

# Notação Canônica para a Citação de Trechos de Normas Estatutárias Jurídicas Brasileiras

Peter de Padua Krauss<sup>1</sup>, João Alberto de Oliveira Lima<sup>2</sup>

<sup>1</sup> Escola Superior de Direito Constitucional (ESDC)  
R. dr. Gabriel dos Santos, 88 – 01231-010 – São Paulo – SP – Brasil

<sup>2</sup> Prodasen / SSDSL  
Av. N-2 Norte – Anexo C do Senado Federal – 70165-900 – Brasília – DF  
Krauss@INTOLABS.com.br, JoaoLima@SENADO.gov.br

**Resumo.** *O uso do padrão Xpointer na recuperação de trechos de textos normativos citados ou referenciados por outros textos, é de grande importância na geração de hipertextos e bases de conhecimento jurídico. Apesar da ausência de padrões, e da complexidade das DTDs que venham a suprir esta lacuna, demonstra-se que é possível fixar-se métodos simples para o redirecionamento de ponteiros. Para esse fim é proposto o conceito de “notação canônica da citação”, que consiste de um pequeno conjunto de regras notacionais e de controle terminológico, com o qual, tanto a construção dos Xpointers a partir das citações expressas em linguagem natural, quanto a conversão de uma DTD qualquer para a forma intermediadora, são facilmente obtidos.*

**Abstract.** *The use of the Xpointer standard for the retrieval of pieces of norm texts cited or referenced by other texts, is of great importance for the generation of hypertexts and juridical knowledge bases. In spite of the absence of standards, and of the complexity of the DTDs to overcome this gap, it can be demonstrated that it is possible to determine simple methods for the redirectioning of pointers. To this end, the concept of "Canonical Citation Form" is proposed, which consists of a small set of notational rules and terminological control, with which the construction of the Xpointers from the citations expressed in natural language, as well as the conversion of any DTD to the intermediary form, are easily obtained.*

## 1. Introdução

Normas Estatutárias Jurídicas Brasileiras (NEJBs) compõem um *corpus* lingüístico regular e com volume bastante significativo em meio digital. As citações (remissões) a estas normas vão desde a auto-citação (anáfora direta) no texto normativo, até citações contidas em textos externos ao *corpus* normativo, mas que prescindem de igual precisão — por exemplo citações às normas em contratos e em jurisprudência.

Normas estatutárias são documentos que estabelecem estatutos, ou seja, não versam sobre um caso específico, e estabelecem regras gerais sobre assuntos determinados. Normas jurídicas brasileiras, são aquelas que existem em decorrência da Constituição Federal da República do Brasil, para fixar obrigações em estatutos originados e mantidos por entidades (agências normativas), que representam os interesses da comunidade. Leis, Decretos, e Emendas Constitucionais são exemplos de NEJBs.

Apesar das tecnologias de marcação compatíveis com a noção de citação — *links*, relacionamentos, etc., doravante “citações eletrônicas” — serem bastante difundidas e consolidadas, seu uso para a citação eletrônica de NEJBs é pouco freqüente. Tanto as NEJBs (que possuem entre si uma densa rede de citações), como outros tipos de conteúdo digital contendo citações a NEJBs, apresentam grande número de aplicações com demanda reprimida para o uso de citações eletrônicas.

Uma NEJB pode ser integralmente citada, p.ex. “...em decorrência da Constituição Federal...”, ou ter seus trechos citados, “... conforme inciso X do artigo 5...”. As citações de trechos são de particular interesse devido à ausência de padrões uniformes para sua notação (ortografia da citação eletrônica).

Se o exemplo acima fosse expresso em HTML através de um link, “... conforme <a href=#5X>inciso X do artigo 5</a> ...”, para um trecho devidamente ancorado (“<a name=5X>”), a elaboração da citação

eletrônica (o *link*) não seria possível sem o conhecimento prévio da notação empregada para rotular o trecho desejado (foi “5X”, e não “5.x” ou “123” ou “art5\_inc10”). Se o exemplo fosse expresso com uma DTD específica, em XML, o padrão XPointer, conforme [Grosso 2003], não seria de utilidade sem o conhecimento prévio (portanto padronização) desta mesma DTD.

A proposta de uma “notação canônica”, que consiste de uma notação “conceitualmente padronizada”, é o ponto de partida para suprir a ausência referida, de padrões uniformes, e a ausência de ferramentas eficientes, e de baixo custo, para o reconhecimento e marcação automáticos dos textos, em linguagem natural, contendo citações a NEJBs.

Darão sustentação à proposta o trabalho, em andamento, de análise sistemática das NEJBs contidas na base do PRODASEN (Processamento de Dados do Senado Federal), e a disponibilização pública, aberta a contribuições, dos dados dos quadros 2 e 3 fixados abaixo — ampliados e revisados, bem como *scripts* demonstrando seu uso no tratamento e composição de Xpointers. Atualmente, o Senado Federal possui normas marcadas em conformidade com a DTD fixada em [Lima 2001].

Posteriormente o padrão decorrente da proposta será utilizado para redirecionar URLs persistentes (PURLs) fixados no domínio do senado para a recuperação de trechos citados. Por exemplo,

```
http://normas.senado.gov.br?doc=constituicao&trecho=a5.x
ou ...&trecho=xpointer(art[idcit='5']/inc[idcit='x'])
```

recuperariam o mesmo trecho de NEJB, no primeiro caso através de seu identificador, no segundo, através de uma expressão xpath; em ambos os casos fazendo uso de mapeamento pela notação canônica proposta.

## 2. Processos envolvidos

Para que haja compatibilidade (interoperabilidade) entre textos marcados contendo citações, e bases de dados contendo NEJBs, pode-se supor que ambos tenham sido elaborados respeitando-se alguns procedimentos, descritos a seguir.

### 2.1. Tradução da citação em LN para notação canônica

As principais operações envolvidas na tradução (“parsing”) da notação expressa em linguagem natural (LN), para uma notação expressa em linguagem controlada (no caso a “notação canônica” proposta) são:

*Normalização ortográfica*: simplificação ortográfica (sem acentos, distinção de maiúsculas, etc.) e a conversão de termos-chave em termos controlados simples e unívocos. *Eliminação e/ou substituição de conectivos*: preposições, artigos, etc. podem ser eliminados sem perda de informação, pois se tornam redundantes depois de isolado o núcleo da citação. Para efeito de normalização da separação entre elementos da citação, tais conectivos podem ser substituídos por um símbolo de separação, por exemplo, vírgula. *Inversão*: em caso de notação reversa, normalizar para notação direta, ou seja, na mesma seqüência (esquerda para direita) em que ocorrem, hierarquicamente, os elementos na norma. *Split*: em caso de composição — tipicamente identificada por plurais, mas pode depender de indícios gramaticais externos ao núcleo — a citação deve ser decomposta em núcleos isolados.

Uma vez reconhecido o corpo da citação e seu núcleo, procede-se ao “parsing”, automático ou assistido, da notação usual, que resultará na notação canônica (Qdr. 1).

O “parsing” realizado por expressões regulares foi proposto por [Dorante 1997], e mostrou-se bastante eficiente. Uma base de NEJBs interligadas por citações é disponibilizada em [DJI 2003], vide “grau.1”. Sistemas e princípios mais rígidos para a redação de citações são já bem difundidos nos Estados Unidos, veja-se [AALL 2002].

**Quadro 1: exemplos de uso das operações de “parsing”.**

Núcleo da citação	Traduções sucessivas, até chegar na notação canônica
Cap. III do Título V	1. Normalização ortográfica: “cap iii do tit v” 2. Eliminação e/ou substituição de conectivos: “cap iii, tit v”

	3. Inversão: “tit v, cap iii”
Art. 5, Alínea b, Inciso XXII	1. Normalização ortográfica: “art 5, alin b, inc xxii”
Artigos 3, 4 e 9	1. Normalização ortográfica: “art_plural 3, 4 e 9” 2. Eliminação e/ou substituição de conectivos: “art_plural 3, 4, 9” 3. Split: “art 3”, “art 4”, “art 9”

**Quadro 2: Normalização ortográfica dos rótulos de elementos citáveis.**

Termos (opcionais)	Expressão regular (perl)	Forma convertida
Alínea, Alin., al., al, ...	al(\. (f in)(\.(nea)))?	Alin
Artigo, Art., art, ...	art(\.(igo))?	Art
Capítulo, Cap., cap, ...	cap(\.(f itulo))?	Cap
Parágrafo, par., par, ...	par(\.(a(\.(grafo)))?)	Par
Título, Tít., tit, ...	tit(\.(ulo))?	Tit
...	...	...

**Quadro 3: tipos de rótulo e formas de contagem.**

Elemento canônico	Tipo no idcit fixado	Início da contagem no contador	Exceção ao incremento +1 normalizada como
alin	letra	elemento-pai	letra “-” cardinal
art	cardinal	norma	cardinal “-” letra
cap	romano	norma ou elemento-pai	
par	cardinal	elemento-pai	cardinal “-” letra
tit	romano	norma ou elemento-pai	
...	...	...	...

NOTA-1: as *tags* foram eleitas tendo em vista apenas a simplicidade e apelo mnemônico.

NOTA-2: o atributo idcit (id de citação) é designação dada na notação canônica para o rótulo/contador do elemento, e seu valor, quando necessário, é obtido através de uma função de conversão (no caso de “par”, p. ex., converte-se “Parágrafo Único” para “1” ordinal, 2º para 2, etc.).

## 2.2. Sumarização para a notação canônica

Uma operação bastante comum efetuada sobre o XML de NEJBs é a listagem dos conteúdos, ou seja, a produção de um “sumário da norma”. Se um sumário foi produzido, digamos em XHTML, então significa que, independente da DTD em que foi expressa a NEJB, existe uma transformação XSLT para esse fim.

A abstração complementar do conceito de “notação canônica” é o sumário. Imaginemos que seja necessário padronizar um XML intermediário para a produção de sumários. A sua DTD seria a mais simples possível, e empregaria “elementos canônicos” do Quadro 3. Assim, seja qual for a DTD empregada para estruturar a norma, se existir uma transformação XSLT através da qual possamos produzir sumários, então podemos garantir que essa mesma norma pode ser citada através da notação canônica. Veja o exemplo no quadro 4 abaixo.

**Quadro 4.1: exemplo de XML de uma NEJB.**

```
<norma tipo="lei" rotulo="166/2003"><ementa id="e1"> ... </ementa>
```

```

<capitulos>
  <capitulo id="e2" rstd="i"> <titulo>Capitulo I - ...</titulo>
    <artigo id="e3" rstd="1">Art.1 - ....</artigo>
    <artigo id="e4" rstd="2">Art.2 - ....</artigo>
    <artigo id="e5" rstd="2-a">Art.2A - ....</artigo>
    ...
  <capitulo id="e11" rstd="II"> <titulo>Capitulo II - ...</titulo>
    <artigo id="e12" rstd="8">Art.8 - ....</artigo>
    ...
</capitulos> ... </norma>

```

**Quadro 4.2: aspecto da transformação XSLT que traduz o XML do quadro 4.1 para o XML do quadro 4.3.**

```

<xsl:template match="capitulo">
  <cap idcit="{@rstd}" idref="{@id}"><xsl:apply-templates
select="secao|artigo"></cap>
</xsl:template>
<xsl:template match="artigo">
  <art idcit="{@rstd}" idref="{@id}"><xsl:apply-templates
select="paragr|inciso"></art>
</xsl:template>

```

NOTA: nesta etapa também é conveniente (e mais confiável) fazer a conversão do rótulo em idcit, acima exemplificada sem a função de conversão por simplicidade.

**Quadro 4.3: aspecto do XML do sumário da norma do quadro 4.1.**

```

<emen idcit="" idref="e1"/>
<cap idcit="i" idref="e2">
  <art idcit="1" idref="e3"/> <art idcit="2" idref="e4"/> <art idcit="2-a"
idref="e5"/>
  ...
</cap>
<cap idcit="ii" idref="e11"><art idcit="8" idref="e12/><art .../>...</cap> ...

```

Tendo em mente tais considerações, adotamos as seguintes definições:

**DTD do sumário da norma citável:** DTD constituída de elementos (canônicos) fixados em conformidade com Quadro 3, com atributos idcit.

**Norma Citável:** uma norma formatada em XML com uma DTD tal que apresente uma transformação XSLT, que permita a tradução da mesma para uma DTD do sumário da norma citável.

### 2.3. Recuperação do trecho de norma

A recuperação do trecho de norma a partir da citação canônica, é realizada conforme exemplos do Quadro 5. Em [PRODASEN 2003] tem-se um importante exemplo de DTD de norma citável já em uso, prevendo, em projeto, a resolução Xpointer.

**Quadro 5: Estratégia para recuperação da informação.**

Citação canônica	Elemento terminal	ID do terminal	XPath para resgatar idref do sumário
tit v, cap 3	cap	e1001	/tit[position()=5]/cap[icdit='3']
art 123	art	e1002	//art[icdit='123']
art 5, inc xx, alin b	alin	e1003	//art[icdit='5']/inc[icdit='xx']/alin[icdit='b']

art 123, par 1	par	e1004	//art[idcit='123']/par[idcit='1']
<p>NOTA: O uso da função XPath <code>position</code> (bem como <code>element</code> no padrão XPointer) apesar de poder ser empregada com segurança na maior parte dos elementos de agrupamento (Títulos, Capítulos, etc.), não pode ser empregada com mesma segurança em artigos e sua hierarquia inferior (incisos, parágrafos, alíneas). As normas de redação (vide [Leis Redação]) também sugerem que exceções à rotulação incremental são válidas para <code>idcit</code>, portanto invalidando o uso de <code>position</code>.</p>			

### 3. Conclusões

Os principais obstáculos ao uso do Xpointer (ou tecnologias similares para o relacionamento entre citação e trecho citado), são a ausência de DTDs (ou XML-Schemas) padronizadas, o que impossibilita a redação do Xpointer em ambientes heterogêneos (ou onde não seja dado conhecer a DTD previamente), e a complexidade das DTDs que venham a suprir esta lacuna, que tornariam a sintaxe do Xpointer igualmente complexa, e pouco atraente.

Demonstrou-se que é possível transpor o problema da ausência de padrões, fixando-se apenas a restrição de “norma canonicamente citável” (para as DTDs portadoras de trechos a serem citados pelo Xpointer).

Evidenciou-se, através da exposição dos processos de tradução (sec.2), que o “custo de adoção do padrão” é bastante baixo. O “padrão proposto”, em si, faz uso indireto de “padrões de fato”: as formas de redação de uma citação (em LN), consagradas pelo uso no meio jurídico, e as [Leis de Redação], que aos poucos têm se imposto nas agências normativas. O custo de “sistematização e manutenção” do padrão proposto corresponde ao custo de revisão sistemática dos quadros 2 e 3, e publicação dos mesmos.

A notação proposta permite que exista a interoperabilidade de citação entre normas que utilizaram diferentes estruturas, ou até mesmo, entre documentos externos (contratos, jurisprudência) e os dispositivos da norma. Uma extensão do conceito, a “notação canônica compacta”, pode ser empregada na padronização de notações para identificadores; já em uso na ESDC para geração de âncoras XHTML uniformizadas.

### 4. Referências

- AALL (2002) “AALL Universal Citation Guide, Version 2.1”, American Association of Law Librarians.  
<http://www.aallnet.org/committee/citation/ucg/index.html>.
- Bray, T. et all, eds. (2000) “Extensible Markup Language (XML) 1.0 (Second Edition)”, World Wide Web Consortium. <http://www.w3.org/TR/REC-xml>
- Clark, J., ed. (1999) “XSL Transformations (XSLT), Version 1.0”, World Wide Web Consortium.  
<http://www.w3c.org/TR/1999/REC-xslt-19991116>.
- DJI (2003) portal “Índice Fundamental do Direito”, [http://www.dji.com.br/diversos/indexp\\_2-legis.htm](http://www.dji.com.br/diversos/indexp_2-legis.htm)
- Dorante, A. (1997) “Investigação de processo de conversão automática de textos estruturados para hiperdocumentos”, dissertação apresentada ao DFI-IFSC-USP.
- Grosso, P. et all, eds. (2003) “XPointer Framework”, W3C Recommendation 25 March 2003.  
<http://www.w3.org/TR/xptr-framework/>
- Leis Redação: Lei Complementar nº 95, de 26 de fevereiro de 1998; Decreto Nº 2.954, de 29 de janeiro de 1999; Lei Complementar nº 107, de 26 de abril de 2001; Decreto nº 4.176, de 28 de março de 2002.
- Lima, João A. O. e Souza, Marco A. M. (2001) “Legislação Brasileira na Internet”, Senado Federal - Prodasen/DSL. [www.interlegis.gov.br/produtos\\_servicos/20020109121259/20020109130716/view](http://www.interlegis.gov.br/produtos_servicos/20020109121259/20020109130716/view)
- PRODASEN (2003), portal da Base de Dados da Constituição da República Federativa do Brasil,  
<http://www.senado.gov.br/bdtextual/const88/const88.htm>
- Thompson, H. et al., eds. (2001) “XML Schema Part 1”, World Wide Web Consortium.  
<http://www.w3.org/TR/xmlschema-1/>