

Tagging CCP Constitutions

Roy Gardner

August 19, 2024

Contents

1	Introduction	2
2	Outline of the Problem	2
3	Recommendations	2
3.1	Tagged constitutions	2
3.2	Untagged constitutions	3
4	Requirements for the ConstituteProject site	3

1 Introduction

Constitutions are structured documents that are serialised as XML. Document hierarchy is represented by nested elements with types:

- `title`
 - `body`
 - `olist`

Tagging of constitution provisions provides the textual evidence for the presence of topics in constitution text.

2 Outline of the Problem

Tagging of `title` elements of constitutions is prevalent. The assumption behind `title` element tagging is that some or all `body` or `olist` sub-elements contain text that matches the topic.

Once a `title` element has been tagged, there are two possible states:

- Every sub-element below a tagged `title` element contains text that matches the topic.
- Only a subset of sub-elements below a tagged `title` element contain text that matches the topic.

Because it is impossible to automatically distinguish between these two states, tagging `title` elements reduces the textual evidence for topics. This has an impact on downstream qualitative and quantitative analyses that rely on this textual evidence.

3 Recommendations

3.1 Tagged constitutions

- If a `title` element has been tagged because one or more of its sub-elements contain substantive text related to a topic, then the topic must be removed from the `title` element and transferred to the relevant `body` or `olist` sub-elements.
- If a `title` element containing substantive textual content related to a topic has been tagged, then the tag can remain. However, any relevant `body` or `olist` sub-elements must also be tagged.

3.2 Untagged constitutions

- Tags must be applied to `body` or `olist` sub-elements.
- If a `title` element contains substantive textual content related to a topic, then the `title` element may also be tagged.

4 Requirements for the ConstituteProject site

Contextualisation of tagged provisions is an important feature of the ConstituteProject site, and implementing the above requirements should have no impact on the ability to display provisions in context. Given that many `body` and `olist` elements are already tagged, the capability of displaying these elements in context must already exist.

If all elements in the same level are tagged with a topic, then the elements are displayed in order and contextualised with the parent `title` element. If only a subset of elements in a level are tagged with a topic, then ellipses can be used to indicate the presence and position of untagged elements.