***Resumen Descriptivo***

Re-empadronamiento es una de las palabras claves que impulsan el desarrollo del presente trabajo de investigación, que tiene por objetivo automatizar el proceso de redistribución de un padrón electoral en base a la distancia óptima del domicilio del elector al lugar de sufragio asignado.

El desarrollo de la elaboración del proyecto de tesis comenzará partiendo del padrón brindado por la Junta Electoral, con el propósito de estandarizar los domicilios y obtener su geo-codificación buscando el mayor grado de exactitud posible. Esta codificación se utilizará luego como insumo de entrada a un algoritmo a diseñar para la reasignación de los lugares de sufragio.

**1.  Área temática**

El presente trabajo es un caso de aplicación de procesos ETL (Extract, Transform and Load), extracción, transformación y carga, de datos espaciales.

**¿Qué es ELT?** La función principal de la integración de datos es obtener los datos de donde se encuentran actualmente, cambiarlos para que sean compatibles en formato con el destino deseado e introducirlos en el sistema de destino. Estos tres pasos se denominan extracción, transformación y carga (ETL). Toda la integración de datos, independientemente de si se realiza por lotes o en tiempo real, de forma sincrónica o asincrónica, física o virtual, gira en torno a estas acciones básicas [1].

En el presente caso los procesos ETL permitirán a partir de los datos de una planilla de cálculo, mediante la estandarización y geo-codificación, obtener una Base de Datos Espacial.

Una colección de datos referidos a objetos de los que se conoce su localización espacial se denomina **Base de Datos Espaciales**. Esta Base de Datos es modelo de la realidad que se ajusta a un cierto fenómeno.

Una base de datos geográfica puede definirse como la colección lógica de IG (Información Geográfica) interrelacionada que se administra y almacena como una unidad, comprendiendo antecedentes sobre la ubicación de las características del mundo real y sobre sus particularidades en relación a su entorno.

Es importante indicar que para el caso del almacenamiento de la IG, se consideran las bases de datos espaciales y temáticas, en las que se almacenan de manera estructurada los objetos cartográficos (posición, tamaño y forma) y sus características no geométricas (atributos alfanuméricos) respectivamente. En algunos sistemas las dos bases de datos están separadas, mientras que en otros se integran en una sola [2].

**2.  Descripción del problema**

De acuerdo al Censo Nacional de Hogares y Vivienda realizado en el año 2010, Junín (B) arrojó una cifra de 87509 habitantes, de los cuales 78535 son los que por ley están posibilitados a votar de acuerdo al padrón publicado oficialmente en el año 2017 compuesto por 27 centros electorales.

La construcción del Padrón está regida por el Código Nacional Electoral. A continuación, se presentan algunos puntos de interés para este trabajo:

* División geográfica. Cada provincia como también la Ciudad Autónoma de Buenos Aires (CABA), conforma un Distrito. Estos, a su vez, están particionados en Secciones y estas se subdividen en Circuitos. La normativa establece que cada persona debe votar en el Circuito Electoral al que corresponde su domicilio legal.
* Mesas de votación. El Padrón de cada Circuito se distribuye en mesas de votación repartidas entre las escuelas. Cada mesa puede recibir un máximo establecido de votantes. El criterio por el cual se decide cuántas mesas poner en cada escuela depende de la capacidad de las infraestructura.
* Asignación de votantes a mesas. Una vez distribuidas y numeradas las mesas, el Padrón de cada Circuito Electoral es ordenado alfabéticamente y asignado en ese orden a las mesas; salvo que una disposición específica de alguna localidad fije algún otro criterio.

Al hacerse la distribución de votantes por orden alfabético, la asignación a escuelas es totalmente aleatoria [3]. Esto podría dar a pensar que se genera una desigualdad en la población que bajo ciertas circunstancias (personas de la tercera edad, personas con problemas de movilidad, personas con escasos recursos económicos), se encuentren impedidas de asistir al lugar del sufragio. En el caso particular de la ciudad de Junín, esta problemática se acentúa por la concentración de los centros electorales asignados mayoritariamente en la zona céntrica.

Para poder dar visibilidad al problema planteado se geo-codificarán los electores del padrón antes mencionado. Luego, se realizará una distribución del padrón en orden a las cercanías del domicilio declarado por cada elector respecto a los centros electorales. Como resultado se espera obtener de manera automática un padrón geo-gráficamente mejor distribuido lo que permitirá el fácil acceso de los votantes a los centros

**3.      Destinatario / Beneficiario**

En primera instancia los destinatarios del presente trabajo son la Junta Electoral y los Partidos Políticos.

En caso de implementarse la distribución del padrón resultante los beneficiarios serán los votantes de la ciudad de Junín.

**4.      Antecedentes**

Inicialmente se realizó una revisión de antecedentes respecto a trabajos realizados sobre re-empadronamiento de electores, focalizando en la identificación de la variable utilizada para la redistribución.

Un primer trabajo es el denominado “Asignación de votantes a centros de votación utilizando técnicas de Programación Matemáticas para minimizar el tiempo total invertido”, realizada por Mauro Giormenti Moravec en la Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales Departamento de Matemáticas, Argentina en el año 2014. Esta investigación tuvo como objetivo el estudio de estrategias alternativas a la designación del votante por orden alfabético, en busca de una asignación más inteligente, que disminuya el tiempo que les consume a los votantes la participación en los comicios.

Un segundo estudio corresponde al llamado “Asignación Eficiente de Votantes a Escuelas mediante Programación Matemática: El Caso de la Ciudad de Pergamino”, llevada a cabo por Jeremías Lenzi en el año 2013; en la misma se desarrolla un modelo lineal buscando minimizar el tiempo promedio invertido por los votantes.

Ambos trabajos utilizan el tiempo como variable de redistribución.

**5.      Justificación**

La elaboración de un padrón distribuido geográficamente con distancia óptima entre el elector y el centro de votación implica la aplicación de conocimientos sobre bases de datos, procesos de ETL, datos geoespaciales, entre otros.

Con esta distribución del padrón se espera facilitar el proceso de votación a los electores mediante la cercanía con el centro electoral.

En principio, se espera que el resultado del trabajo aporte un beneficio a la comunidad de la ciudad de Junín, y a futuro pueda extenderse y ponerse en práctica en otras localidades, ayudando así a que la mayor parte de los ciudadanos puedan realizar su sufragio de una manera más cómoda.

**6.  Diagnóstico**

La República Argentina es un país democrático desde el año 1983, los ciudadanos tienen el deber cada 2 años de participar de las elecciones las cuales están reguladas por el Código Electoral Nacional. El mismo especifica la forma en que estas elecciones deben ser organizadas, desde el armado de las boletas, el traslado de las urnas, las horas de veda, el armado de las listas de los centros electorales, como así también la distribución de los electores de acuerdo al orden alfabético de sus apellidos, dividido en la cantidad máxima de cupos de los centros.

Esta distribución de votantes por orden alfabético a las escuelas es totalmente aleatoria, lo que puede generar en algunos casos que la asignación del centro electoral no sea óptima.

**7.  Objetivos**

**Objetivos General**

* Automatizar el proceso de redistribución de un padrón electoral en base a la distancia óptima del domicilio del elector al lugar de sufragio asignado.

Objetivos Específicos:

* Dar visibilidad al problema de la distribución aleatoria de los electores mediante la geo-codificación del padrón electoral del circuito 444 de la VI Sección Electoral.
* Redistribuir el padrón electoral a partir de un algoritmo que asigne el centro electoral óptimo.

**8.      Metodología de trabajo**

El método que se va a utilizar para la elaboración del proyecto es el “proceso ETL”, ya que se requiere de este para el tratamiento de datos.

Para dar comienzo con el proceso ETL se requerirá que la fuente proveedora envié los padrones con la información de cada elector, abarcando en esté al elector propiamente dicho, con sus respectivos domicilios, entre otros campos; el formato que se estipulará para el envió del padrón es en Excel (.xls).

Una vez obtenido ya el padrón en el cual se irá a trabajar, se analizará cada uno de los campos, identificando que tipo y tamaño de datos poseen, para luego así poder definir y crear una tabla en la base de datos, cargando en está una vez ya creada, el padrón enviado en formato Excel, junto al refinamiento e identificación de los campos.

De la tabla ya obtenida se identificaran cuáles son los centros electores para poder reconocerlos y ser usados más adelante en el re-empadronamiento y obtener las distancias entre estos y los electores.

La depuración de los datos, el agregado de información, la obtención de las coordenadas es parte fundamental del proceso ETL que se llevara a cabo, debido a que esto es el pie para el cumplimiento de los objetivos mencionados, el procedimiento para ejecutar esto es el agregado de información al domicilio enviado, el tipo de información a adicionar hace que el domicilio sea más preciso, sumándole la ciudad, la provincia y el país. Contando con estos domicilios se procederá a la obtención vía web service las coordenadas geográficas de cada uno de los domicilios de los electores, procediendo a ser validados de acuerdo a la parte entera tanto de la latitud y longitud ya que cada ciudad se diferencia una de otra en este sentido.

Todas aquellas coordenadas geográficas que no pudieron ser validadas por algún motivo pasaran par un proceso de normalización de domicilios donde se contemplaran gran parte de los errores de tipeo humano, para posteriormente usar nuevamente el web service y la validación.

El proceso anterior se deberá repetir para los domicilios de los centros electorales.

Con la finalización de lo antes mencionado se podrá tener un padrón con un porcentaje alto de datos geográficos, creando y actualizando la base de datos geoespaciales que nos permitirá hacer cálculos geográficos que necesitaremos para obtener las distintas distancias y obtener la mínima.

Toda la información obtenida en los pasos anteriores será el insumo de un proceso de reasignación de centros electorales que será programado en el lenguaje Python, usando y adaptando el algoritmo de K-Means

~~Como origen al trabajo de desarrollo de la tesis se procede en un principio a dividirlo en dos partes, una de las partes es el desarrollo del modelo teórico y por otro lado es el desarrollo del trabajo de campo. En un principio se va a poner foco en este último para dar pie al desarrollo teórico que conlleva la redacción tanto de esta propuesta como la redacción de la misma tesis.~~

~~Para comenzar con todo el trabajo lo que se necesita son los últimos padrones en los cuales la información es accesible y pública, este padrón viene en dos Excel divididos por el tamaño, el cual está ordenado en orden alfabético y ascendente, con estos archivos en mano se proceden a ser cargados a una base de datos de PostgreSQL.~~

~~Con la obtención de una tabla base con todo el padrón, se procede a hacer una división de tablas, esa división consiste en una tabla para las escuelas, con sus correspondientes domicilios, y la tabla de los electores con todos sus datos correspondientes. Esta división se realizó para una posterior utilización de centralización de los electores de acuerdo a las escuelas.~~

~~A partir de las tablas cargadas en la base de datos lo que corresponde hacer es a partir de los domicilios de los electores que figuran en el padrón por medio de alguna herramienta es la obtención de las latitudes y longitudes de cada uno de los electores, en mi caso se decidió por la utilización de un plugins que se instala en y configura en el software libre en el cual se va a trabajar, que es Pentaho; este plugins solo necesita para su uso la configuración de parámetros de entrada y de salidas, las entradas de las cuales hace uso son la dirección a la cual se desea obtener las latitudes y longitudes, con la precisión de indicar en qué ciudad y provincia es esa dirección ya que pueden existir más de una dirección.~~

~~Las coordenadas geográficas pueden ser corroboradas de acuerdo a la información de cada una de las ciudades; la ciudad de Junín le corresponde una latitud con comienzo de -34 y una longitud de -60. Verificado esto e impactado en la tabla de nuestra base de datos, se procede a realizar una corrección de aquello domicilios en los cuales el plugins no pudo obtener sus coordenadas debido a que las direcciones están mal escritas, después de la corrección se procede a geo-codificarlas.~~

~~Una vez realizado todo lo anterior y con un porcentaje alto geocodificado se pasa a la siguiente fase que es la cargas de las bases geoespaciales que nos va a ser de utilidad más adelante para operar con este tipo de datos, calculando así la geometría de cada dato coordenado, referenciando a la curvatura de la tierra.~~

~~Como conclusión de los procedimientos anteriores se va a tener 2 bases de datos, una base de datos con datos planos y otra con datos geoespaciales para la realización de cálculos de geometría, cada una de estas cuenta con al menos 2 tablas, la de electores y la de escuelas con los campos que sean necesarios almacenar en estas. Ya con todo esto se va a tener un padrón casi a un 90% con domicilios estándares, geo-codificados, tanto el padrón como las escuelas van a ser nuestras entradas junto con otros datos como la cantidad máxima de electores por entidad, etc. de un algoritmo que se va a encargar de devolver todos aquello electores cercanos a estos centros, finalizando así con un padrón distribuido geográficamente en base a las distancia entre centro de elección y el votante.~~

**9.      Desarrollo Propuesto**

Como estudiante de una Universidad Pública en la cual se fomenta desde los principios el uso de Software libre con las ventajas que se tienen sobre estos, se decide a proceder a usar todo tipo de tecnologías de este estilo.

Para el manejo de datos, se utiliza como motor de base de datos PostgreSQL, usando como software de interacción pgAdmin versión 4.2; a su vez como se trabaja con datos geoespaciales se va a utilizar PostGIS versión 2.3.3 todos estos datos desde el lado de las base de datos.

Para la manipulación de flujos de datos, utilizar técnicas y herramientas de ETL, son las más convenientes, por lo cual existen varias herramientas de este estilo, tales como Talend, Hadoop, SAS Data Manager, Oracle Data Integrator, entre otras, pero para el desarrollo de este trabajo se escogió Pentaho Data Integration ETL Versión 7.1.0.0-12 que es el más estable hasta el momento en el cual se comenzó el desarrollo.

Además se utilizó para una visualización del padrón de una forma más real y a escala, con los mapas de la ciudad de Junín el software Qgis versión 2.18.13.

**10.  Cronograma**

**11.  Resultado Esperados**

Cuando se decidió llevar a cabo la propuesta y entrar más en labores de campo, permitiéndonos introducir ya en la visualización propia de los datos de origen, los resultados que se veían eran claros, esperando ver set de datos lo más estándar que fuera posible desde el lado de los domicilios de cada uno de los electores, con direcciones que sean fácil de entender y leer, capaces de ser tomadas en el plugin o en el mismo mapa para ser buscadas, se pretende que se armen set de datos en los cuales están agrupados de forma geográfica en base a un punto de referencia que va a ser los centros electores, y llevando esto a la visualización se puedan apreciar estos grupos armados.

Como se mencionó en un punto arriba se busca que el periodo de elección sea más confortables para cada uno de los electores, siendo rápido, sin necesidad de grandes movimientos a través de la ciudad para recurrir a estos centros electorales.

Y desde un punto más abarcativo y generar poder hacer un cambio y una actualización del actual sistema de empadronamiento a un sistema de empadronamiento más tecnológico e inteligente, pudiéndose llevar a cabo no solo a la ciudad de Junín sino ir en escalas a más de una ciudad; sin embargo todo esto depende de la aceptación en un gran porcentaje de la aceptación del pueblo.

12.  Bibliografía

[1]Fundamentos de bases de datos. Cuarta edición. Abraham Silberschatz.

[2]Fundamentos de las infraestructuras de datos espaciales (IDE). Miguel A. Bernabé-Poveda, Carlos M. Lopez-Vazquez. 2012.

[2] ¿Qué es la geocodificación? ArcMap.

<http://desktop.arcgis.com/es/arcmap/10.3/guide-books/geocoding/what-is-geocoding.htm>

[3]Asignación Eficiente de Votantes a Escuelas mediante Programación Matemática: El Caso de la Ciudad de Pergamino

<http://www2.famaf.unc.edu.ar/institucional/biblioteca/trabajos/601/16875.pdf>

[4]TESIS. Asignación de votantes a centros de votación utilizando técnicas de Programación Matemática para minimizar el tiempo total invertido

<http://cms.dm.uba.ar/academico/carreras/licenciatura/tesis/2014/Mauro_Giormenti_Moravec.pdf>

[5]Código electoral nacional

<http://www.bcnbib.gov.ar/uploads/Codigo-Electoral-Nacional-actualizado-mayo-2017.pdf>

Sistemas de información geográfica. Víctor Olaya 2014.

Agregar libro de Petraglia

Ver normas IEEE o APA para referenciar