

# RSS 0.91 Spec, revision 3

Netscape Communications  
Primary Author: Dan Libby  
July 10, 1999

## Notes

Files must be 100% valid XML. We're trying to move towards a more standard format, and to this end we have included several tags from the popular <scriptingNews> format. We have also ensured that this version is 100% valid XML. We did this by requiring that a DOCTYPE tag be included, and validating each RSS document against that DTD. This means that it is not enough for an RSS document to be "well-formed". It must also be "valid" with respect to its DTD.

No mixed content tags. We are specifically not including any tags that contain mixed content in RSS 0.91. This means that each tag either contains sub-tags only, or text only, not a combination. This is both because we want to keep the format simple, and because our current validation system is not able to handle this type of tag. We also are not allowing any HTML markup beyond the commonly used entities such as &quot; A full list of these are defined in the RSS 0.91 DTD.

New tags for syndication community. Our validator will now allow several new tags through the system, though most of them will not actually be used by Netcenter. However, these may work when syndicating content to other sites. These tags are noted explicitly in the spec as "ignored."

RDF references removed. RSS was originally conceived as a metadata format providing a summary of a website. Two things have become clear: the first is that providers want more of a syndication format than a metadata format. The structure of an RDF file is very precise and must conform to the RDF data model in order to be valid. This is not easily human-understandable and can make it difficult to create useful RDF files. The second is that few tools are available for RDF generation, validation and processing. For these reasons, we have decided to go with a standard XML approach.

## Specification

Tags in alphabetical order.

### <channel>

#### Description

information about a particular channel. Everything pertaining to an individual channel is contained within this tag.

#### Netcenter Usage

Currently displayed on "My Netscape". May use in other locations in the future.

#### Attributes

none

## Sub-elements:

- required:
  - [description](#)
  - [language](#)
  - [link](#)
  - [title](#)
- optional:
  - [copyright](#)
  - [docs](#)
  - [image](#)
  - [item](#)
  - [lastBuildDate](#)
  - [managingEditor](#)
  - [pubDate](#)
  - [rating](#)
  - [skipDays](#)
  - [skipHours](#)
  - [textInput](#)
  - [webMaster](#)

## Examples

See [example 1](#)

---

## <copyright>

### Description

copyright string

### Netcenter Usage

ignored

### Attributes

none

### Sub-elements:

none

## Examples

See [example 2](#)

## <day>

### Description

The day of the week, spelled out in English.

### Netcenter Usage

ignored

### Attributes

none

### Sub-elements:

none

### Examples

See [example 2](#)

---

## <description>

### Description

a plain text description of an item, channel, image, or textinput.

### Netcenter Usage

displayed as appropriate depending on context.

### Attributes

none

### Sub-elements:

none

### Examples

See [example channels](#)

---

## <docs>

### Description

This tag should contain a URL that references a description of the channel.

### Netcenter Usage

ignored

### Attributes

none

### Sub-elements:

none

### Examples

See [example 2](#)

---

## <!DOCTYPE>

### Description

Document Type Identifier. This is an XML tag that identifies where to find the definition for this format. It should follow the [xml](#) tag. The full DTD is [here](#).

### Netcenter Usage

required to ensure document validity

### Attributes

- 1 of these two formats is required:
  - `rss PUBLIC "-//Netscape Communications//DTD RSS 0.91//EN" ""`
  - `rss PUBLIC "-//Netscape Communications//DTD RSS 0.91//EN" "http://my.netscape.com/publish/formats/rss-0.91.dtd"`

### Sub-elements:

none

### Examples

```
<!DOCTYPE rss PUBLIC "-//Netscape Communications//DTD RSS 0.91//EN"
```

```
"http://my.netscape.com/publish/formats/rss-0.91.dtd">
```

```
<!DOCTYPE rss PUBLIC "-//Netscape Communications//DTD RSS 0.91//EN" "">
```

---

## <height>

### Description

Specifies the height of an [image](#). Should be an integer value.

### Netcenter Usage

The value must be between 1 and 400. If omitted, the default value is 31.

### Attributes

none

### Sub-elements:

none

### Examples

See [image](#)

---

## <hour>

### Description

Specifies an hour of the day. Should be an integer value between 0 and 23. See [skipHours](#).

### Netcenter Usage

ignored

### Attributes

none

### Sub-elements:

none

### Examples

See [skipHours](#)

## <image>

### Description

Specifies an image associated with a [channel](#).

### Netcenter Usage

Optionally (user preference) display an image along with the channel content.

### Attributes

none

### Sub-elements:

- required:
  - [url](#)
  - [link](#)
  - [title](#)
- optional:
  - [description](#)
  - [width](#)
  - [height](#)

### Examples

```
<image>  
<url>http://my.site.com/images/1.gif</url>  
<link>http://my.site.com/index.html</link>  
<title>my image alt text</title>  
</image>
```

```
<image>  
<url>http://my.site.com/images/1.gif</url>  
<link>http://my.site.com/index.html</link>  
<title>my image alt text</title>  
<width>120</width>  
<height>200</height>  
</image>
```

---

## <item>

### Description

An item that is associated with a [channel](#). The item should represent a web-page, or subsection within a web page. It should have a unique URL associated with it. Each item must contain a title and a link. A description is optional.

## Netcenter Usage

generates a list of links. The description, if supplied, may optionally be viewed by the user as plain text beneath the link. Also, a maximum of 15 items per channel is enforced at this time.

### Attributes

none

### Sub-elements:

- required:
  - [title](#)
  - [link](#)
- optional:
  - [description](#)

### Examples

```
<item>
<title>Item #1</title>
<link>http://my.site.com/story1/index.html</link>
</item>
```

```
<item>
<title>Item #2</title>
<link>http://my.site.com/story2/index.html</link>
<description>Some stuff about this item</description>
</item>
```

---

## <language>

### Description

Specifies the language of a [channel](#). See [supported language codes](#)

### Netcenter Usage

used to assist user with determining correct page encoding

### Attributes

none

### Sub-elements:

none

## Examples

See [example 1](#)

---

## <lastBuildDate>

### Description

The last time the channel was modified.

### Netcenter Usage

ignored

### Attributes

none

### Sub-elements:

none

## Examples

See [example 2](#)

---

## <link>

### Description

This is a url that a user is expected to click on, as opposed to a [<url>](#) that is for loading a resource, such as an image.

### Netcenter Usage

must start with either "http://" or "ftp://". All other urls are considered invalid.

### Attributes

none

### Sub-elements:

none

## Examples



See [examples](#)

---

## <managingEditor>

### Description

The email address of the managing editor of the site, the person to contact for editorial inquiries

### Netcenter Usage

ignored

### Attributes

none

### Sub-elements:

none

### Examples

See [example 2](#)

---

## <name>

### Description

The name of an object, corresponding to the "name" attribute of an HTML <INPUT> element. Currently, this only applies to [textinput](#).

### Netcenter Usage

generates "name" attribute in html form

### Attributes

none

### Sub-elements:

none

### Examples

See [textinput](#)

## <pubDate>

### Description

Date when channel was published.

### Netcenter Usage

ignored

### Attributes

none

### Sub-elements:

none

### Examples

See [example 2](#)

---

## <rating>

### Description

- recommended links rating agencies:
  - <http://www.w3.org/PICS/raters.htm> (W3 maintained list of rating agency links)
  - RSACi <http://www.rsac.org> (Click on 'register' link)
  - SafeSurf <http://www.safesurf.com/> <http://www.safesurf.com/classify/index.html> (direct)
- User actions:
  - Obtain a rating for your site from a well-known rating agency (eg RSACi, SafeSurf)
  - Copy rating data into RSS file. Include only the data within the 'content=' attribute.
- Expected format:
  - starts with "(PICS-1.1"

### Netcenter Usage

ignored. May use in the future to dynamically decide page rating.

### Attributes

none

### Sub-elements:

none

## Examples

```
Tag obtained from rating agency: <META http-equiv="PICS-Label" content='(PICS-1.1
"http://www.classify.org/safesurf/" 1 r (SS~~000 1))'>
RSS Rating tag: <rating>(PICS-1.1 "http://www.classify.org/safesurf/" 1 r (SS~~000
1))</rating>
```

---

## <rss>

### Description

Identifies begin and end of rss content.

### Netcenter Usage

identifies content type

### Attributes

- required:
  - version (must be 0.91)

### Sub-elements:

- required:
  - [channel](#)

## Examples

```
<rss version="0.91">
<channel>
...
</channel>
</rss>
```

---

## <skipDays>

### Description

A list of <day>s of the week, in English, indicating the days of the week when your channel will not be updated. As with activeHours, if you know your channel will never be updated on Saturday or Sunday, for example

### Netcenter Usage

ignored

## Attributes

none

## Sub-elements:

- required:
  - [day](#)

## Examples

```
<skipDays>  
<day>Saturday</day>  
<day>Sunday</day>  
</skipDays>
```

---

## <skipHours>

### Description

A list of <hour>s indicating the hours in the day, GMT, when the channel is unlikely to be updated. If this sub-item is omitted, the channel is assumed to be updated hourly.

### Netcenter Usage

ignored

## Attributes

none

## Sub-elements:

- required:
  - [hour](#)

## Examples

```
<skipHours>  
<hour>6</hour>  
<hour>7</hour>  
<hour>8</hour>  
<hour>9</hour>  
<hour>10</hour>  
<hour>11</hour>  
</skipHours>
```

---

## <textinput>

### Description

An input field for the purpose of allowing users to submit queries back to the publisher's site. This element should have a title, a link (to a cgi or other processor), a description containing some instructions, and a name, to be used as the name in the HTML tag `<input type=text name="[name]">`

### Netcenter Usage

Displays form for submission back to publisher.

### Attributes

none

### Sub-elements:

- required:
  - [title](#)
  - [link](#)
  - [description](#)
  - [name](#)

### Examples

```
<textinput>
<title>Search Now!</title>
<description>Enter your search &lt;terms&gt;</description>
<name>find</name>
<link>http://my.site.com/search.cgi</link>
</textinput>
```

---

## <title>

### Description

An identifying string for a resource. When used in an item, this is the name of the item's link. When used in an image, this is the Alt text for the image. When used in a channel, this is the channel's title. When used in a textinput, this is the the textinput's title.

### Netcenter Usage

displayed as appropriate depending on context.

### Attributes

none

### **Sub-elements:**

none

### **Examples**

See [examples](#)

---

## **<url>**

### **Description**

Location to load a resource from. Note that this is slightly different from the [link](#) tag, which specifies where a user should be re-directed to if a resource is selected.

### **Netcenter Usage**

must start with either "http://" or "ftp://". All other urls are considered invalid.

### **Attributes**

none

### **Sub-elements:**

none

### **Examples**

See [image](#)

---

## **<webMaster>**

### **Description**

The email address of the webmaster for the site, the person to contact if there are technical problems with the channel.

### **Netcenter Usage**

ignored

### **Attributes**

none

## Sub-elements:

none

## Examples

See [example 2](#)

---

## <width>

### Description

Specifies the width of an [image](#). Should be an integer value.

### Netcenter Usage

The value must be between 1 and 144. If omitted, the default value is 88.

### Attributes

none

## Sub-elements:

none

## Examples

See [image](#)

---

## <?xml?>

### Description

Identifies this as an XML document and specifies encoding. see [w3c](#) Note that this must be on the first line of the document.

### Netcenter Usage

required for XML compliance.

### Attributes

- version: must be "1.0"
- encoding: see [list](#)

**Sub-elements:**

none

**Example usage:**

```
<?xml version="1.0"?> <?xml version="1.0" encoding="utf-8"?>
<?xml version="1.0" encoding="Shift_JIS"?>
```

---

**Example 1 - Simple**

```
<?xml version="1.0"?>
<!DOCTYPE rss SYSTEM "http://my.netscape.com/publish/formats/rss-0.91.dtd">
<rss version="0.91">
<channel>
<language>en</language>
<description>News and commentary from the cross-platform scripting community.</description>
<link>http://www.scripting.com/</link>
<title>Scripting News</title>
<image>
<link>http://www.scripting.com/</link>
<title>Scripting News</title>
<url>http://www.scripting.com/gifs/tinyScriptingNews.gif</url>
</image>
</channel>
</rss>
```

**Example 2 - Complete**

```
<?xml version="1.0"?>
<!DOCTYPE rss SYSTEM "http://my.netscape.com/publish/formats/rss-0.91.dtd">
<rss version="0.91">
<channel>
<copyright>Copyright 1997-1999 UserLand Software, Inc.</copyright>
<pubDate>Thu, 08 Jul 1999 07:00:00 GMT</pubDate>
<lastBuildDate>Thu, 08 Jul 1999 16:20:26 GMT</lastBuildDate>
<docs>http://my.userland.com/stories/storyReader$11</docs>
<description>News and commentary from the cross-platform scripting community.</description>
<link>http://www.scripting.com/</link>
<title>Scripting News</title>
<image>
<link>http://www.scripting.com/</link>
<title>Scripting News</title>
<url>http://www.scripting.com/gifs/tinyScriptingNews.gif</url>
<height>40</height>
<width>78</width>
<description>What is this used for?</description>
</image>
<managingEditor>dave@userland.com (Dave Winer)</managingEditor>
<webMaster>dave@userland.com (Dave Winer)</webMaster>
<language>en-us</language>
<skipHours>
<hour>6</hour>
```



```

<hour>7</hour>
<hour>8</hour>
<hour>9</hour>
<hour>10</hour>
<hour>11</hour>
</skipHours>
<skipDays>
<day>Sunday</day>
</skipDays>
<rating>(PICS-1.1 "http://www.rsac.org/ratingsv01.html" l gen true comment "RSACi North America
Server" for "http://www.rsac.org" on "1996.04.16T08:15-0500" r (n 0 s 0 v 0 l 0))</rating>
<item>
<title>stuff</title>
<link>http://bar</link>
<description>This is an article about some stuff</description>
</item>
<textinput>
<title>Search Now!</title>
<description>Enter your search &lt;terms></description>
<name>find</name>
<link>http://my.site.com/search.cgi</link>
</textinput>
</channel>
</rss>

```

## Example 3 - International

```

<?xml version="1.0" encoding="EuC-JP"?>
<!DOCTYPE rss SYSTEM "http://my.netscape.com/publish/formats/rss-0.91.dtd">
<rss version="0.91">
<channel>
<title>è†@ÿÛë´é´ì´x´è´¸Á´õ</title>
<link>http://www.mozilla.org</link>
<description>è†@ÿÛë´é´ì´x´è´¸Á´õ</description>
<language>ja</language> <!-- tagged as Japanese content -->
<item>
<title>NYð™Áç,»îêÛil5285.25´f´\´ÁÉ´Û´-´ÁÁ´´é´ì´éõ™Ûì;êçõõ%îêÁ£</title>
<link>http://www.mozilla.org/status/</link>
<description>This is an item description...</description>
</item>
<item>
<title>,§±Çç;È§ÛÁÔ•éøó,ÁÉÈ²@ÿè†Ûèâ•±Çì´;íæ-éøè†Á£</title>
<link>http://www.mozilla.org/status/</link>
<description>This is an item description...</description>
</item>
<item>
<title>ÈÛÈ´•ìêÈ§ÁçÈ†È§æâÀè±ŽÈ%Û,ÁçÈ,âÛ¼§Û~íÈùÈÁ£</title>
<link>http://www.mozilla.org/status/</link>
<description>This is an item description...</description>
</item>
<item>
<title>2000,øíŠâ•Áç«\|éÛè¹•Û•çéÛ§ÛÁè†òæó,ÁÉì¼«...æ-õÝéøf,Á£</title>
<link>http://www.mozilla.org/status/</link>
<description>This is an item description...</description>
</item>
</channel>
</rss>

```

# Supported languages

## Why these?

These are the language codes that are accepted by Netcenter. Other language codes may be available [as specified](#) by the [w3c](#), but these are guaranteed to work with most browsers. Netcenter will currently reject other language codes, however other sites may accept them.

## Codes

```
af          # Afrikaans
sq          # Albanian
eu          # Basque
be          # Belarusian
bg          # Bulgarian
ca          # Catalan
zh-cn      # Chinese (Simplified)
zh-tw      # Chinese (Traditional)
hr          # Croatian
cs          # Czech
da          # Danish
nl          # Dutch
nl-be      # Dutch (Belgium)
nl-nl      # Dutch (Netherlands)
en          # English
en-au      # English (Australia)
en-bz      # English (Belize)
en-ca      # English (Canada)
en-ie      # English (Ireland)
en-jm      # English (Jamaica)
en-nz      # English (New Zealand)
en-ph      # English (Phillipines)
en-za      # English (South Africa)
en-tt      # English (Trinidad)
en-gb      # English (United Kingdom)
en-us      # English (United States)
en-zw      # English (Zimbabwe)
fo          # Faeroese
fi          # Finnish
fr          # French
fr-be      # French (Belgium)
fr-ca      # French (Canada)
fr-fr      # French (France)
fr-lu      # French (Luxembourg)
fr-mc      # French (Monaco)
fr-ch      # French (Switzerland)
gl          # Galician
gd          # Gaelic
de          # German
de-at      # German (Austria)
```

de-de # German (Germany)  
de-li # German (Liechtenstein)  
de-lu # German (Luxembourg)  
de-ch # German (Switzerland)  
el # Greek  
hu # Hungarian  
is # Icelandic  
id # Indonesian  
ga # Irish  
it # Italian  
it-it # Italian (Italy)  
it-ch # Italian (Switzerland)  
ja # Japanese  
ko # Korean  
mk # Macedonian  
no # Norwegian  
pl # Polish  
pt # Portuguese  
pt-br # Portuguese (Brazil)  
pt-pt # Portuguese (Portugal)  
ro # Romanian  
ro-mo # Romanian (Moldova)  
ro-ro # Romanian (Romania)  
ru # Russian  
ru-mo # Russian (Moldova)  
ru-ru # Russian (Russia)  
sr # Serbian  
sk # Slovak  
sl # Slovenian  
es # Spanish  
es-ar # Spanish (Argentina)  
es-bo # Spanish (Bolivia)  
es-cl # Spanish (Chile)  
es-co # Spanish (Colombia)  
es-cr # Spanish (Costa Rica)  
es-do # Spanish (Dominican Republic)  
es-ec # Spanish (Ecuador)  
es-sv # Spanish (El Salvador)  
es-gt # Spanish (Guatemala)  
es-hn # Spanish (Honduras)  
es-mx # Spanish (Mexico)  
es-ni # Spanish (Nicaragua)  
es-pa # Spanish (Panama)  
es-py # Spanish (Paraguay)  
es-pe # Spanish (Peru)  
es-pr # Spanish (Puerto Rico)  
es-es # Spanish (Spain)  
es-uy # Spanish (Uruguay)  
es-ve # Spanish (Venezuela)  
sv # Swedish  
sv-fi # Swedish (Finland)  
sv-se # Swedish (Sweden)

tr # Turkish  
uk # Ukranian

## Supported encodings

*Note: these are **not** case sensitive*

IANA standard name	MIME preferred name (if different from IANA)
ANSI_X3.4-1968	US-ASCII
ISO_8859-1:1987	ISO-8859-1
ISO_8859-2:1987	ISO-8859-2
ISO_8859-5:1988	ISO-8859-5
ISO_8859-7:1987	ISO-8859-7
ISO_8859-9:1989	ISO-8859-9
Shift_JIS	
Extended_UNIX_Code_Packed_Format_for_Japanese	EUC-JP
GB2312	
EUC-KR	
Big5	
windows-1250	
windows-1251	
UTF-8	
x-mac-roman	

## DTD

### Location

Public ID: -//Netscape Communications//DTD RSS 0.91//EN  
System ID: <http://my.netscape.com/publish/formats/rss-0.91.dtd>

### The DTD itself

```
<!--
Rich Site Summary (RSS) 0.91 official DTD, proposed.
RSS is an XML vocabulary for describing
metadata about websites, and enabling the display of
"channels" on the "My Netscape" website.
RSS Info can be found at http://my.netscape.com/publish/
XML Info can be found at http://www.w3.org/XML/
copyright Netscape Communications, 1999
Dan Libby - danda@netscape.com
Based on RSS DTD originally created by
Lars Marius Garshol - larsga@ifi.uio.no.
$Id: rss-spec-0.91.html,v 1.1.2.2 2001/11/09 08:10:07 dprusak Exp $
-->
<!ELEMENT rss (channel)>
<!ATTLIST rss
```

```

version      CDATA #REQUIRED> <!-- must be "0.91"> -->
<!ELEMENT channel (title | description | link | language | item+ | rating? | image? | textinput? |
copyright? | pubDate? | lastBuildDate? | docs? | managingEditor? | webMaster? | skipHours? |
skipDays?)*>
<!ELEMENT title (#PCDATA)>
<!ELEMENT description (#PCDATA)>
<!ELEMENT link (#PCDATA)>
<!ELEMENT image (title | url | link | width? | height? | description?)*>
<!ELEMENT url (#PCDATA)>
<!ELEMENT item (title | link | description)*>
<!ELEMENT textinput (title | description | name | link)*>
<!ELEMENT name (#PCDATA)>
<!ELEMENT rating (#PCDATA)>
<!ELEMENT language (#PCDATA)>
<!ELEMENT width (#PCDATA)>
<!ELEMENT height (#PCDATA)>
<!ELEMENT copyright (#PCDATA)>
<!ELEMENT pubDate (#PCDATA)>
<!ELEMENT lastBuildDate (#PCDATA)>
<!ELEMENT docs (#PCDATA)>
<!ELEMENT managingEditor (#PCDATA)>
<!ELEMENT webMaster (#PCDATA)>
<!ELEMENT hour (#PCDATA)>
<!ELEMENT day (#PCDATA)>
<!ELEMENT skipHours (hour+)>
<!ELEMENT skipDays (day+)>
<!--

```

Copied from HTML 3.2 DTD, with modifications (removed CDATA)

<http://www.w3.org/TR/REC-html32.html#dtd>

===== BEGIN =====

-->

<!--

Character Entities for ISO Latin-1

(C) International Organization for Standardization 1986

Permission to copy in any form is granted for use with conforming SGML systems and applications as defined in ISO 8879, provided this notice is included in all copies. This has been extended for use with HTML to cover the full set of codes in the range 160-255 decimal.

-->

<!-- Character entity set. Typical invocation:

```
<!ENTITY % ISOLat1 PUBLIC
```

```
"ISO 8879-1986//ENTITIES Added Latin 1//EN//HTML">
```

```
%ISOLat1;
```

-->

```
<!ENTITY nbsp    "&#160;"> <!-- no-break space -->
```

```
<!ENTITY iexcl  "&#161;"> <!-- inverted exclamation mark -->
```

```
<!ENTITY cent   "&#162;"> <!-- cent sign -->
```

```
<!ENTITY pound  "&#163;"> <!-- pound sterling sign -->
```

```
<!ENTITY curren "&#164;"> <!-- general currency sign -->
```

```
<!ENTITY yen    "&#165;"> <!-- yen sign -->
```

```
<!ENTITY brvbar "&#166;"> <!-- broken (vertical) bar -->
```

```
<!ENTITY sect   "&#167;"> <!-- section sign -->
```

```
<!ENTITY uml    "&#168;"> <!-- umlaut (dieresis) -->
```

```
<!ENTITY copy   "&#169;"> <!-- copyright sign -->
```

```
<!ENTITY ordf   "&#170;"> <!-- ordinal indicator, feminine -->
```

```
<!ENTITY laquo  "&#171;"> <!-- angle quotation mark, left -->
```

```
<!ENTITY not    "&#172;"> <!-- not sign -->
```

```
<!ENTITY shy    "&#173;"> <!-- soft hyphen -->
```

```
<!ENTITY reg    "&#174;"> <!-- registered sign -->
```

```
<!ENTITY macr   "&#175;"> <!-- macron -->
```

```

<!ENTITY deg      "°" > <!-- degree sign -->
<!ENTITY plusmn   "±" > <!-- plus-or-minus sign -->
<!ENTITY sup2     "²" > <!-- superscript two -->
<!ENTITY sup3     "³" > <!-- superscript three -->
<!ENTITY acute    "´" > <!-- acute accent -->
<!ENTITY micro    "µ" > <!-- micro sign -->
<!ENTITY para     "¶" > <!-- pilcrow (paragraph sign) -->
<!ENTITY middot   "·" > <!-- middle dot -->
<!ENTITY cedil    "¸" > <!-- cedilla -->
<!ENTITY sup1     "¹" > <!-- superscript one -->
<!ENTITY ordm     "º" > <!-- ordinal indicator, masculine -->
<!ENTITY raquo    "»" > <!-- angle quotation mark, right -->
<!ENTITY frac14   "¼" > <!-- fraction one-quarter -->
<!ENTITY frac12   "½" > <!-- fraction one-half -->
<!ENTITY frac34   "¾" > <!-- fraction three-quarters -->
<!ENTITY iquest   "¿" > <!-- inverted question mark -->
<!ENTITY Agrave   "À" > <!-- capital A, grave accent -->
<!ENTITY Aacute   "Á" > <!-- capital A, acute accent -->
<!ENTITY Acirc    "Â" > <!-- capital A, circumflex accent -->
<!ENTITY Atilde   "Ã" > <!-- capital A, tilde -->
<!ENTITY Auml     "Ä" > <!-- capital A, dieresis or umlaut mark -->
<!ENTITY Aring    "Å" > <!-- capital A, ring -->
<!ENTITY AElig    "Æ" > <!-- capital AE diphthong (ligature) -->
<!ENTITY Ccedil   "Ç" > <!-- capital C, cedilla -->
<!ENTITY Egrave   "È" > <!-- capital E, grave accent -->
<!ENTITY Eacute   "É" > <!-- capital E, acute accent -->
<!ENTITY Ecirc    "Ê" > <!-- capital E, circumflex accent -->
<!ENTITY Euml     "Ë" > <!-- capital E, dieresis or umlaut mark -->
<!ENTITY Igrave   "Ì" > <!-- capital I, grave accent -->
<!ENTITY Iacute   "Í" > <!-- capital I, acute accent -->
<!ENTITY Icirc    "Î" > <!-- capital I, circumflex accent -->
<!ENTITY Iuml     "Ï" > <!-- capital I, dieresis or umlaut mark -->
<!ENTITY ETH      "Ð" > <!-- capital Eth, Icelandic -->
<!ENTITY Ntilde   "Ñ" > <!-- capital N, tilde -->
<!ENTITY Ograve   "Ò" > <!-- capital O, grave accent -->
<!ENTITY Oacute   "Ó" > <!-- capital O, acute accent -->
<!ENTITY Ocirc    "Ô" > <!-- capital O, circumflex accent -->
<!ENTITY Otilde   "Õ" > <!-- capital O, tilde -->
<!ENTITY Ouml     "Ö" > <!-- capital O, dieresis or umlaut mark -->
<!ENTITY times    "×" > <!-- multiply sign -->
<!ENTITY Oslash   "Ø" > <!-- capital O, slash -->
<!ENTITY Ugrave   "Ù" > <!-- capital U, grave accent -->
<!ENTITY Uacute   "Ú" > <!-- capital U, acute accent -->
<!ENTITY Ucirc    "Û" > <!-- capital U, circumflex accent -->
<!ENTITY Uuml     "Ü" > <!-- capital U, dieresis or umlaut mark -->
<!ENTITY Yacute   "Ý" > <!-- capital Y, acute accent -->
<!ENTITY THORN    "Þ" > <!-- capital THORN, Icelandic -->
<!ENTITY szlig    "ß" > <!-- small sharp s, German (sz ligature) -->
<!ENTITY agrave   "à" > <!-- small a, grave accent -->
<!ENTITY aacute   "á" > <!-- small a, acute accent -->
<!ENTITY acirc    "â" > <!-- small a, circumflex accent -->
<!ENTITY atilde   "ã" > <!-- small a, tilde -->
<!ENTITY auml     "ä" > <!-- small a, dieresis or umlaut mark -->
<!ENTITY aring    "å" > <!-- small a, ring -->
<!ENTITY aelig    "æ" > <!-- small ae diphthong (ligature) -->
<!ENTITY ccedil   "ç" > <!-- small c, cedilla -->
<!ENTITY egrave   "è" > <!-- small e, grave accent -->
<!ENTITY eacute   "é" > <!-- small e, acute accent -->
<!ENTITY ecirc    "ê" > <!-- small e, circumflex accent -->
<!ENTITY euml     "ë" > <!-- small e, dieresis or umlaut mark -->
<!ENTITY igrave   "ì" > <!-- small i, grave accent -->

```

```

<!ENTITY iacute "&#237;"> <!-- small i, acute accent -->
<!ENTITY icirc "&#238;"> <!-- small i, circumflex accent -->
<!ENTITY iuml "&#239;"> <!-- small i, dieresis or umlaut mark -->
<!ENTITY eth "&#240;"> <!-- small eth, Icelandic -->
<!ENTITY ntilde "&#241;"> <!-- small n, tilde -->
<!ENTITY ograve "&#242;"> <!-- small o, grave accent -->
<!ENTITY oacute "&#243;"> <!-- small o, acute accent -->
<!ENTITY ocirc "&#244;"> <!-- small o, circumflex accent -->
<!ENTITY otilde "&#245;"> <!-- small o, tilde -->
<!ENTITY ouml "&#246;"> <!-- small o, dieresis or umlaut mark -->
<!ENTITY divide "&#247;"> <!-- divide sign -->
<!ENTITY oslash "&#248;"> <!-- small o, slash -->
<!ENTITY ugrave "&#249;"> <!-- small u, grave accent -->
<!ENTITY uacute "&#250;"> <!-- small u, acute accent -->
<!ENTITY ucirc "&#251;"> <!-- small u, circumflex accent -->
<!ENTITY uuml "&#252;"> <!-- small u, dieresis or umlaut mark -->
<!ENTITY yacute "&#253;"> <!-- small y, acute accent -->
<!ENTITY thorn "&#254;"> <!-- small thorn, Icelandic -->
<!ENTITY yuml "&#255;"> <!-- small y, dieresis or umlaut mark -->
<!--
Copied from HTML 3.2 DTD, with modifications (removed CDATA)
http://www.w3.org/TR/REC-html32.html#dtd
===== END =====
-->

```

## Proprietary Schema (Validation Rules)

### Explanation

XML currently provides a limited amount of validation via DTD's. However, DTD's do not provide any support for common validation requirements, such as data types, length of strings, number of sub-elements, or pattern matching.

A standard has been proposed to solve this problem. XML Schemas looks like it will do all of this and more. Unfortunately, there are few, if any parsers available today that understand them.

As a proprietary, *interim only solution*, we have developed a very simplistic schema format that performs a second level of validation after the parser has read the XML document into memory. We are listing the schema used to validate RSS 0.91 files, so that there will be no ambiguity when validation fails.

Here are the basic rules:

- Each XML element must be defined by an <Element> tag.
  - Each Element definition must have a unique id attribute and a type attribute.
  - Each Attribute of an Element must be referenced by an <Attrib> tag
  - Each sub-Element of an Element of type container must be referenced by <Contains> tag.
  - Each Element may have a type associated with it. Currently supported types are:
    - container: this Element contains other Elements only.
    - string: this Element contains text data.
    - int: this Element contains an integer.
  - Each string or int Element may contain a matching rule, specified via <Matches>
  - Each string or int Element may specify a minimum and maximum number of characters (or value if type int) via min, max, and exactly.
- Each XML attribute must be defined by an <Attribute> tag.

- Each Attribute definition must have a unique id attribute and a type attribute.
- Each Attribute may be of type string or int.
- Each Attribute may contain a matching rule, specified via <Matches>
- Each Attribute may specify a minimum and maximum number of characters (or value if type int) via min, max, and exactly.
- Each <Contains> and <Attrib> definition must contain a 'ref' attribute that refers to a uniquely defined Element or Attribute with the value of 'ref' as its id.
- Each <Contains> and <Attrib> definition may contain min, max, or exactly attributes to define the number of Elements or Attributes required.
- Each <Matches> must contain a valid regular expression, against which the corresponding Element or Attribute will be evaluated.

## Schema

Here is the schema for RSS 0.91.

```
<?xml version="1.0"?>
<!DOCTYPE Schema PUBLIC "-//Netscape Communications//DTD Schema 1.0//EN"
"http://my.netscape.com/publish/formats/schema-1.0.dtd">
<Schema version="DKHXVF 1.0" root="rss" name="RSS 0.91">
<Element id="rss" type="container">
<Contains ref="channel" exactly="1"/>
<Attrib ref="version" exactly="1"/>
</Element>
<Attribute id="version" type="string">
<Matches>0.91</Matches>
</Attribute>
<Element id="channel" type="container">
<Contains ref="description" exactly="1"/>
<Contains ref="image" min="0" max="1"/>
<Contains ref="item" min="0" max="15"/>
<Contains ref="language" exactly="1"/>
<Contains ref="link" exactly="1"/>
<Contains ref="rating" min="0" max="1"/>
<Contains ref="textInput" min="0" max="1"/>
<Contains ref="title" exactly="1"/>
<Contains ref="copyright" min="0" max="1"/>
<Contains ref="pubDate" min="0" max="1"/>
<Contains ref="lastBuildDate" min="0" max="1"/>
<Contains ref="docs" min="0" max="1"/>
<Contains ref="managingEditor" min="0" max="1"/>
<Contains ref="webMaster" min="0" max="1"/>
<Contains ref="skipHours" min="0" max="1"/>
<Contains ref="skipDays" min="0" max="1"/>
</Element>
<Element id="copyright" type="string" max="100"/>
<Element id="pubDate" type="string" max="100"/>
<Element id="lastBuildDate" type="string" max="100"/>
<Element id="docs" type="string" max="500"/>
<Element id="managingEditor" type="string" max="100"/>
<Element id="webMaster" type="string" max="100"/>
<Element id="skipHours" type="container">
<Contains ref="hour" min="0" max="24"/>
</Element>
<Element id="skipDays" type="container">
<Contains ref="day" min="0" max="7"/>
```



```

</Element>
<Element id="hour" type="int" min="0" max="24"/>
<Element id="day" type="string" min="0" max="10"/>
<Element id="item" type="container">
<Contains ref="title" exactly="1"/>
<Contains ref="link" exactly="1"/>
<Contains ref="description" min="0" max="1"/>
</Element>
<Element id="image" type="container">
<Contains ref="title" exactly="1"/>
<Contains ref="link" min="0" max="1" />
<Contains ref="url" exactly="1"/>
<Contains ref="width" min="0" max="1"/>
<Contains ref="height" min="0" max="1"/>
<Contains ref="description" min="0" max="1"/>
</Element>
<Element id="textinput" type="container">
<Contains ref="title" exactly="1"/>
<Contains ref="link" exactly="1"/>
<Contains ref="description" exactly="1"/>
<Contains ref="name" exactly="1"/>
</Element>
<Element id="title" type="string" min="1" max="100"/>
<Element id="description" type="string" min="1" max="500"/>
<Element id="url" type="string" min="1" max="500">
<Matches>^(http://|^ftp://)</Matches>
</Element>
<Element id="link" type="string" min="1" max="500">
<Matches>^(http://|^ftp://)</Matches>
</Element>
<Element id="language" type="string" min="2" max="5">
<Matches>
^(af |      # Afrikaans
sq |      # Albanian
eu |      # Basque
be |      # Belarusian
bg |      # Bulgarian
ca |      # Catalan
zh-cn |   # Chinese (Simplified)
zh-tw |   # Chinese (Traditional)
hr |      # Croatian
cs |      # Czech
da |      # Danish
nl |      # Dutch
nl-be |   # Dutch (Belgium)
nl-nl |   # Dutch (Netherlands)
en |      # English
en-au |   # English (Australia)
en-bz |   # English (Belize)
en-ca |   # English (Canada)
en-ie |   # English (Ireland)
en-jm |   # English (Jamaica)
en-nz |   # English (New Zealand)
en-ph |   # English (Phillipines)
en-za |   # English (South Africa)
en-tt |   # English (Trinidad)
en-gb |   # English (United Kingdom)
en-us |   # English (United States)
en-zw |   # English (Zimbabwe)
fo |      # Faeroese
fi |      # Finnish

```

fr	# French
fr-be	# French (Belgium)
fr-ca	# French (Canada)
fr-fr	# French (France)
fr-lu	# French (Luxembourg)
fr-mc	# French (Monaco)
fr-ch	# French (Switzerland)
gl	# Galician
gd	# Gaelic
de	# German
de-at	# German (Austria)
de-de	# German (Germany)
de-li	# German (Liechtenstein)
de-lu	# German (Luxembourg)
de-ch	# German (Switzerland)
el	# Greek
hu	# Hungarian
is	# Icelandic
id	# Indonesian
ga	# Irish
it	# Italian
it-it	# Italian (Italy)
it-ch	# Italian (Switzerland)
ja	# Japanese
ko	# Korean
mk	# Macedonian
no	# Norwegian
pl	# Polish
pt	# Portuguese
pt-br	# Portuguese (Brazil)
pt-pt	# Portuguese (Portugal)
ro	# Romanian
ro-mo	# Romanian (Moldova)
ro-ro	# Romanian (Romania)
ru	# Russian
ru-mo	# Russian (Moldova)
ru-ru	# Russian (Russia)
sr	# Serbian
sk	# Slovak
sl	# Slovenian
es	# Spanish
es-ar	# Spanish (Argentina)
es-bo	# Spanish (Bolivia)
es-cl	# Spanish (Chile)
es-co	# Spanish (Colombia)
es-cr	# Spanish (Costa Rica)
es-do	# Spanish (Dominican Republic)
es-ec	# Spanish (Ecuador)
es-sv	# Spanish (El Salvador)
es-gt	# Spanish (Guatemala)
es-hn	# Spanish (Honduras)
es-mx	# Spanish (Mexico)
es-ni	# Spanish (Nicaragua)
es-pa	# Spanish (Panama)
es-py	# Spanish (Paraguay)
es-pe	# Spanish (Peru)
es-pr	# Spanish (Puerto Rico)
es-es	# Spanish (Spain)
es-uy	# Spanish (Uruguay)
es-ve	# Spanish (Venezuela)
sv	# Swedish

```

sv-fi | # Swedish (Finland)
sv-se | # Swedish (Sweden)
tr | # Turkish
uk | # Ukranian
)$
</Matches>
</Element>
<Element id="rating" type="string" min="20" max="500">
<Matches>^\(PICS-1.1</Matches>
</Element>
<Element id="width" type="int" min="1" max="144"/>
<Element id="height" type="int" min="1" max="400"/>
<Element id="name" type="string" min="1" max="20"/>
</Schema>

```

## Schema DTD

Here is the DTD for the schema format.

```

<!--
A DTD for Dan's Kinda Hacky XML Validation Format (DKHXVF)
Basically, this format allows us to enforce some additional rules
that DTD's do not. Specifically, we can:
- specify min and max for number of each child element
- specify a regular expression that text elements and attributes must match
- specify type of text elements and attributes (int, float, string, timestamp)
- specify min and max for any type. (length compare for strings, numeric otherwise)
The hope is that this will allow the rapid creation of new formats, and modification
of existing formats (adding/removing tags, attributes etc), without requiring
code changes in the validation software.
This is not in any way intended to be an alternative to XML schemas. In the
absence of code supporting XML schemas, I created this, but it is meant as
a transitional work only.
For more on XML schemas, see:
http://www.w3.org/1999/05/06-xmlschema-1/ and
http://www.w3.org/1999/05/06-xmlschema-2/
This is also not meant to replace DTDs. There are many things that you can do
with DTDs that you cannot do with this format. For example, you cannot declare
entities with this format. You must do that in the DTD. If you want your
parser to interpret them correctly, you must use a validating parser.
It is possible to use these schemas without DTD validation, however you may run
into problems with entity expansion and other things.
Dan Libby - danda@netscape.com
$Log: rss-spec-0.91.html,v $
Revision 1.1.2.2 2001/11/09 08:10:07 dprusak
Merged for 6.2

Revision 1.1.2.1 2001/10/17 22:25:28 dprusak
NewMyNetscape
Revision 1.1.2.1 2001/05/03 00:44:50 hoangtv
adding DTD definition
Revision 1.4 1999/09/10 03:01:44 jquach
removed comments
Revision 1.3 1999/09/10 03:01:24 jquach
pulled ref to internal file
Revision 1.2 1999/08/07 04:53:02 danda
'cleaning' (removing useful info) for public release
Revision 1.3 1999/08/07 04:52:12 danda

```

```

'cleaning' (removing useful info) for public release
Revision 1.2 1999/07/22 07:09:41 danda
fixing examples, RDF Site Summary -> Rich Site Summary
Revision 1.1 1999/06/09 07:01:29 danda
adding schema and dtd for rss 0.9 and 1.0
-->
<!--
Tag: Schema
Description: Document wrapper.
Sub tags: Element & Attribute
Attributes: version, root, name
Notes:
version must be "DKHXVF 1.0"
root is the document root.
-->
<!ELEMENT Schema (Element | Attribute)*>
<!ATTLIST Schema
version CDATA #FIXED "DKHXVF 1.0"
root CDATA #REQUIRED
name CDATA #REQUIRED>
<!--
Tag: Element
Description: Definition of an allowed element (tag)
Sub tags: Contains, Attrib, Matches
Attributes: id, type, min, max, exactly
Notes: exactly="1" is equivalent to min="1" max="1"
-->
<!ELEMENT Element ((Contains | Attrib)* | Matches?)>
<!ATTLIST Element
id CDATA #REQUIRED
type (int | float | container | string | timestamp) #REQUIRED
min CDATA #IMPLIED
max CDATA #IMPLIED
exactly CDATA #IMPLIED>
<!--
Tag: Contains
Description: Defines rules for a sub-element.
Sub tags: None, this tag must be empty.
Attributes: ref, min, max, exactly
Notes: ref must refer to the 'id' of an element defined elsewhere or the schema
is invalid.
-->
<!ELEMENT Contains EMPTY>
<!ATTLIST Contains
ref CDATA #REQUIRED
min CDATA #IMPLIED
max CDATA #IMPLIED
exactly CDATA #IMPLIED>
<!--
Tag: Attrib
Description: Defines rules for an element attribute.
Sub tags: None, this tag must be empty
Attributes: ref, min, max, exactly
Notes: ref must refer to the 'id' of an Attribute defined elsewhere or the schema
is invalid.
-->
<!ELEMENT Attrib EMPTY>
<!ATTLIST Attrib
ref CDATA #REQUIRED
min CDATA #IMPLIED
max CDATA #IMPLIED

```

```

exactly CDATA #IMPLIED>
<!--
Tag: Attribute
Description: Definition of an allowed attribute
Sub tags: Matches
Attributes: id, type, min, max, exactly
Notes: none
-->
<!ELEMENT Attribute (Matches?)>
<!ATTLIST Attribute
id CDATA #REQUIRED
type (int | float | string | timestamp) #REQUIRED
min CDATA #IMPLIED
max CDATA #IMPLIED
exactly CDATA #IMPLIED>
<!--
Tag: Matches
Description: A regular expression that values will be compared against
Sub tags: None
Attributes: None
Notes: Matches may be used for elements of any type but container, and for attributes.
An example of a useful matching pattern is:
<Matches>^(foo|bar|foobar)$</Matches>
This will allow any values that exactly match "foo", "bar", or "foobar".
Whitespace is allowed in the regex and '#' is used for comments. The following
is valid:
<Matches>
&# # Start of a numeric entity reference, xml escaped &
(?P&lt;char&gt; # xml escaped <, >
[0-9]+[^\0-9] # Decimal form
| 0[0-7]+[^\0-7] # Octal form
| x[0-9a-fA-F]+[^\0-9a-fA-F] # Hexadecimal form
)
</Matches>
which is equivalent to:
<Matches>&#(?P&lt;char&gt;[0-9]+[^\0-9]| 0[0-7]+[^\0-7]| x[0-9a-fA-F]+[^\0-9a-fA-F])</Matches>
For help on regular expressions, see:
http://www.python.org/doc/howto/regex/regex.html or
http://www.ciser.cornell.edu/info/regex.html
-->
<!ELEMENT Matches (#PCDATA)>
<!--
Example of a DKHXVF 1.0 file:
<?xml version="1.0"?>
<!DOCTYPE Schema PUBLIC "-//Netscape Communications//DTD Schema 1.0//EN"
"http://my.netscape.com/publish/formats/schema-1.0.dtd">
<Schema version="DKHXVF 1.0" root="rdf:RDF" name="RSS 0.9">
<Element id="rdf:RDF" type="container">
<Contains ref="channel" exactly="1"/>
<Contains ref="image" min="0" max="1"/>
<Contains ref="item" min="1" max="15"/>
<Contains ref="textinput" min="0" max="1"/>
<Attrib ref="xmlns" exactly="1"/>
<Attrib ref="xmlns:rdf" exactly="1"/>
</Element>
<Attribute id="xmlns" type="string">
<Matches>http://my.netscape.com/rdf/simple/0.9/</Matches>
</Attribute>
<Attribute id="xmlns:rdf" type="string">
<Matches>http://www.w3.org/1999/02/22-rdf-syntax-ns#</Matches>
</Attribute>

```

```
<Element id="channel" type="container">
<Contains ref="link" exactly="1"/>
<Contains ref="title" exactly="1"/>
<Contains ref="description" exactly="1"/>
</Element>
<Element id="item" type="container">
<Contains ref="title" exactly="1"/>
<Contains ref="link" exactly="1"/>
</Element>
<Element id="image" type="container">
<Contains ref="title" exactly="1"/>
<Contains ref="link" exactly="1" />
<Contains ref="url" exactly="1"/>
</Element>
<Element id="textinput" type="container">
<Contains ref="title" exactly="1"/>
<Contains ref="description" exactly="1"/>
<Contains ref="link" exactly="1"/>
<Contains ref="name" exactly="1"/>
</Element>
<Element id="title" type="string" min="1" max="100"/>
<Element id="description" type="string" min="1" max="500"/>
<Element id="url" type="string" min="1" max="500">
<Matches>^(http://|^ftp://)</Matches>
</Element>
<Element id="link" type="string" min="1" max="500">
<Matches>^(http://|^ftp://)</Matches>
</Element>
<Element id="name" type="string" min="1" max="20"/>
</Schema>
-->
```