



Laboratorio Nacional
de Políticas Públicas



CIDE

Metodología para la creación de una base de datos con las notas del Diario Oficial de la Federación.

Versión: 25 de may. de 2019

Esta base de datos contiene las notas del Diario Oficial de la Federación (DOF) que están disponibles en la página web del DOF (<https://www.dof.gob.mx>). Toda la información proviene exclusivamente de lo publicado en esta página y no usamos ninguna otra fuente.

Debido a que la información original viene en formato de página web, es posible que al extraer la información esta no sea idéntica a la publicada en la versión impresa, o que no tenga el mismo formato que la que se publicó en la página. Por tanto, advertimos que esta información se publica con carácter académico, para facilitar la labor de investigadores o personas interesadas en analizar de forma sistemática el DOF en un formato de datos abierto que permita su procesamiento y de ninguna manera debe considerarse como una herramienta que sustituye la consulta de la versión impresa o la publicada en la página web.

Debido a la disponibilidad de información en el sitio web, para los años anteriores a 2013 se cuenta solo con el título de la nota, pero no con el texto. A partir de 2013 las notas tienen tanto título como texto.

Para obtener los datos usamos la técnica de [web scrapping](#) para ingresar a la página de cada edición del diario y extraer la información relevante. A continuación, ilustraremos cuál fue la información que extrajimos y a qué campo de la base de datos corresponde, usando como ejemplo la edición matutina del día 2 de mayo del 2019.

La URL a la edición de una edición del DOF sigue el patrón:

<https://www.dof.gob.mx/index.php?year=2019&month=05&day=02&edicion=MAT>

donde se especifican los parámetros de año (`year=2019`), mes (`month=05`), día (`day=02`) y edición (`edicion=MAT`). Aunque el DOF se publica usualmente de lunes a viernes, según la ley se puede publicar cualquier día, incluido feriados, por lo tanto, buscamos los DOF publicados cada día desde 1917. El DOF puede tener más de una edición al día, por eso el parámetro de edición puede tener dos valores: MAT y VES, que corresponden a la edición matutina y vespertina, respectivamente. Para cada día buscamos ambas versiones y si existen extraemos los datos.

En la Ilustración 1 se puede ver el principio de la página. Aquí identificamos varios elementos que usamos en la base de datos para identificar a cada nota: sección, rama, institución y título. Al hacer clic encima de un título nos lleva a otra página que contiene el texto de esa nota, como se muestra en la Ilustración 2. Este contenido está en formato HTML, por lo que para extraer solo el texto usamos el método `.get_text()` de la librería [BeautifulSoup](#) del lenguaje de programación [Python](#). Al extraer solo el texto podemos perder algunas características que son útiles para mostrar la información, como formato de los encabezados, tablas y gráficas, pero que no se pueden representar en texto plano. Por ello reiteramos que esta base de datos debe ser considerada solo para fines de investigación.

La base de datos resultante la podemos ver en la Ilustración 3. Cada fila corresponde a una nota en el DOF, y las columnas son las variables para esa nota: fecha, edición, sección, rama, institución, título, texto, URL de la edición del DOF y la URL del texto de la nota.

Ilustración 1. Página de una edición digital del DOF



The screenshot shows the digital edition page of the DOF. At the top, there is a search bar and navigation links. The main content area displays the date and edition (02/05/2019 - Edición Matutina) and a table of contents with the following items:

- PRIMERA SECCION** (Annotated with 'sección')
- PODER EJECUTIVO** (Annotated with 'poder')
- SECRETARIA DE SEGURIDAD Y PROTECCION CIUDADANA** (Annotated with 'institución')
- SECRETARIA DE AGRICULTURA Y DESARROLLO RURAL** (Annotated with 'titular')

Each item is followed by a brief description of the content. On the right side, there is a calendar for May 2019 and a sidebar with various utility links such as 'Crear Usuario', 'Búsqueda Avanzada', 'Noticias', and 'Suscripción'.

Ilustración 2. Contenido de una nota



Laboratorio Nacional
de Políticas Públicas

Laboratorio Nacional de Políticas Públicas LNPP



CIDE

CONVENIO de Coordinación que en el marco del Sistema Nacional de Seguridad Pública, celebran el Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública y el Estado de Durango, relativo al Fondo de Aportaciones para la Seguridad Pública de los Estados y del Distrito Federal (FASP) 2019.

Al margen un sello con el Escudo Nacional, que dice: Estados Unidos Mexicanos.- SEGURIDAD.- Secretaría de Seguridad y Protección Ciudadana.- El Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública.

CONVENIO DE COORDINACIÓN QUE EN EL MARCO DEL SISTEMA NACIONAL DE SEGURIDAD PÚBLICA, CELEBRAN POR UNA PARTE EL PODER EJECUTIVO FEDERAL, POR CONDUCTO DEL SECRETARIADO EJECUTIVO DEL SISTEMA NACIONAL DE SEGURIDAD PÚBLICA, EN LO SUCESIVO "EL SECRETARIADO", REPRESENTADO POR SU TITULAR, EL C. FRANCO GERARDO MARCELLO FABBRI VÁZQUEZ, Y POR LA OTRA, EL PODER EJECUTIVO DEL ESTADO LIBRE Y SOBERANO DE DURANGO, EN LO SUCESIVO "LA ENTIDAD FEDERATIVA", REPRESENTADO POR SU GOBERNADOR, EL C. JOSÉ ROSAS AISPURO TORRES, ASISTIDO POR EL SECRETARIO GENERAL DE GOBIERNO, EL C. ADRIÁN ALANÍS QUIÑONES; EL SECRETARIO DE FINANZAS Y DE ADMINISTRACIÓN, EL C. JESÚS ARTURO DÍAZ MEDINA Y EL SECRETARIO EJECUTIVO DEL CONSEJO ESTATAL DE SEGURIDAD PÚBLICA, EL C. OMAR CARRAZCO CHÁVEZ; A QUIENES CONJUNTAMENTE SE LES DENOMINARÁ "LAS PARTES" DE CONFORMIDAD CON LAS DECLARACIONES Y CLÁUSULAS SIGUIENTES:

DECLARACIONES

I. DECLARA "EL SECRETARIADO", A TRAVÉS DE SU REPRESENTANTE QUE:

- I.1 Es un Órgano Administrativo Desconcentrado de la Secretaría de Seguridad y Protección Ciudadana, operativo del Sistema Nacional de Seguridad Pública, con autonomía técnica, de gestión y presupuestal, de conformidad con los artículos 17 de la Ley General del Sistema Nacional de Seguridad Pública, en lo subsecuente "Ley General", y 1 del Reglamento del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública, concatenado con los artículos Octavo, Décimo Tercero y Décimo Cuarto transitorios del Decreto por el que se reforman, adicionan y derogan diversas disposiciones de la Ley Orgánica de la Administración Pública Federal, publicado en el Diario Oficial de la Federación el 30 de noviembre de 2018.



Laboratorio Nacional de Políticas Públicas LNPP



Ilustración 3. Versión en base de datos de las notas de la página

fecha	edicion	seccion	rama	inst	titulo	texto	url_dof	url_texto
2019-05-02	MAT	PRIMERA SECCION	PODER EJECUTIVO	SECRETARIA DE SEGURIDAD Y PROTECCION CIUDADANA	Convenio de Coordinación que en el marco del S...	\n CONVENIO de Coordinación que en el marco ...	https://www.dof.gob.mx/index.php?year=2019&mon...	https://www.dof.gob.mx/nota_detalle.php?codigo...
2019-05-02	MAT	PRIMERA SECCION	PODER EJECUTIVO	SECRETARIA DE SEGURIDAD Y PROTECCION CIUDADANA	Convenio de Coordinación que en el marco del S...	\n CONVENIO de Coordinación que en el marco ...	https://www.dof.gob.mx/index.php?year=2019&mon...	https://www.dof.gob.mx/nota_detalle.php?codigo...
2019-05-02	MAT	PRIMERA SECCION	PODER EJECUTIVO	SECRETARIA DE SEGURIDAD Y PROTECCION CIUDADANA	Convenio de Coordinación que en el marco del S...	\n CONVENIO de Coordinación que en el marco ...	https://www.dof.gob.mx/index.php?year=2019&mon...	https://www.dof.gob.mx/nota_detalle.php?codigo...
2019-05-02	MAT	PRIMERA SECCION	PODER EJECUTIVO	SECRETARIA DE SEGURIDAD Y PROTECCION CIUDADANA	Convenio Específico de Adhesión para el otorga...	\n CONVENIO Específico de Adhesión para el o...	https://www.dof.gob.mx/index.php?year=2019&mon...	https://www.dof.gob.mx/nota_detalle.php?codigo...
2019-05-02	MAT	PRIMERA SECCION	PODER EJECUTIVO	SECRETARIA DE SEGURIDAD Y PROTECCION CIUDADANA	Declaratoria de Emergencia por la presencia de...	\n DECLARATORIA de Emergencia por la presenc...	https://www.dof.gob.mx/index.php?year=2019&mon...	https://www.dof.gob.mx/nota_detalle.php?codigo...