

Luis Fernando Muñoz A00046396

Steven Ma Mei A00130572

Daniel Galvis Torres A00130540

Sebastian Arango A00130532

Fase 1: Identificación del problema

Síntomas y necesidades

- Se necesita obtener información de los grupos de investigación de diferentes ciudades de Colombia obtenidos a partir de la página oficial de datos abiertos del gobierno de Colombia
- Guardar la información de los grupos para su posterior uso
- Debe poder consultarse y generar reportes, preferentemente gráficos, para consolidar la información por regiones, ciudades, áreas de investigación y clasificación
- Se requiere de un mapa que muestre los grupos de manera aleatoria dentro de la ciudad a la que pertenezca
- Es necesario conocer los artículos de desarrollo más usados para las investigaciones en los diferentes grupos

Definición del problema

Para mejorar la capacidad competitiva del país, la CTI en Colombia desea un programa para gestionar la información de los grupos de investigación reconocidos de tal forma en que se pueda visualizar su ubicación geográfica, generar reportes de clasificación y además, desea conocer los artículos para el desarrollo más frecuentados. Para llevar a cabo estas solicitudes, se elaboraron los siguientes requerimientos:

Listado de requerimientos

Requerimientos funcionales

Nombre	R1 – Registrar grupos de investigación.
Resumen	La aplicación debe permitir el ingreso de un nuevo grupo de investigación
Entradas	
Información del grupo a ingresar	
Resultados	
Un nuevo grupo de investigación se ha añadido a la lista	

Nombre	R2 – Permitir la visualización de los datos de los grupos de investigación.
Resumen	Por medio del nombre del grupo, se mostrarán los datos referentes a él

Entradas
Nombre del grupo a buscar
Resultados
Un reporte con los datos del grupo buscado

Nombre	R3 – Actualizar la información de los grupos de investigación.
Resumen	Permitir la modificación de datos de un grupo dado su nombre
Entradas	
Grupo de investigación	
Resultados	
Grupo de investigación actualizado con cambios en los datos	

Nombre	R4 – Filtrar grupos de investigación.
Resumen	R4 – Permitir el filtro de grupos de investigación por áreas de investigación, clasificación, nombre, y fecha de fundación. Los grupos filtrados deben ser visualizados mostrando su ubicación en un mapa.
Entradas	
Región, ciudad, área de investigación o clasificación	
Resultados	
Todos los grupos de investigación correspondientes al filtro de entrada	

Nombre	R5 – Generar reportes estadísticos de la información contenida
--------	--

Resumen	Permitir la generación de reportes estadísticos de la información contenida, a través de gráficos y tablas.
Entradas	
Null	
Resultados	
Reporte estadístico generado gráficamente	

Nombre	R6 – Consultar los artículos más referenciados por los grupos de investigación.
Resumen	Debe mostrarse cuál es el elemento más frecuente en materia de investigación
Entradas	
Null	
Resultados	
Se muestra como mensaje el/los artículos más frecuentados en las investigaciones	

Requerimientos no funcionales

-Los archivos deben de ser persistentes.

Fase 2: Recopilación de información

Definiciones

- **Gmaps.Net:** Es una librería que permite utilizar google maps en Visual Studio
- **Coordenadas:** “Par de magnitudes (latitud y longitud) que sirven para determinar la posición de un punto en la superficie de la Tierra.”¹
- **Grupo de investigación:** “Conjunto de investigadores con una trayectoria común (habitualmente en torno a un jefe de grupo) determinada por un registro de publicaciones conjuntas y por el acceso a fondos de financiación de la investigación.”²

¹ Diccionario de Google

² <http://www.iisgm.com/investigacion/definicion-grupos-investigacion/>

- **Visual Studio:** "Microsoft Visual Studio es un entorno de desarrollo integrado (IDE, por sus siglas en inglés) para sistemas operativos Windows. Soporta múltiples lenguajes de programación, tales como C++..."³
- **Data mining:** "La minería de datos o exploración de datos (es la etapa de análisis de "Knowledge Discovery in Databases" o KDD) es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos."⁴
- **Clustering (data mining):** "Agrupa un conjunto de observaciones en un número dado de clusters o grupos, está basado en la idea de similitud de los grupos."⁵
- **Clasificación (data mining):** Se utiliza para tratar de predecir a qué tipo de clasificación cierto objeto pertenece⁶
- **Asociación (data mining):** Se utiliza para ver la correlación que tienen 2 objetos
- **Reporte:** "Noticia o informe acerca de un hecho o suceso reciente."⁷
- **Artículo científico:** "Un artículo científico es un Informe original, escrito y publicado, que plantea y describe resultados experimentales, nuevos conocimientos o experiencias que se basan en hechos conocidos."⁸
- **Estadística:** "Ciencia que utiliza conjuntos de datos numéricos para obtener, a partir de ellos, inferencias basadas en el cálculo de probabilidades."⁹

Fase 3: Búsqueda de soluciones creativas

Todas las ideas se generaron mediante "BRAINSTORM"

Ideas para la creación de la aplicación:

- (1) Crear una aplicación que permita almacenar y procesar la información de los grupos de investigación. A parte de eso, que la aplicación se vincule con la aplicación web de google maps y que de esta manera cada vez que se quiera obtener la ubicación de un grupo el programa abra Google Maps web
- (2) Crear la aplicación que procese y almacene la información y que cada grupo tenga una imagen de un mapa mostrando la ubicación del grupo
- (3) La aplicación procesa la información y la almacena. Para mostrar la ubicación en el mapa se usará la librería de Google Maps: Gmaps.
- (4) Crear una aplicación que procese y almacene la información y que el grupo haga el diseño de los mapas para la ubicación de los centros de investigación

Ideas para el procesamiento de la información:

Las ideas se harán con base a Data Mining. Las técnicas que hay en el data mining y que se pueden usar para el desarrollo de la aplicación son las siguientes:

- (1) Clasificación de los grupos de investigación por características de más alto nivel.

³ https://es.wikipedia.org/wiki/Microsoft_Visual_Studio

⁴ https://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos

⁵ <http://www.evaluandosoftware.com/tecnicas-data-mining/>

⁶ <https://www.datasciencecentral.com/profiles/blogs/the-7-most-important-data-mining-techniques>

⁷ Diccionario de Google

⁸ <https://www.hotcourseslatinoamerica.com/study-abroad-info/choosing-a-university/como-escribir-un-articulo-cientifico-o-paper-para-la-universidad/>

⁹ Diccionario de Google

(2)Asociación de los grupos de investigación de acuerdo a sus características

(3)Clustering agrupando grupos de investigación de acuerdo a una función de distancia.

Ideas para definir la ubicación aleatoria de los centros de investigación dentro de la ciudad:

(1)Buscar manualmente en Google Maps las coordenadas que rodean las ciudades que están en el reporte y generar coordenadas aleatorias que estén dentro de la zona determinada para ubicar los centros de investigación en sus respectivas ciudades.

(2)Para cada ciudad, tomar por centro la coordenada donde ubica automáticamente el mapa al buscar la ciudad, definir a partir de qué decimal se hará aleatorio la latitud y la longitud. Será un número de decimales definido para las ciudades grandes y otro para las ciudades pequeñas

(3)Para cada ciudad, tomar por centro la coordenada donde ubica automáticamente el mapa al buscar la ciudad, definir a partir de qué decimal se hará aleatorio la latitud y la longitud. El número de decimales aleatorios dependerá del total de centros de investigación que haya en dicha ciudad.

(4)Para cada ciudad, tomar por centro la coordenada donde ubica automáticamente el mapa al buscar la ciudad, definir a partir de qué decimal se hará aleatorio la latitud y la longitud. El número de decimales aleatorios será el mismo para todas las ciudades.

(5)Buscar alguna forma que la aplicación de GMaps retorne una lista (o algo similar) con los puntos pertenecientes a la ciudad, sacando de dicha lista los puntos pertenecientes a la ciudad.

Fase 4: Transición de la formulación de ideas a los diseños preliminares

Ideas descartadas:

Para la creación de la aplicación:

- Se descarta la idea de crear el diseño de los mapas (4), ya que el grupo no está capacitado para cumplir dicha tarea, y es mucho más fácil, óptimo y eficiente utilizar las bibliotecas ya implementadas para manejo de mapas (Como GMaps).

Para definir la ubicación aleatoria de los centros de investigación:

-Se descarta la idea de buscar manualmente las coordenadas en Google Maps (1), ya que en caso de que sean muchas ciudades, sería una tarea bastante dispendiosa, la cual podrían buscarse formas de automatizar procesos similares. Además, si en algún momento se agrega una nueva ciudad, este método se complicaría el proceso.

-Se descarta la idea de buscar alguna forma de que el GMaps retorne una lista con los puntos pertenecientes a la ciudad (5), ya que tras una ardua investigación no se encontró alguna forma de solucionar esto. Seguramente la hay, y esta idea sería la más precisa, pero debido a que es

un prototipo y el tiempo de entrega es corto, la gran precisión que ofrecería este método no compensa su complejidad de implementación.

Para el procesamiento de la información:

-Se descarta la idea de hacerle clustering de alguna forma a los grupos de investigación, ya que se componen de datos que no permiten ser analizados fácilmente por algoritmos de clusterización, y no se obtendrá información valiosa a partir de ello.

Diseños preliminares:

Para la creación de la aplicación:

(1): La aplicación simplemente se encargará de mostrar la información principal de cada uno de los centros de investigación, no mostrará directamente su ubicación. Más bien, habrá un botón para cada uno de los centros de investigación que, con ayuda de bibliotecas, se abrirá el navegador web predeterminado del usuario con Google Maps, mostrando la ubicación del centro de investigación especificado.

(2): La aplicación simplemente se encargará de mostrar la información principal de cada uno de los centros de investigación, no mostrará directamente su ubicación. Más bien, habrá un botón para cada uno de los centros de investigación, el cual abrirá una imagen (archivo .jpg) de una sección de un mapa mostrando la ubicación del centro de investigación seleccionado.

(3): La aplicación se encargará de mostrar la información principal de cada uno de los centros de investigación. Además de esto, se tendrá un mapa interactivo con la ayuda de GMaps, en la cual aparecerá la ubicación de cada uno de los centros de investigación.

Para definir la ubicación aleatoria de los centros de investigación:

(2): Se definirá un arreglo constante con las principales ciudades de Colombia (Bogotá, Medellín, Cali, Barranquilla). Las ciudades que estén dentro de este arreglo, se definirá un número de decimales a partir del cual la latitud y longitud serán aleatorias, teniendo como centro el punto predeterminado por GMaps al ubicar la ciudad. Los centros de investigación estarán dentro de dicho círculo. Para las ciudades que no estén dentro del arreglo, se el proceso es el mismo pero con un radio número de decimales diferente. Para definir el número de decimales aleatorios, se debe calibrar a través de varios intentos para definir los más óptimos.

(3): El proceso es similar al del punto (2). La diferencia es que el número de decimales aleatorios será definido por el total de centros de investigación que tenga la ciudad (Asumiendo que entre más centros de investigación tenga una ciudad, por lo general es porque es más grande). El número de decimales debe ser calculado a partir de una función que crezca rápido al inicio y después de forma lenta (por ejemplo $k * e^{(-x*c)}$ o $k * \ln(x*c)$), siendo x el total de centros de investigación de la ciudad. Las constantes k y c se definen a partir de intentar varias opciones para ver cuál es la más óptima.

(4): La idea de los decimales aleatorios en el que se ubicarán los centros de investigación en las ciudades es la misma que para los anteriores puntos, con la diferencia que el radio el número de decimales aleatorios será el mismo para todas las ciudades. El número se deberá determinar calibrando con varias opciones para determinar cuál es la más óptima.

Para el procesamiento de la información:

(1):La idea de clasificación de grupos de investigación puede llevarse a cabo analizando con mucho más detalle sus características y cómo se relacionan estas con otros grupos de investigación, dando lugar a nuevas clasificaciones de más alto nivel.

(2):La idea de Asociación entre grupos de investigación se puede desarrollar aplicando el concepto de probabilidad condicional, encontrando la probabilidad de encontrar un grupo de investigación con una determinada característica en grupos de investigación cumpliendo con otra característica.

Fase 5: Evaluación y selección de la mejor solución

Criterios de evaluación para el diseño general de la aplicación

Criterio A.La solución es respecto dificultad/beneficio:

[3]La dificultad justifica el beneficio que se obtendrá en la aplicación.

[2]La dificultad no justifica la solución propuesta en su totalidad, puesto que tiene complejidad innecesaria para los propósitos

[1]La dificultad no justifica absolutamente nada el beneficio que se puede obtener.

Criterio B. La solución si cumple con los objetivos y requerimientos planteados:

[3] Cumple con los requerimientos y objetivos

[2] Cumple a medias.

[1] No cumple para nada.

Criterio C. El nivel de aprendizaje por implementar la solución es:

[3]Alto

[2]Intermedio

[1]Bajo

Criterio D. La solución es amigable con el usuario en sentido de que la interfaz es buena, los mapas permiten interactividad, etc..:

[2] La solución es amigable con el usuario.

[1] La solución no es amigable con el usuario.

Soluciones	Criterio A	Criterio B	Criterio C	Criterio D	Sumatoria
(1)	2	1	2	1	7
(2)	2	2	3	1	8
(3)	3	3	3	2	11

Criterios de evaluación para las ideas de procesamiento de la información:

Criterio A: La idea no se ve tan forzada para dar solución:

[2] La idea perfectamente encaja con el objetivo

[1] La idea no encaja muy bien con el objetivo

Criterio B: La idea aporta un muy buen aprendizaje:

[3]Alto

[2]Intermedio

[1]Bajo

Criterio C: La idea es alcanzable durante el tiempo estipulado para el desarrollo

[2] Si

[1] No

Idea	Criterio A	Criterio B	Criterio C	Suma
(1)	1	3	1	5
(2)	2	2	2	6

Criterios de evaluación para la ubicación de los centros de investigación

Criterio A: La idea es sencilla de implementar:

[2]: Se implementa muy fácilmente.

[1]: Requiere cierta complejidad para implementarse.

Criterio B: La idea aporta aprendizaje para manejar GMaps:

[3]: Aporta un buen aprendizaje en la herramienta y en el uso de C#.

[2]: Aporta cierta aprendizaje en la herramienta y en el uso de C#.

[1]: Aporta poco o nulo aprendizaje en la herramienta y en el uso de C#.

Criterio C: La solución es precisa para definir el área donde se pondrán los puntos aleatorios:

[3]: La solución tiene buena precisión para que el área(definida por los decimales aleatorios) no quede muy pequeña o muy grande con respecto a la ciudad.

[2]: La solución es medianamente precisa con el tamaño del área respecto a la ciudad.

[1]: La solución es muy poco precisa para determinar el tamaño del área en la ciudad: Bastantes zonas serán muy pequeñas para su ciudad, mientras otras serán muy grandes.

Criterio D: La solución es realizable dentro de los plazos previstos:

[2]: La solución no toma mucho tiempo de implementar, no tendrá mayor complicación para cumplirse dentro de los plazos.

[1]: La solución se puede cumplir dentro de los plazos, aunque habrá que dedicarle gran parte del tiempo.

[1] Para las ciudades grandes, la solución podría ubicar demasiados centros de investigación por fuera de los límites de la ciudad, en caso que la coordenada definida por el GMap al buscar la ciudad no esté centrado sino que, por el contrario, esté muy recostado hacia un extremo de la ciudad.

Soluciones	Criterio A	Criterio B	Criterio C	Criterio D	Criterio E	Sumatoria
(2)	1	3	2	2	1	9
(3)	1	3	3	1	1	9
(4)	2	2	1	2	3	10



Diagrama de Objetos

