

APORTES DEL TESAURO DEL SISTEMA ARGENTINO DE INFORMÁTICA JURÍDICA A LA TERMINOLOGÍA JURÍDICA

Antonia OSES

Ministerio de Justicia de Argentina

CARACTERÍSTICAS DE LA BASE DE DATOS

La información que ofrece esta base proviene de varios tipos de documentos: jurisprudencia, legislación (que incluye la legislación nacional, provincial, decretos del Poder Ejecutivo, circulares del Banco Central), doctrina de los autores, dictámenes de la Procuración del Tesoro, y Resoluciones del Mercado Común.

LENGUAJE DESCRIPTOR

Los documentos contenidos en esta base de datos están archivados en lenguaje natural y por lo tanto toda comunicación que se entabla con el sistema es a través del lenguaje.

Cuando hablamos de comunicación con el sistema nos referimos al usuario que lo consulta para obtener información, y también al analista, o sea la persona que recibe cada nuevo documento y debe indizarlo y agruparlo con los ya existentes. Por eso el analista no pierde de vista que todo documento que entra nuevo, debe por un lado, agruparse temáticamente con aquellos con los que comparta conceptos relevantes, y por el otro, recuperarse en el momento deseado ya que el sistema es precisamente de "almacenamiento y recuperación de información".

Para la comunicación con el sistema contamos con un **lenguaje controlado, o sea un sistema de comunicación normalizado, inequívoco y homogéneo.**

Cada documento que entra a la base de datos se clasifica temáticamente estableciendo la clase a la que pertenece.

Una vez realizada la clasificación del corpus documental se aborda cada documento haciendo un análisis conceptual del mismo. Es decir se analiza cada tema relevante que trate el documento y se lo traduce a un lenguaje normalizado.

Con el lenguaje controlado describimos el contenido semántico de los documentos. Para soportar este lenguaje descriptor, o vocabulario controlado contamos con un Tesauro Jurídico.

Aclaremos que cuando decimos "**vocabulario controlado**" nos referimos a que:

1. restringimos la definición al campo jurídico.
2. diferenciamos cada descriptor de los otros del Tesouro por relaciones que éste entabla con los demás en el universo documental del sistema.
3. lo clasificamos en una faceta según corresponde jurídicamente.
4. elegimos una forma autorizada, descartando las otras.
5. delimitamos el nivel de especificidad relevante sobre cada tema, teniendo en cuenta para eso el corpus total de documentos.

Cada documento tiene un campo para recuperación compuesto de los descriptores. Para acceder a la información, lo hacemos a través de una interfase que nos introduce en el Tesouro y de allí nos lleva a los documentos.

El vocabulario controlado nos garantiza una comunicación satisfactoria.

1. *Cada descriptor o expresión normalizada cuenta con un respaldo documental adecuado.*
2. *El lenguaje es lo suficientemente específico para lograr un aceptable nivel de precisión.*
3. *Este conjunto de descriptores promueve una indización representativa del contenido de los documentos, así como una respuesta representativa de los requerimientos del usuario.*
4. *Los descriptores eliminan la ambigüedad controlando la sinonimia y la homonimia.*

A) Relaciones entre los términos

1. Relación de sinonimia

Esta relación nos permite encontrar toda la información indizada con un descriptor o con cualquiera de sus sinónimos.

Esto es relevante en una base de datos donde participan todas las provincias, y en cada una de ellas se prefiere un término distinto para denominar un mismo concepto.

2. Género-especie

En esta relación el término genérico expresa una clase de conceptos de la cual el descriptor específico es un miembro.

Las relaciones de Género-especie son transitivas, o sea que, cada vez que el usuario busque en la Base de Datos por un término que tenga específicos, recuperará por default todos aquellos documentos que contengan en la indización cualquiera de los términos específicos.

Si un término tuviera un específico y éste a su vez tuviera otros, recuperará todos por transitividad. Ejemplo: si el usuario busca documentos con la voz "contratos comerciales" y en el Tesouro figura lo siguiente:

descriptor = contratos comerciales
término específico = comodato comercial
término específico = compraventa mercantil
término específico = contrato de agencia
término específico = contrato de comisión
término específico = contrato de concesión

descriptor = compraventa mercantil

up = compraventa comercial
término específico = compraventa de automotores
término específico = compraventa de mercaderías

Se recuperarán todos los documentos en los que aparezca cualquiera de las siguientes voces:

comodato comercial
compraventa de mercaderías
compraventa de automotores
compraventa comercial
compraventa mercantil
contrato de comisión
contrato de concesión
contrato de agencia
contratos comerciales

3. Relación temática

Una de las características fundamentales de una herramienta como el Tesauro es que **expresa los significados de los términos por las relaciones entabladas dentro de la clasificación establecida**. Las relaciones y las jerarquías expresadas en cada faceta constituyen el campo conceptual de cada descriptor.

Para establecer estas asociaciones trabajamos con estadísticas que nos muestran la frecuencia de aparición de un descriptor con otros en el mismo campo. También se trabajan grillas jurídicas elaboradas a partir de la organización del Derecho que hacen los Códigos, las leyes o especialistas en la materia.

4. Términos relacionados documentalmente

En una primera etapa se consideró que con un Tesauro que contiene solo términos pertenecientes al campo jurídico, se lograría una recuperación adecuada.

Siguiendo este criterio, en forma estricta, se podía llegar hasta un determinado nivel de especificidad de los términos jurídicos. Este nivel resulta insuficiente, ya que trae como resultado la recuperación de demasiados documentos.

Ante la inmensa cantidad de información que contiene la base, debieron tomarse ciertas medidas a fin de optimizar la recuperación:

Se estableció una relación directa entre los términos del Tesauro jurídico y un listado alfabético de voces que no pertenecen al Derecho.

Esta relación que explicita la herramienta, nos permite indicar al usuario la relación documental del instituto central de la indización de cada documento, con los términos que clasificamos como "**términos de lenguaje natural**", en el caso de las facetas jurídicas.

La vinculación de un tema específico con los términos denominados "jurídicos", es lo que llamamos "**TRD**".

Ante esto, debieron tomarse ciertas medidas a fin de optimizar la recuperación:

1. incorporación en el campo de descriptores de términos del lenguaje natural no comprendidos en el lenguaje jurídico propiamente dicho.
2. descartar la unificación de Tesauros a través de una clasificación única.
3. estructurar los descriptores en distintos Tesauros.

El sistema cuenta con un soft adecuado para soportar varios Tesoros.

Esto nos permite diferenciar el lenguaje jurídico, del lenguaje natural, de nombres que sólo identifican sin describir, nombres propios: "identificadores".

Pero a su vez se debe hacer una referencia de la relación existente entre el Tesoro jurídico y el Tesoro llamado en este sistema Léxico.

Estas relaciones a diferencia de las otras son puramente documentales, por ello los llamamos "**Términos Relacionados Documentalmente**".

Para ejemplificar planteamos la búsqueda de documentos que contengan el tema de "depósito comercial", ya se sabe que esta figura se encuadra dentro del **Derecho Comercial**. Pero esta voz recupera en la base demasiados documentos. Al ver las relaciones que presenta en el Tesoro vemos que una de sus relaciones es el TRD "muerte de animales". Esta voz no la encuadramos en el Derecho comercial, sino que la relacionamos además con otros temas de otras facetas del Derecho. Pero a la hora de acotar las búsquedas nos sirve el lenguaje que no es de especialidad jurídica.

B) Incorporación de términos al Tesoro

Para la incorporación de un descriptor al Tesoro jurídico se tienen en cuenta varios aspectos:

1. Al ser un Tesoro de recuperación de información, en primer lugar se considera su respuesta documental. Esto es que todos los términos del Tesoro al menos recuperan un documento cuyo tema es descrito por el término.

2. Cuando el término es muy específico se realiza una búsqueda para ver si trae ruido.

Por ejemplo, si busco por "sociedades", y este término fuera el descriptor preferido, me mezclaría en la recuperación de documentos referidos a todo tipo de sociedades, no sólo a las comerciales, sino las civiles, etc. Es por ello que no se autoriza la voz "sociedad". Si se realiza una búsqueda de este tipo el sistema pedirá desambiguar entre todas las sociedades existentes. Los términos autorizados son: "sociedades comerciales", "sociedades civiles", etc.

3. Una vez que se pide "sociedades comerciales", por transitividad trae todos los documentos indizados con las voces específicas de este término.

Esto nos lleva a autorizar voces del tipo "remoción del administrador de la sociedad", en lugar de desdoblado en dos o tres descriptores unitérminos.

Explicitamos esto para aclarar que el sistema de recuperación de información, el soft, y la interfase con el usuario, son determinantes a la hora de autorizar los términos.

4. Otra pauta para priorizar la elección de los términos es su uso más corriente, se prefiere la voz mas usada por los especialistas del Derecho, y no su aparición en los documentos. Para ello contamos con programas que nos registran los términos que utilizan los usuarios para hacer las búsquedas en sistema.

Esta prevalencia de las búsquedas nos lleva a considerar como preferidos términos en plural y singular, o con artículos formantes, preposiciones o las llamadas palabras vacías. Es el caso de los términos "conservación y administración de los bienes", "modificación del estatuto social".

Otro ejemplo es "sociedades comerciales", cuya definición la da la LEY DE SOCIEDADES, en su artículo 1.- "Habrá sociedad comercial cuando dos o más personas en forma organizada, conforme a uno de los tipos previstos en esta ley, se obliguen a realizar aportes para aplicarlos a la producción o intercambio de bienes o servicios participando de los beneficios y soportando las pérdidas". A pesar de su aparición en singular en la ley se priorizó el plural.

Otros ejemplos como éste los proporcionan los siguientes términos que transcribimos con fuentes documentales: "remoción del administrador de la sociedad", que en la ley mencionada, aparece en un sintagma más breve: artículo 129.- "...Los socios disconformes con la remoción del administrador cuyo nombramiento fue condición expresa de la constitución de la sociedad, tienen derecho de receso."

Pero esto dificultaría la recuperación ya que traería ruido por la "remoción del administrador del consorcio", o remoción del administrador de la herencia" por ejemplo.

CONCLUSIONES

Si bien este Tesoro está pensado con fines muy específicos, para uso de esta base de datos en particular, queda en evidencia que al contener documentación jurídica en gran cantidad, al ser consultada por especialistas del Derecho, y al intentar representar cada campo semántico según el universo de documentos, es de un gran aporte a la hora de una herramienta terminológica o una base de datos terminológicos, ya sea por superar los diez mil términos jurídicos, que en el campo de las sociedades comerciales son aproximadamente trescientos, como por los miles de documentos fuente.

BIBLIOGRAFÍA CONSULTADA

- ATCHINSON, Juan. "Thesaurus construction; a practical manual". Aslib, London, 1972.
- UNESCO. "Guidelines for the establishment and development of monolingual thesauri." París, Septiembre, 1973.
- LOPEZ- MUÑIZ GOMEZ, Miguel. Informática jurídica documental. Díaz De Santos S.A., Madrid-España, 1ª edición, 1984.

LABURPENA / RESUMEN / RÉSUMÉ / ABSTRACT

Informatika juridikoko sistema argentinarreko Thesaurusetik terminologia juridikora ekarpenak

Informatika zuzenbidearen esparrura aplikatzean, informatika juridikoaz hitz egiten da. Agirietako informatika juridikoa datu-base juridikoen inguruan dago oinarrituta.

Datu-base horietan hiru agiri-mota aurkitzen dira batez ere, zuzenbidearen hiru iturri nagusien arabera, hain zuzen: jurisprudentzia, legegintza eta autoreen doktrina.

Datu-base batzuk publikoak eta beste batzuk pribatuak dira. Gure herrialdean, datu-base publiko bakarra INFORMATIKA JURIDIKOKO SISTEMA ARGENTINARRA da, Estatuko Justizi Ministerioaren mende dagoena. Gaur egun, base horrek 510.000 agiri baino gehiago ditu.

Lantxo honetan, sistemaren erabiltzaileak eskura duen baliabide linguistiko baten gainean jardungo dugu batez ere: thesaurus juridikoa, informazioa gaika berreskuratzeko aukera ematen duena.

Aportes del Tesouro del sistema argentino de informática jurídica a la terminología jurídica

La aplicación de la informática en el campo del Derecho se denomina Informática Jurídica. La informática jurídica documental, gira alrededor de las llamadas Bases de datos jurídicos.

Las mismas presentan fundamentalmente tres tipos de documentos que responden a las tres fuentes principales del Derecho: jurisprudencia, legislación y doctrina de los autores.

Existen bases de datos públicas o privadas. En nuestro país, la única base de datos pública es el SISTEMA ARGENTINO DE INFORMÁTICA JURÍDICA, dependiente del Ministerio de Justicia de la Nación. Esta base cuenta actualmente con más de 510.000 documentos.

Con este breve trabajo nos referiremos en particular a la herramienta lingüística que ofrece a los usuarios del sistema para la recuperación temática de la información: el Tesouro jurídico.

Contribution du Thesaurus du système argentin d'information juridique à la terminologie juridique

L'application de l'informatique dans le domaine du Droit se dénomme "Informatique Juridique". L'informatique juridique documentaire porte sur ce que l'on a coutume d'appeler les "Bases de données juridiques".

Celles-ci présentent fondamentalement trois types de documents qui répondent aux trois sources principales du Droit : jurisprudence, législation et doctrine des auteurs.

Il existe des bases de données publiques ou privées. Dans notre pays, la seule base de données publique est le SYSTEME ARGENTIN D'INFORMATIQUE JURIDIQUE, dépendant du Ministère de la Justice de la nation. Cette base compte à l'heure actuelle plus de 510.000 documents.

A travers ce bref travail, nous nous référons en particulier à l'outil linguistique qu'elle offre aux utilisateurs du système pour la récupération thématique de l'information : le Thesaurus juridique.

Contributions of the Thesaurus of the argentinian legal IT system to legal terminology

The application of information technology to the field of law is referred to as "legal IT". Documentary legal IT is based on legal databases.

These databases contain basically three types of documents, corresponding to the three main sources of law: jurisprudence, legislation and the doctrine of authors.

There are public and private databases. The only public database in our country is the SISTEMA ARGENTINO DE INFORMATICA JURIDICA ("Argentinian Legal IT System"), which depends on the national Justice Ministry. At present this database has over 510.000 documents.

In this brief study we shall look in particular at a linguistic tool available to users of the system which enables them to recover information by topics: the Legal Thesaurus.