language-Lenguaje de Marcado Geográfico)
 - MathML (Mathematical Markup Language-Lenguaje de Marcado Matemático)
 - RSS (Really Simple Syndication-Sindicación Realmente Simple)
 - SVG (Scalable Vector Graphics-Gráficos Vectoriales Escalables)
 - XHTML (eXtensible HyperText Markup Language-Lenguaje de Marcado de Hipertexto eXtensible)
-

Herramientas básicas de XML

Para trabajar en XML es necesario editar los documentos y luego procesarlos. Así que tenemos dos tipos de herramientas:

- **Editores XML:** Para crear documentos XML hay que usar algún editor XML. Cualquier editor simple de texto, como el bloc de notas de Windows, es

suficiente pero algunos ayudan a crear estructuras y etiquetas de los elementos usados en los documentos.

- **Procesadores XML:** Para interpretar el código XML se puede utilizar cualquier navegador. Los procesadores de XML permiten leer los documentos XML y acceder a su contenido y estructura. Un procesador es un conjunto de módulos de software entre los que se encuentra un parser o analizador de XML que comprueba que el documento cumple las normas establecidas para que pueda abrirse. Estas normas pueden corresponderse con las necesarias para trabajar sólo con documentos de tipo válido o sólo exigir que el documento esté bien formado. El modo en que los procesadores deben leer los datos XML está descrito en la recomendación de XML establecida por W3C.

Estructura

En un documentos XML se distinguen claramente dos partes:

- **Prólogo:** no es obligatorio pero es muy conveniente e informa al intérprete encargado de procesar el documento de todos aquellos datos que necesita para realizar su trabajo. Si existe, debe ir al principio del documento y consta, a su vez, de dos partes:

DECLARACIÓN DE XML

DECLARACIÓN DE TIPO DE DOCUMENTO

Donde se indica la versión de XML utilizada, la codificación empleada para representar los caracteres y se puede incluir una declaración de documento autónomo (standalone).

DECLARACIÓN DE XML

DECLARACIÓN DE TIPO DE DOCUMENTO

Opción que define qué tipo de documento estamos creando y siempre comienza por la cadena:

```
<!DOCTYPE NombreEjemplar ...>
```

El NombreEjemplar ha de ser idéntico al del ejemplar del documento XML en el que se está trabajando.

Documento XML sin prólogo

Falta información esencial para el procesamiento

Declaración XML

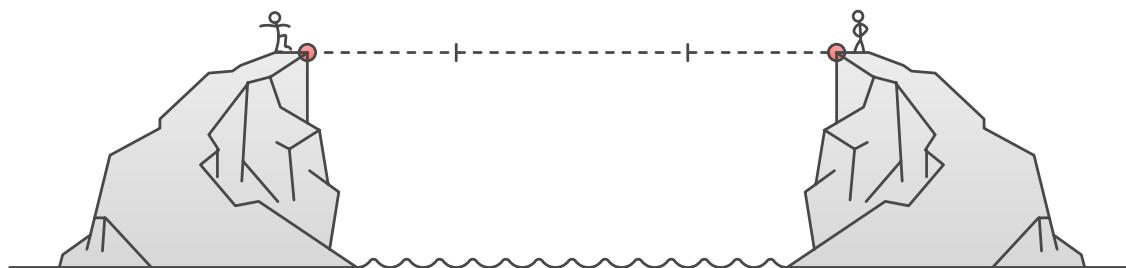
Especificar versión, codificación y autonomía

Declaración de tipo de documento

Definir el tipo de documento XML

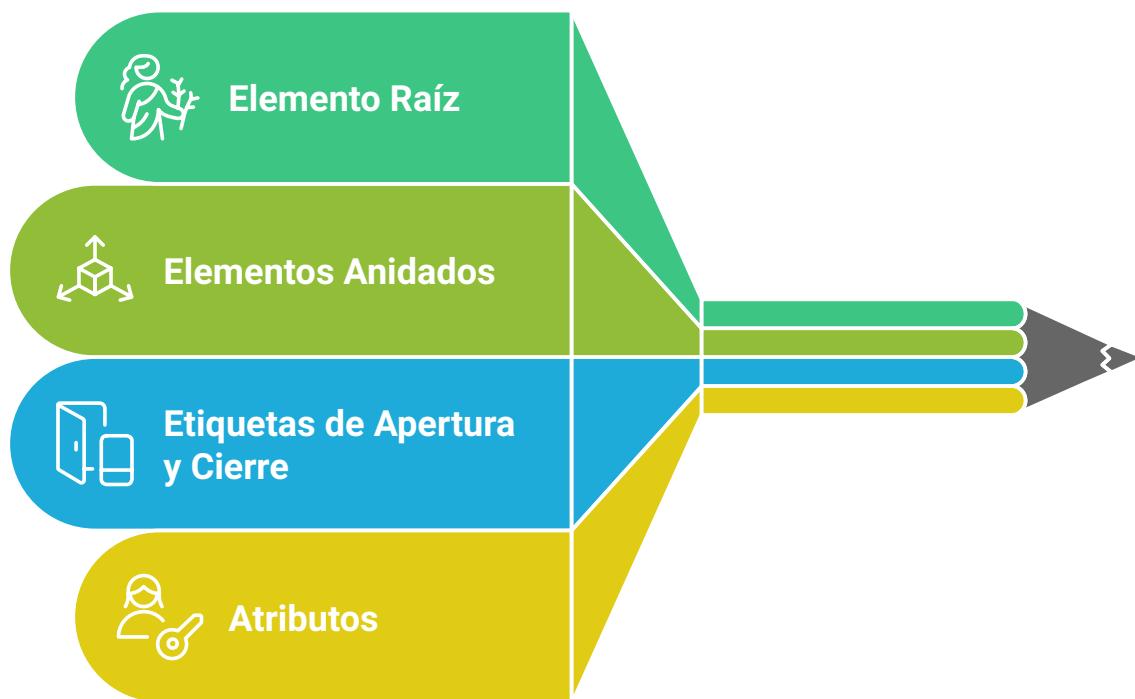
Documento XML con prólogo

Información completa para el procesamiento



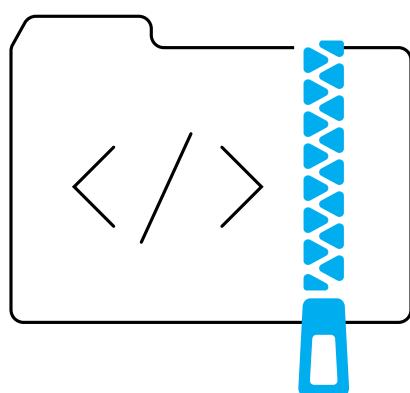
Elaboración propia

- **Ejemplar:** obligatorio y la parte más importante de un documento XML, ya que contiene los datos reales del documento. Es el elemento raíz de un documento XML y es único. Todos los demás elementos de un documento XML están contenidos en el mismo. Es decir, está formado por elementos anidados. Los elementos son los distintos bloques de información que permiten definir la estructura de un documento XML. Están delimitados por una etiqueta de apertura y una etiqueta de cierre. A su vez los elementos pueden estar formados por otros elementos y atributos.



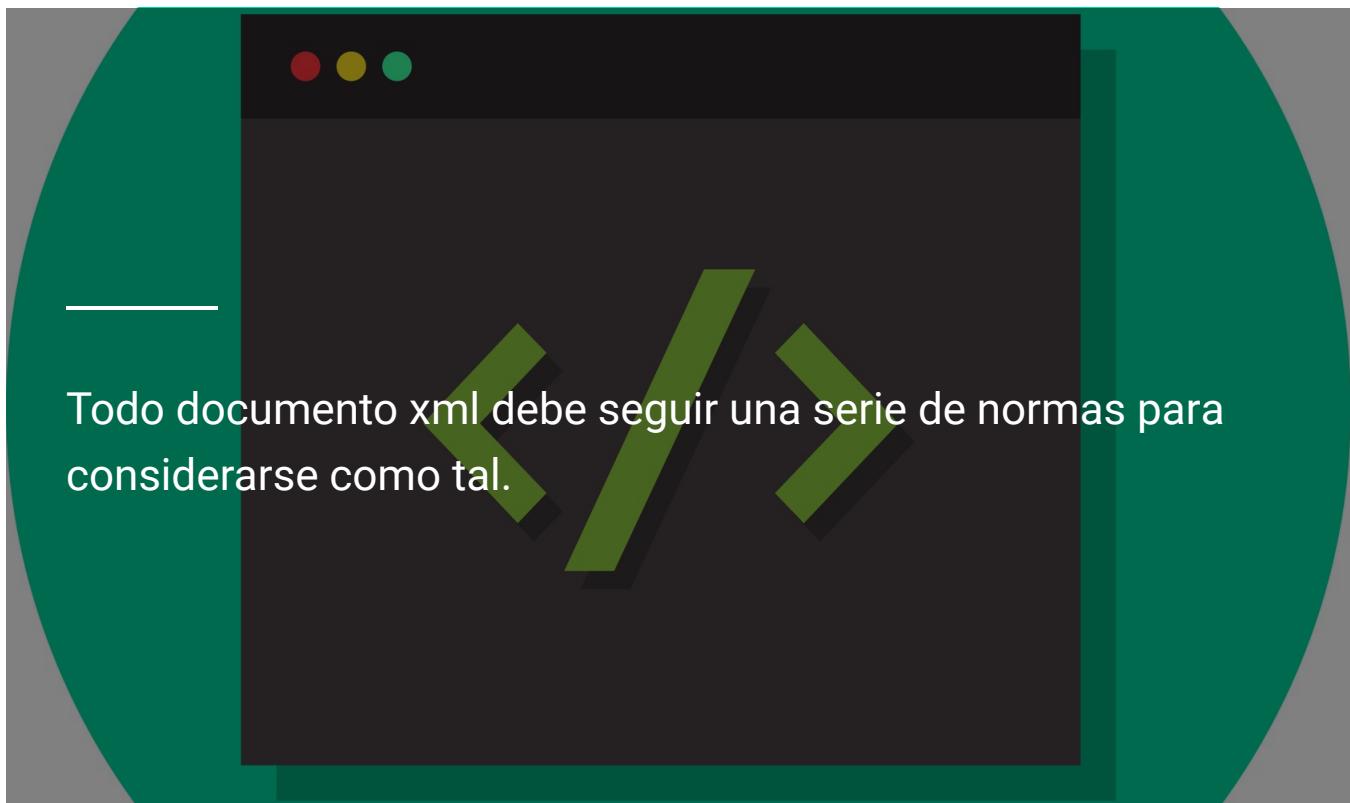
Elaboración propia

El siguiente código es un ejemplo de documento xml donde <catalogo> es el elemento raíz que contiene dos elementos secundarios <libro>, cada uno con sus propios elementos (como <titulo>, <autor>, etc.) y atributos (como id).



Custom code isn't available in PDF format

Sintaxis XML



Sintaxis

La formación de elementos ha de cumplir ciertas normas para que queden perfectamente definidos y que el documento XML pueda ser interpretado por los procesadores XML sin generar ningún error fatal. Dichas reglas son:

- En un documento XML debe existir un elemento raíz, y sólo uno.
- Todos los elementos tienen una etiqueta de inicio y otra de cierre. Si en el documento existen elementos vacíos, se pueden sustituir las etiquetas de inicio y cierre por una de elemento vacío sustituyendo el carácter ">" por "/>". Es decir, <elemento></elemento> puede sustituirse por: <elemento/>
- No puede cerrarse un elemento que contenga algún otro elemento que aún no se haya cerrado.
- Los nombres de las etiquetas de inicio y de cierre de un mismo elemento han de ser idénticos, respetando las mayúsculas y minúsculas. Esto quiere decir que XML es "case sensitive".
- Los nombres de las etiquetas de inicio y cierre pueden ser cualquier cadena alfanumérica que no contenga espacios y no comience ni por el carácter dos puntos, ":" ni por la cadena "xml" ni ninguna de sus versiones en que se cambien mayúsculas y minúsculas ("XML", "XmL", "xML",...).
- Entidades. El contenido de los elementos no puede contener la cadena "]]>" por compatibilidad con SGML. Además no se pueden utilizar directamente algunos caracteres en cuyo se sustituyen por las siguientes cadenas ó entidades predefinidas:
 - > por >
 - < por <
 - & por &
 - " por "
 - ' por '
 - Para utilizar caracteres especiales, como £, ©, ®,... hay que usar las expresiones &#D; o &#H; donde D y H se corresponden respectivamente con el número decimal o hexadecimal correspondiente al carácter que se quiere representar en el código UNICODE. Por ejemplo, para incluir el carácter de Euro, €, se usarían las cadenas € o &x20AC;
- Los nombres de las etiquetas han de ser autodescriptivos para facilitar el trabajo.
- Los atributos deben de ir entre comillas.
- Los documentos XML pueden tener comentarios, que no son interpretados por el interprete XML. Estos se incluyen entre las cadenas "<!--" y "-->". Pueden estar en cualquier sitio en el documento salvo antes del prólogo ó dentro de una etiqueta.

En resumen, un documento XML estará **bien formado** si cumple las reglas sintácticas anteriores.

LMSGI04 XML básico



Marcas ó etiquetas

En XML, las "marcas" generalmente se refieren a las etiquetas que se utilizan para definir los elementos y estructurar el contenido del documento. Las etiquetas en XML pueden ser:

- **Etiqueta de Apertura:** Es el inicio de un elemento y se escribe entre los símbolos `<` y `>`. Contiene el nombre del elemento y, opcionalmente, los atributos del elemento.

Ejemplo: <libro id="1">

- **Etiqueta de Cierre:** Es el final de un elemento y se escribe entre los símbolos `</` y `>`. Contiene el nombre del elemento que coincide con la etiqueta de apertura del mismo elemento.

Ejemplo: </libro>

Entre las etiquetas de apertura y cierre, se puede incluir contenido, como texto, otros elementos XML, comentarios, etc.

Un ejemplo de cómo se usan las etiquetas en XML para definir un elemento simple:

```
<titulo>El Principito</titulo>
```

En este ejemplo, <titulo> es la etiqueta de apertura del elemento "titulo", y </titulo> es la etiqueta de cierre que indica el final del elemento "titulo". Dentro de estas etiquetas, se encuentra el contenido del elemento.

Elementos

Están compuestos por etiquetas de inicio y fin, escritas con el mismo nombre y todo el contenido que exista entre ambas.

Un ejemplo de cómo se usan las etiquetas en XML para definir un elemento que contiene otros:

```
<libro>
  <titulo>El Principito</titulo>
  <autor>Antoine de Saint-Exupéry</autor>
</libro>
```

En este ejemplo, <libro> es la etiqueta de apertura del elemento "libro", y </libro> es la etiqueta de cierre que indica el final del elemento "libro". Dentro de estas etiquetas, se encuentran las etiquetas de los elementos "titulo" y "autor", que contienen texto.

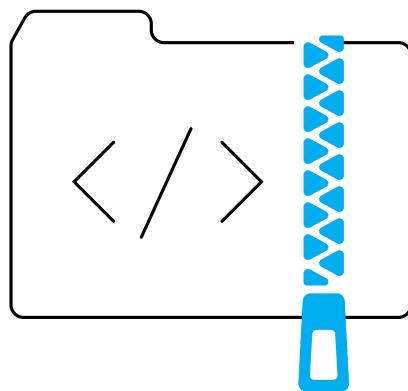
Aunque no es muy normal, pueden existir elementos vacíos, sin contenido alguno. Por ejemplo <direccion></direccion> se puede escribir como <direccion/>.

Atributos

Como en HTML, los elementos XML pueden tener atributos que contienen datos relacionados con un elemento específico.

Por ejemplo: <persona dni="12345678A">.

En la siguiente tabla hay ejemplos que contienen la misma información:



Custom code isn't available in PDF format

A la hora de utilizar atributos hay que tener en cuenta que no pueden contener subestructuras y la expansión para usos futuros no es fácil.

Comentarios

Sirven de ayuda cuando el documento es extenso ó se quiere añadir claridad al documento.

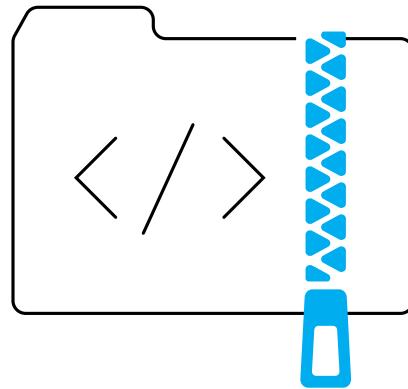
Pueden ir en cualquier sitio excepto dentro de las etiquetas, otros comentarios ó declaraciones.

Su sintaxis es la misma que en HTML comenzando por "<!--" y terminando con "-->".

Sección CDATA

Sirve para añadir contenido que no va a ser procesado al estilo de un comentario. Debe aparecer dentro del elemento raíz de un documento xml comenzando con la cadena de caracteres "<![CDATA[" y terminando con los caracteres "]]>". No se pueden anidar secciones CDATA.

Por ejemplo, el siguiente documento xml:



Custom code isn't available in PDF format

Este fichero XML no parece tener ninguna información de estilo asociada. Se muestra debajo el árbol del documento.

```
-<documento>
  <título>Documento con código HTML</título>
  -<contenido>
    <h1>Título de la página</h1> <p>Este es un párrafo de ejemplo.</p> <ul>
      <li>Elemento 1</li> <li>Elemento 2</li> <li>Elemento 3</li> </ul>
    </contenido>
  </documento>
```

Elaboración propia. Vista en Firefox

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<documento>
  <título>Documento con código HTML</título>
  ▼<contenido>
    <![CDATA[ <h1>Título de la página</h1> <p>Este es un párrafo de ejemplo.
      </p> <ul> <li>Elemento 1</li> <li>Elemento 2</li> <li>Elemento 3</li>
      </ul> ]]>
    </contenido>
  </documento>
```

Elaboración propia. Vista en Chrome

Espacios de nombres

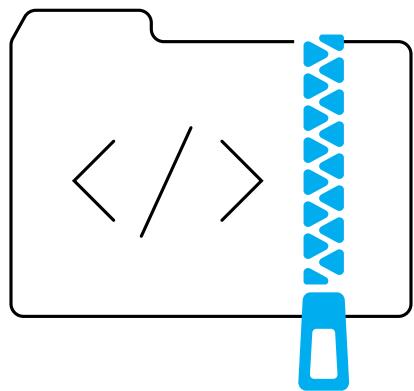
Se utilizan para evitar conflictos de nombres cuando se emplean elementos y atributos con los mismos nombres pero definidos en diferentes contextos.

Espacios de nombres

Supongamos que tenemos un documento XML que contiene información sobre estudiantes y cursos. Con **xmlns** (XML Name Space) podríamos definir un espacio de nombres para distinguir

entre los elementos relacionados con los estudiantes y los elementos relacionados con los cursos.

Por ejemplo, el siguiente documento xml:



Custom code isn't available in PDF format

- Se han definido dos espacios de nombres utilizando los atributos xmlns: estudiantes para los elementos relacionados con estudiantes y cursos para los elementos relacionados con cursos.
- Los elementos como <estudiantes:estudiante> y sus descendientes (<estudiantes:nombre>, <estudiantes:edad>) pertenecen al espacio de nombres estudiantes, mientras que los elementos como <cursos:curso> y sus descendientes (<cursos:nombre>, <cursos:profesor>) pertenecen al espacio de nombres cursos.
- Esto nos permite organizar claramente la información sobre estudiantes y cursos, evitando conflictos de nombres (en este ejemplo, elemento nombre) y facilitando la comprensión del documento XML.

Este fichero XML no parece tener ninguna información de estilo asociada. Se muestra debajo el árbol del documento.

```
--<escuela>
--<estudiantes:estudiante>
  <estudiantes:nombre>Juan Pérez</estudiantes:nombre>
  <estudiantes:edad>20</estudiantes:edad>
</estudiantes:estudiante>
--<estudiantes:estudiante>
  <estudiantes:nombre>María Gómez</estudiantes:nombre>
  <estudiantes:edad>22</estudiantes:edad>
</estudiantes:estudiante>
--<cursos:curso>
  <cursos:nombre>Matemáticas</cursos:nombre>
  <cursos:profesor>Dr. Rodríguez</cursos:profesor>
</cursos:curso>
--<cursos:curso>
  <cursos:nombre>Física</cursos:nombre>
  <cursos:profesor>Dra. Martínez</cursos:profesor>
</cursos:curso>
</escuela>
```

Elaboración propia. Vista en Firefox

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<escuela xmlns:estudiantes="http://ejemplo.com/estudiantes"
  xmlns:cursos="http://ejemplo.com/cursos">
  ▼<estudiantes:estudiante>
    <estudiantes:nombre>Juan Pérez</estudiantes:nombre>
    <estudiantes:edad>20</estudiantes:edad>
  </estudiantes:estudiante>
  ▼<estudiantes:estudiante>
    <estudiantes:nombre>María Gómez</estudiantes:nombre>
    <estudiantes:edad>22</estudiantes:edad>
  </estudiantes:estudiante>
  ▼<cursos:curso>
    <cursos:nombre>Matemáticas</cursos:nombre>
    <cursos:profesor>Dr. Rodríguez</cursos:profesor>
  </cursos:curso>
  ▼<cursos:curso>
    <cursos:nombre>Física</cursos:nombre>
    <cursos:profesor>Dra. Martínez</cursos:profesor>
  </cursos:curso>
</escuela>
```

Elaboración propia. Vista en Chrome

De esta manera, los espacios de nombres en XML nos permiten organizar y estructurar nuestros documentos XML de manera más clara y evitar ambigüedades cuando se utilizan nombres comunes en diferentes contextos.

Autoevaluación

Pregunta

01/03

Las elementos en XML siempre se componen de:

- Etiqueta de cierre
- Contenido
- Etiqueta de apertura
- Metadatos

Pregunta

02/03

Los espacios de nombres en XML:

- Permiten organizar y estructurar nuestros documentos XML de manera más clara.
- No aportan nada y es conveniente evitar su uso.
- Permiten evitar ambigüedades cuando se utilizan nombres comunes en diferentes contextos.

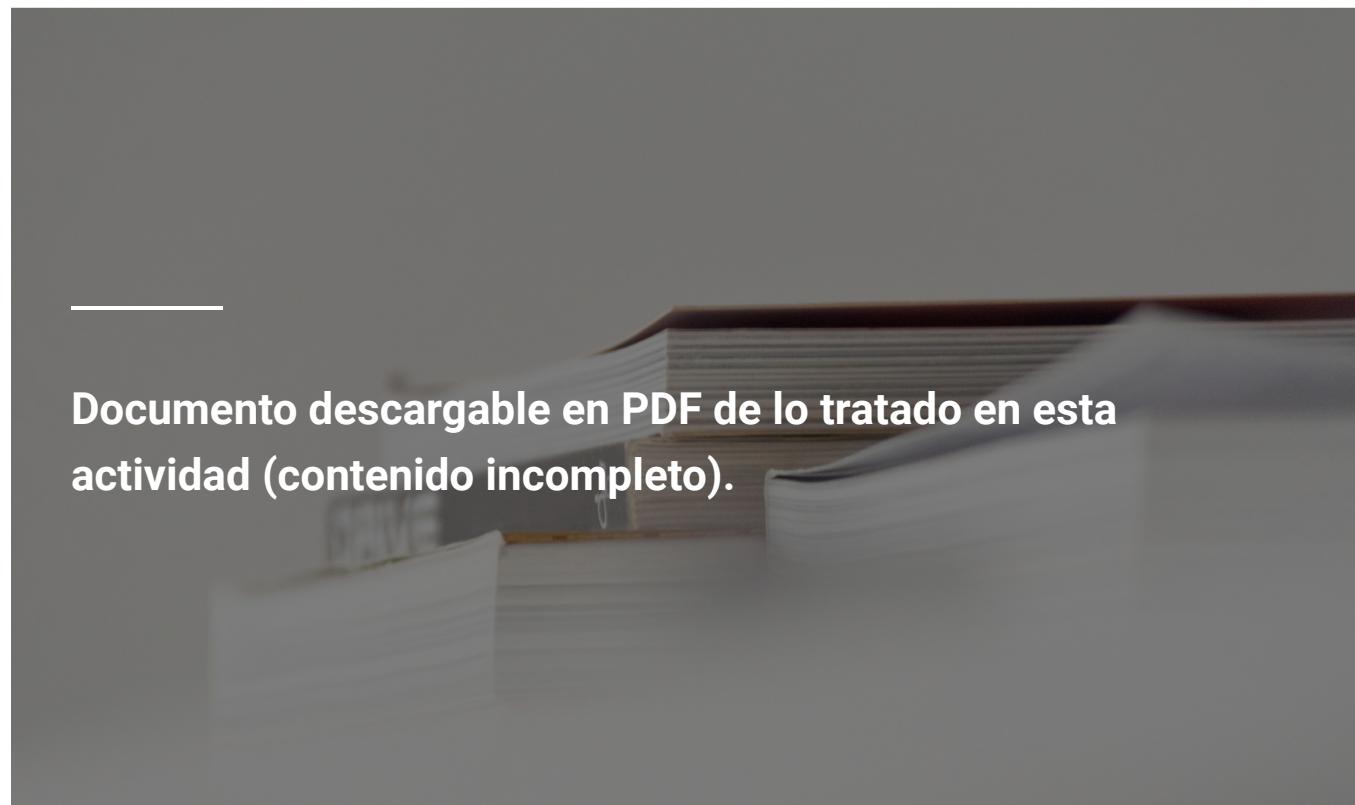
Pregunta

03/03

En un documentos XML se distinguen dos partes:

- Prólogo.
- No es correcto. Tendrá tantas partes como elementos diferentes existan.
- Contenido.
- Ejemplar.

Contenido PDF



El contenido del siguiente documento PDF está incompleto al no ofrecer los enlaces que aparecen en los apuntes ni las funciones de los elementos interactivos de la actividad. Su uso ha de ser por lo tanto, completado por dichos apuntes.

Puedes visualizar y descargar el contenido PDF a través de este enlace

IR AL ENLACE