**法律法规文本标准化**

**（2021年1月26日星期二）**

考虑到法律法规的特殊性，将法律法规文本的特殊性，将法律法规文本的标准化单独做成一个格式标准化类别，名称为**“法律法规文本标准化”**。工作人员不需调整任何参数，直接固定下来。需求如下：

一、主标题

**字体：**宋体、二号、加粗、居中。

**覆盖范围：**整个主标题，至换行符出现。

**规则：**和现有的识别规则相同。

二、副标题

**字体：**楷体GB2312、三号、加粗、居中。

**覆盖范围：**整个副标题，至换行符出现。

**规则：**和现有的识别规则相同。

三、第X编、第X章、目录、序言为一级标题

**字体：**黑体、三号、居中。

**覆盖范围：**（1）第X编、第X章及其后续内容，至换行符出现。（2）目录、序言，仅仅覆盖至目录、序言，没有后续内容。

**规则：**和现有的识别规则不同。不同点如下：

1.有且只有“第X编”“第X章”“目录”“序言”识别为一级标题，此处不再将“一、”等视为一级标题。

2.表现形式为：（1）第一编、第二编……，第一章、第二章……，范围覆盖到100编\100章;（2）目录、序言，必须是单行只有目录、序言两个字的情况，可能的情况是目录两个字中间有若干空格。

3.第X编、第X章和后面的内容之间，需要空两格。此处可能需要判断后面有没有两个空格，有些是没有空格、有些是有一个空格，需要统一为两个空格。如下：

第一章/编国家大力发展科学事业【标准化前】

第一章/编 国家大力发展科学事业【标准化后】

3.“目录”“序言”之间需要空两格。此处可能需要判断后面有没有两个空格，有些是没有空格、有些是有一个空格，需要统一为两个空格。如下：

目录【标准化前】

目 录【标准化前】

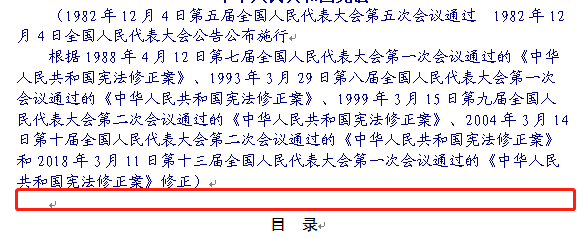
目 录【标准后】

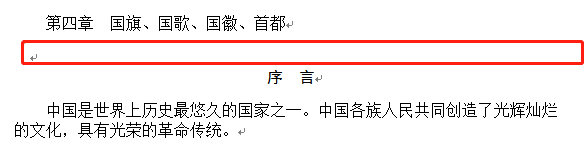
序言【标准化前】

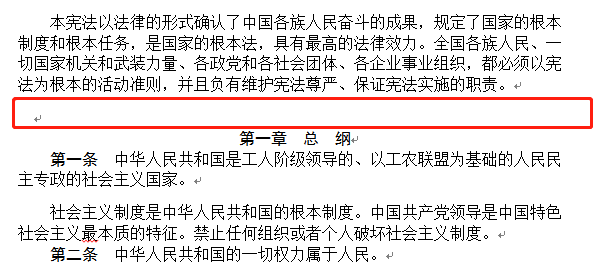
序 言【标准化前】

序 言【标准后】

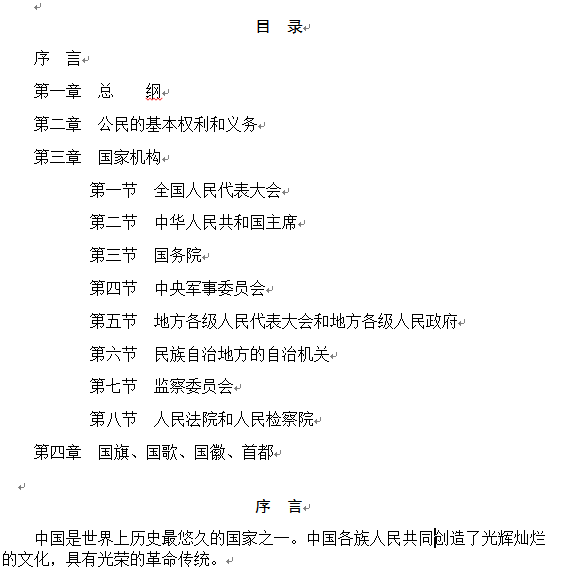
4.向上空一行。一级标题与前述内容之间，也就是向上需要空一行。如下：







5.目录内容的特殊处理规则。这是个难点，要处理目录和其他规则的冲突问题，需要斟酌下如何实现。样式如下图：



目录下内容的规则如下：

1.第X编、序言、第X章等一级标题。

**字体：**仿宋GB2312、三号、缩进2个汉字字符；

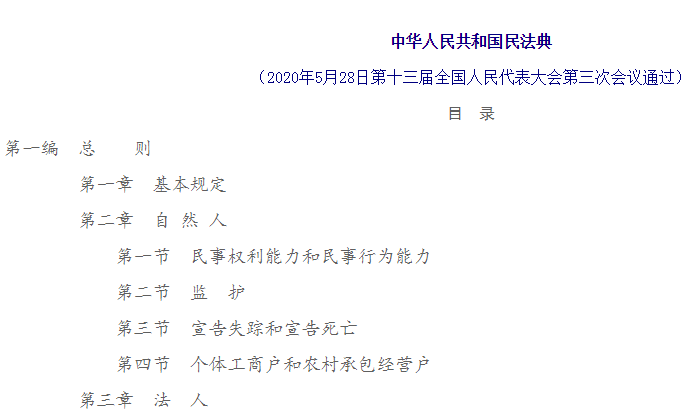
【注】上述情况是指没有“第X编”的情况。

如果有第X编，则第X编缩进2个汉字字符；其他照此类推，序言、第一章等缩进4个汉字字符；二级标题则缩进6个汉字字符。

2.第X节等二级标题。

**字体：**仿宋GB2312、三号、缩进4个汉字字符；

示例如下：



我能够想到的识别处理思路是：

1.将目录内容单独提出来作为一块，适用目录模块的特殊规则。

2.识别规则是“目录”之下，至第二个“序言”“第一章”出现截止，为目录内容。目录下的第一行内容为啥，就截止至到下一个该行内容出现。例如，没有序言，直接是第一章的话，则至“第一章”截止。

四、第X节为二级标题

**字体：**楷体GB2312、三号、加粗、居中。

**覆盖范围：**第X节及其后续内容，至换行符出现。

**规则：**和现有的识别规则不同。不同点如下：

1.通常为第一节、第二节……，后面跟有相关的内容，将节的范围覆盖到50节。不再将“（二）”等视为二级标题。

2.第X节和后面的内容之间，需要空两格。此处可能需要判断后面有没有两个空格，有些是没有空格、有些是有一个空格，需要统一为两个空格。如：

第一节国家大力发展科学事业【标准化前】

**第一节 国家大力发展科学事业【标准版】**

五、第X条为三级标题

**字体：**黑体、三号、缩进2个汉字字符、居左对齐。

**覆盖范围：**第X条。此处和前述不一样，此处仅仅覆盖“第X条”几个字，不覆盖后续的内容。

**规则：**和现有的识别规则不同。不同点如下：

1.通常为第一条、第二条……，最高超过1260条，民法典1260条，将条目范围覆盖到第1500条。注：此处不再将“1.”“第一”“一是”等视为三级标题。

2.第X条和后面的内容之间，需要空两格。此处可能需要判断后面有没有两个空格，有些是没有空格、有些是有一个空格，需要统一为两个空格。如下：

第十二条国家大力发展科学事业。【标准化前】

第十二条 国家大力发展科学事业。【标准版】

六、正文及行间距

字体：仿宋GB2312、三号。

行间距：固定值28磅