

English to Spanish Translation by Trusted Translations, Inc.

Otras Traducciones de W3 en Spanish Translator Services

Conjunto de información XML (Segunda edición)

Este documento es una traducción de la [Recomendación del W3C](#) sobre [Conjunto de información XML \(Segunda edición\)](#)

**La versión inglesa de esta especificación es la única con valor normativo y puede encontrarse en: <http://www.w3.org/TR/xml-infoset>**



## Conjunto de información XML (Segunda edición)

### Recomendación del W3C - 4 de febrero de 2004

**Esta versión:**

<http://www.w3.org/TR/2004/REC-xml-infoset-20040204>

**Última versión:**

<http://www.w3.org/TR/xml-infoset>

**Versión anterior:**

<http://www.w3.org/TR/2003/PER-xml-infoset-20031210>

**Editores:**

John Cowan, [jcowan@reutershealth.com](mailto:jcowan@reutershealth.com)

Richard Tobin, [richard@cogsci.ed.ac.uk](mailto:richard@cogsci.ed.ac.uk)

Por favor, consulte la sección de [erratas](#) de este documento, que puede incluir algunas correcciones normativas.

Vea también las [traducciones](#).

[Copyright](#) ©1999-2004 W3C® (MIT, ERCIM, [Keio](#)), todos los derechos reservados. Se aplican las reglas de [responsabilidad](#), [marca registrada](#), [utilización de documentos](#) y [licencias de software](#) del W3C.

---

## Resumen

La presente especificación ofrece un conjunto de definiciones para usar en otras especificaciones que deban hacer referencia a la información contenida en un documento XML.

## Estado del presente documento

*Esta sección describe el estado del presente documento al momento de su publicación. El presente documento puede ser reemplazado por otros. Una lista de las publicaciones actuales del W3C y la última revisión del presente informe técnico puede hallarse en el [Índice de informes técnicos del W3C](#) en <http://www.w3.org/TR/>.*

Este documento es una [Recomendación](#) del W3C. Ha sido revisado por miembros del W3C y otras partes interesadas, y ha sido avalado como Recomendación del W3C por el Director. Es un documento estable y puede ser empleado como material de referencia o citado como referencia normativa en otro documento. El papel del W3C en la creación de la Recomendación es poner atención en la especificación y fomentar su amplia implementación. Esto mejora la funcionalidad y la interoperabilidad de la Web.

El presente documento actualiza el conjunto de información a fin de cubrir [XML 1.1](#) y

[Namespaces 1.1](#); aclara las consecuencias de ciertas clases de invalidez; y corrige algunos errores tipográficos. Es un producto de la [Actividad del W3C en XML](#). La versión en inglés de esta especificación es la única con valor normativo. Sin embargo, para ver las traducciones del presente documento, visite <http://www.w3.org/2003/03/Translations/byTechnology?technology=xml-infoset>.

En la [página de divulgación pública de derechos de propiedad intelectual](#) del Grupo de Trabajo es posible hallar documentación sobre propiedad intelectual que puede ser pertinente para esta recomendación.

Si encuentra errores en este documento, por favor informe de ellos a [www-xml-infoset-comments@w3.org](mailto:www-xml-infoset-comments@w3.org) (se encuentran disponibles [archivos](#) públicos). La lista de erratas para esta edición se encuentra disponible en <http://www.w3.org/2001/10/02/xml-infoset-errata.html>.

## Índice

### [1 Introducción](#)

### [2 Ítems de información](#)

#### [2.1 Ítem de información de documento](#)

#### [2.2 Ítems de información de elemento](#)

#### [2.3 Ítems de información de atributo](#)

#### [2.4 Ítems de información de instrucción de procesamiento](#)

#### [2.5 Ítems de información de referencia a entidad no expandida](#)

#### [2.6 Ítems de información de carácter](#)

#### [2.7 Ítems de información de comentario](#)

#### [2.8 Ítem de información de declaración de tipo de documento](#)

#### [2.9 Ítems de información de entidad no analizada](#)

#### [2.10 Ítems de información de notación](#)

#### [2.11 Ítems de información de espacio de nombres](#)

### [3 Conformidad](#)

#### [Apéndice A: Referencias](#)

#### [Apéndice B: Requerimientos de provisión de información de XML \(informativo\)](#)

#### [Apéndice C: Ejemplo \(informativo\)](#)

#### [Apéndice D: Lo que no está incluido en el conjunto de información](#)

#### [Apéndice E: Esquema RDF \(informativo\)](#)

---

## 1 Introducción

La presente especificación define un conjunto de datos abstracto llamado **Conjunto de información XML (Infoset)**. Su objetivo consiste en ofrecer un conjunto coherente de definiciones para usar en otras especificaciones que deban hacer referencia a la información contenida en un documento XML bien formado [\[XML\]](#).

Esta especificación no pretende ser exhaustiva; para decidir la inclusión de un ítem de información o una propiedad, el criterio principal empleado fue la utilidad que pudieran tener para futuras especificaciones. Tampoco constituye un conjunto de información mínima que deba ser devuelta por un procesador de XML.

Un documento XML posee un conjunto de información si el documento está bien formado y satisface las restricciones en materia de espacios de nombres que se describen [más adelante](#). Que el documento XML sea válido no es una condición para que tenga un conjunto de información.

Los conjuntos de información pueden crearse mediante otros métodos distintos al análisis de un documento XML, no descritos en la presente especificación. Véase [Conjuntos de información sintéticos](#), más adelante.

El conjunto de información de un documento XML está compuesto por cierta cantidad de **ítems de información**; el conjunto de información de todo documento XML bien formado contiene al menos un ítem de información de [documento](#) y varios otros. Un ítem de información es una descripción abstracta de cierta parte de un documento XML y cada ítem de información tiene asociado un conjunto de **propiedades** con nombre. En la presente especificación, los nombres de las propiedades se muestran entre corchetes, **[asi]**. La lista de los tipos de ítems de información se encuentra en la [sección 2](#).

La especificación del conjunto de información XML no exige ni favorece una interfaz o clase de interfaces en particular. Por razones de claridad y sencillez, la presente especificación presenta el conjunto de información en la forma de un árbol modificado, pero no es obligatorio presentar los conjuntos de información XML por medio de estructuras de árbol. Otros tipos de interfaces, entre ellas (sin que esta lista sea exhaustiva) interfaces basadas en eventos e interfaces basadas en consultas, podrán también presentar información compatible con el conjunto de información XML.

Los términos "conjunto de información" e "ítem de información" son de significado similar a los términos genéricos "árbol" y "nodo", según éstos se emplean en informática. Sin embargo, en esta especificación se emplea el primer par de términos para reducir el riesgo de confusión con otros modelos de datos concretos. No existe una correspondencia biunívoca entre los ítems de información y el DOM o el "árbol" y los "nodos" del modelo de datos XPath.

En esta especificación, el significado de las palabras "debe", "debería" y "puede" es el indicado en [\[RFC2119\]](#), con la salvedad de que no aparecen en mayúsculas.

## Versiones de XML

Diferentes versiones de la especificación de XML pueden indicar diferentes reglas para el análisis de los documentos. Por definición, el conjunto de información de un documento XML es el que se obtiene al analizar el documento según las reglas de la especificación cuya versión se corresponde con la del documento. Si un documento no indica su número de versión, se considerará que corresponde a la versión 1.0. Cuando un procesador de XML acepta un documento de una versión que no comprende, puede ocurrir que no sea capaz de producir el conjunto de información correcto.

## Espacios de nombres

Los documentos XML que, a pesar de estar técnicamente bien formados, no sean compatibles con lo indicado en [\[Namespaces\]](#), no se considera que tengan un conjunto de información propiamente dicho. Es decir, esta especificación no define un conjunto de información para aquellos documentos que tienen nombres de elementos o atributos donde los dos puntos se emplean en modos distintos a los prescritos en [\[Namespaces\]](#).

Además, esta especificación no define un conjunto de información para aquellos documentos que emplean referencias URI relativas en las declaraciones de espacio de nombres. Esto es así de conformidad con la decisión del Grupo de Interés Plenario de XML del W3C, descrito en [\[Relative Namespace URI References\]](#).

El valor de una propiedad [namespace name] es el valor normalizado del correspondiente atributo de espacio de nombres, sin que el procesador le aplique ulteriores mecanismos de escape de URI.

## Entidades

Un conjunto de información describe el documento XML al que corresponde con las referencias a entidades ya expandidas, es decir, representadas por los ítems de información correspondientes al texto de reemplazo. Sin embargo, en varias circunstancias el procesador no puede realizar esta expansión. Una entidad puede no estar declarada o ser irrecuperable. Un procesador no validador tal vez no lea todas las declaraciones, y aún cuando lo hiciera,

puede ocurrir que no expanda todas las entidades externas. En estos casos, para representar la referencia a la entidad se emplea un ítem de información de [referencia a entidad no expandida](#).

## Manejo de fin de línea

Los valores de todas las propiedades en el conjunto de información toman en cuenta la normalización de fin de línea descrita en [\[XML\]](#), 2.11 "Manejo de fin de línea".

## URIs base

Varios ítems de información tienen una propiedad [base URI] o [declaration base URI]. Estas propiedades se calculan de acuerdo a lo descrito en [\[XML Base\]](#). Obsérvese que la recuperación de un recurso puede suponer una redirección en el nivel del analizador (por ejemplo, en un resolvidor de entidades) o en un nivel inferior, en cuyo caso el URI base es el URI final empleado para recuperar el recurso después de todas las redirecciones.

El valor de estas propiedades no refleja aquellos mecanismos de escape de URI que puedan ser necesarios para la recuperación del recurso, pero puede incluir caracteres con mecanismos de escape si esos caracteres fueron especificados en el documento o devueltos por un servidor en el caso de una redirección.

En algunos casos (por ejemplo, un documento leído desde una cadena o una canalización) las reglas descritas en [\[XML Base\]](#) pueden llevar a que un URI base sea dependiente de la aplicación. En estos casos, la presente especificación no define el valor de la propiedad [base URI] o [declaration base URI].

Al resolver URI relativos, debería priorizarse la propiedad [base URI] por sobre los valores de los atributos xml:base, que pueden ser incoherentes en el caso de [conjuntos de información sintéticos](#).

## "Desconocido" y "Sin valor"

En ocasiones, algunas propiedades pueden contener el valor **desconocido** o **sin valor**, en cuyo caso se dice que el valor de la propiedad es desconocido o que la propiedad no tiene un valor, respectivamente. Estos valores son distintos entre sí y respecto de cualquier otro valor. En particular, son distintos de la cadena vacía, del conjunto vacío y de la lista vacía, cada uno de los cuales sencillamente no contiene miembros. En esta especificación no se utiliza el término **nulo**, ya que en algunas comunidades tiene connotaciones particulares que no siempre coinciden con las que aquí se pretenden.

## Incoherencias resultantes de la invalidez

Como ya hemos mencionado, un documento XML no necesita ser válido para tener un conjunto de información. Sin embargo, ciertas clases de invalidez afectan los valores asignados a algunas propiedades. Por ejemplo, puede ocurrir que falten declaraciones de entidades, notaciones, elementos y atributos. Las notaciones y los elementos pueden estar declarados varias veces (la presencia de varias declaraciones para entidades y atributos es válida). Un ID puede estar indefinido o definido varias veces. Cuando fuera pertinente, se señalará explícitamente en las definiciones de los ítems de información que se dan más adelante.

## Conjuntos de información sintéticos

La presente especificación describe el conjunto de información que resulta del análisis de un documento XML. Los conjuntos también pueden construirse por otros medios; por ejemplo, mediante el uso de una API como el DOM o por la transformación de un conjunto de información existente.

Un conjunto de información que corresponda a un documento real debe necesariamente ser coherente de diversas maneras; por ejemplo, la propiedad [in-scope namespaces] de un elemento debe ser coherente con las propiedades [namespace attributes] del elemento y sus ancestros. Para conjuntos de información contruidos por otros medios esta coherencia puede no ser cierta. En tal caso, no habrá un documento XML que se corresponda con el conjunto de información, y para serializar el conjunto será necesario resolver las incoherencias (por ejemplo, imprimiendo declaraciones de espacios de nombres que se correspondan con los espacios de nombres en ámbito).

## 2 Ítems de información

Un conjunto de información puede contener hasta once tipos diferentes de ítems de información, que se detallan en las secciones siguientes. Cada ítem de información contiene propiedades. Para facilidad de referencia, a cada propiedad se le asigna un nombre, indicado **[así]**. Para cada ítem de información se ofrecen enlaces a una definición y/o sintaxis en la Recomendación XML 1.0 [\[XML\]](#).

### 2.1 Ítem de información de documento

**Definición XML:** [documento](#) (Sección 2, Documentos)

**Sintaxis XML:** [1] [Documento](#) (Sección 2.1, Documentos XML bien formados)

En cada conjunto de información existe exactamente un **ítem de información de documento**, a partir de cuyas propiedades se accede a todos los demás ítems, ya sea directamente o indirectamente a través de las propiedades de otros ítems de información.

El ítem de información de documento tiene las siguientes propiedades:

1. **[children]** Una lista ordenada de ítems de información hijos, en el orden en que aparecen en el documento. La lista contiene exactamente un ítem de información de [elemento](#). La lista también contiene un ítem de información de [instrucción de procesamiento](#) para cada instrucción de procesamiento exterior al elemento de documento y un ítem de información de [comentario](#) para cada comentario exterior al elemento de documento. Se excluyen las instrucciones de procesamiento y los comentarios que estén dentro de la DTD. Si existe una declaración de tipo de documento, la lista también contiene un ítem de información de [declaración de tipo de documento](#).
2. **[document element]** El ítem de información de [elemento](#) que corresponde al elemento de documento.
3. **[notations]** Un conjunto no ordenado de ítems de información de [notación](#), uno por cada notación declarada en la DTD. Si una notación tiene más de una declaración, esta propiedad no tiene un valor.
4. **[unparsed entities]** Un conjunto no ordenado de ítems de información de [entidad no analizada](#), uno por cada entidad no analizada que se declare en la DTD.
5. **[base URI]** El URI base de la entidad del documento.
6. **[character encoding scheme]** El nombre del esquema de codificación de caracteres en el cual se expresa la entidad del documento.
7. **[standalone]** Una indicación de si el documento es o no es autónomo (*standalone*). Esta propiedad surge de la declaración opcional de autonomía en la declaración XML, al principio de la entidad del documento, y no tiene un valor si no existe esa declaración.
8. **[version]** Una cadena que representa la versión XML del documento. Esta propiedad se deriva de la declaración XML que puede estar presente al principio de la entidad del documento, y no tiene un valor si no existe una declaración XML.
9. **[all declarations processed]** Estrictamente hablando, esta propiedad no es parte del conjunto de información del documento, sino más bien una indicación respecto de si el procesador ha leído la DTD completa. El valor de la propiedad es booleano. Si fuera falso, entonces ciertas propiedades (que se indican en sus descripciones, a continuación) pueden tener valor desconocido. Si fuera verdadero, dichas propiedades nunca son

desconocidas.

## 2.2 Ítems de información de elemento

**Definición XML:** [elemento](#) (Sección 3, Estructuras Lógicas)

**Sintaxis XML:** [39] [Elemento](#) (Sección 3, Estructuras Lógicas)

Existe un **ítem de información de elemento** para cada elemento que aparece en el documento XML. Uno de los ítems de información de elemento es el valor de la propiedad [document element] del ítem de información de documento y se corresponde con la raíz del árbol de elementos, mientras que todos los demás ítems de información de elemento son accesibles recorriendo recursivamente la propiedad [children].

Un ítem de información de elemento tiene las siguientes propiedades:

1. **[namespace name]** El nombre del espacio de nombres, si lo hubiera, del tipo de elemento. Si el elemento no pertenece a un espacio de nombres, esta propiedad no tiene un valor.
2. **[local name]** La parte local del nombre de tipo de elemento, sin incluir un prefijo de espacio de nombres y los dos puntos que le siguen.
3. **[prefix]** El prefijo de espacio de nombres del nombre del tipo de elemento. Si el nombre no incluye un prefijo, esta propiedad no tiene un valor. Obsérvese que las aplicaciones que tienen en cuenta el espacio de nombres deberían usar el nombre del espacio de nombres y no el prefijo para identificar los elementos.
4. **[children]** Una lista ordenada de ítems de información de hijos, en el orden en que aparecen en el documento. Esta lista contiene ítems de información de [elemento](#), [instrucción de procesamiento](#), [referencia a entidad no expandida](#), [carácter](#) y [comentario](#), uno por cada elemento, instrucción de procesamiento, referencia a una entidad externa no procesada, carácter de datos y comentario que aparezca inmediatamente dentro del elemento actual. Si el elemento está vacío, esta lista no tiene miembros.
5. **[attributes]** Un conjunto no ordenado de ítems de información de [atributo](#), uno por cada atributo de este elemento (especificados o tomados por defecto de la DTD). Este conjunto no incluye declaraciones de espacios de nombres. Si el elemento no tiene atributos, este conjunto no tiene miembros.
6. **[namespace attributes]** Un conjunto no ordenado de ítems de información de [atributo](#), uno por cada declaración de espacio de nombres de este elemento (especificadas o tomadas por defecto de la DTD). Las declaraciones de la forma `xmlns=""` y `xmlns:name=""`, que anulan las declaraciones predeterminadas de espacio de nombres y prefijos, respectivamente, cuentan como declaraciones de espacio de nombres. La anulación de declaración de prefijo se agregó en [Namespaces in XML 1.1](#). Por definición, todos los atributos de espacio de nombres (incluidos los que llevan el nombre `xmlns`, cuya propiedad [prefix] no tiene un valor) tienen un URI de espacio de nombres igual a `http://www.w3.org/2000/xmlns/`. Si el elemento no tiene declaraciones de espacio de nombres, este conjunto no tiene miembros.
7. **[in-scope namespaces]** Un conjunto no ordenado de ítems de información de [espacio de nombres](#), uno por cada espacio de nombres vigente para este elemento. Este conjunto siempre contiene un ítem con el prefijo `xml`, que está implícitamente ligado al espacio de nombres `http://www.w3.org/XML/1998/namespace`, pero no contiene un ítem con el prefijo `xmlns` (que se usa para declarar espacios de nombres), ya que una aplicación no puede encontrar nunca un elemento o un atributo con ese prefijo. El conjunto incluye los ítems de espacio de nombres correspondientes a todos los miembros de [namespace attributes], salvo los que representen declaraciones de la forma `xmlns=""` o `xmlns:name=""`, que no declaran un espacio de nombres sino que anulan el espacio de nombres y los prefijos predeterminados. Al resolver los prefijos de nombres calificados, debería darse prioridad a esta propiedad por sobre la propiedad [namespace attributes] (pueden ser incoherentes en el caso de [conjuntos de información sintéticos](#)).
8. **[base URI]** El URI base del elemento.
9. **[parent]** El ítem de información de documento o de elemento que contiene a este ítem de

información en su propiedad [children].

## 2.3 Ítems de información de atributo

**Definición XML:** *atributo* (Sección 3.1, *Etiquetas de apertura, Etiquetas de cierre y Etiquetas de elemento vacío*)

**Sintaxis XML:** [41] *Atributo* (Sección 3.1, *Etiquetas de apertura, Etiquetas de cierre y Etiquetas de elemento vacío*)

Existe un **ítem de información de atributo** para cada atributo (especificado o tomado por defecto) de cada elemento en el documento, incluidos los que son declaraciones de espacio de nombres. Sin embargo, estos últimos aparecen como miembros de la propiedad [namespace attributes] de un elemento, y no como su propiedad [attributes].

Los atributos declarados en la DTD que no tengan valor predeterminado y que no estén especificados en la etiqueta de apertura del elemento no se representan como ítems de información de atributo.

Un ítem de información de atributo tiene las siguientes propiedades:

1. **[namespace name]** El nombre del espacio de nombres, si lo hubiera, del atributo. De lo contrario, esta propiedad no tiene un valor.
2. **[local name]** La parte local del nombre del atributo, sin incluir un prefijo de espacio de nombres ni los dos puntos que le siguen.
3. **[prefix]** El prefijo de espacio de nombres del nombre del atributo. Si el nombre no incluye un prefijo, esta propiedad no tiene un valor. Obsérvese que las aplicaciones que tienen en cuenta el espacio de nombres deberían usar el nombre del espacio de nombres y no el prefijo para identificar los atributos.
4. **[normalized value]** El valor normalizado del atributo (véase [3.3.3 Normalización del valor del atributo \[XML\]](#)).
5. **[specified]** Un indicador que señala si este atributo fue especificado en la etiqueta de apertura de su elemento o se tomó por defecto de la DTD.
6. **[attribute type]** Una indicación del tipo declarado para este atributo en la DTD. Los valores legítimos son ID, IDREF, IDREFS, ENTITY, ENTITIES, NMTOKEN, NMTOKENS, NOTATION, CDATA y ENUMERATION. Si el atributo no tiene una declaración, esta propiedad no tiene un valor. Si no se ha leído ninguna declaración, pero la propiedad [all declarations processed] del ítem de información de documento es falsa (de modo que puede haber una declaración que no haya sido leída), entonces el valor de esta propiedad es desconocido. Las aplicaciones deberían tratar los valores "sin valor" y "desconocido" como equivalentes de CDATA. El valor de esta propiedad no depende de la validez del valor del atributo.
7. **[references]** Si el tipo de atributo es ID, NMTOKEN, NMTOKENS, CDATA o ENUMERATION, esta propiedad no tiene un valor. Si el tipo de atributo es desconocido, el valor de esta propiedad es desconocido. En los otros casos (es decir, si el tipo de atributo es IDREF, IDREFS, ENTITY, ENTITIES o NOTATION), el valor de esta propiedad es una lista ordenada de los ítems de información de [elemento](#), [entidad no analizada](#) o [notación](#) a los que se hace referencia en el valor del atributo, en el orden en el que aparecen allí. En este caso, si el valor del atributo es sintácticamente inválido, esta propiedad no tiene un valor. Si el tipo es IDREF o IDREFS y alguno de los ID no aparece como valor del atributo ID en el documento, o si el tipo es ENTITY, ENTITIES o NOTATION y para alguna de las entidades o la notación no se ha leído ninguna declaración, entonces esta propiedad no tiene un valor o su valor es desconocido, dependiendo de si la propiedad [all declarations processed] del ítem de información de documento es verdadera o falsa. Si el tipo es IDREF o IDREFS y alguno de los ID aparece como valor de más de un atributo ID en el documento, o si el tipo es NOTATION y hay más de una declaración para la notación, entonces esta propiedad no tiene un valor.
8. **[owner element]** El ítem de información de elemento que contiene a este ítem de información en su propiedad [attributes].

## 2.4 Ítems de información de instrucción de procesamiento

**Definición XML:** [instrucción de procesamiento](#) (Sección 2.6, *Instrucciones de procesamiento*)

**Sintaxis XML:** [16] [IP](#) (Sección 2.6, *Instrucciones de procesamiento*)

Hay un **ítem de información de instrucción de procesamiento** para cada instrucción de procesamiento que aparece en el documento XML. No se consideran instrucciones de procesamiento la declaración XML y las declaraciones textuales de las entidades externas analizadas.

Un ítem de información de instrucción de procesamiento tiene las siguientes propiedades:

1. **[target]** Una cadena que representa la parte de la instrucción de procesamiento que corresponde a su destino (un nombre XML).
2. **[content]** Una cadena que representa el contenido de la instrucción de procesamiento, excluidos el destino y todo espacio en blanco que le siga inmediatamente. Si no hay tal contenido, el valor de esta propiedad será una cadena vacía.
3. **[base URI]** El URI base de la IP. Obsérvese que si un conjunto de información se serializa en la forma de un documento XML, no será posible conservar el URI base de ninguna de las instrucciones de procesamiento que aparecen originalmente en el nivel superior de una entidad externa, ya que no existe para las instrucciones de procesamiento una sintaxis correspondiente al atributo `xml:base` de los elementos.
4. **[notation]** El ítem de información de [notación](#) indicado por el destino. Si no existe una declaración para una notación con ese nombre, o si hay más de una, esta propiedad [all declarations processed] del ítem de información de documento es falsa (de modo que puede haber una declaración que no haya sido leída), entonces el valor de esta propiedad es desconocido.
5. **[parent]** El ítem de información de documento, de elemento o de declaración de tipo de documento que contiene a este ítem de información en su propiedad [children].

## 2.5 Ítems de información de referencia a entidad no expandida

**Definición XML:** Sección 4.4.3, [Incluido en caso de validación](#)

Un **ítem de información de referencia a entidad no expandida** sirve como marcador de posición mediante el cual un procesador de XML puede señalar el lugar donde no ha expandido una entidad analizada externa. Hay uno de estos ítems de información para cada referencia a una entidad general externa no expandida dentro del contenido de un elemento. Un procesador de XML validador, o no validador que procese un documento válido y lea todas las entidades generales externas, jamás generará ítems de información de referencia a entidad no expandida.

Un ítem de información de referencia a entidad no expandida tiene las siguientes propiedades:

1. **[name]** El nombre de la entidad a la que se hace referencia.
2. **[system identifier]** El identificador de sistema de la entidad, como aparece en la declaración de la entidad, sin ningún mecanismo adicional de escape de URI aplicado por el procesador. Si la entidad no tiene una declaración, esta propiedad no tiene un valor. Si no se ha leído ninguna declaración, pero la propiedad [all declarations processed] del ítem de información de documento es falsa (de modo que puede haber una declaración que no haya sido leída), entonces el valor de esta propiedad es desconocido.
3. **[public identifier]** El identificador público de la entidad, normalizado según lo descrito en [4.2.2 Entidades externas \[XML\]](#). Si la entidad no tiene una declaración, o la declaración no incluye un identificador público, esta propiedad no tiene un valor. Si no se ha leído ninguna declaración, pero la propiedad [all declarations processed] del ítem de

información de documento es falsa (de modo que puede haber una declaración que no haya sido leída), entonces el valor de esta propiedad es desconocido.

4. **[declaration base URI]** El URI base en relación al cual debe resolverse el identificador del sistema (es decir, el URI base del recurso dentro del cual se encuentra la declaración de la entidad). Esta propiedad es desconocida o no tiene valor en las mismas circunstancias que la propiedad [system identifier].
5. **[parent]** El ítem de información de elemento que contiene a este ítem de información en su propiedad [children].

## 2.6 Ítems de información de carácter

**Sintaxis XML:** [2] [Carácter](#) (Sección 2.2, Caracteres)

Hay un **ítem de información de carácter** para cada carácter de datos que aparece en el documento, ya sea literalmente, como referencia de carácter o dentro de una sección CDATA.

Cada carácter es un ítem de información lógicamente separado, pero las aplicaciones XML tienen libertad para agrupar los caracteres en secuencias más largas cuando fuera necesario o deseable.

Un ítem de información de carácter tiene las siguientes propiedades:

1. **[character code]** El código del carácter según el estándar ISO 10646 (en el intervalo de 0 a #x10FFFF, aunque no todos los valores en este intervalo son códigos de carácter XML válidos).
2. **[element content whitespace]** Un valor booleano que indica si el carácter es espacio en blanco dentro del contenido del elemento (véase [XML](#), 2.10 "Manejo de espacios en blanco"). Obsérvese que para los procesadores de XML validadores es *obligatorio* suministrar esta información. Si no existe una declaración para el elemento contenedor, o si hay más de una, esta propiedad no tiene un valor para los caracteres de espacio en blanco. Si no se ha leído ninguna declaración, pero la propiedad [all declarations processed] del ítem de información de documento es falsa (de modo que puede haber una declaración que no haya sido leída), entonces el valor de esta propiedad es desconocido para los caracteres de espacio en blanco. La propiedad es siempre falsa para caracteres que no sean espacio en blanco.
3. **[parent]** El ítem de información de elemento que contiene a este ítem de información en su propiedad [children].

## 2.7 Ítems de información de comentario

**Definición XML:** [comentario](#) (Sección 2.5, Comentarios)

**Sintaxis XML:** [15] [Comentario](#) (Sección 2.5, Comentarios)

Hay un **ítem de información de comentario** para cada comentario XML contenido en el documento original, salvo los que aparecen en la DTD (que no están representados).

Un ítem de información de comentario tiene las siguientes propiedades:

1. **[content]** Una cadena que representa el contenido del comentario.
2. **[parent]** El ítem de información de documento o de elemento que contiene a este ítem de información en su propiedad [children].

## 2.8 Ítem de información de declaración de tipo de documento

**Definición XML:** [declaración de tipo de documento](#) (sección 2.8, Prólogo y Declaración de tipo de documento)

**Sintaxis XML:** [28] [declaración de tipo de documento](#) (sección 2.8, Prólogo y Declaración

*de tipo de documento)*

Si el documento XML tiene una declaración de tipo de documento, entonces el conjunto de información contiene un único **ítem de información de declaración de tipo de documento**. Obsérvese que las entidades y las notaciones se incluyen como propiedades del ítem de información de documento, no del ítem de información de declaración de tipo de documento.

El ítem de información de declaración de tipo de documento tiene las siguientes propiedades:

1. **[system identifier]** El identificador de sistema del subconjunto externo, como aparece en la declaración DOCTYPE, sin ningún mecanismo adicional de escape de URI aplicado por el procesador. Si no existe un subconjunto externo, esta propiedad no tiene un valor.
2. **[public identifier]** El identificador público del subconjunto externo, normalizado según lo descrito en [4.2.2 Entidades externas \[XML\]](#). Si no existe un subconjunto externo o éste no tiene identificador público, esta propiedad no tiene un valor.
3. **[children]** Una lista ordenada de ítems de información de [instrucción de procesamiento](#) que representan las instrucciones de procesamiento que aparecen en la DTD, en el orden en que aparecen en el documento original. Los ítems del subconjunto interno de la DTD aparecen antes de los del subconjunto externo.
4. **[parent]** El ítem de información de documento.

## 2.9 Ítems de información de entidad no analizada

**Definición XML:** [entidad](#) (sección 4, Estructuras físicas)

**Sintaxis XML:** [71] [GEDecl](#) (sección 4.2, Entidades)

Existe un **ítem de información de entidad no analizada** para cada entidad general no analizada declarada en la DTD.

Un ítem de información de entidad no analizada tiene las siguientes propiedades:

1. **[name]** El nombre de la entidad.
2. **[system identifier]** El identificador de sistema de la entidad, como aparece en la declaración de la entidad, sin ningún mecanismo adicional de escape de URI aplicado por el procesador.
3. **[public identifier]** El identificador público de la entidad, normalizado según lo descrito en [4.2.2 Entidades externas \[XML\]](#). Si la entidad no tiene un identificador público, esta propiedad no tiene un valor.
4. **[declaration base URI]** El URI base en relación al cual debe resolverse el identificador del sistema (es decir, el URI base del recurso dentro del cual se encuentra la declaración de la entidad).
5. **[notation name]** El nombre de notación asociado con la entidad.
6. **[notation]** El ítem de información de [notación](#) designado por el nombre de notación. Si no existe una declaración para una notación con ese nombre, o si hay más de una, esta propiedad no tiene un valor. Si no se ha leído ninguna declaración, pero la propiedad [all declarations processed] del ítem de información de documento es falsa (de modo que puede haber una declaración que no haya sido leída), entonces el valor de esta propiedad es desconocido.

## 2.10 Ítems de información de notación

**Definición XML:** [notación](#) (sección 4.7, Notaciones)

**Sintaxis XML:** [82] [Declaración de notación](#) (sección 4.7, Notaciones)

Hay un **ítem de información de notación** para cada notación declarada en la DTD.

Un ítem de información de notación tiene las siguientes propiedades:

1. **[name]** El nombre de la notación.
2. **[system identifier]** El identificador de sistema de la notación, como aparece en la declaración de la notación, sin ningún mecanismo adicional de escape de URI aplicado por el procesador. Si no se indicó un identificador de sistema, esta propiedad no tiene un valor.
3. **[public identifier]** El identificador público de la notación, normalizado según lo descrito en [4.2.2 Entidades externas \[XML\]](#). Si la notación no tiene identificador público, esta propiedad no tiene un valor.
4. **[declaration base URI]** El URI base en relación con el cual debe resolverse el identificador de sistema (es decir, el URI base del recurso dentro del cual se encuentra la declaración de la notación).

## 2.11 Ítems de información de espacio de nombres

Cada elemento en el documento tiene un **ítem de información de espacio de nombres** para cada espacio de nombres en cuyo alcance se encuentre dicho elemento.

Un ítem de información de espacio de nombres tiene las siguientes propiedades:

1. **[prefix]** El prefijo cuya ligadura describe este ítem. Sintácticamente, es la parte del nombre del atributo que está a continuación del prefijo `xmlns:`. Si el nombre de atributo es sencillamente `xmlns`, de modo que es la declaración del espacio de nombres predeterminado, esta propiedad no tiene un valor.
2. **[namespace name]** El nombre del espacio de nombres al que está ligado el prefijo.

## 3 Conformidad

Puesto que el propósito del conjunto de información es brindar un conjunto de definiciones, la conformidad es una propiedad de las especificaciones que usan estas definiciones y no de sus implementaciones.

Las especificaciones referidas a conjuntos de información deben:

- Indicar los ítems de información y propiedades necesarios para implementar la especificación. (Indirectamente, esto impone requisitos de conformidad a los procesadores empleados para implementar la especificación).
- Especificar lo que deba hacerse con los otros ítems de información y propiedades (por ejemplo, pasarlos sin modificaciones).
- Señalar toda información requerida de un documento XML que no esté definida por el conjunto de información.
- Señalar cualquier diferencia (algo que debería evitarse) en el uso de los términos definidos por esta especificación.

Si una especificación permite la construcción de un conjunto de información con incoherencias, según lo descrito precedentemente en [Conjuntos de información sintéticos](#) puede describir el modo de resolver dichas incoherencias, y debería hacerlo si especifica modos de serializar el conjunto de información.

## Apéndice A: Referencias

### Referencias normativas

#### ISO/IEC 10646

ISO (Organización Internacional para la Normalización). *ISO/IEC 10646-1:2000. Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane* e *ISO/IEC 10646-2:2001. Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 2: Supplementary Planes*, incluidas todas las enmiendas, sustituciones por ediciones nuevas o ampliaciones que

surjan de agregar partes nuevas. [Ginebra]: Organización Internacional para la Normalización. (Visite <http://www.iso.ch> para acceder a la última versión).

#### **Espacios de nombres**

*Namespaces in XML*, W3C, eds. Tim Bray, Dave Hollander y Andrew Layman. 14 de enero de 1999. Disponible en <http://www.w3.org/TR/REC-xml-names>.

#### **Namespaces 1.1**

*Namespaces in XML 1.1*, W3C, eds. Tim Bray, Dave Hollander, Andrew Layman, Richard Tobin. 4 de febrero de 2004. Disponible en <http://www.w3.org/TR/xml-names11>.

#### **RFC2119**

*Key words for use in RFCs to Indicate Requirement Levels*, ed. S. Bradner. Marzo de 1997. Disponible en <http://www.ietf.org/rfc/rfc2119.txt>.

#### **XML**

*Extensible Markup Language (XML) 1.0 (Third Edition)*, W3C, eds. Tim Bray, Jean Paoli, C.M. Sperberg-McQueen, Eve Maler, François Yergeau. 4 de febrero de 2004. Disponible en <http://www.w3.org/TR/REC-xml>.

#### **XML 1.1**

*Extensible Markup Language (XML) 1.1*, W3C, eds. Tim Bray, Jean Paoli, C.M. Sperberg-McQueen, Eve Maler, John Cowan, François Yergeau. 4 de febrero de 2004. Disponible en <http://www.w3.org/TR/xml11>.

#### **XML Base**

*XML Base*, W3C, ed. Jonathan Marsh. Febrero de 2000. Disponible en <http://www.w3.org/TR/xmlbase>.

## Referencias informativas

#### **DOM**

*Document Object Model (DOM) Level 1 Specification*, W3C, eds. Vidur Apparao, Steve Byrne, Mike Champion, et al. 1 de octubre de 1998. Disponible en <http://www.w3.org/TR/REC-DOM-Level-1>.

#### **XPointer-Liaison**

*XPointer-Information Set Liaison Statement*, W3C, ed. Steven J. DeRose. 24 de febrero de 1999. Disponible en <http://www.w3.org/TR/NOTE-xptr-infoset-liaison>.

#### **Referencias URI relativas en espacios de nombres**

*Results of W3C XML Plenary Ballot on relative URI References in namespace declarations, 3-17 July 2000*, W3C, eds. Dave Hollander, C. M. Sperberg-McQueen. 6 de septiembre de 2000. Disponible en <http://www.w3.org/2000/09/xppa>.

#### **Esquema RDF para el conjunto de información XML**

*RDF Schema for the XML Information Set*, W3C, ed. Richard Tobin. 6 de abril de 2001. Disponible en <http://www.w3.org/TR/xml-infoset-rdfs>.

## Apéndice B: Requerimientos de provisión de información de XML (informativo)

Aunque la Recomendación XML [XML] se ocupa ante todo de la sintaxis de XML, también incluye algunos requisitos concretos de provisión de información para los procesadores de XML.

Estos requerimientos incluyen notificar de los errores, que están fuera del alcance de esta especificación, y suministrar información sobre el documento. Todos los requerimientos de XML para la provisión de información del documento se han integrado en esta especificación; los números entre paréntesis remiten a las secciones de la Recomendación XML:

1. Un procesador de XML siempre debe suministrar todos los caracteres en un documento que no sean parte del marcado de la aplicación (2.10).
2. Un procesador validador de XML debe informar a la aplicación de todos los datos de carácter en el documento que sean espacio en blanco dentro del contenido de un elemento (2.10).
3. Un procesador de XML debe normalizar los fines de línea a LF antes de pasarlos a la aplicación (2.11).

4. Un procesador de XML debe normalizar el valor de los atributos según las reglas descritas en la cláusula 3.3.3 antes de pasarlos a la aplicación.
5. Un procesador de XML debe pasar a la aplicación los nombres e identificadores externos (identificadores de sistema, identificadores públicos o ambos) de las notaciones declaradas (4.7).
6. Cuando el nombre de una entidad no analizada aparece como valor explícito o predeterminado de un atributo ENTITY o ENTITIES, un procesador de XML debe suministrar a la aplicación los nombres, identificadores de sistema y (si están presentes) identificadores públicos tanto de la entidad como de su notación (4.6, 4.7).
7. Un procesador de XML debe pasar a la aplicación las instrucciones de procesamiento (2.6).
8. Un procesador de XML (que necesariamente será del tipo no validador) que no incluya el texto de reemplazo de una entidad analizada externa en lugar de una referencia a entidad debe notificar a la aplicación haber reconocido la entidad, pero no haberla leído (4.4.3).
9. Un procesador validador de XML debe incluir el texto de reemplazo de una entidad en lugar de una referencia a entidad (5.2).
10. Un procesador de XML debe suministrar el valor predeterminado de los atributos declarados en la DTD para un tipo de elemento dado que no aparezcan en la etiqueta de apertura del elemento (3.3.2).

## Apéndice C: Ejemplo (informativo)

En el siguiente documento XML de ejemplo:

```
<?xml version="1.0"?>
<msg:message doc:date="19990421"
                xmlns:doc="http://doc.example.org/namespaces/doc"
                xmlns:msg="http://message.example.org/"
>Phone home!</msg:message>
```

El conjunto de información de este documento XML contiene los siguientes ítems de información:

- Un ítem de información de [documento](#).
- Un ítem de información de [elemento](#) cuyo nombre del espacio de nombres es "http://message.example.org/", su parte local es "message" y su prefijo es "msg".
- Un ítem de información de [atributo](#) cuyo nombre del espacio de nombres es "http://doc.example.org/namespaces/doc", su parte local es "date", su prefijo es "doc" y su valor normalizado es "19990421".
- Tres ítems de información de [espacio de nombres](#) para los espacios de nombres http://www.w3.org/XML/1998/namespace, http://doc.example.org/namespaces/doc y http://message.example.org/.
- Dos ítems de información de [atributo](#) para los atributos de espacio de nombres.
- Once ítems de información de [carácter](#) para los datos de carácter.

## Apéndice D: Lo que no está incluido en el conjunto de información

La siguiente información no está representada en la versión actual del conjunto de información XML (esta lista no pretende ser exhaustiva):

1. Los modelos de contenido de elementos, a partir de las declaraciones ELEMENT en la DTD.
2. El agrupamiento y orden de las declaraciones de atributo en las declaraciones ATTLIST.
3. El nombre de tipo de documento.
4. El espacio en blanco exterior al elemento de documento.
5. El espacio en blanco inmediatamente siguiente al nombre de destino de una instrucción de procesamiento.
6. El hecho de que los caracteres estén representen mediante referencias de caracteres.

7. La diferencia entre las dos formas de un elemento vacío: `<foo` y `<foo></foo>`.
8. El espacio en blanco dentro de las etiquetas de apertura (distinto del espacio en blanco significativo en los valores de atributo) y de las etiquetas de cierre.
9. La diferencia entre terminaciones de línea CR, CR-LF y LF.
10. El orden de los atributos dentro de una etiqueta de apertura.
11. El orden de las declaraciones dentro de la DTD.
12. Los límites de las secciones condicionales dentro de la DTD.
13. Los límites de las entidades de parámetro dentro de la DTD.
14. Los comentarios en la DTD.
15. La ubicación de las declaraciones (ya sea en subconjuntos internos o externos o en entidades de parámetros).
16. Toda declaración ignorada, incluidas aquellas que estén dentro de una sección condicional IGNORE, así como las declaraciones de entidades y atributos que han sido ignoradas por existir declaraciones previas que las anulan.
17. El tipo de comillas (simples o dobles) usadas para encerrar los valores de los atributos.
18. Los límites de las entidades analizadas generales.
19. Los límites de las secciones marcadas como CDATA.
20. El valor predeterminado de los atributos declarados en la DTD.

## Apéndice E: Esquema RDF (informativo)

Véase en [Esquema RDF para el conjunto de información XML](#) una caracterización formal del conjunto de información.