



## **Sistema soporte a la decisión para el agrupamiento de clientes de BRASCUBA S.A.**

### ***Decision support system for the clustering of BRASCUBA S.A.'s customers***

**Annalay Escalera-Fariñas<sup>I</sup>, María Gulnara Baldoquín-de la Peña<sup>II</sup>**

<sup>I</sup> Instituto Superior Politécnico José Antonio Echeverría, Cujae. Facultad de Ingeniería Industrial. La Habana, Cuba.

E-mail: annalay@ind.cujae.edu.cu

<sup>II</sup> Pontificia Universidad Javeriana de Cali, Facultad de Ingeniería, Cali, Colombia.

E-mail: mgulnara@javerianacali.edu.co

*Recibido: 14/04/2011*

*Aprobado: 24/01/2012*

#### **RESUMEN**

Se presenta un sistema soporte a la decisión para el agrupamiento de clientes en zonas de distribución, a partir de un problema logístico real de la empresa BRASCUBA S.A., relacionado con la distribución y comercialización de cigarrillos. Se muestran las funcionalidades fundamentales y pantallas principales del mismo. Se utilizó un conjunto de métodos elaborados para agrupar clientes, debido a que no es posible utilizar métodos clásicos existentes de agrupamiento por requerimientos adicionales del problema presentado, de restricciones de capacidad en los clientes a agrupar. El resultado es una aplicación *desktop*, diseñada para ser utilizada sobre plataforma Windows, con la cual el usuario puede interactuar de manera simple a través de un sistema gestor y generador de diálogos. Se describen la metodología y notaciones usadas en el diseño del sistema. Los directivos de BRASCUBA valoraron positivamente el trabajo desarrollado, que puede ser utilizado por otras empresas con problemas de agrupamiento similares.

**Palabras clave:** sistema soporte a la decisión, agrupamiento, distribución.

#### **ABSTRACT**

*A decision support system is presented for clustering clients in distribution areas, due to a real logistical problem of BRASCUBA S.A. company, related to the distribution and commercialization of cigarettes. The fundamental functionalities, as well as the main screens of the system, are shown. A group of methods elaborated for customers' clustering was used, because it is not possible to use the existent classic methods of cluster, due to the additional requirements that the organization's problem presents, of capacity of each cluster. The result is a desktop application, designed to be used on platform Windows, with which the user can interact in a simple way through a system agent and generator of dialogues. The methodology and notations used in the design of the system are described. The directive of BRASCUBA gave positive valuations of the work, which can be used by other companies with similar cluster's problems.*

**Key words:** decision support system, clustering, distribution.

## I. INTRODUCCIÓN

La empresa mixta cubana-brasilera BRASCUBA S.A. tiene como objetivo la producción de cigarrillos y su comercialización en el mercado nacional. La empresa abarca las ventas en todo el territorio nacional a todas las cadenas de tiendas, centros gastronómicos, hoteles y otros puntos de ventas que comercializan productos en divisa; siendo el objetivo de la empresa llegar con la distribución directa a cada punto de venta. La cantidad de clientes en la actualidad, tan solo en la provincia La Habana, sobrepasa los 1000 y debe incrementarse.

Algunas de las características asociadas a la logística de distribución de la empresa, son las siguientes:

1. El número de visitas que debe realizarse a cada punto de venta depende del volumen de ventas mensual en el lugar, pudiendo ser semanal, bisemanal, quincenal o mensual.
2. Existen 3 tipos de visita a realizar en los puntos de venta, no siempre para descargar mercancías, lo cual implica tiempos de estadía diferentes.
3. La empresa cuenta con un conjunto de carros para llegar a los puntos de venta, siendo una restricción que cada carro tiene asociado un vendedor fijo y que cada punto debe ser siempre visitado por el mismo vendedor.

Para llegar con la distribución directa a cada punto de venta de acuerdo a las entregas requeridas en un mes, se hace necesario resolver 3 subproblemas diferentes, pero relacionados: particionar el conjunto de clientes por zonas de distribución, planificar los días de entrega a cada punto de venta en el mes y buscar las rutas de distribución diarias en cada zona, de acuerdo a los puntos a visitar en el día. El primer subproblema conlleva resolver un problema de agrupamiento, pero no clásico, teniendo en cuenta restricciones impuestas como la capacidad de los clientes; entendiendo como tal el número de visitas a realizar en el mes, que en la actualidad puede oscilar entre 1 y 8. Un buen agrupamiento no solo debe tener cercanos los clientes en una misma zona, sino que debe satisfacer que cada zona no sobrepase una capacidad máxima. Por otro lado, existen restricciones adicionales de puntos que deben ser visitados por el mismo vendedor, lo que significa que deben estar situados en la misma zona de distribución; por ejemplo, si están asociados a la misma casa matriz. Esto conlleva a resolver no solo un problema de agrupamiento con restricciones de capacidad, sino con conocimiento adicional del problema, que debe ser integrado a los métodos de agrupamiento utilizados. Un mal agrupamiento, debido a que exista una zona muy "cargada" de visitas a realizar en un mes, puede provocar que no sea posible en dicha zona lograr una planificación mensual en el tiempo de trabajo estipulado para ello.

La literatura es amplia en aplicaciones y métodos para la solución de problemas de agrupamiento. El algoritmo de las  $k$ -medias, también conocido como  $k$ -means en inglés, es uno de los más utilizados en aplicaciones científicas [1]. En la literatura existen diferentes versiones de este método. Entre ellas pueden citarse: *COP-KMeans* [2], 2 variantes semi-supervisadas de  $k$ -medias, llamadas *Seeded-KMeans* y *Constrained-KMeans* [3], versiones de *iKMeans* que difieren en la definición de las distancias y centroides [1], y  $k$ -means integrado con la distancia de Jaccard [4].

Algunos de los métodos de agrupamiento son sensibles a la partición inicial que se tome [2; 3; 4; 5], por lo que si se proveen buenos puntos iniciales o semillas, los mismos pueden ser muy robustos. Deng y Bard (2011) usan una metaheurística híbrida para hacer agrupamientos con restricciones de capacidad en los *clusters* a obtener, a partir de un grafo con  $p$  vértices (objetos a agrupar), donde no necesariamente todos los vértices están conectados entre sí, no coincidiendo esto con las características del problema abordado [5].

Pacheco y Beltrán (2001) proponen un método ávido-aleatorio de 2 fases para para la creación de *clusters*, proporcionando de manera "inteligente" las semillas iniciales [6]. Taillard (2003) desarrolla 3 métodos heurísticos para la obtención de *clusters* no jerárquicos: *candidate list search* (CLS), *local optimization* (LOPT) y *descomposition-recombination* (DEC); uno de ellos para el agrupamiento de grandes volúmenes de datos, en los que tiene en cuenta cómo seleccionar "buenas" semillas iniciales [7].

Berkhin (2006) hace un estudio de técnicas de Minería de Datos para problemas de agrupamiento [8].

En particular, los *softwares* encontrados para la construcción de *clusters*, como el MINITAB, no tienen en cuenta la restricción de capacidad en los puntos a agrupar. El volumen de clientes que ya tiene la empresa, y el aumento que se prevé en el futuro cercano, hace que la misma no pueda continuar resolviendo el problema de forma manual, aún con un personal experto. Por otro lado, no se ha encontrado en la literatura *software* que aborden este problema como tal, aunque algunos resuelven problemas similares, pero a un alto costo [9; 10; 11; 12]. Lo anterior justifica

## SISTEMA SOPORTE A LA DECISIÓN PARA EL AGRUPAMIENTO DE CLIENTES DE BRASCUBA S.A.

la necesidad de desarrollar una herramienta informática que le permita a la empresa agrupar sus clientes por zonas de distribución, determinar los clientes a visitar cada día de la semana para cada una de dichas zonas y obtener las rutas a seguir por sus vehículos, cada día, para cubrir la entrega de sus productos a los clientes.

En este trabajo se presenta, del sistema soporte a la decisión para la solución del problema planteado, el subsistema que corresponde a la solución de una de las fases del problema: la partición del total de clientes de la empresa en La Habana, en un conjunto de zonas de distribución.

### II. MÉTODOS

El trabajo desarrollado conllevó las siguientes tareas:

1. Entrevistas a directivos y vendedores de BRASCUBA, así como participación como observadores en la distribución de mercancías en diferentes zonas y diferentes días, para definir el problema a resolver.
2. Búsqueda de *softwares* relacionados con el tema (referenciados en epígrafe anterior)
3. Estudio de algoritmos no clásicos de agrupamientos (referenciados en epígrafe anterior).
4. Estudio de reconocidos modelos de procesos [13; 14; 15; 16].

Como resultado de lo anterior se desarrolló un modelo matemático para la construcción de agrupamientos con las restricciones impuestas por la empresa. El modelo corresponde a un problema de Programación Binaria Lineal Cuadrática, que por la dimensión del mismo se hace prácticamente imposible resolver de manera exacta con *software* existentes para este tipo de problemas. De ahí que se desarrollaran diversos algoritmos heurísticos para obtener agrupamientos con restricciones de capacidad en sus puntos. La diferencia entre los algoritmos la determinan diversos métodos de selección de puntos iniciales (semillas) en la construcción de los *clusters*, así como cuándo se realiza la actualización de los centroides (centro geométrico de cada *cluster*) de los diferentes agrupamientos. Para ello se decidió seguir una filosofía similar a la propuesta por Pacheco y Beltrán (2001) [6] y por Taillard (2003) [7]. Los algoritmos son adaptaciones del conocido método *k-means* con el uso de centroides, lo que le proporciona una clara interpretación física al problema abordado (agrupamiento en zonas geográficas), aportando el centro geométrico de cada *cluster*.

La descripción de dichos algoritmos, así como el diseño de experimentos desarrollados para fundamentar la efectividad de dichos métodos, se presentan en el trabajo de Escalera (2011) [17]. La medida de distancia (que permite medir la similitud entre los objetos a agrupar) asumida en esta primera versión del sistema, es la distancia euclidiana, la cual se define como en la expresión 1:

$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2} \quad (1)$$

Las restricciones adicionales sobre clientes que deben estar situados en la misma zona de distribución, se tienen en cuenta mediante una transformación de conjuntos de clientes (coordenadas geográficas de los mismos) y sus capacidades en clientes ficticios, con nuevas coordenadas y capacidades, que permite reducir el problema a un problema solo de restricción de capacidades.

Para el análisis y diseño del sistema se decidió utilizar, de los más reconocidos modelos de procesos en el desarrollo de *software*, el Proceso Unificado de *Rational* (RUP, por sus siglas en inglés) así como el Lenguaje Unificado de Modelado (UML, por sus siglas en inglés). Para la implementación se utilizó *c#* como lenguaje de programación, y una versión Express de Visual Studio.

Puesto que el sistema desarrollado se sustenta en un conjunto de algoritmos heurísticos para la construcción de los agrupamientos, la confiabilidad de los mismos es un aspecto importante. Una exitosa prueba de un algoritmo (programa), asegura que para todas las instancias (por ejemplo: juegos de datos) a las que se les aplica el programa, ofrece los resultados de salida adecuados. En general, es una tarea sumamente difícil. Muchas veces se confunde la prueba con la experimentación, que es el arte de crear muestras de datos sobre las cuales correr un programa. Una prueba dice más que cualquier número finito de experimentos, pero es más difícil.

Como una alternativa a la prueba de los algoritmos, los mismos fueron programados por 2 personas diferentes, utilizando lenguajes de programación y estructuras de datos distintas, desarrollando de forma paralela el diseño de experimentos confeccionado para fundamentar la eficiencia de los métodos. Esto permitió encontrar y corregir errores de programación muy difíciles de encontrar en un corto tiempo de experimentación, llegando al final a resultados satisfactorios en cuanto a la confiabilidad de los algoritmos.

### III. RESULTADOS

La herramienta desarrollada para la construcción de agrupamientos permite, entre sus opciones:

1. Cargar la base de datos de los clientes a agrupar.
2. Mostrar los datos de los clientes de forma tabular, así como representar geográficamente los clientes, teniendo en cuenta las coordenadas de su localización.
3. Presentar estadísticas sobre las capacidades de los clientes, así como de las distancias entre los mismos, teniendo en cuenta sus coordenadas geográficas.
4. Definir parámetros en la construcción de los *clusters*, como son:
  - 4.1. Especificar o no la cantidad de *clusters* a construir.
  - 4.2. Especificar o no la capacidad promedio máxima de cada *cluster*, o una capacidad máxima definida para cada *cluster* (si hay una cantidad prefijada).
  - 4.3. Definir cuáles puntos específicos deban (o no) estar en un mismo *cluster*.
  - 4.4. Definir un valor D tal que si 2 puntos se encuentran a una distancia menor que D, sea aconsejable que pertenezcan al mismo *cluster*, si no existen otras restricciones "duras" que se lo impidan.
  - 4.5. Seleccionar entre varios de los métodos desarrollados para agrupamiento con restricciones de capacidad en los puntos a agrupar.
5. Brindar diferentes opciones de soluciones, gráficas y analíticas.
6. Salvar las soluciones obtenidas por el sistema.

Dentro de los requisitos no funcionales del sistema pueden mencionarse:

El sistema debe ser capaz de dar una respuesta al usuario en la menor brevedad posible (en general, agrupamientos con más de 1000 nodos en segundos y menos de 3 minutos), lo cual fue tenido en cuenta desde el diseño de los algoritmos.

El sistema fue diseñado de manera que los errores que pueden ser provocados por la entrada incorrecta de datos por parte del usuario sean mínimos, siguiendo una estrategia preventiva. No obstante, en los casos en que no es posible su prevención, la validación de los datos se realiza a nivel de código, mostrando los errores en una ventana que contiene un texto describiendo la situación que dio origen al error y qué hacer para evitar que vuelva a ocurrir.

El sistema requiere de un procesador Pentium III o superior, con 256 Mb de memoria de acceso aleatorio RAM, como mínimo.

#### Principales pantallas del sistema

Las figuras que se presentan a continuación corresponden a pantallas de la interfaz de usuarios que permiten dar una visión de las posibilidades del sistema.

En la figura 1 se presenta la pantalla principal del sistema. En el menú que se visualiza en la parte superior se encuentran las opciones que éste brinda, que darán solución a los 3 subproblemas que conforman el problema general: el agrupamiento de clientes por zonas (Zonas), la planificación de las visitas a clientes en un período (Planificación) y la confección de rutas diarias (Ruteo). En rojo aparecen sombreadas las opciones de los menús Datos y Zonas, que corresponden con el subproblema abordado en este trabajo.

Los datos de los clientes a agrupar, pueden "cargarse" en uno de 2 formatos: Open Office Excel o texto. El sistema también brinda la posibilidad de salvar las soluciones que se van obteniendo, en formato Excel de Open Office.

Una vez que son cargados los datos (en formato texto o Excel), éstos pueden ser mostrados en la ventana principal, en forma de tabla (como se muestra en la figura 2) y en forma gráfica (como se puede observar en la figura 3); pudiéndose ver, además, información estadística sobre los datos cargados, como se muestra en la figura 4.

## SISTEMA SOPORTE A LA DECISIÓN PARA EL AGRUPAMIENTO DE CLIENTES DE BRASCUBA S.A.

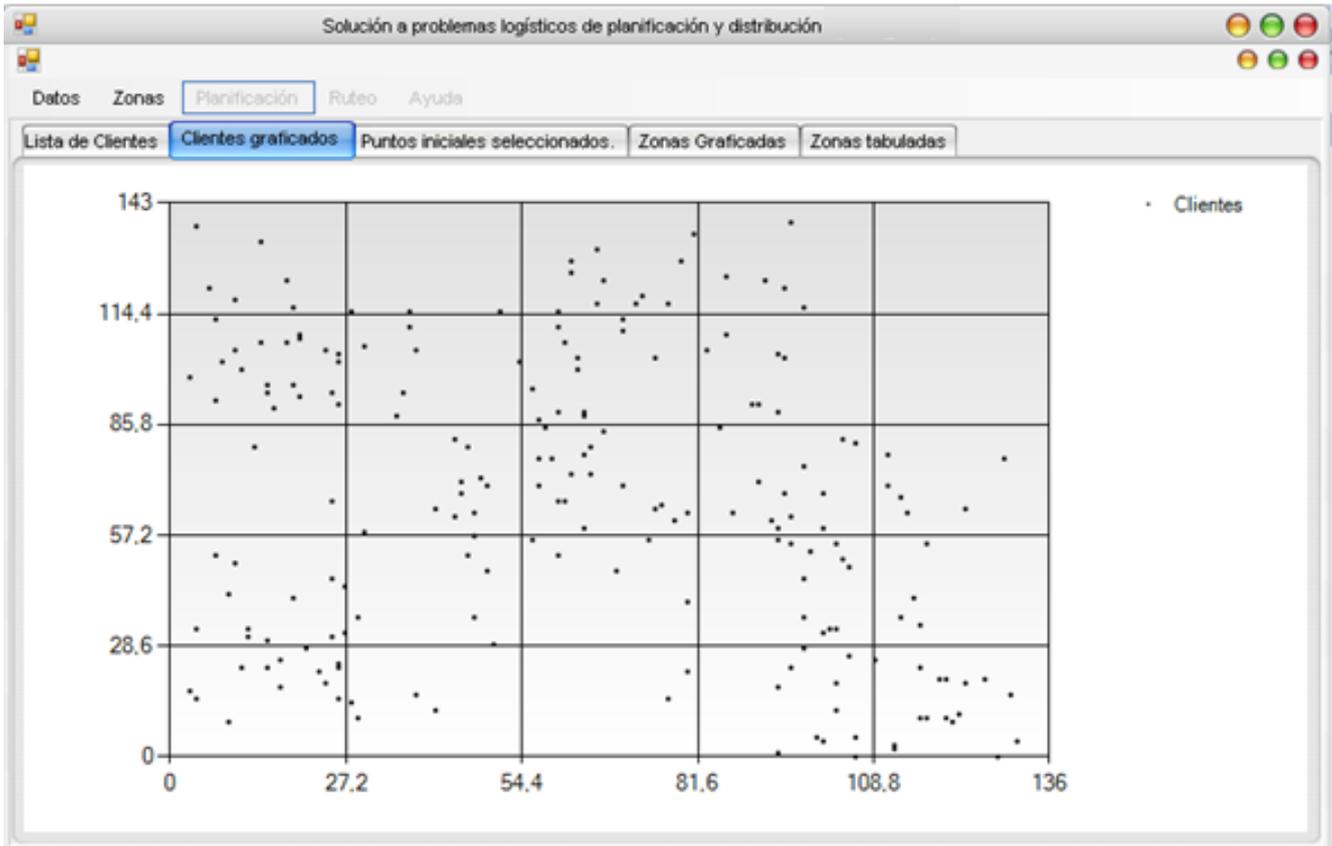


**Figura 1.** Opciones del sistema.

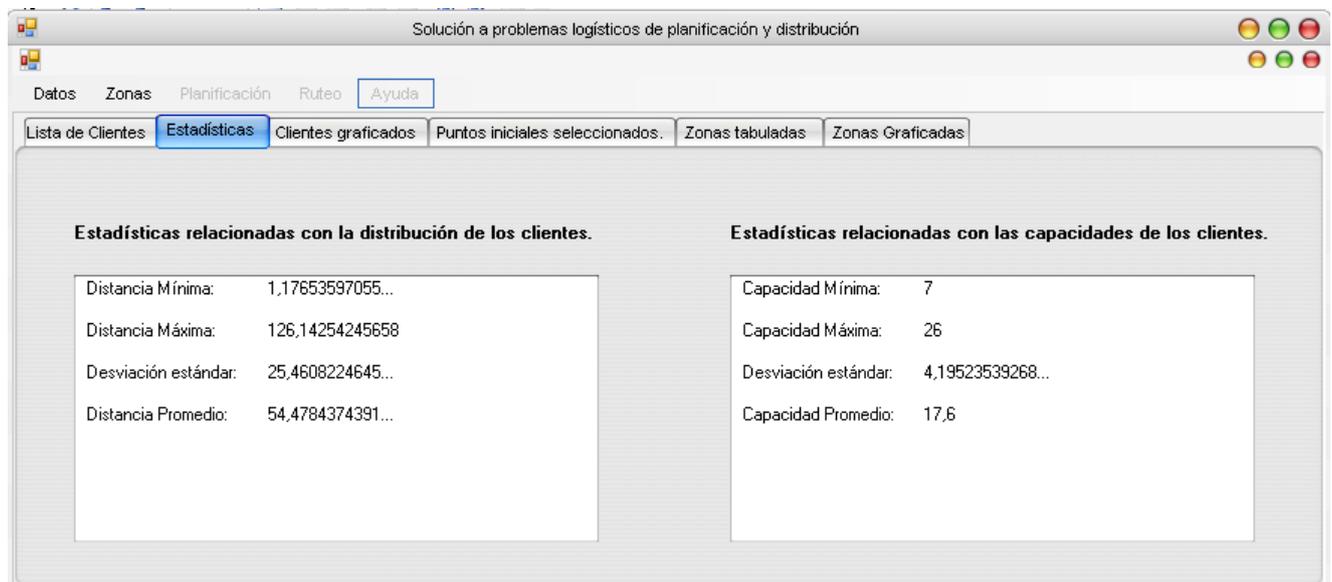
The screenshot shows the same window as Figure 1, but with the "Lista de Clientes" option selected. Below the menu bar, there are five sub-tabs: "Lista de Clientes", "Clientes graficados", "Puntos iniciales seleccionados.", "Zonas Graficadas", and "Zonas tabuladas". The "Lista de Clientes" tab is active, displaying a table with the following data:

No. Cliente	Coordenada X	Coordenada Y	Demanda	Es Casa matriz	Casa matriz
4900	70	70	0	False	-1
929	33	78	20	False	-1
8204	59	52	20	False	-1
7558	10	137	30	False	-1
8278	4	28	10	False	-1
2195	25	26	20	False	-1
3349	86	37	10	False	-1
6009	1	109	10	False	-1
493	6	135	40	False	-1
2304	32	79	20	False	-1
2858	24	26	20	False	-1
2894	86	36	20	False	-1
979	95	35	10	False	-1
2815	63	50	10	False	-1
2884	100	106	10	False	-1
4672	99	112	20	False	-1
2793	36	135	10	False	-1
1594	57	59	10	False	-1
4984	8	124	10	False	-1

**Figura 2.** Datos presentados en forma tabular.



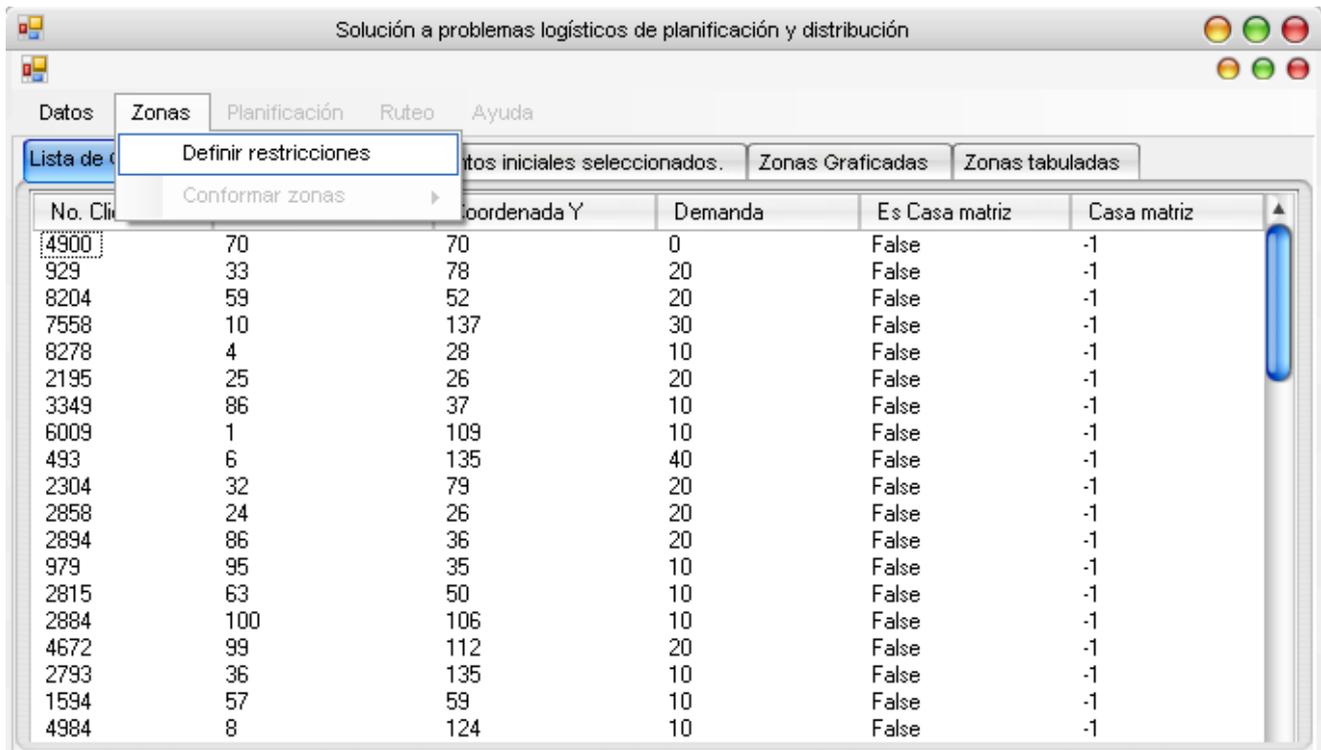
**Figura 3.** Datos presentados en forma gráfica.



**Figura 4.** Estadísticas sobre los datos de los clientes.

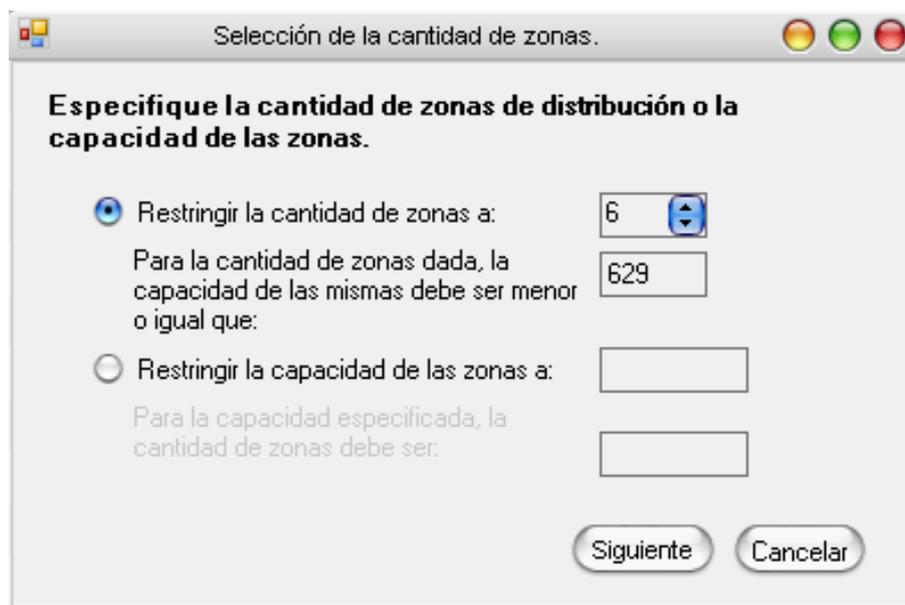
Las funcionalidades del sistema que se encuentran en el menú Zonas, mostrado en la figura 5, se encuentran deshabilitadas mientras no existan datos cargados en el sistema.

## SISTEMA SOPORTE A LA DECISIÓN PARA EL AGRUPAMIENTO DE CLIENTES DE BRASCUBA S.A.

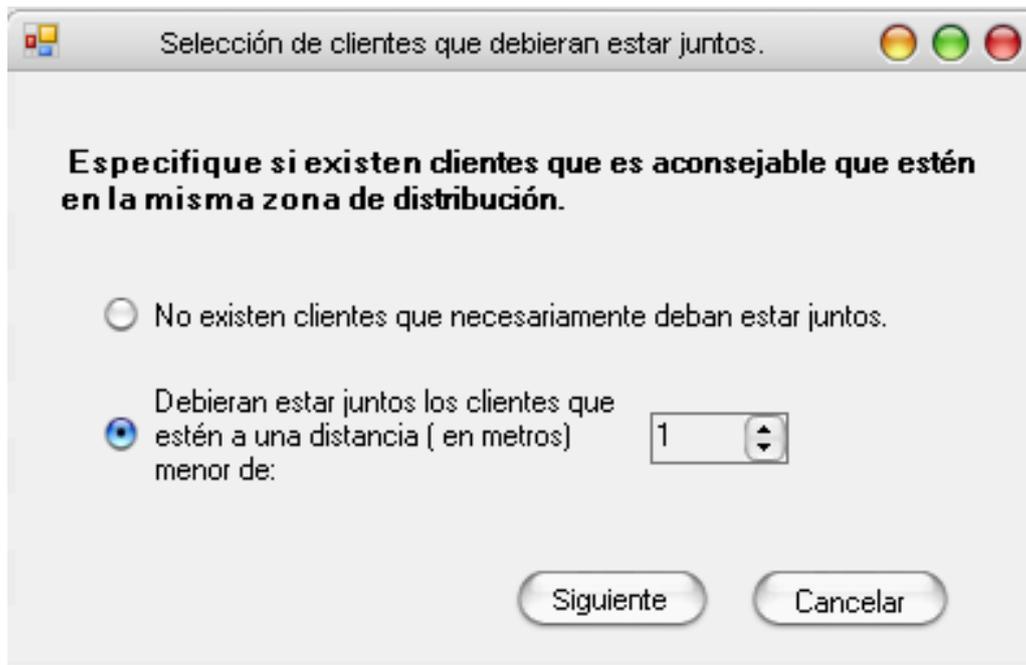


**Figura 5.** Submenú Zonas.

Una de las opciones del menú Zonas es la definición de la cantidad de zonas de distribución (en cuyo caso, el sistema le ofrece la capacidad máxima promedio que debe tener cada zona) o las capacidades promedios máximas de las mismas, de manera que todas queden lo más uniformemente posible en cuanto a capacidad; lo cual se observa en la figura 6. En el último caso, el sistema le indica la cantidad de zonas de distribución que debe tener. Otra opción posible es definir un valor numérico que posibilite que clientes cercanos, a una distancia menor de dicho valor, traten de ubicarse en la misma zona (como se expone en la figura 7), siempre que las capacidades de las zonas lo permitan, por lo que no se considera una restricción fuerte para el problema.

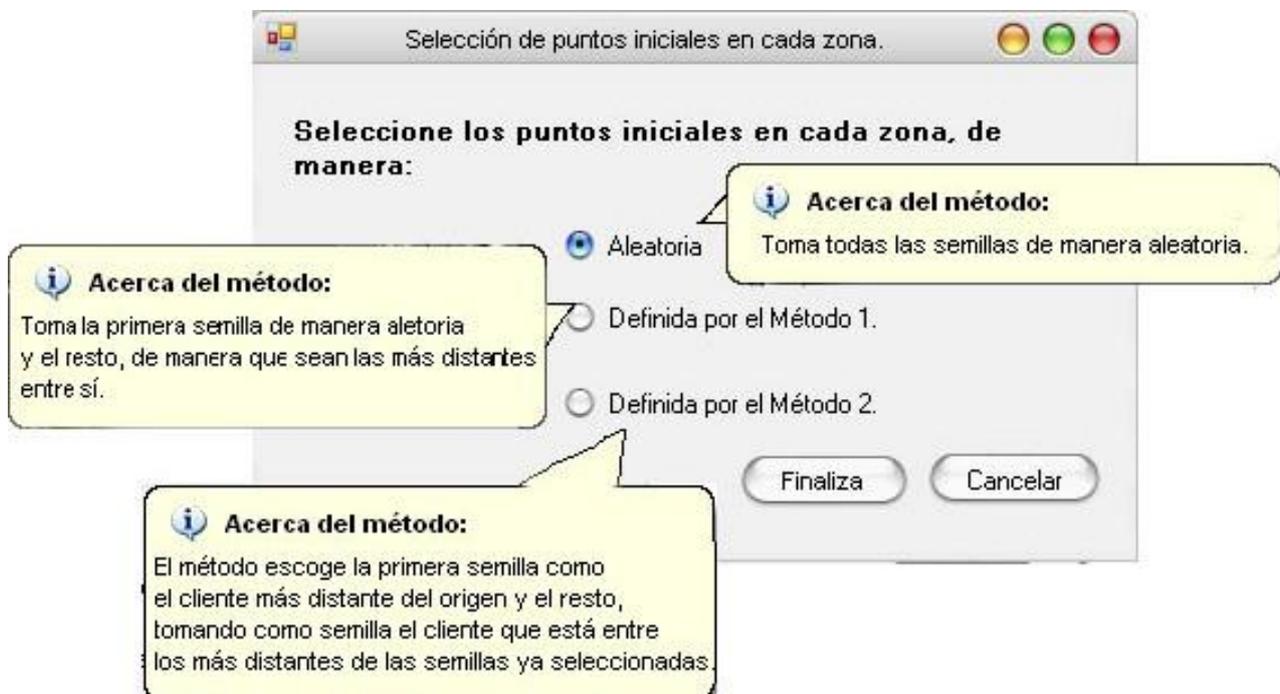


**Figura 6.** Pantalla para especificar la cantidad de zonas o la capacidad máxima permitida para cada zona.



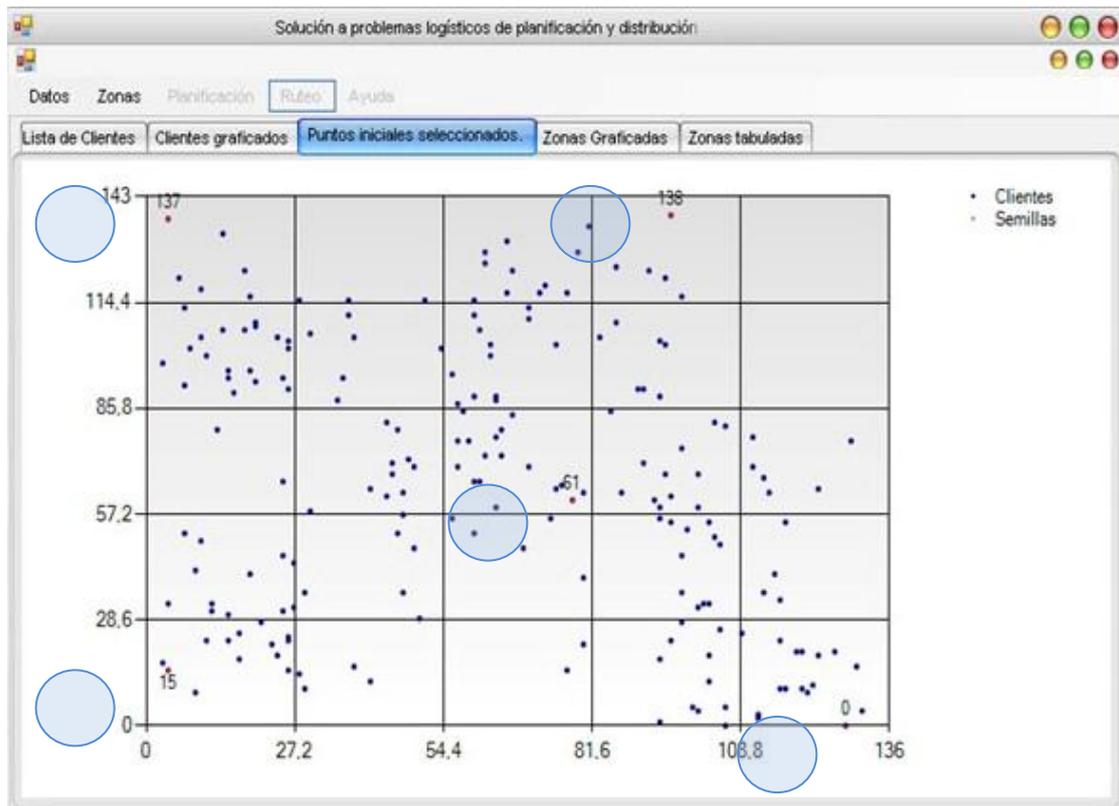
**Figura 7.** Pantalla para especificar si existen clientes que debieran estar juntos por la distancia existente entre ellos.

La figura 8 muestra cómo el sistema permite seleccionar un método para escoger los puntos iniciales o semillas, aunque el sistema le ofrece uno por defecto. El sistema permite visualizar gráficamente las semillas seleccionadas por el método escogido por el usuario, como se observa en la figura 9, permitiéndole al usuario valorar la calidad de las mismas.



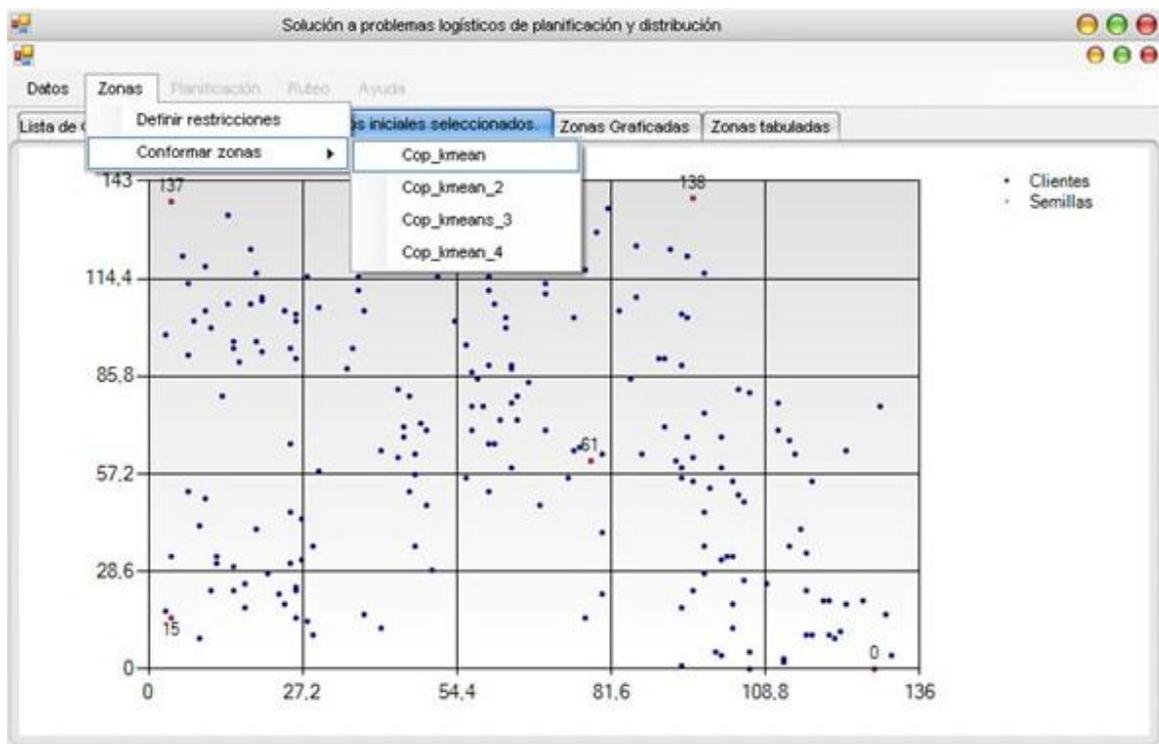
**Figura 8.** Pantalla para seleccionar el método para obtener las semillas.

## SISTEMA SOPORTE A LA DECISIÓN PARA EL AGRUPAMIENTO DE CLIENTES DE BRASCUBA S.A.



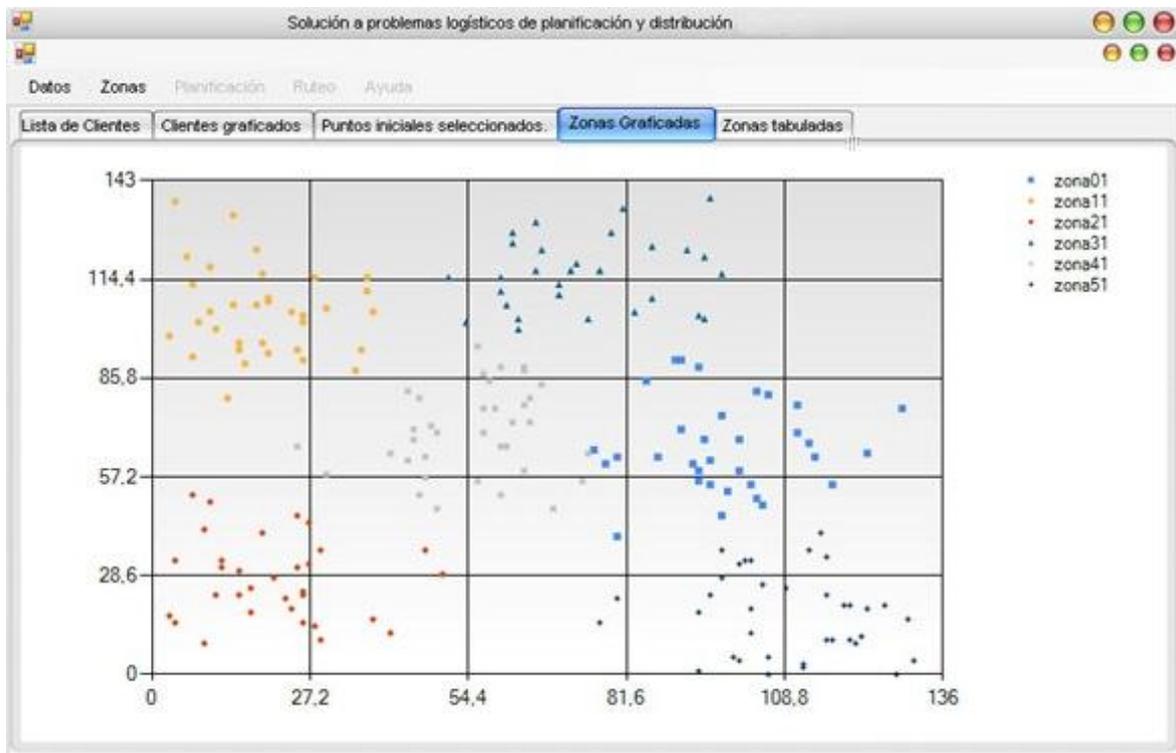
**Figura 9.** Representación gráfica de las semillas seleccionadas.

Una vez que el usuario selecciona las semillas iniciales, puede proceder a conformar las zonas de distribución, escogiendo para ello uno de los métodos implementados por el sistema, como se aprecia en la figura 10, aunque el sistema le ofrece uno por defecto de acuerdo a los parámetros que se fijaron.



**Figura 10.** Opción que permite seleccionar el método para conformar las zonas.

La solución obtenida puede mostrarse de forma gráfica como se observa en la figura 11, y/o de forma tabular como en la figura 12. El sistema le permite al usuario salvar la solución obtenida (figura 12) en una tabla de Open Office Excel. En caso de que la solución no satisfaga los requerimientos del cliente, el mismo puede volver a ejecutar los pasos descritos anteriormente.



**Figura 11.** Gráfico de la solución obtenida.

No. Zona	No. Cliente	Suma de la...	Capacidad	X del Centr...	Y del Centr...
Zona0	4647	805,778302...	670		
Zona0	786	805,778302...	670		
Zona0	2893	805,778302...	670		
Zona0	5450	805,778302...	670		
Zona0	331	805,778302...	670		
Zona0	8	805,778302...	670		
Zona0	1142	805,778302...	670		
Zona0	334	805,778302...	670		
Zona0	3516	805,778302...	670		
Zona0	5550	805,778302...	670		
Zona0	855	805,778302...	670		
Zona0	1471	805,778302...	670		
Zona0	829	805,778302...	670		
Zona0	2946	805,778302...	670		
Zona0	5734	805,778302...	670		
Zona0	2958	805,778302...	670		
Zona0	2956	805,778302...	670		
Zona0	790	805,778302...	670		
Zona0	6004	805,778302...	670		
Zona0	6022	805,778302...	670		
Zona0	268	805,778302...	670		
Zona0	6111	805,778302...	670		
Zona 0		805,778302...	670	109,25	17,3611111...
Zona 1		584,243347...	610	19,2647058...	104,941176...
Zona 2		707,294243...	620	21,125	27,46875
Zona 3		586,079838...	580	74,3793103...	115,931034...
Zona 4		970,782222...	650	55	70,138888...
Zona 5		899,258037...	640	98,6969696...	65,8484848...
Distancia total				56971,2316...	

**Figura 12.** Visualización tabular de la solución.

#### IV. DISCUSIÓN

El sistema propuesto presentado en la sección anterior fue validado teniendo en cuenta 2 aspectos fundamentales: el comportamiento de los algoritmos utilizados y la efectividad del funcionamiento

## SISTEMA SOPORTE A LA DECISIÓN PARA EL AGRUPAMIENTO DE CLIENTES DE BRASCUBA S.A.

del sistema, en términos de la funcionalidad de la interfaz de usuarios y confiabilidad del sistema, según los criterios de los usuarios del mismo.

Para validar el sistema teniendo en cuenta los algoritmos utilizados, se combinaron datos reales de BRASCUBA en la provincia La Habana con una selección de ejemplos con situaciones muy similares, teniendo en cuenta las capacidades reales de la empresa. Los resultados obtenidos fueron valorados como adecuados, según todas las restricciones que la empresa impone para hacer las agrupaciones de clientes por zonas.

Para la validación de la efectividad del funcionamiento del sistema, los propios usuarios hicieron uso del mismo con algunos ejemplos introducidos por ellos, luego de haber recibido un manual de usuario. El mismo satisfizo sus expectativas en términos de: rendimiento (tiempos de respuesta), tratamiento de errores, integridad (confiabilidad de resultados), modularidad, flexibilidad, sencillez de la interfaz de usuario.

### V. CONCLUSIONES

1. El sistema soporte a la decisión desarrollado ofrece una respuesta satisfactoria a la solución de un problema logístico real y complejo de una empresa de distribución de cigarrillos en Cuba.
2. Los algoritmos no clásicos de agrupamiento desarrollados, en los cuales se soporta el sistema para dar solución al problema planteado, no se encuentran entre los *software* más utilizados de construcción de *clusters*.
3. La empresa BRASCUBA S.A. debe completar la base de datos de las localizaciones geográficas de todos sus clientes, adquirir un Sistema de Información Geográfica a acoplar al sistema, que permita actualizaciones eficientes de sus bases de datos; así como debe lograr la implementación de las otras fases del problema que presenta. 📖

### VI. REFERENCIAS

1. MING-TSO, M.; MIRKIN, B., «Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads» *Journal of Classification*, 2010, vol. 27, no. 1, pp. 3-40, ISSN 1432-1343.
2. WAGSTAFF, K.; CARDIE, C. et al., «Constrained k-means clustering with background knowledge», en *Proceedings of the Eighteenth International Conference on Machine Learning, First Asian Himalayas International Conference 2001*, [consulta: 2011-04-01]. Disponible en: <doi: 10.1109/AHICI.2009.5340335>;
3. BASU, S.; BANERJEE, A.; MOONEY, R., «Semi-supervised clustering by seeding», en *19th International Conference on Machine Learning Sydney (Australia), 2002*, [consulta: 2011-03-14]. Disponible en: <<http://www.connex.lip6.fr>>
4. SHAMEEM, M. U. S.; FERDOUS, R., «An efficient k-means algorithm integrated with Jaccard distance measure for document clustering», en *AH-ICI 2009, First Asian Himalayas International Conference 2009*, [consulta: 2011-03-20]. Disponible en: <[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5340335&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5340335&tag=1)>
5. DENG, Y.; BARD, J., «A reactive GRASP with path relinking for capacitated clustering», *Journal of Heuristics* [en línea], 2011, vol. 17, no. 2, pp. 119-152 [consulta: 2011-03-20], ISSN 1381-1231. Disponible en: <doi: 10.1007/s10732-010-9129-z>;
6. PACHECO, J.; BELTRÁN, M., «Análisis de nuevos métodos de clasificación. Un ejemplo ilustrativo de su uso en la agrupación de los municipios de Castilla y León» *Estadística Española*, 2001, vol. 43, no. 148, pp. 209-224, ISSN 0014-1151.
7. TAILLARD, E., «Heuristic Methods for Large Centroid Clustering Problems» *Journal of Heuristics*, 2003, vol. 9, no. 1, pp. 51-73, ISSN 1572-9397.
8. BERKHIN, P., «Survey of Clustering Data Mining Techniques», *Grouping Multidimensional Data: Recent Advances in Clustering* [en línea], 2006, pp. 25-71 [consulta: 2011-03-14], Disponible en: <<http://ebookbrowse.com/survey-of-clustering-data-mining-techniques-pdf-d342285823>>
9. EOM, S.; KIM, E., «A survey of decision support system applications (1995–2001)» *Journal of the Operational Research Society*, 2006, vol. 57, no. 11, pp. 1264–1278, ISSN 0160-5682.
10. KOHLER, U.; KREUTER, F., *Data Analysis Using Stata*, 2nd. ed., Texas, Stata Press, 2009, ISBN: 978-1-59718-046-7.
11. MARQUES DE SÁ, J. P., *Applied Statistics Using SPSS, STATISTICA, MATLAB and R*, 2nd. ed., Porto, Springer, 2007, ISBN: 978-3-540-71971-7.

12. PARTYKA, J.; HALL, R., «On the road to connectivity» *OR/MS Today*, 2010, vol. 37, no. 1, pp. 42-49, ISSN: 1085-10386.
13. CURTIS, B.; HEFLEY, W. E.; MILLER, S. A., *Capability Maturity Model (P-CMM) Version 2.0 Technical Report CMU/SEI-2001-MM-01* Software Engineering Institute, 2001 -, publ. - [consulta: 2011-03-15]. Disponible en: <<http://www.sei.cmu.edu/library/abstracts/reports/01mm001.cfm>>.
14. HUMPHREY, W. S., *A Discipline for Software Engineering*, Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc., 1995, ISBN 9780201546101.
15. HUMPHREY, W. S., *Managing Technical People: Innovation, Teamwork and the Software Process*, Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc., 1998, ISBN 9780201545975.
16. JACOBSON, I.; BOSCH, G.; RUMBAUGH, J., *El Proceso Unificado de Desarrollo de Software*, Madrid, Addison Wesley Iberoamericana, 2000, ISBN 74-7829-036-2.
17. ESCALERA, A., «Sistema soporte a la decisión para el agrupamiento en zonas de distribución de los clientes de BRASCUBA S.A.», [tesis de maestría], La Habana, Instituto Superior Politécnico José Antonio Echeverría, Cujae, Departamento de Matemática General, Facultad Ingeniería Industrial, 2011.