Atom Bidirectional Attribute
draft-snell-atompub-bidi-08.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.
Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.
Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."
The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.
The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.
This Internet-Draft will expire on December 11, 2009.

Copyright Notice

Abstract

This document adds a new attribute to the Atom Syndication Format used to indicate the base directionality of directionally-neutral characters.

Table of Contents

---

1. Introduction   [TOC](#)

This document updates the Atom Syndication Format [RFC4287] (Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," December 2005.) by adding a new "dir" attribute used to define the base directionality of directionally-neutral characters contained within an Atom document.

1.1. Namespace   [TOC](#)

The XML Namespaces URI [W3C.REC-xml-names-19990114] (Hollander, D., Layman, A., and T. Bray, "Namespaces in XML," January 1999.) for the Atom Syndication Format [RFC4287] (Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," December 2005.) is:

    http://www.w3.org/2005/Atom

1.2. Notational Conventions   [TOC](#)

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, [RFC2119] (Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," March 1997.).

The Atom Syndication Format [RFC4287] (Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," December 2005.) is specified using terms from the XML Infoset [W3C.REC-xml-infoset-20040204] (Tobin, R. and J. Cowan, "XML Information Set (Second Edition)," February 2004.). This specification uses a shorthand form for two commons terms: The phrase "Information Item" is omitted when naming Element and Attribute Information Items. Therefore, when this specification uses the term "element," it is referring to an Element Information Item in Infoset terms. Likewise, when it uses the term "attribute," it is referring to an Attribute Information Item. Portions this specification are illustrated with fragments of a non-normative RELAX NG Compact schema [RELAXNG] (Clark, J., "RELAX NG Compact Syntax," December 2001.). However, the text of this specification provides the sole definition of conformance.

---

2. The "dir" Attribute   TOC

The "dir" attribute specifies the base direction of directionally-neutral text [ISO10646] (International Organization for Standardization, "ISO/IEC 10646:2003: Information Technology - Universal Multiple-Octet Coded Character Set (UCS)," December 2003.) in an Atom document. Possible values for the attribute are "ltr" and "rtl" indicating "left-to-right" and "right-to-left" respectively, or an empty string indicating that no base-direction is specified. If a dir attribute is not provided, the value MUST be assumed to be an empty string. The attribute can appear on any element in an Atom document.

```
atomCommonAttributes =
  attribute xml:base { atomUri }?,
  attribute xml:lang { atomLanguageTag }?,
  attribute dir { "ltr" | "rtl" | "" }?,
  undefinedAttribute*
```

The direction specified by "dir" applies to elements and attributes whose values are specified as being "Language-Sensitive" as defined by Section 2 of [RFC4287] (Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," December 2005.). The direction specified by the attribute is inherited by descendant elements and attributes and may be overridden.
Values other than "ltr", "rtl" and "" MUST be ignored and processed as if the dir attribute was not present; Atom processors MUST NOT stop processing or signal an error. The value of the attribute is not case-sensitive.
The "dir" attribute is an extension to the Atom vocabulary that will be treated as unknown foreign markup by existing Atom processors that have not been explicitly implemented to support the "dir" attribute. As per

the rules specified in [RFC4287] (Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," December 2005.), such processors are required to ignore unknown foreign markup and continue processing as if the markup does not exist.
Example atom:feed with right-to-left directionality

```
<?xml version="1.0" ?>
<feed xmlns="http://www.w3.org/2005/Atom" dir="rtl">
  <title>&#x0679;&#x067A;&#x067B;</title>
  ...
</feed>
```

If an Atom document contains bidirectional text, the Unicode Bidirectional Algorithm [UAX9] (Davis, M., "Unicode Standard Annex #9: The Bidirectional Algorithm," September 2006.) SHOULD be used to render that text. Because consumers of Atom documents vary broadly in the way they display text, the "ltr" and "rtl" values do not necessarily open an additional level of embedding or override with respect to the bidirectional algorithm. Consuming applications that render bidirectional text are responsible for determining the appropriate level of embedding. If the dir attribute value is "rtl", Atom processors that display affected text MAY choose to right-align that text as per the rules described in Section 8 of [W3C.REC-html401-19991224] (Hors, A., Jacobs, I., and D. Raggett, "HTML 4.01 Specification," December 1999.).
When Atom Text Constructs or the atom:content elements contain bidirectional text and the type attribute value is either "html" or "xhtml", the bidirectional markup mechanisms specific to each format SHOULD be used. The value of the "dir" attribute defines the base directionality of Language-Sensitive text within Text Constructs and atom:content elements regardless of the specific type of content each may contain.
Example atom:feed with bidirectional XHTML:

```
<?xml version="1.0" ?>
<feed xmlns="http://www.w3.org/2005/Atom" dir="rtl">
  ...
  <title type="xhtml">
    <div xmlns="http://www.w3.org/1999/xhtml">
      <p dir="rtl">&#x0679;&#x067A;&#x067B;</p>
    </div>
  </title>
  ...
</feed>
```

Unicode bidirectional control characters MAY be used within attributes and element values to indicate the directionality of text or to modify the default operation of the Bidirectional Algorithm. Implementers are

reminded that unexpected results could occur when using both the "dir" attribute and the Unicode control characters within a single document. When an Atom document contains a mixture of Bidi attributes, Unicode control characters or other forms of bidirectional markup, the most specific mechanism is given the highest precedence when determining how to render bidirectional text. For instance, if the dir attribute specifying left-to-right ordering is used on an atom:title element containing Unicode control characters that specify a right-to-left ordering, then the text will be displayed as right-to-left as specified by the control characters.

2.1. Direction Guessing

In Atom documents that do not contain a "dir" attribute, it is possible to apply heuristics to guess the base directionality of text in the document. Such heuristics can take into consideration the in-scope language context established by the use of the xml:lang attribute or an analysis of the directional properties of the Unicode characters used within the text. Such guessing algorithms can produce reasonably acceptable results in many cases but cannot be guaranteed to produce correct results in every case. For this reason, explicit determination of text direction using the "dir" attribute is preferred over any guessing algorithm.

For compatibility with existing Atom documents that rely on direction guessing, user agents MAY perform direction guessing in documents that do not contain a "dir" attribute but they SHOULD NOT do so when an in-scope "dir" attribute is provided.

3. Security Considerations

The security considerations discussed in [RFC4287] (Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," December 2005.) Section 8 apply.

4. IANA Considerations

No IANA actions are required by this document.

5. References

5.1. Normative References

[ISO10646]	International Organization for Standardization, "ISO/IEC 10646:2003: Information Technology - Universal Multiple-Octet Coded Character Set (UCS)," December 2003.
[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).
[RFC4287]	Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," RFC 4287, December 2005 (TXT, HTML, XML).
[UAX9]	Davis, M., "Unicode Standard Annex #9: The Bidirectional Algorithm," September 2006.
[W3C.REC-xml-infoset-20040204]	Tobin, R. and J. Cowan, "XML Information Set (Second Edition)," World Wide Web Consortium Recommendation REC-xml-infoset-20040204, February 2004 (HTML).
[W3C.REC-xml-names-19990114]	Hollander, D., Layman, A., and T. Bray, "Namespaces in XML," World Wide Web Consortium FirstEdition REC-xml-names-19990114, January 1999 (HTML).

5.2. Informative References

[RELAXNG]	Clark, J., "RELAX NG Compact Syntax," December 2001.
[W3C.REC-html401-19991224]	Hors, A., Jacobs, I., and D. Raggett, "HTML 4.01 Specification," World Wide Web Consortium Recommendation REC-html401-19991224, December 1999 (HTML).

Appendix A. Acknowledgements

The author gratefully acknowledges the feedback from the Atom Publishing Format and Protocol Working Group.

Author's Address

	James M Snell
Email:	jasnell@us.ibm.com
URI:	http://www.ibm.com