jindw / xmldom Public

Code Issues 93 Pull requests 45 Actions Projects Wiki ...

New issue

Jump to bottom

Missing error for XML documents with multiple root element nodes #150

Open markgollnick opened this issue on Jan 19, 2016 · 6 comments

markgollnick commented on Jan 19, 2016

Bug Description:

xmldom allows incoming documents to have multiple root element nodes. This appears to me to be a violation of the W3C DOM Level 2 Core Specification:

> [...] Each document contains zero or one doctype nodes, **one root element node,** and zero or more comments or processing instructions; the root element serves as the root of the element tree for the document. [...]
>
> — *from "What the Document Object Model is" (emphasis mine)*
>
> **The root node is the unique node that is not a child of any other node.** All other nodes are children or other descendants of the root node.
>
> — *from the Glossary (emphasis mine)*

However, the spec *also* says this:

> [...] However, **the DOM does not specify that documents must be implemented as a tree or a grove,** nor does it specify how the relationships among objects be implemented. **The DOM is a logical model that may be implemented in any convenient manner.** [...]
>
> — *from "What the Document Object Model is" (emphasis mine)*
>
> The children of a `DocumentFragment` node are zero or more nodes representing the tops of any sub-trees defining the structure of the document. `DocumentFragment` nodes do not need to be *well-formed XML documents* (although they do need to follow the rules imposed upon **well-formed XML parsed entities, which can have multiple top nodes**). For example, a `DocumentFragment` might have only one child and that child node could be a `Text` node. Such a structure model represents neither an HTML document nor a well-formed XML document.
>
> — *from § 1.2 / Fundamental Interfaces / Interface DocumentFragment (emphasis mine)*

In light of this, for the record, **I actually don't dislike the fact that xmldom can parse such documents.** This comes with a few reservations (outlined below) because it seems like the current behavior is contrary to what the specs (both W3C DOM Level 2 Core *and* XML 1.1) assert ought to be the case.

Bug Reproduction:

The following code:

```
var DOMParser = require('xmldom').DOMParser;
var xmlData = '<?xml version="1.0" encoding="UTF-8"?>\n' +
'<root>\n' +
'  <branch girth="large">\n' +
'    <leaf color="green" />\n' +
'  </branch>\n' +
'</root>\n' +
'<root>\n' +
'  <branch girth="twig">\n' +
'    <leaf color="gold" />\n' +
'  </branch>\n' +
'</root>\n';
var xmlDOM = new DOMParser().parseFromString(xmlData);
console.log(xmlDOM.toString());
```

...produces the following output:

```
<?xml version="1.0" encoding="UTF-8"?><root>
  <branch girth="large">
    <leaf color="green"/>
  </branch>
</root>
<root>
  <branch girth="twig">
    <leaf color="gold"/>
  </branch>
</root>
```

In contrast to this, libxmljs — which relies on libxml2 — refuses to parse such documents, opting to throw an error instead:

```
var xml = require('libxmljs');
var xmlDoc = xml.parseXmlString(xmlData);
Error: Extra content at the end of the document
```

Firefox behaves in a similar way, and refuses to parse the document:

```
console.log(
  new XMLSerializer().serializeToString(
    new DOMParser().parseFromString(xmlData, 'text/xml')));
```

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="chrome://global/locale/intl.css" type="text/css"?>
<parsererror xmlns="http://www.mozilla.org/newlayout/xml/parsererror.xml">XML Parsing Error: junk after document element
Location: about:blank
Line Number 7, Column 1:<sourcetext>&lt;root&gt;
^</sourcetext></parsererror>
```

Chrome goes a *little* bit farther in that it is at least willing to parse/render the first root element node:

```
<?xml version="1.0" encoding="UTF-8"?>
<root>
  <parsererror xmlns="http://www.w3.org/1999/xhtml" style="display: block; white-space: pre; border: 2px solid #c77; padding: 0 1em 0 1em; margin: 1em; background-color: #fdd; color: black">
    <h3>This page contains the following errors:</h3>
    <div style="font-family:monospace;font-size:12px">error on line 7 at column 1: Extra content at the end of the document
</div>
    <h3>Below is a rendering of the page up to the first error.</h3>
  </parsererror>
  <branch girth="large">
    <leaf color="green"/>
  </branch>
</root>
```

Expectations/Recommendations/Discussion:

- I expected the above document to fail to parse in xmldom's DOMParser implementation, but it didn't. In light of the specs, this seems like it might be a "false-positive" bug. However, I think there *is* some utility in being able to parse multiple documents in a single stream, *if* the API can be re-tooled just a bit. More on this below.
- In order to be more compliant with the specs, I think that if more than one root node is present in an XML stream — especially if they appear without corresponding `<?xml?>` declarations — then the DOMParser should, bare-minimum, simply opt to not support such streams, and should raise an error informing the consumer that the given XML stream is invalid.
- **Alternatively,** if it is desired to keep the ability to parse multiple documents in a single stream — which I think would be useful functionality to retain: the alternative of requiring users to manually split XML documents crammed into a single stream before they can pass them into xmldom seems counter-intuitive to me, as that should be the job of the parser — then perhaps the DOMParser API could be re-tooled to return an *array* or a *list* of multiple document objects, each one with a singular root element node (rather than a *single* document object with *multiple root elements,* as is the case today), but *only in the event that multiple root elements are detected within the input stream.* I think that this behavior would help xmldom to be more compliant with the specs, *and* it would help to prevent consumers from accidentally generating poorly formed XML output from their parsed document objects, since each parsed document object would have only *one* root node, as the specs indicate ought to be the case.

Conclusion:

On one hand, **I think it is a useful thing to be able to parse streams containing multiple XML documents.**

On the other hand, to do that silently without issuing even so much as a warning to consumers — especially when the specs say that root nodes should be unique — seems, well, odd. **It seems like the current behavior is more of an artifact or a side-effect of xmldom’s current architecture than it is an intentional aspect of its design,** hence my writing this up as a bug rather than a feature request. (If I’m mistaken, do let me know!)

Environment/Versions:

- xmldom v0.1.21

👍 2 👎 1 ❤️ 1

markgollnick commented on Jan 19, 2016 Author

Sorry if this has already been documented elsewhere. I searched around and didn’t see this behavior being discussed anywhere, so I figured I’d file it here and see where it went.

frumioj commented on Oct 14

Ultimately responsible for this security bug: https://nvd.nist.gov/vuln/detail/CVE-2022-39299

frumioj referenced this issue in node-saml/passport-saml on Oct 14

`Merge pull request from GHSA-m974-647v-whv7` 8b7e3f5

karfau commented on Oct 14 • edited

@frumioj xmldom is currently maintained at https://github.com/xmldom/xmldom.
What would you suggest that we change in the behavior/implementation (as in what should be the preferred behavior in the context of the issue you linked, what assumption was made, which xmldom didn't fulfill)?
Feel free to involve/ping me in an already ongoing discussions elsewhere or to kick off the process following our security policy.
Thx for the information.

karfau commented 28 days ago

Fixes have been provided for `@xmldom/xmldom` in the versions 0.7.7, 0.8.4 and 0.9.0-beta.4.
For more details see GHSA-crh6-fp67-6883

MahmoudNafiseh mentioned this issue 25 days ago

update @xmldom/xmldom package expo/expo-cli#4585

Closed

alxndrstffn mentioned this issue 24 days ago

Upgrade to xmldom 0.8.5 (CVE-2022-39299) mswjs/interceptors#296

Closed

alvaldes mentioned this issue 23 days ago

Dependency Error: xmldom allows multiple root nodes in a DOM Sesame-Disk/cool-storage-microfrontend#39

Open

meganwalker-ibm commented 12 days ago

GHSA-crh6-fp67-6883 notes that 0.7.7 is both affected and fixed. Can the impacted version ranges be clarified?
Affected versions: <=0.7.7
Patched versions: ~0.7.7

I guess that should probably be < not <=?

👍 1

karfau commented 12 days ago • edited

Thank you for letting me know, you guessed correctly, I fixed it.

Assignees

No one assigned

Labels

None yet

Projects

None yet

Milestone

No milestone

Development

No branches or pull requests

4 participants