RUHR-UNIVERSITÄT BOCHUM

Bridging the Gap: Secure and lossless conversion of XML data structures to the JSON format

Bachelor thesis ■ March 30, 2017 – June 29, 2017

May 23, 2017

Advisors: Dennis Felsch & Paul Rösler

Jan Holthuis

hgi **Lehrstuhl für Netz- und Datensicherheit**

hg NDS

Last time we learned that...

- During conversion, we want to make sure that document content and logical structure stays the same...
- ...but the physical structure is not important.
- Thus we can use the XPath Data Model and Canonicalization for abstraction
- I gave some conversion critera examples:
    - Comments are *not* important ($\rightarrow$ XML 1.0, Sec 2.5)
    - Element Order *is* important ($\rightarrow$ XPath 1.0, Sec 5)
    - Arribute Order is *not* important ($\rightarrow$ XML 1.0, Sec 2.5)

hg NDS

Progress made in the last two weeks
Progress Report I

- Spent a lot of time reading W3C recommendations
- Added more conversion criteria
- ... and continued writing my thesis.

hg NDS

Progress made in the last two weeks
Progress Report II

- Based on my current conversion critera, *no existing solution* allows lossless XML→JSON→XML roundtrips. . .

- Based on my current conversion critera, *no existing solution* allows lossless XML→JSON→XML roundtrips. . .
- . . . but I found a promising candidate that *almost* gets it right: **JsonML**

1 Quick Recap

2 Progress report

3 Currently leading solution

4 Next steps

hg NDS

- JsonML = "JSON Markup Language"
- Originally written in JavaScript, but there are also other implementations (e.g. Java: `org.json.JSONML`)
- Fairly mature: Github repo dates back to November 2006
- Nicely documented grammar (Backus-Naur form)
- Downsides:
  - Resulting *JSON* is not really "friendly"
  - No support for XML Processing Instructions (no tested solution supports them) → I already wrote a patch to fix this

JsonML
Example Transformation

```
<albums>
  <?my-pi with some data ?>
  <album catno="ARGO LP-667">
    <artist>The Ahmad Jamal Trio</artist
    ↪ >
    <title>Jamal At The Pershing/Vol. 2<
    ↪ /title>
    <recording>Recorded <date>January
    ↪ 17, 1958</date>, in live
    ↪ performance at the <location>
    ↪ Pershing Lounge, Chicago</
    ↪ location>.</recording>
  </album>
</albums>
```

```
[ "albums", "\n  ", [
  "album", { "catno": "ARGO LP-667" }, "\n    ",
  [ "artist", "The Ahmad Jamal Trio" ], "\n    ",
  [ "title", "Jamal At The Pershing/Vol. 2" ], "\
    ↪ n    ",
  [ "year", "1960" ], "\n    ",
  [ "recording",
    "Recorded ",
    [ "date", "January 17, 1958" ],
    ", in live performance at the ",
    [ "location", "Pershing Lounge, Chicago" ],
    "."
  ], "\n  "
], "\n" ]
```

xmljson-Badgerfish
Example Transformation (for comparison)

```xml
<albums>
  <?my-pi with some data ?>
  <album catno="ARGO LP-667">
    <artist>The Ahmad Jamal Trio</artist
      ↪ >
    <title>Jamal At The Pershing/Vol. 2<
      ↪ /title>
    <recording>Recorded <date>January
      ↪ 17, 1958</date>, in live
      ↪ performance at the <location>
      ↪ Pershing Lounge, Chicago</
      ↪ location>.</recording>
  </album>
</albums>
```

```json
{ "albums": {
    "album": {
      "@catno": "ARGO LP-667",
      "artist": { "$": "The Ahmad Jamal Trio" },
      "title": { "$": "Jamal At The Pershing/Vol.
      ↪  2" },
      "year": { "$": 1960 },
      "recording": {
        "$": "Recorded",
        "date": { "$": "January 17, 1958" },
        "location": { "$": "Pershing Lounge,
      ↪ Chicago" }
      }
    }
} }
```

Overview

1 Quick Recap

2 Progress report

3 Currently leading solution

4 Next steps

Next Steps

- Overcome my Writer's Block ;-)
- Complete checking tool unittests
- Look for more conversion criteria and recheck everything (Did I miss any edge cases?)
- ???
- Profit!

Thanks!

Questions?

Reach out via email:

- **Jan Holthuis**
  jan.holthuis@rub.de