

Manual Técnico de INLEXPO

RICARDO E. MONGE

Instituto de Investigaciones Lingüísticas

2 de julio de 2015

El presente manual documenta y especifica los aspectos técnicos y de desarrollo que rigen el funcionamiento del sistema de manipulación de diccionarios que el Instituto de Lexicografía de la Facultad de Letras de la Universidad de Costa Rica utiliza para la creación, gestión y edición de sus obras lexicográficas.

El documento está dividido en tres secciones, la primera trata sobre la organización de la base de datos, la segunda de la organización de los archivos del sistema y la tercera sobre las contraseñas y accesos para el personal desarrollador y técnico al sistema.

Diseño y funcionamiento de la base de datos

El almacenamiento de información lexicográfica resulta ser todo un reto para los sistemas computacionales tradicionales, por variedad de razones, pero entre las más importantes (y que fueron las que establecieron las bases para el diseño de la base de datos completamente nuevo) están:

1. Las entradas lexicográficas son estructuras jerárquicas, muy variables. En un caso, una entrada puede tener una simple definición (acepción); pero en otro caso, puede haber ejemplos, referencias para varias acepciones y variantes de la entrada, cada una con sus acepciones, ejemplos y referencias.
2. El texto de las distintas secciones (ejemplos, acepciones, traducciones, etc.) puede variar mucho en longitud y puede tener materiales multimedia asociados (imágenes, grabaciones de audio de pronunciaciones, etc.).

3. La plataforma requiere flexibilidad para colocar nuevos tipos de información, pues no será solamente para un solo tipo de material lexicográfico.

La característica (1) indica que la base de datos debe poder almacenar información de una estructura jerárquica de una forma eficiente, tomando en cuenta de que la jerarquía resultante no es muy profunda (a lo más, son cuator o cinco niveles de jerarquía). Se implementa mediante una tabla en la base de datos con referencia al nodo *padre* y se declaran distintos tipos de nodos (y según esos tipos, se interpreta el nivel en la jerarquía).

La característica (2) indica que en el aspecto del contenido para cada *nodo* de la jerarquía debe ser ampliable y flexible, con posibilidad de almacenar tanto texto (sin límite de longitud y formato) así como materiales binarios adjuntos. De la misma forma, si fuera necesario agregar nuevos tipos de contenido, debe ser una operación relativamente sencilla, para cumplir con la característica (3). Se implementa mediante una tabla que enlaza los distintos contenidos a un elemento de la jerarquía.

Tablas de la base de datos

El nombre y el propósito de cada una de las tablas de la base de datos se describe en el cuadro 1 y en los subsiguientes se describe el detalle de los campos de cada una y su propósito para el almacenamiento seguro de la información lexicográfica.

Las tablas *content*, *content_types* y *content_choice_options* colaboran juntas para obtener la flexibilidad necesaria para la colocación de material lexicográfico diverso. A manera de ejemplo, la definición y el ejemplo de una cierta acepción de una entrada lexicográfica tienen cada uno un registro en *content*, diferenciable mediante el *content_type* (1 es para definiciones, 2 es para ejemplos). De la misma manera, para la asignación de las categorías gramaticales (*content_type* es 5) y marcas lingüísticas (*content_type* entre 10 y 16), pero los datos de las opciones permitidas (las categorías y marcas en sí) se obtienen de *content_choice_options*. Esto hace que el sistema sea sumamente flexible e independiente de las necesidades lexicográficas.

En la versión de INLEXPO desarrollada a lo largo del I Semestre del año 2015, no se utilizan las referencias cruzadas; pero la plataforma de base de datos está preparada para su incorporación en versiones posteriores. La idea es que relaciones como las equivalencias de traducción en otros idiomas,

Nombre de la tabla	Propósito
content	Guarda los datos particulares de acepciones, ejemplos y demás contenidos de una entrada lexicográfica
content_choice_options	Contiene las opciones para la selección de categorías gramaticales y marcas lingüísticas
content_types	Contiene los distintos tipos de contenidos que pueden enlazarse a una entrada lexicográfica
crossref	Almacena los datos de una referencia cruzada entre entradas lexicográfica
crossref_type	Contiene los distintos tipos de referencias cruzadas
dictionary	Contiene los diccionarios y obras lexicográficas del sistema
entry	Almacena los datos propios de una entrada lexicográfica (sin importar su nivel jerárquico)
entry_type	Contiene los detalles de los niveles jerárquicos de las entradas
languages	Permite la separación de material según el idioma, así como la especificación de referencias cruzadas para la traducción
users	Contiene la información de los usuarios que pueden autenticarse para realizar modificaciones a las entradas lexicográficas

Cuadro 1: Tablas del sistema INLEXPO

Columna	Tipo	Propósito
id	int	Identificador de contenido
entry_id	int	ID de entrada asociada
lang	char(2)	Referencia al idioma de este contenido
content_type	int	Tipo de contenido
source	varchar(150)	Texto de la fuente
content_text	varchar(9000)	Texto del contenido
content_int	int	Opción seleccionada del contenido
content_attachment	varchar(200)	Nombre del archivo adjunto asociado

Cuadro 2: Columnas de la tabla `content`

Columna	Tipo	Propósito
lang	char(2)	Referencia al idioma del tipo de contenido
content_type_id	int	Tipo de contenido
content_choice_id	int	Identificador de la opción
content_value	varchar(100)	Detalle de la opción
content_value_abbr	varchar(50)	Abreviatura de la opción

Cuadro 3: Columnas de la tabla `content_choice_options`

Columna	Tipo	Propósito
type_id	int	Identificador de tipo de contenido
lang	char(2)	Referencia al idioma del tipo de contenido
label	varchar(40)	Nombre del tipo de contenido
parent	int	Si el contenido es jerárquico, se guarda el identificador del tipo de contenido <i>padre</i>
required	bit	Indica si el campo es opcional o no
control	int	Tipo de visualización asociada. Ver cuadro 4.
max	int	Indica si hay un máximo de contenidos para una entrada lexicográfica dada

Cuadro 4: Columnas de la tabla `content_types`

Valor	Propósito
1	Un campo de texto de una línea, sin formato
2	Casillas de selección única, una por opción
3	Dos campos de texto de una línea, sin formato, para el contenido y la fuente

Cuadro 5: Tipos de visualización para `control`

Columna	Tipo	Propósito
src	int	Identificador de entrada que origina la referencia cruzada
tgt	int	Identificador de entrada que recibe la referencia cruzada
cross_ref_type	int	Tipo de referencia cruzada
icon	varchar(10)	Referencia al símbolo que debe utilizarse para representar la referencia cruzada

Cuadro 6: Columnas de la tabla `crossref`

Columna	Tipo	Propósito
type_id	int	Identificador del tipo de referencia cruzada
type_name	varchar(20)	Nombre en pantalla del tipo de referencia cruzada

Cuadro 7: Columnas de la tabla `crossref_type`

Columna	Tipo	Propósito
id	int	Identificador de diccionario
name	varchar(100)	Título del diccionario
author	varchar(20)	Identificador del usuario propietario del diccionario

Cuadro 8: Columnas de la tabla `dictionary`

términos asociados o similares se establezcan como referencias cruzadas entre entradas (o entre acepciones).

Por su naturaleza, la estructura de la tabla que almacena los diccionarios (ver cuadro 5) no hace referencia al idioma; pues las características de lengua corresponden a las entradas lexicográficas que se han ido creando y sus referencias cruzadas.

Las tablas `entry` y `entry_type` también forman una simbiosis que permite la creación de estructuras jerárquicas muy flexibles. Por el momento, hay tres tipos de entradas (*lema*, *acepción*, *sublema*) y pueden colocarse *acepciones* dentro de *lemas* y *sublemas*, así como colocar *sublemas* dentro de los *lemas*. La especificación del tipo de entrada y el identificador de la entrada *contenedora* se hace en registros de la tabla `entry`.

Organización de archivos del sistema

El sistema INLEXPO desarrollado se compone de varios archivos con variedad de funciones y carpetas asociadas a componentes adicionales para la manipulación de archivos y recursos externos. El cuadro 13 documenta los archivos y carpetas y su propósito en el funcionamiento de INLEXPO, que están en el servidor en la carpeta `/var/www/inlexpo15/`.

Columna	Tipo	Propósito
id	int	Identificador de la entrada lexicográfica
d_id	int	Identificador del diccionario que contiene la entrada
lang	char(2)	Referencia al idioma de la entrada
type	int	Tipo de entrada y nivel jerárquico
head	varchar(50)	Palabra(s) de la entrada lexicográfica
parent	int	Si la entrada pertenece a otra, se guarda el identificador de la entrada <i>contenedora</i>
number	int	Para acepciones, el número de acepción.

Cuadro 9: Columnas de la tabla entry

Columna	Tipo	Propósito
type_id	int	Identificador del tipo de entrada
type_name	varchar(20)	Nombre del tipo de entrada
permitted_children	varchar(60)	Listado de tipos de entradas que pueden colocarse dentro de este tipo de entradas
permitted_content	varchar(60)	Listado de tipos de contenido que se pueden colocar en este tipo de entradas

Cuadro 10: Columnas de la tabla entry_type

Columna	Tipo	Propósito
id	char(2)	Identificador de idioma
name	varchar(50)	Nombre en pantalla del idioma

Cuadro 11: Columnas de la tabla languages

Columna	Tipo	Propósito
login	varchar(20)	Login del usuario
name	varchar(100)	Nombre real del usuario
lastname	varchar(100)	Apellido del usuario
email	varchar(100)	Correo electrónico del usuario
level	int	Tipos de permisos.
password	varchar(100)	Contraseña, cifrada con SHA1

Cuadro 12: Columnas de la tabla users

Archivo	Tipo	Propósito
js/jquery-2.1.3.min.js	Javascript	Librería necesaria para el funcionamiento de INLEXPO
vendor/phpoffice/*	PHP	Librería para leer y escribir documentos de Word, para la funcionalidad de exportación de entradas
ajax.php	PHP	Procesador de peticiones de visualización y actualización de la base de datos
export.doc.php	PHP	Componente que produce la versión en Word del diccionario seleccionado
export.html.php	PHP	Componente que produce una versión visible en el navegador (HTML) del diccionario seleccionado
inlexpo.png	Imagen	Logotipo del proyecto INLEXPO
log.php	PHP	Componente que procesa el inicio de sesión
style.css	Hoja de estilos	Hoja CSS que especifica el formato y colores de la página
ui.js	Javascript	Contiene los comandos Javascript que permite la interacción del usuario con el sistema y la base de datos subyacente

Cuadro 13: Detalle de archivos y carpetas de INLEXPO

