

ANOTHER TEI ODD PROCESSOR

Syd Bauman, Helena Bermúdez-Sabel, Martin Holmes, David Maus



OUTLINE

- a few words about TEI
- what's ODD
- and why we need another processor
- the work so far



TEXT ENCODING INITIATIVE

- 1980s: CoP with the goal to develop guidelines for encoding machine-readable texts of interest to the humanities and social sciences
- What the TEI is today: its own ecosystem
 - technical standard
 - toolchain
 - conference series, mailing list, journal, repository, wiki, 33 GitHub repositories...



THE TEI GUIDELINES

The *TEI Guidelines* and its schemas are generated from a single XML resource that contains:

1. Descriptive prose
2. Examples of usage
3. Formal declarations for components of the TEI Abstract Model:
 - elements and attributes
 - modules
 - classes and macros



ODD: ONE DOCUMENT DOES IT ALL

- TEI's own built-in schema language also used for customizing the TEI language and documenting the customizations (Bauman 2019)
- ODD chaining: create a customization derived from other ODD customizations
- ODD as a literate programming language for XML schemas separate from TEI (e.g. Internationalization Tag Set)



TOOLS

- **TEI Stylesheets** (to transform TEI XML documents to various formats and to/from a few formats, including the conversion of a customization ODD to a customized RELAX NG schema with embedded Schematron rules)
- **Roma** (web application for ODD creation and customization)



THE TEI STYLESHEETS

- Written by the late Sebastian Rahtz
- “The TEI Stylesheets are a work of genius” (Holmes 2021: 3)
- “Sebastian was such a genius he really found it impossible to imagine how dim the rest of us actually are” (Holmes 2021: 3)
- “I consider the Stylesheets as broken beyond repair. They work to a certain degree but they are, from my point of view, effectively unmaintainable.” David Maus, commenting on a PR, 2020-11



ATOP: ANOTHER TEI ODD PROCESSOR

- Mission: new processor ODD → RNG+Schematron
- (Custom documentation is being tackled by a different subgroup linked to TEIPublisher)



SOFTWARE DESIGN

- Future maintenance informs every decision
- Modular programming
- Decision documentation
- Well documented test suites from the start



SOME OF THE ESTABLISHED GUARDRAILS

- Strict file naming convention
- Strict naming conventions for variables, parameters, functions, and templates
- Mandatory use of @as and of xd : elements
- Strict definition of what subset of TEI ODD will be processed (and not)
- Schematron rules and XSpec scenarios



THE WORK SO FAR

- Our GitHub repo: <https://github.com/TEIC/atop>
- Detailed Wiki



Requirements of a base ODD

David Maus edited this page on Apr 15 · 1 revision

A base ODD is an ODD that readily transpiles to a RelaxNG schema without resolving external specifications. A Schematron schema checks if an ODD satisfies these requirements before it is fed to the transpiler.

References

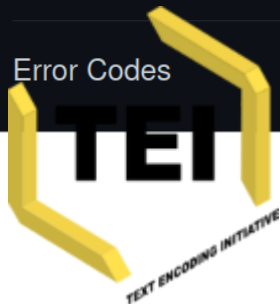
An `attRef` with only a `@class` attribute must reference an existing `classSpec` with a `@type` of `atts`.

An `attRef` with only a `@name` attribute must reference an existing `attDef` within a containing `attList`.

An `attRef` with both a `@class` and a `@name` attribute must reference an existing `attDef` inside an existing `classSpec` with a `@type` of `atts`.

See also

Error Codes



THE WORK SO FAR

- Our GitHub repo: <https://github.com/TEIC/atop>
- Detailed Wiki
- Early stages



REFERENCES

- Bauman, Syd (2019): “A TEI Customization for Writing TEI Customizations”. *Journal of the Text Encoding Initiative*, Issue 12. [doi:10.4000/jtei.2573](https://doi.org/10.4000/jtei.2573)
- Holmes, Martin (2021): “Are the TEI Stylesheets really broken beyond repair?” (Conference presentation.) *Text Encoding Initiative Conference 2021* (online). 2021-10-25.



THANK YOU!

