Tutorial: Invisible Markup

General information:

Title of the tutorial: Invisible Markup

Organizers' and Presenters' names, affiliation, contact information, and brief bio:

Steven Pemberton, CWI, Amsterdam; steven.pemberton@cwi.nl; +31 624 671 668

Steven Pemberton is a researcher at CWI, the Dutch national research centre for mathematics and computing science. At university his tutor was Dick Grimsdale, who built the world's first transistorised computer and whose tutor was Alan Turing, making Pemberton a grand-tutee of Turing.

Pemberton co-designed the programming language that Python was based on, he was the first user of the European open internet in 1988, he organised workshops at the very first Web Conference of this series in 1994, and was involved with W3C from the very beginning, chairing the first workshop on stylesheets, chairing the HTML working group, co-designing CSS, HTML, XHTML, RDFa, XForms and several others. He currently chairs two groups at W3C, and is co-organiser of an annual conference on declarative technologies. In 2022 he received the ACM SIGCHI Lifetime Practice Award.

Abstract

1-2 paragraphs suitable for inclusion in the conference registration material.

Humans are good at identifying implicit structure in notations. People can deduce the structure of a date like 30 April 2023 with no help. Computers on the other hand need extra information, and that is why we have markup languages, like `<date><day>23</day><month>May</month></month><year>2023</year></date>` to make the structure explicit. *Invisible Markup* is a method of automatically discovering the structure in notations, and adding markup.

It doesn't matter what the input form is, whether CSV, or JSON, CSS, bibliographic entries, family tree information, or countless other examples, the output is a consistently marked up result, currently XML being the target, as the most general of the available markup languages. For instance the input might be

```
[spec] Steven Pemberton (ed.), Invisible XML Specification, invisiblexml.org, 2022,
       https://invisiblexml.org/ixmlspecification.html
```

and while you have a lot of control over the details, the output could be

```
<biblioentry>
   <abbrev>spec</abbrev>
   <editor>
      <firstname>Steven</firstname>
      <surname>Pemberton</surname>
   </editor>
   <title>Invisible XML Specification</title>
   <publisher>invisiblexml.org</publisher>
   <pubdate>2022</pubdate>
   <link href='https://invisiblexml.org/ixml-specification.html'/>
</biblioentry>
```

The Invisible Markup language ixml was formally released as a standard in the Summer of 2022. This hands-on tutorial introduces you to the principles, and takes you through how to use it.

Topic and relevance:

A description of the tutorial topic, providing a sense of both the scope of the tutorial and depth within the scope, and a statement on why the tutorial is important and timely, how it is relevant to the Web Conference, and why the presenters are qualified for a high-quality introduction of the topic.

We are often obliged for different reasons to represent data in some way or another, but in the end those representations are all of the same abstraction; there is no essential difference between the JSON

```
{"temperature": {"scale": "C"; "value": 21}}
```

and an equivalent XML

```
<temperature scale="C" value="21"/>
```

or

```
<temperature>
    <scale>C</scale>
    <value>21</value>
</temperature>
```

or indeed

```
temperature: 21°C
```

since the underlying abstractions being represented are the same.

What ixml does is takes a representation of data (typically with implicit structure), uses a description of the format of that data to recognise the data's structure, creates an internal representation of the data, now with the structure made explicit, which can then be used for multiple purposes, including creating an external representation with explicit structure.

ixml is a newly released standard from a community group at W3C. The fact that there are already 7 implementations either ready or in preparation shows the wider commuunity interest.

The tutorial introduces the whole language and techniques; the presenter is the originator of the technology, and a well-known speaker.

Style and Duration:

Please indicate whether this will be a lecture-style or hands-on tutorial. In the case of the latter, please indicate the equipment needs for participants (e.g., pre-installed Jupyter notebook with specific packages).

This is a hands-on tutorial, length 3 hours, and uses a rapid-fire sequence of exercises, each consisting of 5 minutes of presentation, followed by 5 minutes of coding. This maximises attendees' concentration, thanks to the recurrent switching of activities. People should bring a laptop with a browser, but no installation is required.

Audience:

A description of the intended audience, prerequisite knowledge, and the expected learning outcomes.

Anybody who knows a little about a markup language such as HTML will be fine attending this tutorial. At the end of the tutorial they will be able to write notation descriptions that allow them to convert input notations into marked up versions.

Previous editions:

If the tutorial was given before, where and when was it presented? Please give details on the number of attendees, and how the proposed tutorial differs or builds on the previous ones. If possible, provide a link to the material of the previous version of the tutorial.

This tutorial has been given twice before: at Declarative Amsterdam November 2021, and XML Prague in June 2022.

The tutorial is viewable at http://www.cwi.nl/~steven/ixml/tutorial/
The slides at https://www.cwi.nl/~steven/Talks/2022/06-11-ixml/
A video at https://www.youtube.com/watch?v=_qo5CWEvYfs&t=18263s

Both times the attendees numbered around 70 (a mixture of live and online)

The current tutorial builds on experience of giving the tutorial twice, adapting material and exercises, and adjusting it to the final released version of the specification.

Tutorial material:

What materials will be provided to attendees of the tutorial? Are there any copyright issues?

The attendees will have access to the tutorial via the link for work in their own time after the conference. The tutorial contains worked examples of all exercises. There are no copyright issues.

Equipment:

Indicate any additional equipment needed (if any) in the room. The standard equipment includes an LCD projector, a single projection screen and microphones.

None.

Video teaser:

A Video teaser, up to 3 minutes, is required at the time of submission. The video can be hosted on any video sharing platform (e.g., YouTube) or any file sharing service (e.g., WeTransfer, Dropbox) and the link to the video MUST be included in the proposal.

May I propose you watch the first three minutes of https://www.youtube.com/watch?v=_qo5CWEvYfs&t=18263s

Organization details:

Tutorial organizers are required to provide a backup plan that overcomes the potential occurrence of technical problems, e.g., pre-recorded lectures, self-paced exercises. Please also describe how the tutorial will work in an online or hybrid setting. The tutorial presenter(s) will be responsible for making sure that the slides and any material needed for the tutorial are made available online in advance for attendees. For tutorials that introduce or use standards or software, the tutorial must be based on the latest version of the standards and software.

Both previous times, the tutorial was given hybrid. There is a pre-recorded version of the tutorial available (see above), and the slides are already online. The tutorial is based on the latest version of the standard. The tutorial is also designed for self-study use.