Universidad de la Sierra Sur

Análisis exploratorio del fenómeno del acceso a la información pública gubernamental federal de México, usando métodos de ciencia de datos: periodo 2003 a 2015

TESIS
Para obtener el título de:
Maestro en Gobierno Electrónico

Presenta:
Hermelando Cruz Pérez

Bajo la dirección del
Dr. Sergio Coria Olguín

Miahuatlán de Porfirio Díaz, Oaxaca, septiembre de 2019

Tesis desarrollada por el Ing.Hermelando Cruz Pérez, egresado de la maestría en Gobierno Electrónico, bajo la dirección del comité Tutorial:

Director: Dr. Sergio Rafael Coria Olguín.

Asesor: Dr. Christian Arturo Cruz Meléndez.

Asesor: Dr. Guadalupe Gabriel Durán Férman.

Tesis presentada en Examen Profesional el 24 de septiembre de 2019, ante el siguiente Jurado:

Presidente: Dr. Víctor Alberto Gómez Pérez

Secretario: Mtra. Nina Martínez Cruz

Vocal: Dr. Sergio Rafael Coria Olguín

Suplente: Dr. Diego Soto Hernández

Suplente: Dr. Arisaí Darío Barragán López

AGRADECIMIENTOS

En primer lugar expreso mi agradecimiento al director de tesis, Dr. Sergio Coria Olguín, por el apoyo, paciencia y confianza brindadas en este trabajo, a mi comité tutorial el Dr. Christian Arturo Cruz Meléndez y el Dr. Guadalupe Gabriel Durán Férman por su apoyo y disponiblidad. A mis padres y hermanos quienes a lo largo de toda mi vida han apoyado y motivado mi formación académica, creyeron en mi en todo momento y no dudaron de mis habilidades.

RESUMEN

El acceso a la información pública gubernamental en México es un fenómeno relativamente reciente y poco investigado; quizá, aún menos que el fenómeno de la transparencia, con el cual llega a confundirse. Por ello, el objetivo de esta investigación es analizar y descubrir patrones relevantes en las características de las solicitudes de información pública federal, de la ubicación de los solicitantes y en las dependencias gubernamentales accedidas. El objeto de estudio específico es la base de datos de las solicitudes generadas en el período de 2003 a 2015, dirigidas al gobierno federal mediante la plataforma informática que está a cargo del Instituto Nacional de Acceso a la Información y Protección de Datos Personales (INAI). Esta base contiene aproximadamente veinte atributos, correspondientes a más de un millón 200 mil solicitudes. La metodología de investigación propuesta tiene enfoque cuantitativo, con alcance descriptivo y exploratorio. Las técnicas aplicadas son: estadística descriptiva y aprendizaje automático. De este último, se usa específicamente un algoritmo ya conocido para generar automáticamente árboles clasificadores, con el fin de descubrir y representar las interacciones entre los atributos de las solicitudes. Una aportación sustantiva de esta investigación es el hallazgo de un conjunto de patrones de tipos uni-variable y multi-variable en estos atributos. Algunos de los patrones más generales son: 1) la cantidad anual de solicitudes presenta una tendencia ascendente, y 2) los sectores gubernamentales que han recibido la mayoría de las solicitudes son: seguridad social, hacienda y crédito público, y educación pública. Las entidades federativas que han tenido las mayores tasas de solicitudes por cada cien mil habitantes son: Distrito Federal (CDMX), Morelos y Estado de México. Se presentan también otros patrones más específicos. Además de los patrones hallados, que constituyen una contribución teórica, otra aportación es la propuesta metodológica aplicada para descubrirlos y representarlos. Adicionalmente, el cálculo de tasas de solicitudes por cada cien mil habitantes ofrece una perspectiva complementaria para el entendimiento del fenómeno. En forma práctica, los resultados pueden aprovecharse para revisar y mejorar leyes, políticas, trámites y procesos de acceso a la información pública mediante el uso de las tecnologías de información y comunicación.

Palabras clave: gobierno abierto, gobierno electrónico, minería de datos, árboles clasificadores.

ABSTRACT

Access to federal government public information in Mexico is a relatively recent and understudied phenomenon; perhaps, it has been less investigated than the transparency phenomenon, with which it can be confused. Therefore, the aim of this research is to analyze and discover significant patterns in the features of: federal public information requests, location of requesters, and referred agencies. The specific subject matter of the study is the database of requests generated in the period from 2003 to 2015, addressed to the federal government by means of the computing platform in charge of the Mexican Institute for Information Access and Personal Data Protection (INAI). This database contains, approximately, twenty attributes, corresponding to over one million two hundred thousand requests. The proposed research methodology has a quantitative approach with descriptive and exploratory scope. The applied techniques include: descriptive statistics and machine learning. In regards to the latter a known algorithm for automatically generating classification trees is specifically used to discover and represent interactions among the request attributes. A significant contribution from this research is the finding of a set of uni-variate and multi-variate patterns in these attributes. A number of the most general patterns are: 1) the annual quantity of requests shows a rising trend, and 2) the government sectors receiving most of the requests are: social security, treasury and public credit, and public education. The States having the highest rates of requests per one hundred thousand inhabitants are: Federal District (CDMX), Morelos and State of Mexico. Other more specific patterns are also presented. In addition to the discovered patterns, which are a theoretical contribution, another contribution is the methodological proposal applied for discovering and representing them. Besides, the calculation of rates of requests per one hundred thousand inhabitants offers a complementary perspective for understanding the phenomenon. A practical application of the results can be to use them as input for reviewing and improving laws, policies, procedures and processes of access to public information by means of information and communication technologies.

Key words: open government, electronic government, data mining, classification trees.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

LISTA DE ABREVIATURAS

Aguascalientes	Ags
Baja California	BC
Baja California Sur	BCS
Campeche	Camp
Chiapas	Chih
Chihuahua	Chis
Coahuila	Coah
Colima	Col
Ciudad de México	DF
Durango	Dgo
Guanajuato	Gro
Guerrero	Gto
Hidalgo	Hgo
Jalisco	Jal
México	Mex
Michoacán	Mich
Morelos	Mor
Nayarit	Nay
Nuevo León	NL
Oaxaca	Oax
Puebla	Pue
Querétaro	QR
Quintana Roo	Qro
San Luis Potosí	Sin
Sinaloa	SLP
Sonora	Son
Tabasco	Tab
Tamaulipas	Tamps
Tlaxcala	Tlax
Veracruz	Ver
Yucatán	Yuc
Zacatecas	Zac

INTRODUCCIÓN

El presente trabajo tiene como principal objetivo analizar las solicitudes de información pública gubernamental, contenidas en la base de datos del Instituto Nacional de Transparencia Acceso a la Información y Protección de Datos Personales (INAI), correspondientes al periodo 2003 a 2015, aplicando técnicas de minería de datos, que incluyen estadística descriptiva y aprendizaje automático. Junto con ello, se propone una metodología para el análisis de la base de datos.

La importancia de estudiar este tema radica en que el acceso a la información pública federal en México es un tema de interés en el área de conocimiento del gobierno electrónico, específicamente respecto a la relación entre gobierno y ciudadano. El motivo es que recientemente se ha hecho uso de las plataformas digitales denominadas Sistema Informatizado de Solicitudes de Información (SISI) e Infomex para automatizar y agilizar el proceso de acceso a la información pública de las dependencias gubernamentales. Además, la información en formatos abiertos, generada por el gobierno, se ha aprovechado poco debido al desconocimiento del potencial de las técnicas y herramientas de la minería de datos que permiten encontrar patrones y comportamientos relevantes.

Al descubrir patrones estadísticos en la base de datos de solicitudes de información procesadas por el INAI, se puede generar conocimiento innovador acerca del comportamiento de los solicitantes y de los otorgantes de información gubernamental federal. Este conocimiento le sirve al propio gobierno, a la comunidad de investigación, al sector empresarial y a la sociedad en general. Se puede generar un análisis más profundo del fenómeno del acceso a la información y recomendar acciones de política pública para mejorar e incrementar el acceso a la información, contribuyendo la rendición de cuentas.

Esta tesis consta de seis capítulos, como se describe a continuación. En el primer capítulo, se plantea la problemática que dio origen a esta investigación, la descripción del objeto de estudio, las preguntas de investigación, los objetivos y la justificación. El segundo capítulo aborda los antecedentes de la investigación, donde se analizan algunos trabajos previos relacionados con aspectos de metodología. El tercer capítulo se dedica al marco teórico, que incluye principalmente conceptos de gobierno electrónico, gobierno abierto, acceso a la información pública, datos abiertos, minería de datos y modelos estadísticos. El cuarto capítulo describe la metodología, el enfoque de la investigación y las técnicas de análisis de datos. En el quinto capítulo se presentan los resultados, incluyendo los basados en estadística descriptiva y los basados en aprendizaje automático. Finalmente, en el capítulo sexto se presentan las conclusiones, se proponen algunas recomendaciones de política pública y se sugiere trabajo de investigación a futuro.