

Taller 2: GMaps y Datos Abiertos

Juan Pablo Herrera (A00358465)

Jhon Edward Mora (A00355710)

Gabriel Cruz Libreros (A00056020)

Febrero 2020.

Universidad Icesi

Departamento de Tecnologías de Información y Comunicaciones (TIC)

Proyecto Integrador I

Fase 1

Identificación del problema

Descripción del contexto problemático (Causas y síntomas)

United States Travel Association (USTA) es una organización sin ánimo de lucro que representa la industria del comercio y turismo en Estados Unidos, generando más de 2.5 trillones de dólares anuales. La organización posee diversos programas, entre ellos uno con el objetivo de promover los viajes aéreos al ciudadano estadounidense común. Este programa ha estado activo por un tiempo, pero los resultados que ha mostrado no han sido los deseados. Por tanto, la Junta Directiva de USTA ha decidido reestructurar el programa, esperando cumplir con los resultados planeados desde un principio.

Como parte del proceso reestructural, la USTA ha decidido empezar un estudio en los aeropuertos de diversas ciudades y estados, con el fin de identificar los destinos más visitados por los viajeros. Sin embargo, debido a la gran cantidad de datos que fueron recolectados, la tarea de analizarlos y obtener conclusiones de estos se ha vuelto extremadamente compleja, impidiendo el objetivo inicial de este estudio. Una de estas tareas, la clasificación de los destinos más populares para los viajeros, es de hecho la más importante para esta organización, pues es en base a esta información que podrá determinar las decisiones a tomar para mejorar la promoción de viajes aéreos.

Con esto en mente, la USTA ha optado por solicitar una aplicación que pueda mostrar la información relevante de los estudios realizados. Esta aplicación deberá mostrar en un mapa los aeropuertos visitados, además de permitir ordenar los destinos de acuerdo a su popularidad y mostrar de manera gráfica dichos resultados en el mapa. Con esto se espera que

el análisis de los datos reunidos y la toma de decisiones de la organización respecto a estos se pueda realizar de la manera más eficaz y cómoda posible.

Identificación de necesidades

Las siguientes fueron las necesidades a solucionar identificadas:

- Se requiere leer los datos recolectados por la USTA.
- Se necesita ordenar los destinos por su popularidad.
- Se requiere mostrar todos los aeropuertos registrados en los datos en el mapa.
- Deberá ser posible filtrar dichos aeropuertos respecto a ciertos criterios:
 - Mostrar únicamente los aeropuertos más populares en el mapa.
 - Mostrar únicamente los aeropuertos menos populares en el mapa.

Definición del problema

La USTA, una organización representativa del comercio y turismo de Estados Unidos, ha recolectado datos de los diferentes aeropuertos y viajes con el fin de reestructurar su plan de promoción al turismo para el ciudadano estadounidense promedio. Sin embargo, como el volumen de datos que reunió con este fin es bastante alto, la realización de tareas como consultas o búsquedas es algo bastante ineficaz e impide el objetivo inicial del estudio. Es por esto que la organización requiere una aplicación capaz de procesar y analizar los datos ya reunidos para facilitar la toma de decisiones que ayuden a incentivar el turismo.

Fase 2

Recopilación de la información necesaria

Marco teórico

Las siguientes son definiciones que se consideraron importantes en el desarrollo de la solución para este problema:

- Bases de Datos

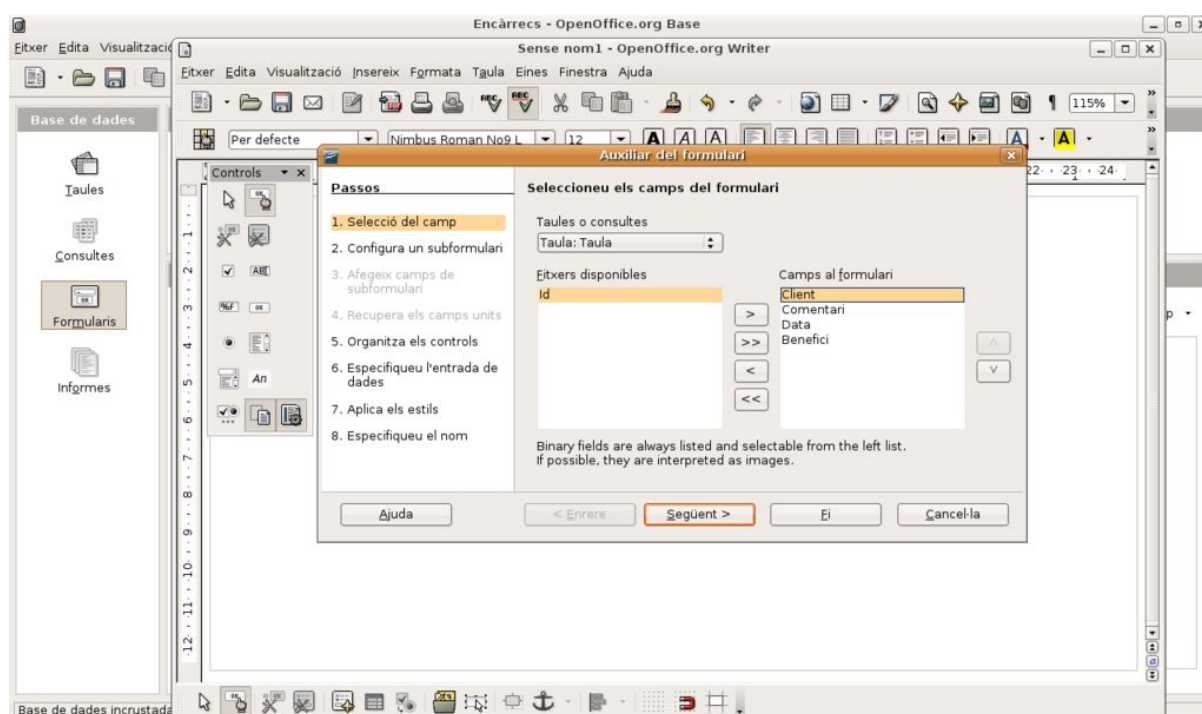
Se entiende por base de datos el conjunto de datos reunidos bajo un contexto específico que se almacenan de manera sistemática y organizada para su uso posterior en diversas aplicaciones.

Su concepto se origina desde tiempos antiguos, viniendo de las bibliotecas, en donde se almacenaban libros ordenados por sus contenidos. Con el tiempo y la llegada de los registros civiles de ciudadanos, la necesidad de almacenar y clasificar datos empezó a surgir, y las tarjetas perforadas empezaron a sistematizar esta tarea. Esto dio origen a la creación y almacenamiento de datos sistematizados. Finalmente, con la transición a la computación, las bases de datos pasaron a cumplir los mismos objetivos de las tarjetas perforadas respecto al manejo de datos, con lo que se encontró una manera más eficaz para solucionar estos inconvenientes.

Las bases de datos son clasificadas en diferentes tipos, dependiendo del contexto que manejen o la utilidad de la misma. Existen bases de datos continuas, estáticas, jerárquicas, de red, etcétera. Además, sobre estas se pueden realizar operaciones como consultas, búsquedas u organización respecto a un criterio. Sin embargo, debido al tamaño que éstas suelen tener, las operaciones antes mencionadas tienden a ser realizadas de manera más efectiva utilizando diversos programas de computadora. Estos programas reciben el nombre de Sistema de

Gestión de Base de Datos, y realizan estas tareas a través de interfaces de consulta conocidas como métodos de acceso.

En una base de datos, las entradas son toda información individual que se encuentre pertinente para registrar. Dichas entradas se clasifican por campos, en los que se guarda la información respectiva para cada entrada. Como ejemplo, en una base de datos que desee registrar información sobre los usuarios de cierto medio de transporte, se puede considerar una entrada cada uno de los usuarios que se registraron en la base de datos, y los campos de esta entrada serían su tiempo de uso, distancia de transporte, etc.



OpenOffice.org, un Sistema de Gestión de Bases de Datos

Para mayor practicidad, los programas de computación generan bases de datos a partir de la información que estos procesan y crean, mediante la creación de archivos de texto planos que registran las entradas como una línea de texto separada por campos, cuyo separador es un caracter especial determinado por el programa.

- Mapas satelitales

Con la introducción de la tecnología satelital, varios gobiernos empezaron a experimentar a impulsar los límites de esta tecnología. Con ella el desarrollo de las fotos satelitales, que fueron el inicio de la tecnología de mapas digitales que poseemos hoy en día. Los mapas satelitales consisten en varias capas de información para poder ser lo más precisos posible.

Las empresas proveedoras de los servicios de mapas utilizan distintos métodos de recolección de datos para poder generarlos. Empiezan tomando fotografías del mundo y países específicos para tener un boceto de la figura y la escenografía del país. Luego, a partir de cámaras montadas en carros, se recorren las calles de la ciudad o provincia con el fin de recolectar información sobre las calles y su infraestructura. Con esta información se crea un mapa con vías y carreteras, al igual que algunos edificios o infraestructura notable. Por último, con ayuda de usuarios y el GPS de los celulares móviles, se llenan datos más pequeños, así como información específica de lugares menos comunes. Toda esta información permite dar una visión precisa y actual de la geografía del mundo.

Estado del arte

Diversas soluciones se han dado antes para este tipo de problemas. Las siguientes son algunas de estas:

- **Google Maps**

Es el servicio de mapas satelitales perteneciente a Google. Tiene una funcionalidad que le permite a los usuarios buscar aeropuertos cercanos, al igual que clasificarlos por cercanía, popularidad y otros criterios.

- **FlightRadar24**

El sitio web permite ver el tráfico actual de más de 180,000 vuelos a nivel mundial, mostrando la trayectoria de los vuelos, resaltando su aeropuerto inicial y de llegada.

- **Trip Advisor**

La famosa aplicación de viajes tiene una función llamada “Mi Mapa”, la cual le permite al usuario seleccionar puntos de interés en un mapa y crear la ruta más deseable para recorrer los lugares.

Fase 3

Búsqueda de soluciones creativas

Se llegaron a las siguientes propuestas a partir del método de Lluvia de Ideas:

- Desarrollar una hoja de cálculo programada que permita el manejo y procesamiento de la base de datos.
- Implementar una estructura de datos que permita organizar las entradas en la base de datos respecto a sus campos, de manera ordenada y eficaz. Después, con los órdenes ya establecidos, utilizar librerías externas que permitan la visualización de dicha base de datos.
- Utilizar un lenguaje de programación estadístico especializado en bases de datos para generar el análisis y mostrar los resultados mediante gráficas y tablas.
- Desarrollar una aplicación que realice un conteo individual en cada una de las entradas, permitiendo así encontrar los destinos más y menos populares para los viajeros.
- Realizar una aplicación que seleccione destinos al azar y los muestre como los destinos más y menos populares.
- Utilizar algoritmos de ordenamiento aplicados sobre la base de datos y realizar un conteo de entradas iguales y consecutivas sobre los campos de destino, encontrando los destinos más y menos populares mediante un conteo simple.

Fase 4

Transformación de las ideas a los diseños preliminares

En este apartado se realiza un análisis sobre las ideas planteadas en el anterior.

Alternativa 1 - Hoja de cálculo

Las hojas de cálculo permiten operaciones de búsqueda y ordenamiento de datos bajo distintos criterios, además de manejar bases de datos en texto plano como una tabla de datos, en las que cada fila representa una entrada y las columnas representan los campos de información. Sin embargo, a la hora de mostrar los resultados buscados de manera gráfica, representados en un mapa, las soluciones se verían obligadas a buscar un proveedor de mapas externo que muestre las ubicaciones encontradas en las búsquedas, lo cual no resultaría cómodo para las investigaciones.

Alternativa 2 - Estructura de Datos

En esta alternativa se busca implementar una estructura de datos que permita organizar la información respecto a los campos que la compongan. Se pensó en una estructura similar a un mapa de datos, en la que cada entrada del mapa corresponde a un valor de un campo de la base de datos. Después, bastaría con encontrar la entrada del mapa de mayor y menor cantidad de elementos almacenados, con lo que se encontrarán los destinos más y menos populares de los viajeros.

Alternativa 3 - Lenguaje de Programación Estadístico

R es un lenguaje y un entorno de programación enfocado a la estadística. En este lenguaje de programación se encuentran diversas funciones y operaciones que permitirían realizar tareas como encontrar el valor más y menos repetido de un conjunto de datos de manera eficaz, pues el diseño completo del lenguaje de programación es para este tipo de tareas.

Alternativa 4 - Conteo Individual de Datos

Esta alternativa es una de las más sencillas, pues consiste en contar individualmente cada una de las entradas iguales en el campo de destino dentro de la base de datos. Sin embargo, debido al volumen de esta, una tarea como contar individualmente cada una de las entradas resulta ineficaz e incómoda debido al tiempo de procesamiento que conlleva. Es por esta razón que esta alternativa es descartada como una solución para este problema.

Alternativa 5 - Selección de Destinos al Azar

Bastaría con leer dos datos al azar en el campo de destino del vuelo, los cuales serían mostrados como el destino más y menos popular del viajero promedio. Sin embargo, debido a que los datos que requiere la organización son importantes para su toma de decisiones, un dato al azar podría llevar al fracaso del programa en su reestructuración. Por esta y otras razones que resultan del sentido común, esta alternativa es descartada como una solución para este problema.

Alternativa 6 - Algoritmos de Ordenamiento

Aplicar algoritmos de ordenamiento sobre los datos haría que el conteo de entradas sea bastante sencillo. Sin embargo, debido a que los algoritmos de ordenamiento requieren de un tiempo de cómputo bastante elevado y el volumen de datos que se maneja es bastante alto, la solución no resultaría cómoda y eficaz para las tareas requeridas. Es por esto que esta alternativa es descartada como una solución para este problema

Fase 5

Evaluación de la mejor solución

En la siguiente sección se plantean criterios importantes para la selección de la solución más apropiada:

Alcance de la solución

Este criterio será evaluado en una escala desde 0 hasta 5, en la que un 0 será otorgado cuando la solución no pueda abarcar el volumen de datos reunidos en el problema. Una calificación entre 0 (excluyente) y 2.5 (incluyente) indican que la solución maneja con problemas el volumen de datos reunidos. Calificaciones superiores a 2.5 y menores a 5 serán dadas a soluciones que puedan manejar la totalidad de datos reunidos con problemas leves; y la calificación de 5 será dada a soluciones que puedan fácilmente manejar todos los datos necesarios para resolver el problema dado.

Eficiencia de la solución

Este criterio será evaluado con una escala desde 0 hasta 5, en la que un 0 será otorgado cuando la solución no pueda resolver el problema en lo absoluto. Una calificación entre 0 y 2.5 indican una solución que resuelve el problema entregando datos de manera desorganizada. Calificaciones superiores a 2.5 y menores a 5 serán dadas a soluciones que entreguen información puntual y clara y la calificación máxima será otorgada a una solución que pueda entregar información puntual y clara en un contexto que facilite la toma de decisiones para la organización.

Facilidad de implementación

Este criterio será evaluado con una escala desde 0 hasta 5, en la que un 0 será otorgado cuando la solución requiere un manejo de herramientas que se encuentre totalmente por fuera

de nuestra área de experiencia. Calificaciones entre 0 y 2.5 indican que la solución requiere el uso de herramientas que se encuentran por fuera de nuestra área de experiencia pero que requerirían un esfuerzo medio para poder ser manejadas. Calificaciones superiores a 2.5 y menores a 5 indican que la solución requiere del uso de herramientas que no se encuentran dentro del área de experiencia, pero que su curva de aprendizaje es nivelada y de baja inclinación o que puede ser fácilmente homologadas con conocimientos ya adquiridos. Una calificación de 5 se otorgará a las soluciones que requieran el uso de herramientas ya conocidas y manejadas por el grupo de trabajo.

La siguiente tabla resume los resultados de la evaluación.

Tabla 1

Rúbrica de evaluación de las soluciones:

Solución	Criterio de Alcance de la Solución (40%)	Criterio de Eficiencia de la Solución (20%)	Criterio de Facilidad de Implementación (40%)	Total (100%)
Hoja de Cálculo	4	3	2	3
Estructura de Datos	4.5	4.5	5	4,7
Lenguaje de Programación Estadístico	5	5	0	3

Como se observa, la solución elegida a realizarse será la implementación de una Estructura de Datos capaz de analizar los grupos de destinos en las entradas de la base de datos, pues es la que mejor calificación obtuvo bajo el análisis de los criterios estipulados.

Síntesis reflexiva

Inicialmente se planteó un problema en el que una compañía generó cantidades grandes de datos pero que no planeó con antelación la manera en la que estos podrían ser usados. Debido a esto, la información que sería necesitada de esos datos no ha sido posible de usar para su propósito inicial, el cual es encontrar los destinos más y menos populares para los viajeros. Es de ahí que surge la necesidad de una herramienta que permita manejar estos volúmenes de datos, permitiendo conocer datos puntuales e información importante de la ya obtenida. Investigaciones sobre bases de datos y mapas satelitales se realizaron, lo que permitió entender diversas características respecto a sus formas y funcionamientos, además de revisiones sobre trabajos anteriores que han solucionado ya este tipo de problemáticas. A la hora de abordar la fase de generación de ideas, se encontró que ya han existido diversas soluciones a este problema, lo que produjo dificultades para realizar ideas prometedoras. Sin embargo, se lograron producir ideas que resultaban prometedoras junto a otras que no aparentaban serlo tanto. Lo anterior nos llevó a los diseños previos de la solución, en la que se descartaron algunas ideas y se desarrollaron otras. Finalmente, se estipulan algunos criterios que ayudaron a determinar la mejor solución entre las ideas obtenidas, dándole prioridades a soluciones que puedan manejar la totalidad de datos y que fueran fácilmente implementables. Es así como se llegó a una solución que implementa una estructura de datos conocida, mostrando los resultados de manera gráfica en un mapa soportado por GMaps.

Lista de referencias

- What is a database?. (2020). Consultado el 4 de Febrero 2020, de <https://www.oracle.com/database/what-is-database.html>
- D. Foote, K. (2017). A Brief History of Database Management - DATAVERSITY. Consultado el 4 de Febrero 2020, de <https://www.dataversity.net/brief-history-database-management/>
- Oostenrijk, A. (2016). GMap.NET Beginners Tutorial: Maps markers, polygons and routes (updated for VS2015 and GMap.NET 1.7). Consultado el 4 de Febrero 2020, de <http://www.independent-software.com/gmap-net-beginners-tutorial-maps-markers-polygons-routes-updated-for-vs2015-and-gmap1-7.html>
- C# GMap | Geocoding and Reverse Geocoding | GMap - XIV. (2020). [Video]. India. Consultado el 4 de Febrero 2020, de <https://www.youtube.com/watch?v=07ypcnNhZ-k&t>
- Nightingale, R. (2017). How Does Google Maps Work?. Consultado el 4 de Febrero 2020, de <https://www.makeuseof.com/tag/technology-explained-google-maps-work/>