

平成 18 年度 情報工学コース卒業研究報告要旨

外山 研究室	氏 名	村 瀬 裕 城
卒業研究題目	法令文書の XML 文書型定義と自動マークアップ	
<p>法令を解釈・運用する法律実務や、法令の起草・改廃などを行う法制執務においては、法令文書を参照することが必要である。従来、こうした作業は知識を持つ専門家が手作業によって行ってきた。しかし、現行法律だけで約 1,800 本あり、加えて毎年約 200 本もの法律が新たに施行されるため、専門家の負担は大きい。その負担を小さくするために、法令文書の電子化が進められているが、現時点では、単にテキスト化しているだけである。</p> <p>ところで、法令文書はもともと論理的な構造を持ち、法令の名称である「題名」、実質的な規定を集めた「本則」、付随的な規定を集めた「附則」といった部分に分けることができる。さらに、規定は「章」「節」「条」「項」などの単位でまとめられている。</p> <p>そこで、法令文書をテキスト化するだけでなく、論理的な構造を保って電子化することによって、より効率的な法令文書の参照・利用が可能となる。それにより、専門家の負担も軽減される。本研究では、法令文書の持つ構造を、XML 文書型定義 (DTD) を用いて形式化するとともに、その DTD に従った自動マークアップ・システムを作成する。</p> <p>DTD を作成するためには、構造要素や属性を定義するとともに、構造要素間の階層関係・兄弟関係を特定する必要がある。法令文書の専門書には構造要素やそれらの関係が記載されているが、それだけでは DTD を作成するには不十分である。そこで、本研究では、一部改正法に着目した。一部改正法は、法令を改正するときに行われる法令であり、改正対象のどの部分をどのように改正するかを記述した『改め文』からなる。改め文は一定の書式と表現で書かれており、改正対象となる規定を指定する際に使われる表現は、法令文書が持つ構造要素を表していると考えられる。例えば、「第二条第二項第六号を次のように改める。」という改め文から「条」「項」「号」の構造を獲得できる。この点に着目し、平成元年から平成十六年までの一部改正法に出現した全ての改め文計 76,159 文を利用し、正規表現を用いてその中から構造要素を抽出した。結果として 39 種類の構造要素を獲得し、そのうち 32 種類を DTD の構造要素として採用した。しかし、それだけでは法令文書を形式化するには不十分であったため、55 種類の構造要素を追加し、計 87 種類の構造要素を定義した。また、各構造要素間の親子関係については、改め文における構造要素の出現順に着目した。たとえば、先程の改め文からは、「条」-「項」-「号」の親子関係が特定できる。結果として 39 組の親子関係を獲得した。各要素間の兄弟関係の決定については専門書の情報を利用した。このようにして、法令文書用 DTD を作成した。</p> <p>また膨大な数の法令文書をマークアップするためには多大なコストがかかる。そのコストを軽減するために、自動マークアップ・システムを作成した。その際、本研究では、法令文書の書式と構造情報に着目した。法令文書の書式は、長年の慣習により決められている。本システムでは、行頭から何字目で書き出すかという配字と、章名・条名などの定型表現を利用した。また、構造情報については、本研究で作成した法令文書用 DTD によって各構造要素の階層関係を明らかにした。この書式と構造情報を BNF で記述し、Ruby で実装された yacc である racc を用いて、パーサとして自動マークアップ・システムを実装した。また、本システムに実際の法令文書 12 本を入力として与えたところ、いずれも正しくマークアップできた。</p>		