XEP-0394: Message Markup

Konstzntin Kozlov
mailto:yagiza@yandex.ru
xmpp:yagiza@isgeek.info

Marvin Wißfeld
mailto:xmpp@larma.de
xmpp:jabber@larma.de

Status	Type	Short Name
Experimental	Standards Track	markup

This specification provides an alternative to XHTML-IM with rigid separation of content and markup information, improving the resilience against spoofing and injection attacks.

Legal

Copyright

This XMPP Extension Protocol is copyright © 1999 – 2024 by the [XMPP Standards Foundation](https://xmpp.org) (XSF).

Permissions

Permission is hereby granted, free of charge, to any person obtaining a copy of this specification (the "Specification"), to make use of the Specification without restriction, including without limitation the rights to implement the Specification in a software program, deploy the Specification in a network service, and copy, modify, merge, publish, translate, distribute, sublicense, or sell copies of the Specification, and to permit persons to whom the Specification is furnished to do so, subject to the condition that the foregoing copyright notice and this permission notice shall be included in all copies or substantial portions of the Specification. Unless separate permission is granted, modified works that are redistributed shall not contain misleading information regarding the authors, title, number, or publisher of the Specification, and shall not claim endorsement of the modified works by the authors, any organization or project to which the authors belong, or the XMPP Standards Foundation.

Warranty

NOTE WELL: This Specification is provided on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE.

Liability

In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall the XMPP Standards Foundation or any author of this Specification be liable for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising from, out of, or in connection with the Specification or the implementation, deployment, or other use of the Specification (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if the XMPP Standards Foundation or such author has been advised of the possibility of such damages.

Conformance

This XMPP Extension Protocol has been contributed in full conformance with the XSF's Intellectual Property Rights Policy (a copy of which can be found at <https://xmpp.org/about/xsf/ipr-policy> or obtained by writing to XMPP Standards Foundation, P.O. Box 787, Parker, CO 80134 USA).

Contents

1 Introduction

Currently, XHTML-IM (XEP-0071) [1] or ad-hoc text-based formats are used to provide styling and semantic information in messages sent over the XMPP network.
These approaches have several drawbacks, including but not limited to:

- Lack of standardisation (ad-hoc text-based formats), and thus interoperability.

- Lack of extensibility (ad-hoc text-based formats).

- Pollution of <body/> with markup information (ad-hoc text-based formats), possibly reducing accessiblity.

- Possibility of sending different textual content in the marked-up version vs. the plain-text version (XHTML-IM), allowing for spoofing attacks.

- Difficult to sanitize potentially malicious input (XHTML-IM mostly) (see e.g. Security Issues with XHTML-IM (again) [2]), leading to injection attacks.

2 Requirements

- Textual data and markup metadata MUST be separated strictly.

- There MUST only be a single source of truth for the text associated with each content language in a message.

- The markup specification MUST be extensible in order to support more complex use-cases in the futurue.

- The markup SHOULD convey semantic information, if possible, as opposed to stylistic information.

- Entities SHOULD be able to cherry-pick a subset of the markup which is suitable for their presentation (for example, a terminal-based client may support inline emphasis and strike through, but no block-level markup).

- The specification MUST NOT require server support.

- Messages using this markup MUST NOT reduce readability for text-to-speech engines and other accessibility technologies.

- Messages using this markup MUST NOT reduce readability for people with color vision deficiencies.

- Requirements on the contents of the <body/> MUST NOT be imposed.

---

[1] XEP-0071: XHTML-IM <https://xmpp.org/extensions/xep-0071.html>.

[2] Security Issues with XHTML-IM (again) <https://mail.jabber.org/pipermail/standards/2017-October/0
33546.html>.

3 Glossary

OPTIONAL.

4 Use Cases

4.1 Transport inline markup

Inline markup is declared with the <span/> element.

```
<message>
  <body>There is really no reason to worry.</body>
  <markup xmlns="urn:xmpp:markup:0">
    <span start="9" end="15">
      <emphasis/>
    </span>
  </markup>
</message>
```

The following child elements are defined for <span/>:

- <emphasis/>: The spanned range is emphasized. Suggested rendering: italics.

- <strong/>: The spanned range is strongly emphasized. Suggested rendering: bold.

- <code/>: The spanned range is some kind of machine code. Suggested rendering: monospaced.

- <deleted/>: The spanned range has been deleted. Suggested rendering: striked through.

The start and end attributes define the range at which the span is applied. They are in units of unicode code points in the character data if the body element. The first affected codepoint is the one at start (where the first codepoint of a message has index 0) and the last affected codepoint is the one just before end. The above example could render in HTML as:
There is *really* no reason to worry.

4.2 Transport code blocks

Code blocks are declared with the <bcode/> element:

```
<message>
  <body>Just run this command:
$ cowsay XMPP is awesome.</body>
  <markup xmlns="urn:xmpp:markup:0">
```

```
      <bcode start="23" end="48" language="bash"/>
  </markup>
</message>
```

The start and end attributes work just like for <span/>.

The optional language attribute allows to specify the programming or markup language used in the code block. TODO: What are valid language names?

The suggested rendering of code blocks is as block-level elements with monospaced font. The above example could render in HTML as:

Just run this command:

$ cowsay XMPP is awesome.

4.3 Transport itemized lists

Itemized lists are declared with the <list/> and elements:

```
<message>
  <body>This XEP supports many things:
* inline markup
* code blocks
* lists
* and possibly more!</body>
  <markup xmlns="urn:xmpp:markup:0">
    <list start="31" end="89" ordered="false">
      <li start="31"/>
      <li start="47"/>
      <li start="61"/>
      <li start="69"/>
    </list>
  </markup>
</message>
```

The start and end attributes of <list/> define the scope of the list. The start of the elements denote the start of a new list item. A list item continues until the end of the list or the start of the next list item. The first in a <list/> MUST have a start value equal to the start value of the <list/>.

The optional ordered attribute may be set to true to indicate the list is ordered and displaying it with numerical or alphabetical ordering instead of bullets is suggested.

The above example could render in HTML as:

This XEP supports many things:

- * inline markup

- * code blocks

- * lists

- * and possibly more!

4.4 Transport blockquotes

A block quote is declared with a <bquote/> element:

```xml
<message>
  <body>He said:
&gt; Thou shalt not pass!
and raised his hand.</body>
  <markup xmlns="urn:xmpp:markup:0">
    <bquote start="9" end="32"/>
  </markup>
</message>
```

In addition, References (XEP-0372) [3] or a similar mechanism MAY be used to attribute the origin of the quote. The above example could render in HTML as:

He said:

> Thou shalt not pass! and raised his hand.

A nested quotation can be created by adding two <bquote/> elements where the start/end range is nested. If plain text quotation markers are used, the start of the blockquote MUST be placed at the first quotation marker of the *outer* quote.

```xml
<message>
  <body>&gt; He said:
&gt;&gt; Thou shalt not pass!
&gt; and raised his hand.

Isn't this from some famous movie?</body>
  <markup xmlns="urn:xmpp:markup:0">
    <bquote start="0" end="57"/>
    <bquote start="11" end="34"/>
  </markup>
</message>
```

The above example could render in HTML as:

> He said: >> Thou shalt not pass! > and raised his hand. Isn't this from some famous movie?

5 Business Rules

- Spans MUST NOT overlap with each other.

- Spans MUST NOT overlap with the boundaries of a block-level markup element, but MAY be fully contained within a block-level markup element.

---

[3]XEP-0372: References <https://xmpp.org/extensions/xep-0372.html>.

- Block level markup elements MUST NOT overlap with each others boundaries.

- There MUST NOT be a <markup/> element in a <message/> without corresponding <body/>. Note that there may be one <markup/> elements with appropriate xml:lang attribute value for each <body/>, if the message contains multiple <body/> elements.

- The start and end attributes operate on unicode code points in the XML character data of the corresponding <body/> element.

- Entities MUST silently ignore elements and attributes (arbitrarly deep) in <markup/> which they do not understand; this allows for future extensions of the markup without breaking existing implementations.

6 Implementation Notes

OPTIONAL.

7 Accessibility Considerations

Entities are encouraged use the semantic information to make the presentation of the textual content more precise, for example by applying spoken emphasis to passages marked with an <emphasis/> <span/>.

8 Internationalization Considerations

Since a message may have multiple <body/> elements in different languages, there MAY be multiple <markup/> elements, one for each of the <body/> elements. There is no requirement to include a <markup/> element for each language.

9 Security Considerations

REQUIRED.

10 IANA Considerations

This document requires no interaction with the Internet Assigned Numbers Authority (IANA) [4].

---

[4]The Internet Assigned Numbers Authority (IANA) is the central coordinator for the assignment of unique parameter values for Internet protocols, such as port numbers and URI schemes. For further information, see

11 XMPP Registrar Considerations

This specification defines the following XML namespaces:

- urn:xmpp:markup:0

12 XML Schema

REQUIRED for protocol specifications.

13 Acknowledgements

Thanks to Georg Lukas and Emmanuel Gil Peyrot for feedback on the initial idea of this XEP.

<http://www.iana.org/>.