Customizing Akoma Ntoso

Prof. Fabio Vitali
University of Bologna

Purpose of this talk

- To list the actions that need to be taken to customize Akoma Ntoso to the specifics of parliamentary documents of your country.

The Document Architecture

- Akoma Ntoso describes two different but connected families of document formats
  - The **General Schema**: One vocabulary and minimal set of constraints that all documents must comply to.
  - **Detailed Schemas**: A set of stricter schemas. They provide more constraints over the same vocabulary to enforce the rules of specific national Pa
  - rliaments.
- All documents satisfying one of the Detailed Schemas must also satisfy the General Schema.
- The General Schema is absolutely descriptive, and is meant to be applied consistently across all adopting countries.
- The Detailed schemas are country-specific, and can be more prescriptive, as long as the legal drafting office can impose forms and structures to the Parliament itself.

Current classes of documents

- Bills and Acts
  (laws being drafted and in force)
  - National legislation
  - Local legislation (regions, provinces, cities, etc.)
  - Non-Parliamentary legislation (decrees, authorities, regulations, etc.)
- Minutes
  (detailed logs of meetings with attribution of utterances)
  - Parliamentary meetings
  - Committees meetings
  - City council meetings
- Other documents
  (of various forms and structures)
  - Attachments
  - Weekly and daily agendas
  - …

Current organization of documents

- Every document class is organized in five parts:
  - Metadata (not part of the content)
  - Introductory material
  - Body of the document
  - Conclusion of the document
  - Attachments
- Each part may be composed of subparts.
  - E.g.: body of acts if a hierarchy of sections and subsections
  - E.g.: introduction of acts is composed of preface (recital?) and preamble
- Names and hierarchies of parts
  - In the generic schema are very flexible
  - In the detailed schema can be more rigorous

How to organize work

- Categories for which there is a need for customization
  - Document types
  - Document sources
  - Hierarchy and containment
  - Attachments
  - References
  - Semantically relevant subparts
- You take notes and propose
  - Specific vocabulary
  - Additional rules
  - Exceptions
  - Frequency of exceptions
  - Special cases

Specifying document types

- The *generic document* type is extremely flexible and can be used to describe just any kind of document, from novels to recipes to technical manuals to legislative documents.
- Why bother with detailing specific document types?
  - To provide better drafting rules
  - To provide better understanding of internal structure
  - To provide increased uniformity of structure
  - To provide more precise internal references

Specifying document types

- Hierarchical documents
  - Meta, preface, preamble, hierarchy, conclusions, attachments.
  - Hierarchy is composed of part, subtitle, section, subsection, paragraph, chapter, article, clause.
  - Order of containment not relevant at this stage
  - Q1: Besides bills & acts, what other named document type follows the same overall organization in your country?
- Debate structures
  - Meta, preface, debate, conclusions, attachments
  - Debate is composed of subdivisions organizing speeches, questions and answers.
  - Order and containment not relevant at this stage
  - Q2: Besides minutes, what other named document type follows the same overall organization in your country?
- Open structures
  - Just about every possible structure available
  - Q3: What named document type must be considered that has no fixed structure and can take whatever form?
  - Q4: Is there any other organized structure for some document type that it is worth considering for Akoma Ntoso?

Document sources

Q5: Document sources

1. What is the authoritative publication source for each document type?
2. What is the frequency of publication, numbering schema, official name?
3. Are there different series and how are they called and what are they for?
4. What other type of content do they publish? Separately or interleaved with managed document types?

Hierarchy and containment

- The Generic schema does not provide any required sequence of containment of parts and part names. The Detailed Schemas may impose more constraints
- Q6: Is the following vocabulary adequate?
  - part, subtitle, section, subsection, paragraph, subparagraph, chapter, article, clause
- Q7: Is there a required or frequent sequence of containment?
  - E.g.: part/subtitle/section/subsection/paragraph/subparagraph
- Q8: Actual documents follow such sequence
  - Rigorously (no exceptions: document would be rejected)
  - Most always or frequently (exceptions now and then, but rare and despised)
  - Rarely (would not surprise anyone if found. 10% or more of documents are exceptions)
- Q9: Which parts are numbered and how?
  - Numbering is arabic/roman/alfabetic
  - Numbers start always at 1 for each subpart/numbering is global for the document

Attachments (Schedules)

- Q10: Are tables, schemas, images placed inline or in attachments?
- Q11: How are attachments introduced?
  - A title, a number, an introductory text
  - Introductory material belongs to the host document or to the attachment?
- Q12: Can attachments have attachments?
- Q13: How are attachments referred to within the document? From another document?

References

- Q14: How are whole documents referenced? Is there a standard syntax?
- Q15: Is referencing done by name, number, date, all of the above?
- Q16: How are parts of documents referenced (e.g. section 5(1) of such and such act)? Is there a standard syntax?
- Q17: How often such standard syntax is ignored by legislator? How many exceptions to syntax?
  - Never (no exceptions: document would be rejected)
  - Very rarely (exceptions now and then, but rare and despised - would raise an eyebrow)
  - Often (would not surprise anyone if found. 10% or more of references are exceptions)
- Q18: How about deixis? (e.g.: “next section”, “above-mentioned act”, etc.)

Semantically relevant inlines

- Inlines are part of the document that are not structurally relevant (they appear within paragraphs) but may have importance semantically.
- Akoma Ntoso requires that at least all references are correctly marked up. But other parts can and possibly should.
  - Defined terms (a.k.a. Interpretation), sums, dates, individuals, organizations, offices, places, times, etc.
  - Parts in the preface: title, number, date, proponent, purpose, etc.
- Q19: What inlines of the document you think relevant to identify systematically?
- Q20: How often in documents appear inlines that are NOT semantically relevant, but have different typographical aspect (e.g., bold, italic)

Metadata

- Metadata for identification
  - Publication and further publications details
  - Enactment date
  - Official identifier
  - Alias (secondary names)
  - Keywords (taken from a thesaurus such as Eurovoc)
- Metadata for lifecycle
  - Individuals
  - Roles (Speaker of the house, ministries, etc.)
  - Organizations
  - Other documents
- Metadata for preservation
  - Editors' names and their roles
  - Storage details and digital signatures
  - Versioning information and metadata versioning

Next: Metadata (2)

Metadata (2)

- Metadata for consolidation
  - Affecting documents / affected documents
  - Date of application of amendment
  - Type of amendment
- Metadata for provision analysis
  - Types of provisions
  - Attributes of provisions

- Q21: Are we forgetting other types of metadata?

Conclusions

- Before trying to create a Detailed Schema for individual countries, we need to write down how different countries write and describe common document types
  - These 21 questions are in that direction
- Before extending the General Schema and the Detailed Schema towards new document types, we need to find similarities and differences from exiting types.