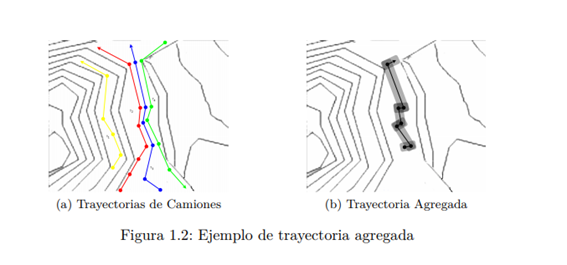
**Guayaquil, 6 de abril de 2018**

**INFORME BÚSQUEDA DINÁMICA DE PATRONES SOBRE TRAYECTORIAS**

El presente informe tiene como objetivo dar a conocer los puntos más importantes de la tesis denominada Búsqueda Dinámica de Patrones sobre Trayectorias, la misma que presenta un nuevo concepto de trayectoria agregada, la misma que muestra algoritmos para obtenerlas de manera dinámica. Una trayectoria agregada no es más que una nueva trayectoria que se obtiene a partir de un conjunto de trayectorias que siguen un patrón similar de comportamiento, y no solo muestra el patrón de comportamiento, sino que además nos indica la densidad de ese patrón.

Un ejemplo que encontramos es que una empresa de extracción de minerales cuya planificación de rutas (trayectorias) de extracción se realiza de manera dinámica. La información de camiones y otros vehículos de extracción se obtiene de manera continua lo que permite generar nuevas rutas de extracción. Por ejemplo, la Figura 1.2(a) muestra cuatro trayectorias pertenecientes a camiones en faena de extracción. Podemos entonces, generar una trayectoria agregada en base a las trayectorias de extracción que siguen el mismo patrón de comportamiento, lo que podría entregar información de algún área que se encuentra saturada por camiones.



Una Trayectoria Agregada (AT) representa entonces, la agregación de trayectorias en base a patrones de comportamiento con una densidad relevante. Para lograr obtenerla es necesario cumplir con dos fases. Primero es la Obtención de clústeres de segmentos de trayectorias diferentes y segundo es la Agregación de clústeres similares.

Es muy importante definir una medida de similitud para comparar dos trayectorias, puesto que el principal problema cuando agregamos trayectorias es la imprecisión de datos espacio-temporales.

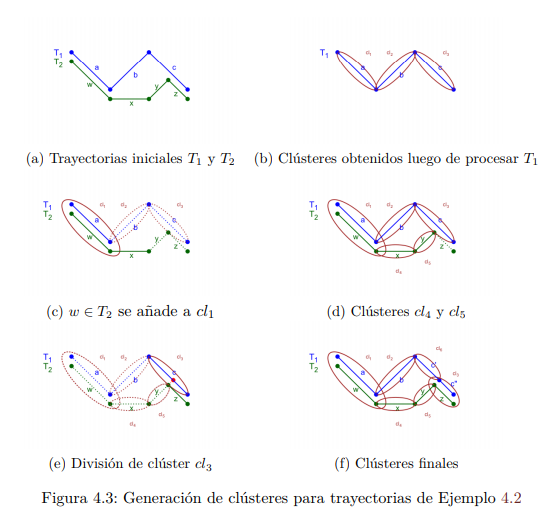
Para efectos de este trabajo, se utiliza la medida distancia euclidiana.

Durante la revisión de la tesis también se pudo detectar las relaciones entre segmentos:

Intuitivamente, dos segmentos S1 y S2 se relacionan totalmente cuando tanto sus puntos iniciales como sus puntos finales son cercanos, espacial y temporalmente, y la dirección de ambos segmentos es semejante, es decir, el ángulo formado entre ellos es menor a 90º.

Sin embargo, dos segmentos se relacionan parcialmente cuando, sin estar relacionados totalmente y teniendo la misma dirección, un extremo de un segmento es cercano, espacial y temporalmente, a algún punto (no necesariamente extremo) del otro.

Por otra parte, en el área de las bases de datos espacio-temporales, regularmente se utiliza el término clustering como el proceso de agrupar conjuntos de objetos espaciales (y/o temporales) que tienen características comunes y/o comportamientos similares, los cuales se conocen como clústeres (clusters).

En esta investigación, se utilizó el concepto de clúster de segmentos de trayectorias para referirse a un conjunto de segmentos SS = S1, S2, …, Sn de diferentes trayectorias (no necesariamente consecutivos) que se relacional de manera total.

Explicación: Considerando el conjunto de trayectorias de la Figura 4.3(a) compuesto de trayectorias {T1, T2}, se asume que la primera trayectoria a procesar es T1, se procede como sigue:

1. Cada uno de los segmentos de T1 da origen a un clúster, generándose, los clústeres cl1, cl2 y cl3 en Figura 4.3(b).
2. Luego se considera la trayectoria T2. Para cada uno de los segmentos de T2 se analiza la relación de estos con segmentos de los clústeres ya computados, es decir, se determina si son admisibles de ser añadidos a un clúster previamente computado.

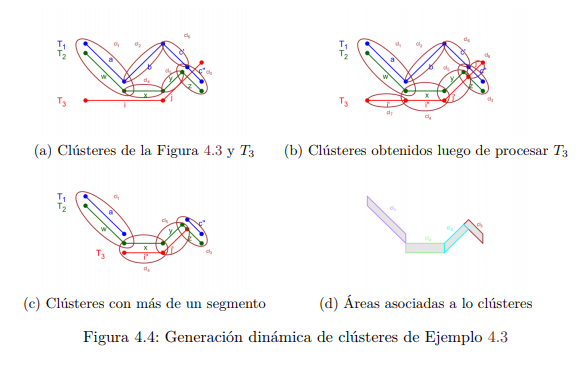
Consideremos el segmento w 2 T2. Se compara w con el clúster cl1, es decir, se determina el tipo de relación (total o parcial) entre w y el único segmento a de cl1 (4.3(b)). Como a y w se relacionan totalmente, se procede a añadir el segmento w al clúster cl1 tal como lo muestra la Figura 4.3(c).

1. Luego, se considera como segmento entrante al segmento x 2 T2. En primer lugar, se compara con cl1 4.3(c), pero como cl1 ya contiene un segmento perteneciente a la trayectoria T2, se continua con el siguiente clúster. Dado que x no se relaciona de manera total ni parcial con los segmentos en los clústeres cl2 y cl3, por lo tanto, se genera un nuevo clúster cl4 conformado por el segmento x, el cual se ilustra en la Figura 4.3(d).
2. Luego, se considera el segmento y 2 T2 y se compara con el clúster cl1. Dado que cl1 ya contiene un segmento de la trayectoria T2 no es posible añadirlo a él (lo mismo sucede cuando se compara con el clúster cl4). Este segmento no se añade a ninguno de los restantes clústeres dado que no se relaciona ni total ni parcialmente con los segmentos en ellos. Como y no puede ser añadido a ningún clúster, se crea un nuevo clúster cl5 conformado por el segmento y, el cual se ilustra en la Figura 4.3(d).
3. Finalmente se procesa el segmento z 2 T2, el cual no puede ser añadido al clúster cl1 puesto que éste ya contiene un segmento de la trayectoria T2. Este segmento tampoco puede ser añadido al clúster cl2 ya que z no se relaciona ni total ni parcial con el segmento b del clúster cl2. Sin embargo, z se relaciona parcialmente con el segmento c del clúster cl3. La forma de relación corresponde a la forma 2 de la Definición 4.7. Por lo tanto, se divide el clúster cl3 en dos. Se obtiene el clúster cl6 que contiene el nuevo segmento c0 (sub-segmento de c obtenido desde STARTp(c) hasta la proyección de STARTp(z) en c) y se actualiza el clúster cl3 con el sub-segmento c00 que se obtiene computando la proyección de STARTp(z) hasta ENDp(c), esto se ilustra en la Figura 4.3(e). El segmento z se añade al nuevo clúster cl3 como se muestra en la Figura 4.3(f).

Además, para que se puedan formar un conjunto de clústeres de segmentos de trayectorias deben compartir las siguientes características:

* Cada clúster de segmentos dentro de un clúster agregado debe contener por lo menos dos segmentos.
* El área asociada a cada clúster de segmentos incluido en un clúster agregado intersecta al menos un punto el área (o áreas) de otro clúster (otros clústeres) de segmento(s) en el mismo clúster agregado.
* Los clústeres de segmentos dentro del clúster agregado cumplen con un nivel de compatibilidad mayor a un umbral ∆J definido. La compatibilidad corresponde a un porcentaje de trayectorias en común entre los clústeres de segmentos dentro de un clúster agregado.

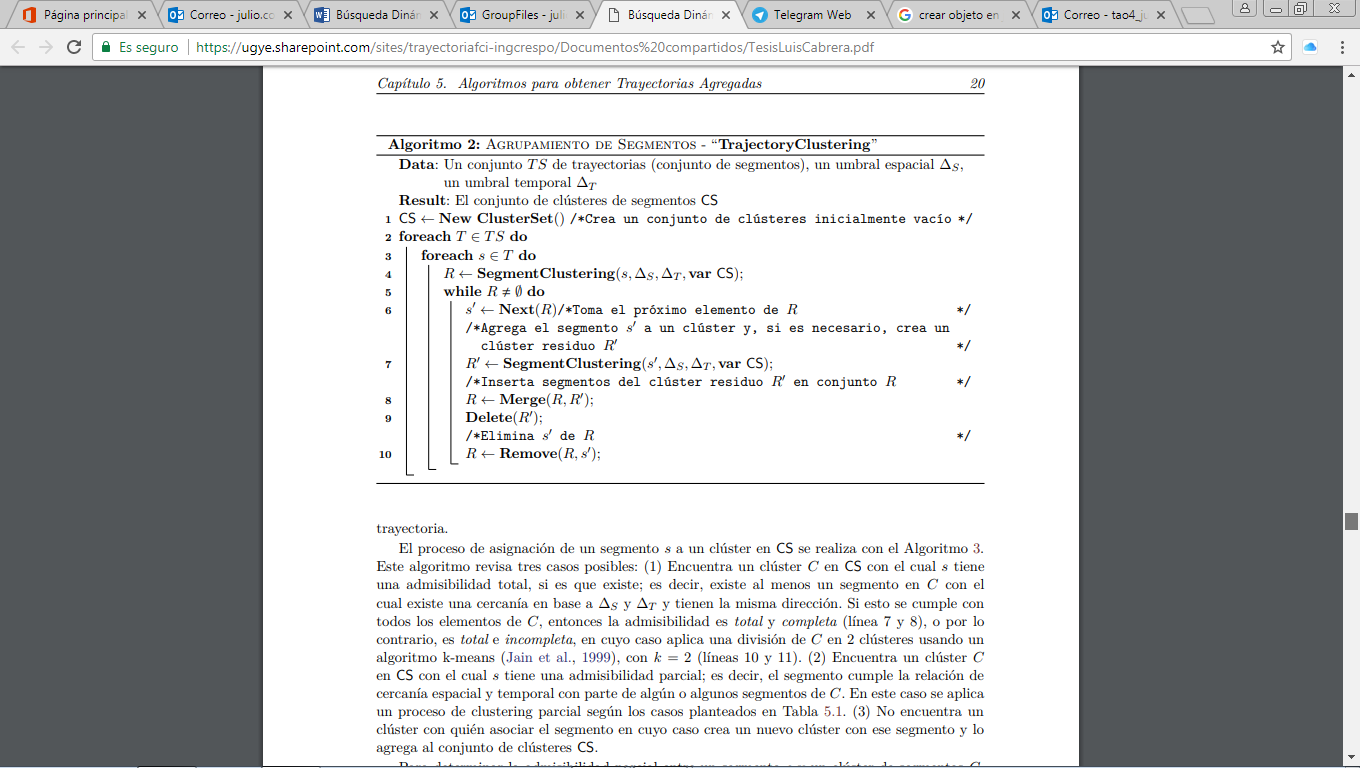
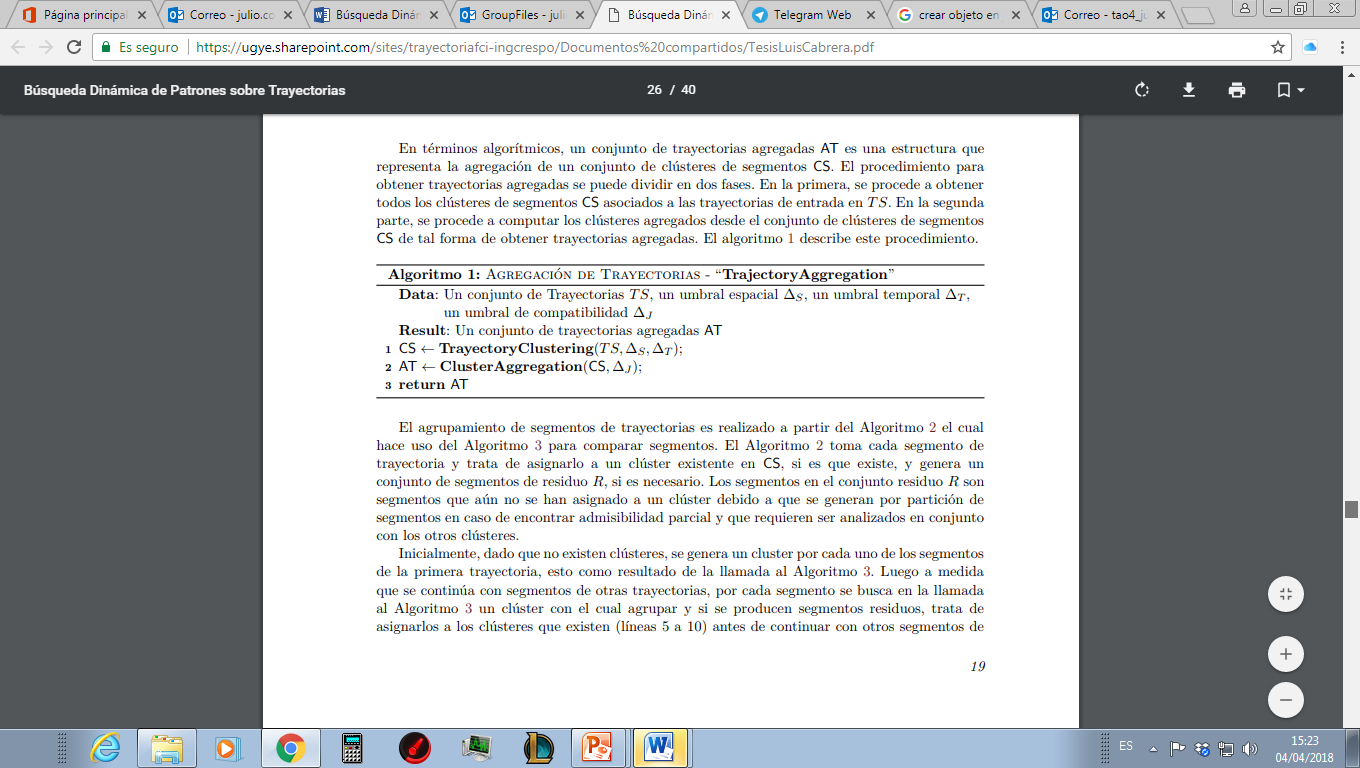
Generación Dinámica de Clústeres de Segmentos

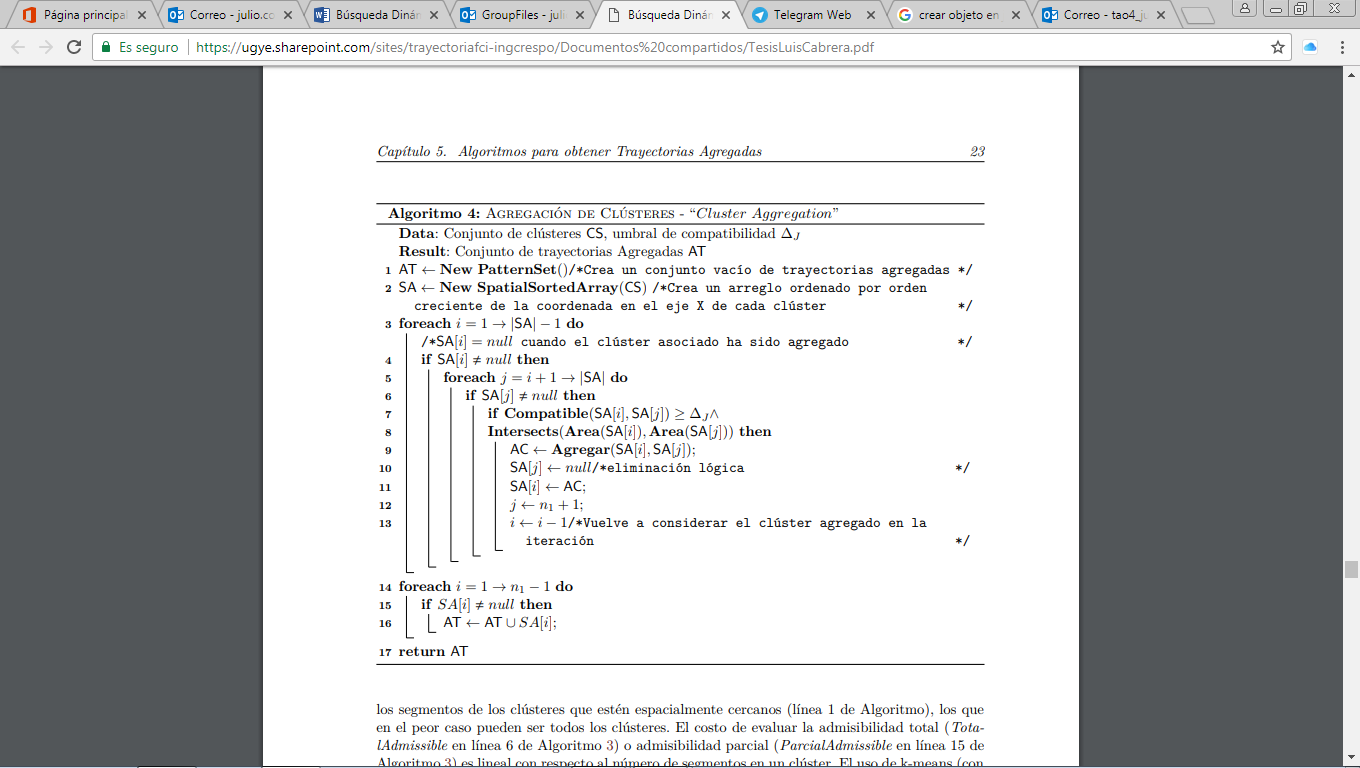
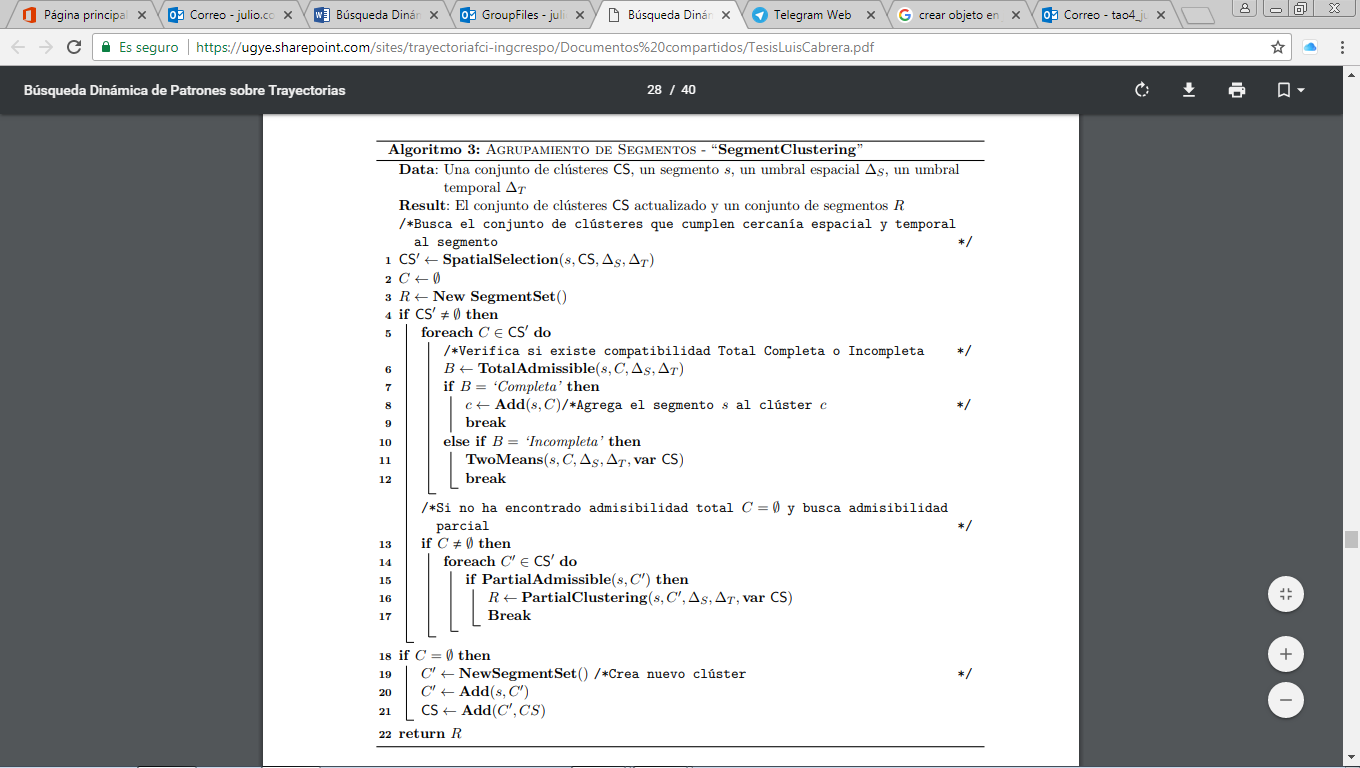
El dinamismo en la generación de clústeres de segmentos de trayectorias se entiende como la capacidad de permitir que, dado un conjunto de clústeres de segmentos CS ya computados, se es capaz de actualizar los clústeres de segmentos ante la llegada de nuevas trayectorias, sin partir con todo el procesamiento desde cero. Es importante mencionar que solo considera la inserción de nuevas trayectorias, pero no la eliminación de trayectorias.

Explicación: Se considera los clústeres de segmentos de la Figura 4.3(f) y la inserción de una nueva trayectoria T3 a los clústeres existentes (Figura 4.4(a)). El segmento i de T3 es particionado en dos sub-segmentos, i0 queda en el clúster cl7, i00 se incorpora al clúster cl4. Lo mismo sucede con el segmento j que es divido en dos sub-segmentos que se ubican en los clústeres cl5 y cl8, respectivamente. La Figura 4.4(b) muestra los ocho clústeres de segmentos resultantes al ingresar la nueva trayectoria.

Para computar los nuevos clústeres agregados, se descartan los clústeres de segmentos cl2, cl6, cl7 y cl8 ya que contienen un único segmento. La Figura 4.4(c) muestra los clústeres que pueden ser agregados, cl1, cl3, cl4 y cl5. La Figura 4.4(d) muestra el área asociada a cada clúster (cierre convexo de los puntos de cada uno de los segmentos en el clúster). Si el nivel de compatibilidad establecido (∆J) entre clústeres se cumple, entonces estos cuatro clústeres podrían ser agregados, dado que sus áreas se intersectan y tienen más de un segmento.

Algoritmo para obtener Trayectorias Agregadas.

Parámetro de entrada: n segmentos de Trayectorias y como parámetro de Salida: m número de clústeres. Pseudocódigo:



El análisis del algoritmo de agrupamiento de trayectorias indica que un segmento es comparado a lo más una vez con algún otro segmento, lo que da un número de comparaciones polinomial respecto al número de segmentos.

En la primera, se procede a obtener todos los clústeres de segmentos CS asociados a las trayectorias de entrada en T S. En la segunda parte, se procede a computar los clústeres agregados desde el conjunto de clústeres de segmentos CS de tal forma de obtener trayectorias agregadas. El agrupamiento de segmentos de trayectorias es realizado a partir del Algoritmo 2 el cual hace uso del Algoritmo 3 para comparar segmentos. El Algoritmo 2 toma cada segmento de trayectoria y trata de asignarlo a un clúster existente en CS, si es que existe, y genera un conjunto de segmentos de residuo R, si es necesario.

El proceso de asignación de un segmento s a un clúster en CS se realiza con el Algoritmo 3. Este algoritmo revisa tres casos posibles: (1) Encuentra un clúster C en CS con el cual s tiene una admisibilidad total. (2) Encuentra un clúster C en CS con el cual s tiene una admisibilidad parcial; es decir, el segmento cumple la relación de cercanía espacial y temporal con parte de algún o algunos segmentos de C. En este caso se aplica un proceso de clustering parcial según los casos planteados en Tabla 5.1. (3) No encuentra un clúster con quién asociar el segmento en cuyo caso crea un nuevo clúster con ese segmento y lo agrega al conjunto de clústeres CS.

Para determinar la admisibilidad parcial entre un segmento s y un clúster de segmentos C, se compara s respecto a los segmentos en C, con su consecuente acción que realiza la función PartialClustering en línea 16 del Algoritmo 3. El Algoritmo 4 genera trayectorias agregadas desde un conjunto de clústeres de segments CS y en base a la medida de compatibilidad definida en el parámetro ∆J.

El Algoritmo 4 toma clústeres en CS y los compara de pares de manera que un clúster es comparado a lo más con todos los otros clústeres. Se nota que cada vez que se agregan dos clústeres, este nuevo clúster reemplaza al clúster en la posición de ordenamiento al clúster original con coordinada en el eje horizontal menor. Aunque se tiene un nuevo clúster, se eliminan los clústeres agregados. En consecuencia, la agregación no altera el máximo número de comparaciones posteriores posibles entre clústeres, dando un costo superior de O(m2), con m el número de clústeres en CS, y donde el costo de evaluar la compatibilidad entre clústeres se considera constante.

Doy por culminado este informe para el uso que se estime conveniente.

Atentamente,

Jordy Andrés Castro López.

Practicante comunitarias Proyecto FCI