Glossary Markup Language
GlossML 1.0 Specification

24 August 2009

Latest version:

Editor:

Rodolfo M. Raya <mrmraya@maxprograms.com>

Abstract

This document defines the Glossary Markup Language (GlossML). The purpose of this vocabulary is to allow the storage of glossaries in XML format, facilitating data exchange.

Status

This document constitutes the first official public release of GlossML. Comments may be sent to tech@maxprograms.com.

Table of Contents

1. Introduction

A glossary is defined as a list of terms in a special subject, field or area of usage, with accompanying definitions. Glossary Markup Language (GlossML) is an XML vocabulary designed for facilitating the exchange of glossaries using XML format.

GlossML can describe monolingual as well as multilingual glossaries used in translation processes.

1.1 File Structure

The root of a GlossML glossary is the <glossary> element and the main language used in the glossary is declared in the required srclang attribute. A <glossary> element contains one optional <comment> and one or more <glossentry> elements. A <glossentry> element contains a term and its optional translations into one or more languages. Terms are stored in <langentry> elements, which contain the term text in one <term> element and an optional definition in a <definition> element.

See Appendix A for a hierarchical view of a GlossML tree. Appendix C contains a sample document.

1.2 XML Conformance

GlossML is an XML vocabulary, defined using an XML Schema (see Appendix B). All glossaries written in GlossML format must be XML-compliant files.

Language codes used in xml:lang attributes must be valid languages defined in RFC 4646.

1.3 GlossML namespace

The XML namespace for GlossML is "http://www.maxprograms.com/gml".

GlossML can be embedded in other XML vocabularies that support extension via namespaces. For example, a GlossML glossary can be included in an XLIFF (XML Localization Interchange File Format) document in this way:

```
<?xml version="1.0" encoding="UTF-8" ?>
<xliff version="1.2" xmlns="urn:oasis:names:tc:xliff:document:1.2"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="urn:oasis:names:tc:xliff:document:1.2 xliff-core-1.2-transitional.xsd"
 xmlns:gls="http://www.maxprograms.com/gml">
 <file datatype="javalistresourcebundle" original="capi.properties" source-language="en">
  <header>
   <gls:glossary version="1.0" srclang="en">
     ... GlossML data ...
   </gls:glossary>
  </header>
  <body>
     ... XLIFF data ...
  </body>
 </file>
</xliff>
```

1.4 Extensibility

All elements of a GlossML glossary may be extended with attributes from a foreign XML namespace, provided that the namespace is properly declared in the GlossML file and that the corresponding XML Schema is made available for validation purposes.

Applications that support GlossML are not required to understand and use data stored in attributes from foreign namespaces and can safely ignore them.

1.5 GlossML File Extension

GlossML documents use the `.gls` extension in file names. No other extension is recommended by this specification.

2. Detailed Specifications

This section contains descriptions of all elements and attributes of the Glossary Markup Language vocabulary defined in the XML Schema included in Appendix B.

2.1 GlossML Elements

The elements that form the GlossML vocabulary are: <glossary>, <comment>, <glossentry>, <langentry>, <definition> and <term>.

A detailed description of each element follows. See Appendix A for a hierarchical relationship view.

<glossary>

Glossary root. Holds all elements of a GlossML glossary.

Contains:

Zero or one <comment> element followed by one or more <glossentry> elements.

Required Attributes:

version, srclang

Optional Attributes:

Attributes from any namespace declared in the GlossML file.

<comment>

Comment at <glossary> or <glossentry> level.

Contains:

Plain text.

Optional Attributes:

from, attributes from any namespace declared in the GlossML file.

<glossentry>

Glossary entry.

Contains:

Zero or one <comment> element followed by one or more <langentry> elements.

Optional Attributes:

Attributes from any namespace declared in the GlossML file.

\<langentry\>

A term in the glossary, expressed in the language specified by the required <u>"xml:lang"</u> attribute.

Contains:

One <u>\<term\></u> element followed by zero or one <u>\<definition\></u> element.

Required Attributes:

<u>xml:lang</u>

Optional Attributes:

Attributes from any namespace declared in the GlossML file.

\<term\>

Term text in the language indicated in the <u>"xml:lang"</u> attribute of the parent <u>\<langentry\></u> element.

Contains:

Plain text.

Optional Attributes:

Attributes from any namespace declared in the GlossML file.

\<definition\>

Optional definition in plain text for the term stored in the <u>\<term\></u> element contained in the parent <u>\<langentry\></u> element.

Contains:

Plain text.

Optional Attributes:

<u>source</u>, attributes from any namespace declared in the GlossML file.

2.2 GlossML Attributes

The attributes defined in the the GlossML grammar (see <u>Appendix B</u>) are: <u>from</u>, <u>source</u>, <u>srclang</u> and <u>version</u>. The <u>xml:lang</u> attribute defined in <u>XML Standard</u> is also used in GlossML.

---

from

Indicates the author of a comment.

Used in: <comment>

---

source

Indicates the source of a definition.

Used in: <definition>

---

srclang

Indicates the main language of the glossary.

Value description:

- The string "*all*", meaning that any language declared in <langentry> elements can be considered the main language.
- Any valid language code defined by RFC 4646.

Used in: <glossary>

---

version

Indicates the version of the GlossML vocabulary to which a document conforms.

Value description: fixed text.

Default value: "1.0"

Used in: <glossary>

---

xml:lang
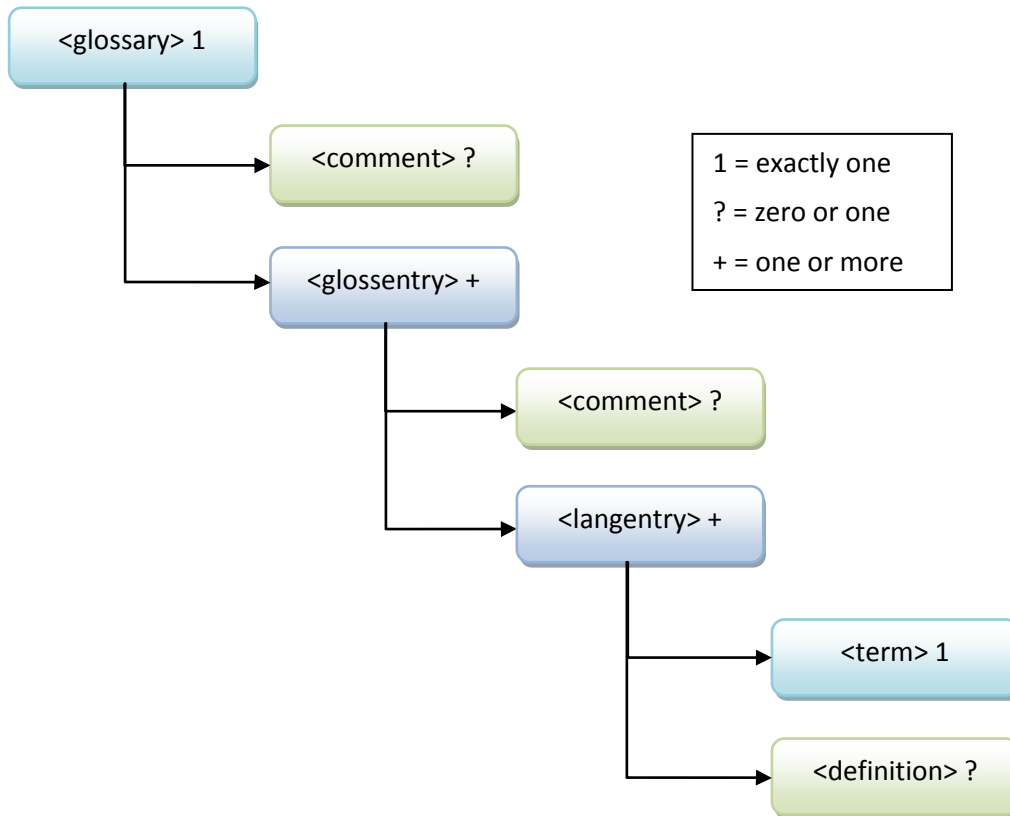
Indicates the language of the term described in a <langentry> element.

Value description: any valid language code defined by RFC 4646.

Used in: <langentry>

Appendix A. GlossML Element Tree

The following diagram indicates the hierarchical relationship of GlossML elements.

Appendix B. XML Schema

The XML Schema for GlossML is as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- ====================================================== -->
<!-- Copyright © 2009 Maxprograms                           -->
<!-- All rights reserved.                                   -->
<!-- Revision: 7D9-1-F                                      -->
<!-- ====================================================== -->
<xs:schema xmlns:gml="http://www.maxprograms.com/gml"
    targetNamespace="http://www.maxprograms.com/gml"
    xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
    <xs:import namespace="http://www.w3.org/XML/1998/namespace"
    schemaLocation="http://www.w3.org/2001/xml.xsd" />
    <xs:element name="glossary">
        <xs:complexType mixed="false">
            <xs:sequence>
                <xs:element minOccurs="0" maxOccurs="1"  ref="gml:comment" />
                <xs:element minOccurs="1" maxOccurs="unbounded" ref="gml:glossentry" />
            </xs:sequence>
            <xs:attribute name="version" use="required">
                <xs:simpleType>
                    <xs:restriction base="xs:string">
                        <xs:enumeration value="1.0"/>
                    </xs:restriction>
                </xs:simpleType>
            </xs:attribute>
            <xs:attribute name="srclang" use="required" />
            <xs:anyAttribute namespace="##any" processContents="lax" />
        </xs:complexType>
    </xs:element>
    <xs:element name="comment">
        <xs:complexType mixed="true">
            <xs:attribute name="from"  />
            <xs:anyAttribute namespace="##any" processContents="lax" />
        </xs:complexType>
    </xs:element>
    <xs:element name="glossentry">
        <xs:complexType mixed="false">
            <xs:sequence>
                <xs:element minOccurs="0" maxOccurs="1"  ref="gml:comment" />
                <xs:element minOccurs="1" maxOccurs="unbounded" ref="gml:langentry" />
            </xs:sequence>
            <xs:anyAttribute namespace="##any" processContents="lax"/>
        </xs:complexType>
    </xs:element>
    <xs:element name="langentry">
        <xs:complexType mixed="false">
            <xs:sequence>
                <xs:element minOccurs="1" maxOccurs="1" ref="gml:term" />
                <xs:element minOccurs="0" maxOccurs="1" ref="gml:definition" />
            </xs:sequence>
            <xs:attribute  ref="xml:lang" use="required" />
            <xs:anyAttribute namespace="##any" processContents="lax" />
        </xs:complexType>
    </xs:element>
    <xs:element name="definition">
        <xs:complexType mixed="true">
            <xs:attribute name="source" use="optional" />
```

```
                <xs:anyAttribute namespace="##any" processContents="lax" />
            </xs:complexType>
        </xs:element>
        <xs:element name="term">
            <xs:complexType mixed="true">
                <xs:anyAttribute namespace="##any" processContents="lax" />
            </xs:complexType>
        </xs:element>
    </xs:schema>
```

Appendix C. Sample document

```xml
<?xml version="1.0" encoding="UTF-8"?>
<glossary version="1.0" srclang="en-US"
    xmlns="http://www.maxprograms.com/gml"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.maxprograms.com/gml GlossML.xsd">
    <comment>Simple bilingual glossary</comment>
    <glossentry>
        <langentry xml:lang="en-US">
            <term>structure</term>
            <definition source="Merriam Webster">the manner in which
                something is constructed</definition>
        </langentry>
        <langentry xml:lang="es">
            <term>estructura</term>
        </langentry>
    </glossentry>
    <glossentry>
        <comment from="RMR">This entry doesn't refer to statistics as a
            science</comment>
        <langentry xml:lang="en-US">
            <term>statistic</term>
            <definition source="Merriam Webster">a numerical fact or datum,
                esp. one computed from a sample</definition>
        </langentry>
        <langentry xml:lang="es">
            <term>estadística</term>
            <definition source="Larousse">cuadro numérico de un hecho que se
                presta a la estadística</definition>
        </langentry>
    </glossentry>
</glossary>
```