

Projeto LexML Brasil – Resolução de Identificadores

URN Resolver

Versão 0.6 (maio/2008)

João Alberto de Oliveira Lima (Senado/PRODASEN – joaolima@senado.gov.br)

Fernando Ciciliati (Senado/Interlegis – fernandociciliati@interlegis.gov.br)

Projeto LexML Brasil – Resolução de Identificadores	1
1. Introdução	2
2. Requisitos Gerais	2
3. Princípios do serviço de resolução	2
3.1. Arquitetura geral do sistema	2
3.2. O catálogo para a resolução	2
3.3. Fases do processo de resolução.....	3
3.4. Tipologia dos resolvedores	3
3.5. Comportamento do resolvedor.....	4
3.6. Auxílio à redação ou à pesquisa e navegação.....	5
4. Referências	6
4.1. Bibliográficas	6
4.2. Sites.....	6

1. Introdução

2. Requisitos Gerais

As implementações de programas deverá atender os seguintes requisitos:

- uso de software livre em todas as camadas;
- interfaces atendendo aos padrões de

3. Princípios do serviço de resolução

3.1. Arquitetura geral do sistema

O serviço de resolução tem a função de associar a um nome uniforme (URN) as respectivas localizações (URL) na rede (internet).

O sistema global se baseia em uma arquitetura distribuída, sobre dois componentes fundamentais: uma cadeia de informações no DNS (Domain Name System) e um conjunto de serviços de resolução de URN para URL, cada um deles com competência sobre um domínio específico do espaço de nomes.

Por meio dos registros NAPTR do DNS (descritos no RCF 2168 [3]), o client identifica as características (protocolo, porta, site) do serviço em condições de associar à URN em questão as correspondentes URLs, e de permitir assim o acesso ao documento.

O serviço de resolução se compõe de dois elementos: uma base de conhecimentos (constituída por um catálogo ou um conjunto de regras de transformação) e um software de consulta a essa base.

3.2. O catálogo para a resolução

A implementação de um catálogo, construído sobre uma base de dados relacional, oferece uma maior flexibilidade para a resolução de nomes uniformes incompletos ou imprecisos, evento provável no campo jurídico, e ainda mais freqüente se o nome uniforme do documento referenciado for construído automaticamente (por meio de um analisador léxico/sintático ou parser) a partir das referências em linguagem natural (português).

Cada autoridade emitente será responsável pela implementação do seu respectivo catálogo, ou pelo menos pela parte que trata seus próprios documentos, atribuindo a cada recurso o nome uniforme adequado. Em especial, providenciará a completa identificação de cada ato, atribuindo-lhes eventualmente uma numeração interna convencionada.

Além das localizações na rede (URL), o catálogo poderia conter uma série de outros dados de grande interesse para o usuário, dentre os quais várias meta-informações (título, dados de publicações, matéria tratada, data de vigência, etc.) e

ainda os relacionamentos ativos e passivos intercorrentes com outros documentos (modificações, delegações, abrogações, conversões, etc.).

No catálogo, os nomes alternativos (alias) de um documento são estreitamente ligados entre si, têm em comum as outras informações (título, URL, relacionamentos, etc.) e, no processo de resolução, são tratados exatamente da mesma forma.

Também os identificadores das diversas versões de um ato são estreitamente vinculados ao nome do ato original, compartilham com este uma série de informações (ex.: título) mas possuem também as suas próprias (ex.: início e fim de vigência).

Esse vínculo estreito vale ainda para os identificadores dos diversos comunicados relativos ao mesmo ato, mesmo tendo um conjunto distinto de informações compartilhadas e próprias (ex.: codificação de publicação).

3.3. Fases do processo de resolução

Em geral o processo de resolução deveria operar-se por meio das seguintes fases:

- normalização do nome uniforme a ser resolvido, consistindo na transformação para a forma canônica de alguns componentes. Esta fase, particularmente útil no caso de construção automática do nome, se ocupa, por exemplo, de expandir as siglas e as abreviações, de uniformizar as nomenclaturas das instituições, de padronizar os tipos de documentos, etc.;
- interrogação ao catálogo com o nome uniforme normalizado e extração das informações relacionadas a este, que, além da URL, poderiam ser o título, a matéria, os relacionamentos, etc.

3.4. Tipologia dos resolvedores

Com a transmissão ao processo de resolução de um nome estendido, pode-se imaginar, pelo menos num período transitório, dois tipos de resolvedores, com base na capacidade de interpretação do espaço de nomes. Estes dois tipos de resolvedores devem conviver, isto é, devem ser capazes de tratar, se bem que com níveis de precisão distintos, os nomes pertencentes a ambos os espaços. Se pode assim distinguir:

Resolvedor Base

É o processo capaz de resolver apenas os nomes uniformes pertencentes ao espaço base. O seu comportamento, no caso dos nomes estendidos, é sinteticamente este:

- ignora a parte que descreve a propriedade do documento;
- extrai o nome base e o resolve (encontra as URLs associadas);
- adiciona às URLs o identificador de partição como fragmento (isto é, converte o "\$" em "#") para permitir os links internos;
- devolve ao solicitante as URLs assim completadas.

Resolver Estendido

É o processo capaz de resolver também os nomes uniformes pertencentes ao espaço estendido. Esse é portanto capaz de:

- devolver não apenas as URLs dos documentos inteiros mas também de partições específicas (por ex.: por meio de uma consulta em banco de dados);
- tratar os metadados associados aos documentos e devolver como resposta as URLs dos atos de acordo com as propriedades solicitadas.

A convivência dos dois tipos distintos de resolver apresenta ao usuário o sério problema de ter que conhecer as características do resolver utilizado para uma dada referência. A transparência necessária se obtém agindo simultaneamente sobre duas frentes:

- cada resolver deve indicar claramente qual parte do nome uniforme é capaz de resolver;
- cada documento deve reportar claramente as suas propriedades (ex.: texto original, texto vigente da data 'x' à data 'y', etc.) e permitir a navegação através das varias versões existentes (seja avante ou para trás). De tal modo, o usuário poderá de alguma forma encontrar, de forma manual, e não automática, a versão desejada.

3.5. Comportamento do resolver

Idealmente, o resolver deveria buscar no catálogo o nome uniforme com correspondência completa àquele informado.

Considerando a possibilidade de ocorrência de nomes imprecisos e/ou incompletos provenientes das referências, o resolver deveria, no caso de falta de correspondência, buscar no catálogo todos os nomes que "contenham" aquele informado, em outras palavras, todas as URNs das quais o nome informado constitua, elemento por elemento, uma substring.

Este comportamento é extremamente útil para o sucesso na busca dos documentos objeto de remissões, posto que, freqüentemente, a citação não contém todos os elementos para construir o nome uniforme canônico: isso se dá, por exemplo, com os atos citados apenas com número e ano; com atos numerados ou não, citados apenas pela data, etc. Tal funcionamento permite a simplificação e a automatização (para um programa ou para a redação) da construção do nome uniforme a partir da remissão. Obviamente, a redação deveria resolver os casos de ambiguidade na referência (mais de um documento associado).

De fato, os elementos do nome são considerados como se fossem compostos por diversos campos elementares, com valores posicionais internos (ex.: autoridades, datas) ou sem (ex.: números identificadores do ato); o confronto deveria assim se dar campo a campo, e resultar como positivo se todos os campos fornecidos encontrarem correspondência em um ou mais nomes uniformes do catálogo. Por ex.: o registro do

catálogo com data do ato "2000-06-13" deveria resultar em confronto (posicional) positivo mesmo com as datas informadas "2000" e "2000-06"; o registro com identificador do ato "aipa-cr-35" deveria dar resultados positivos mesmo com os números "35", "cr-35" ou "aipa-35" (não posicionais).

3.6. Auxílio à redação ou à pesquisa e navegação

Sobre o comportamento standard do resolvedor podem ser construídos software de auxílio à pesquisa e navegação ou à redação de novos documentos.

Tais ferramentas poderiam, idealmente, buscar fornecer de algum modo o acesso aos recursos já existentes (destacando as simplificações executadas), de acordo com regras de comportamento determinadas em função da operação desenvolvida.

A título de exemplo, o software de apoio poderia, no caso de uma pesquisa com resultado negativo, eliminar do nome fornecido partes de maior detalhamento (ex.: especificação do ato, articulações do órgão, etc.) e repetir a busca de nomes um pouco mais genéricos, até obter um resultado positivo ou atingir o limite de significatividade ligado ao domínio em questão.

Será, depois, tarefa do usuário identificar, dentre as respostas fornecidas, aquela correspondente ao caso de interesse.

4. Referências

4.1. Bibliográficas

- [3] M. Mealling, R. Daniel, "The Naming Authority Pointer (NAPTR) DNS Resource Record", RFC 2915, September 2000.
- [4] R. Daniel, "A Trivial Convention for using HTTP in URN", RFC 2169, June 1997.
- [5] P.L. Spinosa, "Identificazione delle risorse tramite URNs (Uniform Resource Names) e gestione dei riferimenti normativos", [Cap. 5 Studio di fattibilità per la realizzazione do projeto "Accesso alle norme in rete"], in: Informatica e Diritto, 2000, 1, pp. 107-131

4.2. Sites

<http://www.nir.it/>

Projeto Norme in Rete (Itália)

<http://www.akomantoso.org/>

Projeto Akoma Ntoso (Nações Unidas/África)