

**O THESAURUS EXPERIMENTAL DA CONSTITUIÇÃO, PROJETO PILOTO  
DESENVOLVIDO PELA SUBSECRETÁRIA DE ANÁLISE DO SENADO FEDERAL**

**Maria Riza Baptista Dutra\***  
**Octaciano Nogueira\***  
**Yamil de Souza Dutra\***

O montante e a diversidade das normas legais a serem implantadas no Sistema de Recuperação de Dados desenvolvido pelo PRODASEN e pela Subsecretaria de Análise do Senado Federal indicaram a necessidade de utilizar-se, para análise e posterior recuperação dessa legislação, um instrumento de controle de linguagem. Optou-se pela criação de um “THESAURUS”. A inexistência de “THESAURI” específicos para este campo e de pessoal treinado em sua estruturação determinou a execução de um Projeto Piloto, utilizando-se a Emenda Constitucional I de 1969, que abrange um universo amplo em informações, e possui uma estrutura interna definida. Assim, os processos, rotinas e documentação exigidos pelo trabalho resultaram em treinamento de pessoal, criação de um documento capaz de ser utilizado como paradigma, e, especialmente, adaptação do serviço ao “software” disponível.

*A palavra parece ser, cada vez mais, um instrumento muito ambíguo para que possa constituir um meio adequado de pesquisa. (Barão Édouard de Houtart. Relatório sobre o Serviço de Processamento de Dados do Senado – PRODASEN, janeiro de 1973, p.5)*

\* Subsecretária de Análise do Senado Federal.

## 1 – A ESTRUTURA ORGANIZACIONAL

### 1.1 – Implantação do PRODASEN e Criação da Secretaria de Informação

Em novembro de 1972, como resultado da Reforma Administrativa por que passou a Secretaria do Senado Federal, e também como consequência da instalação de um Centro de Processamento de Dados, criou-se, na estrutura organizacional desta casa do Congresso, a Secretaria de Informação, destinada a operar o sistema de informações a ser desenvolvido com base no tratamento automatizado de dados de natureza bibliográfica e jurídica.

A Secretaria de Informação, que é responsável pela manutenção, atualização e desenvolvimento de arquivos do Serviço de Processamento de Dados do Senado (PRODASEN), compõe-se dos seguintes órgãos:

#### SUBSECRETARIA DE ANÁLISE

Seção de Referência Legislativa

Seção de Pesquisa

Seção de Administração

#### SUBSECRETARIA DA BIBLIOTECA

Seção Administrativa

Seção de Processos Técnicos

Seção de Periódicos

Seção de Referência Bibliográfica

#### SERVIÇO DE CONTROLE DE INFORMAÇÕES

#### 1.1.1 – Subsecretaria de Análise

À Subsecretária de Análise compete, sumariamente, manter e desenvolver os arquivos; relacionados com a informação jurídica, estruturando métodos sistemáticos para a recuperação -de informações no campo da **jurisprudência** e da **legislação** (referência legislativa), além de outros correlates, como os pronunciamentos dos Senadores e autoridades selecionadas.

#### 1.1.2 – Subsecretaria da Biblioteca

A Subsecretária da Biblioteca mantém e desenvolve os arquivos automatizados de informação bibliográfica, em seu sentido mais amplo, abrangendo livros e publicações periódicas, entre os quais os de doutrina, no campo do Direito.

De tal sorte que a articulação dos dois sistemas permite a recuperação de informações jurídicas - nos campos da **doutrina, da jurisprudência, e da legislação**, podendo vir a se constituir num amplo repositório no campo do Direito.

### **1.1.3 – Serviço de Controle de Informações**

O Serviço de Controle de Informações, se incumbiu de operar os terminais do sistema “on line”, atendendo os usuários que não possuem seus próprios Terminais.

### **1.2 – Acervo Material Existente no Computador**

Ao ser instituída a Secretaria de Informação estavam arquivados no computador do PRODASEN os dados do fichário manual que constituía o acervo do Serviço de Referência Legislativa da Biblioteca do Senado e que passaram a compor aproximadamente 68.000 documentos. Tais dados abrangiam, genericamente, as ementas de normas legais até o nível de Decreto e a respectiva indexação feita ao longo de vários anos, com maior ou menor grau de precisão na análise dos respectivos textos. A transcrição desses dados foi feita sem qualquer responsabilidade da Secretaria de Informação ou de sua Subsecretaria de Análise, daí resultando, entre outros inconvenientes:

- Elevada porcentagem de erros decorrentes de enganos na transcrição ou na correção dos-textos;
- Falta de padronização da indexação utilizada, tendo em vista a sua elaboração, ao longo de muitos anos, por diferentes analistas que utilizaram critérios variáveis;
- Não aproveitamento das potencialidades do sistema a ser utilizado, já que a indexação do arquivo manual foi feita sem a preocupação de seu futuro aproveitamento em processo automatizado.

Além desse acervo documental, foram transcritos, igualmente, textos completos de algumas leis, como por exemplo a Emenda Constitucional nº 1/69, os Regimentos do Senado, da Câmara, do Congresso e a Consolidação das Leis do Trabalho.

### **1.3 – Trabalho Inicial da Subsecretaria de Análise**

Dispondo desse acervo, a rigor não inteiramente sistematizado, coube à Subsecretaria de Análise desenvolver um trabalho inicial, a operacionalizar o sistema dentro de critérios que lhe permitissem estabelecer padrões de indexação para uso por processo automatizado; de controle da linguagem usada na indexação; e de transcrição integral ou resumida dos textos legais arquivados.

Paralelamente foram realizadas pesquisas quanto às peculiaridades do programa standard adquirido pelo PRODASEN, o sistema “STAIRS”, ao qual tinha que ser adaptado o sistema de informações jurídicas a ser desenvolvido pela Subse-

cretaria de Análise. Nesse período a Subsecretária de Análise recebeu assistência esporádica de especialistas de outros países e de técnicos nacionais.

Nos anos de 1973 e 1974 foi realizada a revisão dos arquivos existentes, no sentido de padronizar a linguagem utilizada na indexação; definiu-se a estrutura do THESAURUS a ser montada em caráter experimental e adaptou-se o método de indexação ao programa “STAIRS”. Com o objetivo de manter atualizados os arquivos de referência legislativa, adotou-se uma norma experimental, ainda em uso, para a indexação dos documentos transcritos integral ou resumidamente.

O arquivo inicial de referência legislativa — LEGI (legislação) — foi mantido em sua forma original, estruturando-se um novo data base que evoluiria progressivamente até à extinção do anterior. Este novo arquivo, denominado NJUR (Normas Jurídicas), pretende abranger todas as normas legais, na medida em- que sua indexação for sendo padronizada.

Montou-se um THESAURUS experimental relativo a um texto legal completo, a partir do qual pretende-se criar um instrumento de acesso ao Sistema de Informações Jurídicas, servindo, ao mesmo tempo, de meio para desenvolver o “know-how” necessário à plena utilização das potencialidades do programa “STAIRS”.

## **2 – SISTEMA “STAIRS” COMO INSTRUMENTO**

### **2.1 – Funcionamento do Programa**

Como assinalou o Barão Edouard de Houtart, em seu Relatório datado de janeiro de 1973, sobre o sistema implantado pelo Senado, o STAIRS é “*um espetacular programa standard de documentação*”, apresentando, no entanto, algumas deficiências no tratamento da documentação jurídica. Um dos exemplos citados neste aspecto é a impossibilidade de fazer evoluir os fichários documentários paralelamente à evolução do THESAURUS. Entretanto, o PRODASEN está desenvolvendo um “software” que permitirá a superação deste problema.

Para que se possa melhor compreender o interrelacionamento do sistema STAIRS com a forma de indexação desenvolvida, visando a aproveitar suas potencialidades, é imprescindível descrever, ainda que sumariamente, a apresentação e a operação do mesmo.

#### **2.1.1 – Sistema Interrogável, “On Line”**

O equipamento de que dispõe o PRODASEN é constituído, basicamente, de um computador de grande capacidade, modelo IBM 370-158, de 1024K, a disco

e fita, dispondo de terminais de tela e de impressão. É um sistema “on line”, segundo o qual qualquer usuário autorizado pode interrogar diretamente o centro, valendo-se de operadores específicos, e desde que conheça os arquivos abertos a consulta. No que diz respeito à legislação, que é o material de interesse da Subsecretária de Análise, a informação pode ser obtida diretamente de um **texto integral**, de um **sumário** ou de uma **ementa**, recebendo, cada um, um tratamento específico, tanto no que respeita à indexação quanto à busca de informação.

### 2.1.2 – O Sistema de Informações Jurídicas

O Sistema de Informações Jurídicas é uma fase recente da concepção do sistema. Pretende incluir informações no campo da doutrina, da jurisprudência e da legislação, no sentido de tornar mais abrangente possível o acervo da documentação jurídica oferecido ao usuário. Compõe-se, atualmente, de dez arquivos distintos, com a seguinte caracterização:

SIGLA	DENOMINAÇÃO	CONTEÚDO	RESPONSÁVEL
BIBL PERI	Livros Periódicos	Doutrina	Subsecretária da Biblioteca
DISC JURI NJUR THES	Discursos Jurisprudência Normas Jurídicas Thesaurus	Jurisprudência Legislação Dicionário	Subsecretária de Análise
MATE 1 COMS	Matéria Comissões	Proposições do Senado Federal, Comissões Permanentes do Senado Federal e do Congresso Nacional	Sinopse do Senado Federal
MATE 2 REMI	Matéria Matéria	Proposições da Câmara dos Deputados, Proposições da Câmara dos Deputados de 1967 a 1971	Sinopse da Câmara dos Deputados

Este quadro pretende deixar claro que o THESAURUS é um instrumento auxiliar do sistema, na medida em que pode-se ter acesso às informações contidas nos diferentes arquivos, **sem o auxílio** do THESAURUS. Em outras palavras, as in-

formações disponíveis nos diversos “data base” que compõem o sistema podem ser obtidas recorrendo-se diretamente a cada um deles. Quando houver o THESAURUS definitivo, a ele se recorrerá sempre que o acesso direto se revelar ineficiente.

Nestas condições, o THESAURUS experimental da Constituição serve, **apenas**, como instrumento auxiliar de um texto entre os milhares de documentos que compõem o sistema.

Futuramente, todos os indexadores utilizados no sistema estarão automaticamente incluídos no THESAURUS sob a forma de descritores autorizados (parágrafo use), consignando-se igualmente os não autorizados (parágrafo não use), que, no entanto, farão remissão direta, aos descritores autorizados do sistema. De tal sorte que a consulta ao THESAURUS indicará todas as formas alternativas de indexação dos documentos, dispensando-se os operadores especialistas, já que o objetivo do sistema é ser “aberto”, permitindo a consulta a todo e qualquer usuário leigo.

### 2.1.3 – As Funções do Programa “STAIRS”

A operação do sistema é feita através de uma série de funções específicas, cujo uso permite que se obtenham diferentes tipos de informações entre as que se encontram arquivadas. Assim, por exemplo, a função SELECT permite **selecionar** ou **ordenar** as normas jurídicas ou a jurisprudência, segundo critérios que podem ser escolhidos pelo usuário. É possível, por exemplo, selecionar toda a legislação de um período dado; pode-se, com a mesma função, selecionar todas as decisões judiciais de um determinado relator, e assim sucessivamente.

No da pesquisa por assuntos, utiliza-se a função SEARCH, cujos **operadores** permitem maior ou menor grau de especificidade na recuperação da informação, podendo-se partir de um grau de alta generalidade, até um grau muito preciso de especificidade. São esses operadores que virtualmente orientam o sistema de indexação em uso, razão por que fazemos do mesmo mais detalhada descrição.

Esta função “SEARCH” possui quatro operadores distintos: **ADJ, WITH, SAME e AND**. Utilizando-se o primeiro obtém-se a coordenação de dois descritores **na mesma ordem** que a estabelecida na pergunta; obtém, com o uso do segundo, a coordenação de dois descritores pertencentes à mesma sentença; usando-se o terceiro obtém-se a coordenação de dois descritores pertencentes ao **mesmo parágrafo**, e utilizando-se o último, de dois descritores pertencentes ao **mesmo documento**. De acordo com o seguinte esquema sumário, em que A e B representam dois indexadores:

**A ADJ B** – A e B de forma adjacente e nesta mesma ordem;

**A WITH B** – A e B pertencentes e contidos na mesma sentença;

**A SAME B** – A e B pertencentes e contidos no mesmo parágrafo;  
**A AND B** – A e B pertencentes e contidos no mesmo documento.

É o uso dos diferentes operadores acima indicados que permite ao mesmo tempo (a) evitar falsas coordenações, utilizando-se a forma direta na pergunta, que deve corresponder à forma direta da indexação; (b) dar melhor especificidade à resposta, utilizando-se um maior número de indexadores na pergunta.

A existência desses operadores, porém, exige uma forma específica de indexação que leve em conta as diferenças de cada um.

#### **2.1.4 – Apresentação do Arquivo NJUR**

O arquivo NJUR (Normas Jurídicas), ao qual deve referir-se oportunamente o THESAURUS, é apresentado em dois campos distintos;

**CAMPO FORMATADO**, que contém informações padronizadas em espaços pré-determinados;

**CAMPO LIVRE**, que contém informações não padronizadas em espaços limitados.

Tanto os campos formatados quanto os não-formatados são divididos em parágrafos numéricos. Embora o acesso ao usuário seja livre em qualquer deles, os campos formatados permitem o emprego de um maior número de **funções** que os campos não-formatados.

No caso do arquivo NJUR, são os seguintes os parágrafos usados, segundo os campos e as informações contidas:

CAMPO	PARÁGRAFO	CONTEÚDO
F O R M A T A D O	001	Tipo de norma legal Número Letra Data Artigo Parágrafo Item Alínea Inciso Órgão de Publicação Data de Publicação Página de Publicação Coluna de Publicação
L I V R E	200	Transcrição - Texto integral Transcrição - Sumário Transcrição - Ementa
	250	Parágrafo nominal (transcrição de nomes próprios reveridos na norma legal)
	600	Parágrafo de Vide. (Trata-se de legislação modificadora).
	800	Indexação (através de sentenças lógicas em ordem direta).
	899	Catálogo (assuntos principais para elaboração de catálogo sistemático).

Verifica-se, desse quadro, que quanto maior for o campo de pesquisa utilizado, maior o risco de se obterem informações irrelevantes, pela superveniência de falsas coordenações. Em outras palavras, o recall será tanto maior quanto mais amplo for o campo de pesquisas; ao contrário, será maior a precisão quanto mais específico for o campo da pesquisa.

Nestas condições, utilizando-se as possibilidades das diferentes funções e dos diferentes operadores, chegou-se a um processo de indexação que leva em consideração exatamente essas peculiaridades.

Como se assinala no quadro acima, a indexação é feita através de sentenças lógicas, colocadas na ordem direta. Os indexadores de uma mesma sentença são separados entre si por vírgulas; e as sentenças, por sua vez, separam-se por pontos. Cada sentença constitui um desdobramento dentro do mesmo parágrafo (800), discriminada por numeração sucessiva, de 800 a 899, comportando, portanto, 99 sentenças. O exemplo abaixo pode dar uma idéia mais precisa da forma utilizada:

PARÁGRAFO	LINHA	INDEXAÇÃO	SENTENÇA
800	01	Fixação, limite, permanência, Oficial General, Exército, Serviço Ativo.	1.a
800	02		
800	03		
801	01	Requisitos, inclusão, oficial superior, Exército, Quadro de Acesso, promoção.	2.a
801	02		
801	03		

De tal sorte que, conforme se discrimina no item 2.1.3, o uso dos operadores ali indicados permite coordenar desde as informações contidas numa mesma sentença da indexação, até informações contidas num dado parágrafo ou num determinado documento, o que dá idéia da amplitude e ao mesmo tempo do grau de especificidade da resposta que se poderá obter, na medida em que, ao mesmo tempo: a) se limita o campo da pesquisa c b) se aumenta a especificidade da pergunta.

É a combinação dos dois fatores simultaneamente que permite que se obtenha um alto **recall** com índice razoável de **precisão**.

### 2.1.5 – Apresentação do THESAURUS

Da mesma forma que os diferentes arquivos – um dos quais sumariamente descritos acima – o THESAURUS usado pelo Programa STAIRS, e desenvolvido pelo Senado, possui características próprias, de apresentação. Ele se divide em parágrafos numéricos, com denominações e conteúdos específicos. Suas informações, no entanto, não estão contidas em campos formatados, mas sim em campos livres, podendo o usuário ter acesso a qualquer delas. A forma sumária de sua apresentação pode ser assim descrita:

PARÁGRAFO	DENOMINAÇÃO	CONTEÚDO
001	USE	Descritor autorizado.
100	NOTA EXPLICATIVA	Sentido restritivo que o descritor possui na indexação.
200	NÃO USE	Sinônimo de USE. Descritor não autorizado.
600	VER TAMBÉM	Descritor autorizado, de sentido correlato na indexação.
700	COMBINE	Descritor autorizado, que, associado ao principal (use), forma um novo conceito, em geral mais específico.
300	SUPERIOR	Descritor de maior abrangência semântica, e, conseqüentemente, mais genérico.
500	INFERIOR	Descritor de menor abrangência semântica, e, conseqüentemente, mais específico.

Isto significa que, mesmo na hipótese do usuário recorrer a um termo não autorizado, encontrará o termo autorizado correspondente que deverá ser usado. O THESAURUS lhe indicará, ao mesmo tempo, o termo mais genérico imediatamente superior e o termo mais específico imediatamente inferior. De tal sorte que poderá, sucessivamente, ir subindo ou descendo na escala de indexação, em busca de maior recall ou em busca de maior precisão. Paralelamente, poderá alargar sua pesquisa em sentido horizontal, abrangendo a cadeia de correlação existente entre os indexadores. Se combinar ou associar o descritor autorizado que está indicado no parágrafo USE (001), com outro também autorizado constante do parágrafo COMBINE (700), poderá saber, de antemão, que existem documentos que indicam especificamente a informação daí resultante, e na forma prescrita, isto é, o principal adjacente a outro principal.

## **2.2 – Arquivo NJUR e THESAURUS**

A descrição das possibilidades do programa STAIRS e da forma de apresentação dos diferentes arquivos tem o objetivo de mostrar que a Subsecretaria de Análise, ao receber recursos pré-determinados de “software”, aos quais teve que adaptar o acervo de informações transferidos de um arquivo manual sem completa sistematização, viu-se na contingência de desenvolver um trabalho de natureza experimental, na mesma oportunidade em que adotava medidas correlativas para dar operacionalidade ao sistema. Assim é que, enquanto simultaneamente corrigia e adaptava o acervo inicial, no sentido de mantê-lo atualizado, através de sua Seção de Referência Legislativa, criando uma norma experimental de indexação, dentro das possibilidades do Programa STAIRS, desenvolveu, paralelamente, através de sua Seção de Pesquisa, um trabalho também experimental de análise de textos integrais, com o objetivo de tentar criar um Thesaurus específico para a legislação brasileira, a ser estruturado com base nos indexadores autorizados, decorrente da análise de cerca de 80.000 documentos.

É preciso deixar claro, portanto, que o THESAURUS experimental da Constituição não é ainda o instrumento de acesso a toda a legislação e à jurisprudência que compõem o Sistema de Informações Jurídicas, tendo servido, apenas, como instrumental de pesquisa para a futura estruturação de um Thesaurus definitivo, a ser montado com base nos descritores extraídos do texto das normas jurídicas e da jurisprudência que compõem o arquivo NJUR.

De tal maneira que a Subsecretária de Análise dispõe, atualmente, de duas listas básicas de descritores, não inteiramente consistentes entre si.

A primeira, referente aos textos legais arquivados no data base NJUR, constante de descritores autorizados, seus respectivos sinônimos (descritores não autorizados), notas explicativas e descritores correlatos. A hierarquização desses inde-

xadores formará, oportunamente, o THESAURUS da legislação e da jurisprudência disponíveis.

A segunda, relativa ao texto da Constituição, constante de conceitos já devidamente hierarquizados. É o THESAURUS experimental objeto desta comunicação.

### 2.3 – Problemas da Indexação de Textos Jurídicos

A indexação de três textos legais analisados na íntegra e integralmente também transcritos no computador serviu de base à elaboração do THESAURUS experimental. Três grupos de técnicos, trabalhando distintamente e sem obediência a critérios uniformes, analisaram simultaneamente a legislação do trabalho, a legislação referente ao Direito Tributário e a Emenda Constitucional nº 1/69. **A indexação desses textos não obedeceu ao sistema padronizado, posteriormente estabelecido como norma, para a indexação da legislação ordinária que constitui o arquivo NJUR, não -se levando em consideração as “funções” peculiares do programa STAIRS, nem os “operadores” específicos dessas funções.**

Levou-se em consideração, tratando-se de um trabalho experimental, que era necessário textar a indexação através de conceitos que pudessem ser utilizados em qualquer sistema ou programa, independentemente de suas peculiaridades. Teve-se presente, além disso, a singularidade de que, no caso do texto transcrito integralmente, e acessível ao usuário, tomava-se desnecessário utilizar os conceitos explícitos no próprio texto, como indexadores.

Consideraram-se contudo certas particularidades da pesquisa relativa à referência legislativa.

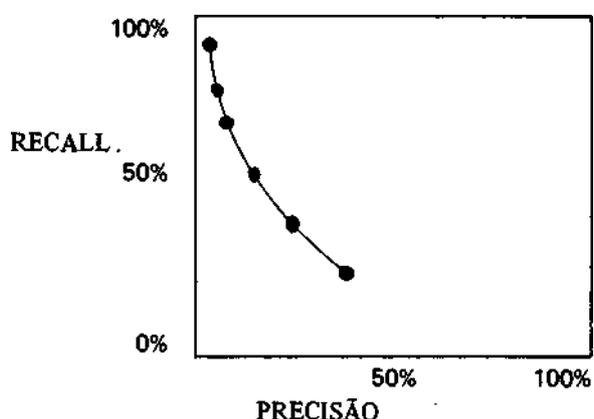
Diferentemente da pesquisa bibliográfica, que jamais é exaustiva, pela impossibilidade de se obter uma lista completa de obras, por mais amplo que seja o acervo disponível, a pesquisa dos textos legais deve, necessariamente, ser exaustiva. Isto porque a indicação de 9 entre 10 textos que discipline determinada matéria pode ser **inteiramente irrelevante**, se o texto não recuperado for o vigente e os 9 recuperados estiverem, por exemplo, tácita ou expressamente revogados. No entanto, uma pesquisa bibliográfica que indique 9 entre 10 documentos relevantes pode ser considerada de excelente nível.

Pode-se afirmar, como conseqüência, que um recall igual a 100 seria sempre desejável num sistema de informações jurídicas. E quanto maior a generalidade do conceito, maior será o recall obtido. Entretanto, num volume muito expressivo de documentos, esse grau de generalidade pode ser um empecilho decisivo à obtenção de informação, razão por que toma-se indispensável que o sistema permita, se-

gundo o desejo do usuário, variar de um elevado e desejado índice de recall a um grau de precisão que atenda às suas necessidades.

O sistema utilizado pela Subsecretária de Análise, que abrange da expressão genérica de alguns conceitos que constituem o catálogo sistemático, até o grau de especificidade que permite situar um determinado dispositivo dentro de um texto legal, procura, exatamente, contornar os inconvenientes da equação gráfica estabelecida por Lancaster \* relativamente à relação inversa existente entre ambos os indicadores – **precision e recall**, não só estabelecendo um equilíbrio desejável entre ambos, como também permitindo, segundo queira o usuário, variar entre as duas alternativas.

Levou-se em conta, paralelamente, a impossibilidade de se utilizar a análise mecânica do texto jurídico, pelo menos na fase atual, embora tenham sido feitas experiências com relações de palavras listadas pelo computador, em razão de sua relevância numérica. As dificuldades neste caso decorrem não só dos conceitos implícitos, como também das peculiaridades da linguagem de cunho legal que, além de não seguir o padrão conceitual de natureza doutrinária, utiliza, por questão de elegância, terminologia típica nem sempre redutível a um padrão aceitável para a indexação controlada.



\* Lancaster, F.W. Vocabulary Control for Information Retrieval. Washington, DC, Information Resources Press, 1972.

### 3. O THESAURUS DA CONSTITUIÇÃO

Para a montagem de um THESAURUS experimental escolheu-se o texto integral da Emenda Constitucional nº 1/69 já transcrito no computador, tendo-se em vista não apenas o amplo limite de abrangência relativamente ao campo do Direito, como também o fato de permitir sistematização mais racional, na medida em que se trata de um texto jurídico já estruturado internamente.

O sistema de indexação, contudo, **não utilizou as potencialidades permitidas pelos diferentes operadores do Programa “STAIRS”**, tendo-se empregado, experimentalmente, conceitos não necessariamente articulados entre si, o que leva a um índice relativamente alto de falsas coordenações, quando se faz a recuperação **pesquisando-se, simultaneamente, no texto e na indexação.**

Os problemas postulados neste trabalho experimental, portanto, referem-se, basicamente, às dificuldades usualmente encontradas na estruturação dos conceitos.

Partiu-se de uma estrutura ideal e teórica, em que o descritor de maior hierarquia é constituído pela palavra THESAURUS, à qual se subordinam todos os demais conceitos do primeiro nível de hierarquia. Embora não haja norma específica para a estruturação, adotou-se o preceito já utilizado nos Thesauri facetados, dividindo-se os conceitos listados automaticamente, segundo o campo da especialidade traçada na própria estrutura interna da lei.

Assim é que, se partirmos do conceito mais genérico no campo jurídico, o de “Direito”, vamos encontrar, como se mostra no exemplo anexo, esse descritor imediatamente abaixo do conceito de maior hierarquia que é, no exemplo teórico, o descritor “thesaurus”. O nível de hierarquia imediatamente inferior a “Direito” será uma de suas divisões teóricas, “Direito Financeiro” ao qual pode-se chegar, também, partindo-se do descritor “Finanças Públicas”.

O segundo nível da hierarquia inferior a “Direito Financeiro” é “Direito Tributário”, ao qual se subordina diretamente o descritor **“Sistema Tributário”**. De tal sorte que um pesquisador que pretenda chegar ao descritor “Direito Tributário”, ignorando se este é um descritor autorizado, poderá atingi-lo tanto pesquisando o descritor **Direito Financeiro** quanto o descritor **Sistema Tributário**.

Este primeiro exemplo serve para mostrar que, embora não se tenha estabelecido uma norma técnica para a estruturação do Thesaurus, procurou-se seguir, a partir dos descritores autorizados dois princípios básicos:

- a) no caso de textos jurídicos, a hierarquização se produz pela maior ou menor **abrangência semântica** do termo;

- b) paralelamente, seguiu-se a estrutura interna do texto, tendo em vista a hierarquização natural da linguagem jurídica, desde que não haja dúvidas quanto ao grau de **generalidade ou especificidade** dos termos.

Na medida em que desce a escala hierárquica, cresce esse grau de especificidade. Assim, por exemplo, no caso em espécie, abaixo do descritor “**Sistema Tributário**”, vamos encontrar os descritores “**Alíquota**”, “**Fazenda Pública**”, “**Contribuinte**”, “**Tributos**” e “**Contribuições Especiais**”.

A escala hierárquica não possui limitações práticas, podendo-se chegar, como no exemplo mostrado, a situações como a do descritor “**Tributos**”, ao qual se subordina outro mais específico “**Imposto**”, com as 17 especificações de Impostos previstos na Constituição.

Isto poderia indicar que, para se chegar a um determinado “descritor” de relativo grau de especificidade, como é o caso de “Imposto sobre minerais”, por exemplo, o pesquisador teria que partir de um conceito genérico como é o caso de Direito Financeiro, Direito Tributário ou Sistema Tributário. Na realidade, usando a palavra “Imposto” o usuário encontrará no vídeo um descritor superior mais abrangente como “Tributos” e 17 outros mais específicos, que são as diferentes modalidades de Impostos previstos na Constituição.

Se, no entanto, estiver pesquisando o que prescreve a Constituição relativamente a minerais poderá encontrar, por associação, **Imposto Sobre Minerais**. Da mesma maneira que, se estiver buscando Imposto Sobre Minerais, saberá, recorrendo ao parágrafo “associado”, que aparece no vídeo, que além de Imposto Sobre Minerais existe um conceito mais genérico que no caso seria “**Mineral**”.

Isto indica que, além de utilizarmos o sistema já conhecido de um “Thesaurus” facetado que abrange praticamente todo o Direito positivo brasileiro, procuramos, paralelamente, aproveitar a potencialidade do programa “STAIRS” que permite, pela similitude das palavras, chegar a conceitos semelhantes que guardam qualquer forma de relacionamento **jurídico ou semântico**, com o tema pesquisado. E isto se deve, como já assinalamos, à necessidade revelada pela indexação legislativa de assegurar um “recall” praticamente igual a 100%. Mesmo que o nível de precisão possa ser prejudicado, na hipótese de haver um número relativamente alto de preceitos legais relativamente à matéria.

É preciso ter em conta, no entanto, que este sistema experimental, que está sendo usado como meio de pesquisa, obedece, em última análise, aos conceitos já tradicionais em indexação de que a pesquisa será tanto mais precisa quanto maior for a possibilidade de pós-coordenação.

Assim é que, se combinarmos o descritor “Alíquota”, que tanto se aplica a “Tributos” quanto a “contribuintes”, a um descritor específico como “ICM”, que equivale a Imposto de Circulação de Mercadorias, vamos poder obter os limites ou as alíquotas máximas permitidas para o referido Imposto.

A grande dificuldade observada quanto ao inter-relacionamento dos conceitos que guardam, similitude semântica ou jurídica, deveu-se à particularidade de que há conceitos suficientemente genéricos ou abrangentes que se aplicam, de forma quase indistinta, a todo e qualquer conceito que se pretenda pesquisar.

É o caso, por exemplo, do termo ou descritor “Competência”, que, embora tenha conceituação específica em Direito Administrativo, é usado generalizadamente, na legislação, como sinônimo perfeito de “Atribuições”.

Partindo-se da premissa de que o sistema é aberto a usuários não especialistas, adotou-se a forma de conceituá-lo não segundo a doutrina estabelece, mas de acordo com a linguagem coloquial que não distingue entre **competência e atribuição**. Neste caso “competência” passou a ser um descritor autorizado e “atribuição” um **descritor não-autorizado**. Se o usuário recorrer ao termo conceitualmente correto, mas não autorizado, como é o caso de “atribuição”, ele fatalmente chegará ao descritor autorizado “competência”, na medida em que o primeiro aparecerá automaticamente como **não-use** e competência **como use**.

Ainda assim, entretanto, competência tanto se refere a competência tributária, quanto a competência administrativa. Pode-se, na linguagem controlada do THESAURUS, combinar competência com a palavra contribuinte, como com a palavra alíquota. Tais termos, que não podem ser hierarquizados, nem por sua abrangência jurídica (no caso de textos estruturados internamente), nem por sua abrangência semântica, foram conceituados como “especificadores”, isto é, descritores autorizados, cuja combinação com conceitos específicos praticamente não tem limitação.

#### 4. CONCLUSÃO

Este dicionário controlado e hierarquizado que se estabeleceu para a Constituição não é, evidentemente, uma solução para a pesquisa, em sistemas automatizados. Em muitos casos a discrepância entre conceitos jurídicos não consagrados leva à adoção de uma linguagem artificial que deve ser necessariamente aprimorada. Mas tem servido, em última análise, como instrumento essencial de pesquisa, a fim de que se possa chegar a uma forma aceitável na padronização da linguagem jurídica, aplicável, indistintamente, à legislação, à jurisprudência e à doutrina, que constituem os campos essenciais de nosso interesse.

## **Abstract**

The PRODASEN and the Analysis Office from the Federal Senate built a system of data recovery indicating the necessity of using an instrument of language control, to get these informations. It was chosen a thesaurus system. Since there was no such specific thesaurus to this field and no trained personal, it was decided to use a Pilot Project, using the Constitutional amendment I, 1969.