

LOCATION OF INFORMATION OF PERSONAL CONTACTS USING TECHNOLOGIES OF WEB

LOCALIZACIÓN DE DATOS DE CONTACTOS PERSONALES UTILIZANDO TÉCNICAS DE MINERÍA WEB Y REDES SOCIALES

Recibido: 14 de julio 2015 - aceptado: 25 de noviembre 2015

Rafael Cabezas¹
Universidad Simón Bolívar

Keywords:

social networks,
location data,
development,
implementation.

Abstract

In this paper show the results of our Research Project With Regard to Location Data Contacts of Persons scam Specific Features of society using social networks , task de este Project is to show A Guide Organization search results for the name of the person using web mining, con Tools Help Many tales as Pipl , YAHOO , GOOGLE , DUCKDUCK , Among Others, If paragraph , laying the foundations of Como itself are the Different Development and implementations in this area , availing of related queries in different databases designed and Research Methodology para fines Similar preferring implemented in Developing Countries items . Identification of the work performed and / or methodologies used by other researchers, air THEIR respective Findings and Conclusions.

Palabras clave:

Redes sociales,
ubicación, datos,
desarrollo,
implementación.

Resumen

En este artículo se muestran los resultados de nuestro proyecto de investigación con relación a la localización de datos de contactos de personas con características específicas de la sociedad utilizando las redes sociales, La tarea de este proyecto consiste en mostrar una guía de organización de resultados de búsqueda para el nombre de la persona utilizando la minería web, con la ayuda de varias herramientas tales como PIPL, YAHOO, GOOGLE, DUCKDUCK, entre otras, para sí, sentar bases de cómo se encuentran los diferentes desarrollos e implementaciones en esta área, valiéndonos de consultas de artículos relacionados en diferentes bases de datos e investigación de metodologías diseñadas para fines similares prefiriendo las implementadas en países en desarrollo. Se realizó la identificación de los trabajos y/o metodologías empleados por otros investigadores, con sus respectivos hallazgos y conclusiones.

¹ Ingeniero de Sistemas, Magister en Gobierno IT, profesor investigador en la Universidad Simón Bolívar – Email: rcabeza1@unisimonbolivar.edu.co

*Este artículo es asociado al proyecto llamado: localización de datos de contactos personales utilizando técnicas de minería web y redes sociales

I. INTRODUCCIÓN

Hoy en día, la Búsqueda de personas en la web es una actividad muy popular en línea que está recibiendo cada vez más apoyo en búsquedas especializadas como son los motores de búsqueda de personas. Además del motor de búsqueda de Yahoo ofrece un servicio de búsqueda de personas, así como pipl, Spokeo, duckduck etc. cuya actividad principal es buscar personas en la web. Un problema importante que enfrenta todos estos motores de búsquedas es cuando el nombre de persona es muy común debido al hecho de que puede referirse a diferentes individuos, entonces la idea es que en esta guía la persona pueda encontrar a quien busca por medio de un grupo específico, por ejemplo la institución educativa, la ciudad, o amistades en común.

Las bases investigativas para la realización de un proyecto en la universidad Simón Bolívar que tiene como fin localizar datos de personas con características específicas de la sociedad utilizando las redes sociales, sirviendo como herramienta para reconocer la actualidad de los proyectos y metodologías que ya se están aplicando y desarrollando en otras partes del mundo; una investigación basada en el localizar datos de personas con características específicas de la sociedad utilizando las redes sociales podría generar un gran impacto satisfaciendo una gran necesidad para muchas personas en el mundo.

APLICACIONES SIMILARES

Esta investigación está enfocada en la ubicación de personas para ello inicialmente vamos a utilizar el estudio que se vienen realizando en cuanto a la MINERÍA WEB, los cuales se están utilizando para la toma de decisiones y se han convertido en una herramienta de gran importancia para el mejoramiento de la red. La minería de datos se distribuye en tres paradigmas los cuales son: minería de contenidos, minería estructurada y minería de uso, es de gran importancia la minería de uso la cual trata la iteración de la personas con la web. La minería de contenidos y de estructuras se obtiene de la recolección de datos por medio MOTORES DE BÚSQUEDA que son la forma principal que permite a los usuarios de Internet encontrar sitios con información. Existen tres tipos de motores de búsqueda, los basados en crawlers, los basados en índices o directorios, y las metas motores de búsqueda. Estos se diferencian por la forma como organizan la información y los enlaces a las páginas que se encuentran en Internet [1]. Las búsquedas se priorizaran en la Webs 2.0, las cuales son

actualizadas constantemente por los contenidos de los usuarios, entre las que se encuentran las REDES SOCIALES (Facebook, Twitter, Instagram, entre otras.), que según el concepto actual; se refieren a la que utilizamos virtualmente. Pero si entramos a analizar en sí, el concepto tradicional de las redes sociales el cual es la interacción de personas con un interés común, la tendencia a buscar apoyo de otras personas empezando por la familia, amistades de barrio, trabajo etc; y que con el uso de las TIC y la expansión del internet se generó una nueva forma de comunicación mundial trayendo consigo el concepto que hoy en día asociamos a las redes sociales. Siendo así, considerado como uno de los fenómenos sociales más interesantes del siglo. Se estima que una de cada siete personas utiliza redes sociales en el mundo, es por ello, que se convierte en una herramienta importante para desarrollar el estudio propuesto. A la fecha se han venido realizando importantes investigaciones con respecto a la búsqueda de personas en la red. Se destaca el utilizado en la Universidad Carlos III de Madrid (UC3M) donde han desarrollado un importante algoritmo llamando SoSaCO (que es) para la búsqueda en redes sociales basado en el comportamiento de las hormigas para encontrar comida. Los detalles se publican en la revista Applied Intelligence[2].

Una de las principales cuestiones técnicas en el ámbito de las redes sociales, cuyo uso cada vez está más generalizado, consiste en localizar la cadena de referencia que lleva de una persona a otra, de un nodo a otro.

El mayor reto que se plantea en este ámbito es el enorme tamaño de estas redes y que la respuesta debe ser rápida, dado que el usuario final espera resultados en el menor tiempo posible. Y la solución parece venir de la mano del algoritmo SoSACO, que acelera la búsqueda de caminos entre dos elementos pertenecientes al grafo, es decir, la representación mediante nodos y enlaces de las relaciones de una red social [2].

El funcionamiento de SoSACO se inspira en el comportamiento que ha perfeccionado a lo largo de miles de años uno de los insectos más disciplinados del planeta a la hora de buscar comida.

En general, **los algoritmos de colonias de hormigas imitan cómo estas son capaces de encontrar el camino entre el hormiguero y la fuente de alimento**, mediante la deposición y seguimiento de un rastro químico denominado feromonas.

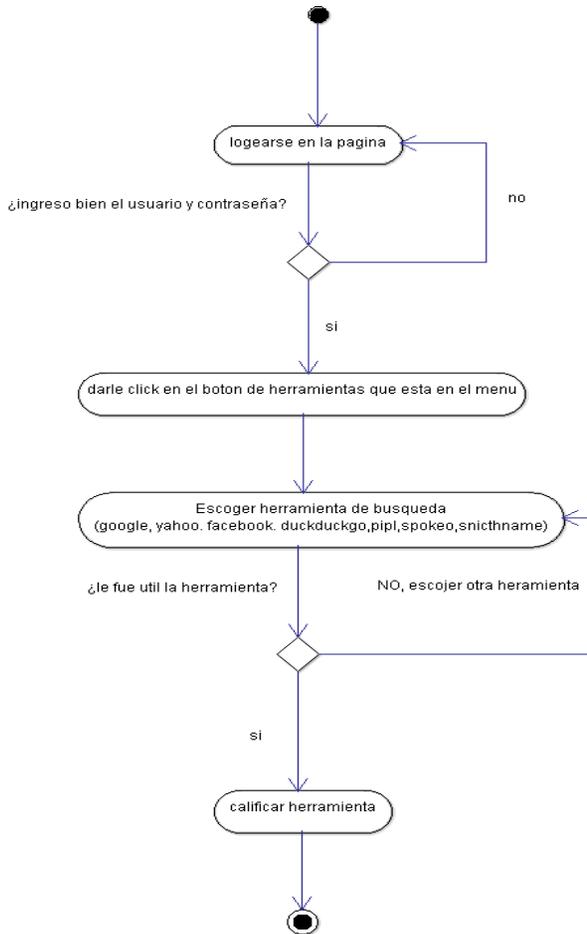


Fig 1. Algoritmo de colonias de Hormigas

"En este estudio se incorporan además otros rastros olorosos para que las hormigas puedan seguir tanto la feromona como el aroma de la comida, con lo que consiguen encontrar la fuente de alimento de forma mucho más rápida", aclaran los autores.

"Los primeros resultados muestran que la aplicación del algoritmo a redes sociales reales consigue obtener una respuesta óptima en muy poco tiempo (decenas de milisegundos)", indica Jessica Rivero, coautora del trabajo [2].

También muchas personas han querido demostrar el concepto de la teoría de *los seis grados de separación* que fue descrita en el año de 1930 por el escritor húngaro Frigyes Karinthy, en donde argumentaba que cualquier persona en el planeta podía estar conectada a otra persona a través de una cadena de eslabones o conocidos que no tenía más de cinco intermediarios.

La idea se basaba en que el número de conocidos crecía exponencialmente con el número de enlaces en la cadena, y que sólo son necesarios un pequeño número de enlaces para que el conjunto de conocidos se convierta en la población humana entera.

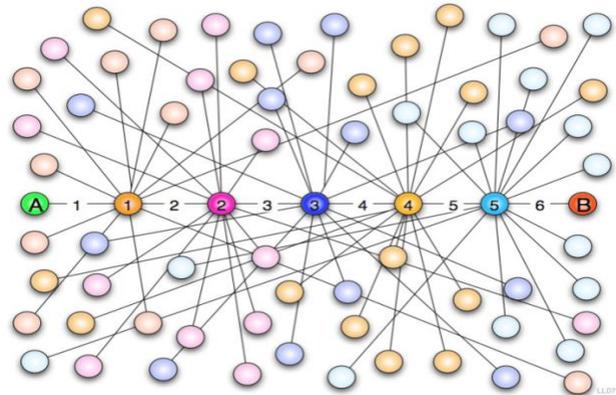


Fig 2. Anatomía

Es el ejemplo en donde en el año del 2011 la empresa Facebook realizó un estudio denominado "Anatomy of Facebook" con todos los usuarios activos de su página en ese momento 721.000.000 miembros (alrededor del 10% de la población mundial) y se analizó el conjunto de amigos en común, para sacar el promedio de cuántos eslabones hay entre cualquier usuario y otro cualquiera. De esta prueba se excluyó a celebridades y famosos. Los resultados mostraron que el 99,6% de pares de usuarios estuvieron conectados por 5 grados de separación. Esta es la prueba más cercana de la teoría a la fecha de hoy y da un resultado aproximado de 4,75 eslabones. [3]

También se realizó una investigación de BÚSQUEDA POR CONTEXTO PARA LA WEB en la cual su autor describe que "La Búsqueda por Contexto surge como la herramienta apropiada para explotar el conocimiento de la interacción web y usuario. Ella define modelos que describen los patrones de uso y caracterizan los perfiles de los usuarios de la web. Nosotros proponemos un Sistema Inteligente de Búsqueda por Contexto compuesto por una ontología para modelar el marco conceptual de búsqueda en la Internet, un mecanismo de razonamiento basado en reglas para usar la ontología e inferir patrones de comportamiento sobre Internet, y meta datos para almacenar instancias de información de la Internet. El Sistema de Búsqueda obtiene un conjunto de criterios del usuario cuando accede al sitio, y en base a esos criterios (preferencias y áreas de interés del usuario) y a la información solicitada por él en un momento dado, establece que será bus-cado, para tratar de hacer más eficaz ese proceso [4].

En la **MINERÍA DE DATOS** continuamente se utilizan las siguientes técnicas. Como ya se ha comentado, las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados. Las técnicas más representativas son:

Redes neuronales. Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida.

II. OTROS TIPOS DE BUSQUEDA

El perceptrón: El caso más sencillo de Red Neuronal (RN) es el que presenta una sola neurona de cómputo. A esta estructura se le denomina perceptrón y su estudio resulta obligado antes de profundizar en redes neuronales más complejas [5].

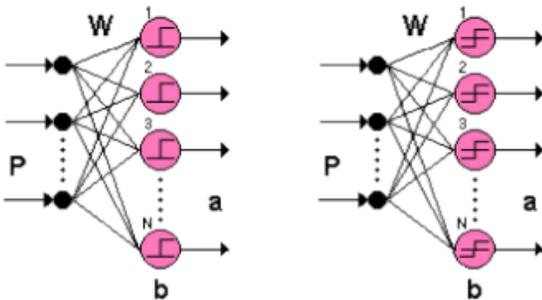


Fig 3. Conexiones del perceptrón

El perceptrón multicapa

Un Perceptrón multicapa es una red con alimentación hacia delante, compuesta de varias capas de neuronas entre la entrada y la salida de la misma, esta red permite establecer regiones de decisión mucho más complejas que las de dos semiplanos, como lo hace el Perceptrón de un solo nivel.

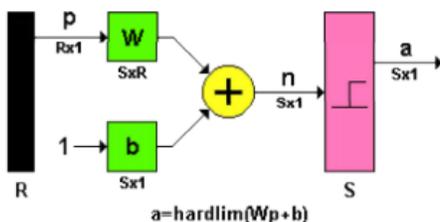


Fig 4. Notación compacta para la red tipo perceptrón

Notación compacta para la red tipo Perceptrón. La salida de la red está dada por:

Donde:

W: Matriz de pesos asignada a cada una de las entradas de la red de dimensiones $S \times R$, con S igual al número de neuronas, y R la dimensión del vector de entrada

p: Vector de entradas a la red de dimensiones $R \times 1$

b: Vector de ganancias de la red de dimensiones $S \times 1$

Capacidades del Perceptrón multicapa con dos y tres capas. En la segunda columna se muestra el tipo de región de decisión que se puede formar con cada una de las configuraciones, en la siguiente se indica el tipo de región que se formaría para el problema de la XOR, en las dos últimas columnas se muestran las regiones formadas para resolver el problema de clases mezcladas y las formas más generales para cada uno de los casos.

Estructura	Tipo de región de decisión	Problema del OR-Exclusivo	Clases lin. no separables	Formas más generales
Una capa	Zonas separadas por hiperplanos			
Dos capas	Zonas convexas			
Tres capas	Zonas de complejidad arbitraria			

El Perceptrón básico sólo puede establecer dos regiones separadas por una frontera lineal en el espacio de entrada de los patrones; un Perceptrón con dos capas, puede formar cualquier región convexa en este espacio. Las regiones convexas se forman mediante la intersección de regiones formadas por cada neurona de la segunda capa, cada uno de estos elementos se comporta como un Perceptrón simple, activándose su salida para los patrones de un lado del hiperplano, si el valor de los pesos de las conexiones entre las neuronas de la segunda capa y una neurona del nivel de salida son todos igual a 1, y la función de salida es de tipo hardlim, la salida de la red se activará sólo si las salidas de todos los nodos de la segunda capa están activos, esto equivale a ejecutar la función lógica AND en el nodo de salida, resultando una región de decisión intersección de todos los semiplanos formados en el nivel anterior. La región de decisión resultante de la intersección será una región convexa con un número de lados a lo sumo igual al número de neuronas de la segunda capa [6].

Regresión lineal: Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente

en espacios multidimensionales donde puedan relacionarse más de 2 variables.

El análisis de regresión es una técnica estadística para investigar la relación funcional entre dos o más variables, ajustando algún modelo matemático. La regresión lineal simple utiliza una sola variable de regresión y el caso más sencillo es el modelo de línea recta. Supóngase que se tiene un conjunto de n pares de observaciones (x_i, y_i) , se busca encontrar una recta que describa de la mejor manera cada uno de esos pares observados. [7].

Árboles de decisión: Dentro del área de la AI, y más exactamente dentro de la sub-área del aprendizaje automático se encuentran los árboles de decisión. Son utilizados principalmente con propósitos de clasificación, pero son también útiles para descubrir características de los datos que no son directamente visibles. Por este motivo, los árboles de decisión son importantes tanto en aplicaciones de clasificación como de minería de datos. Un árbol de decisión es en esencia un modelo predictivo, esto es, permite que las observaciones acerca de las características de un elemento conduzcan a conclusiones acerca de un valor objetivo. La técnica de aprendizaje automático que permite la inducción de un árbol de decisión a partir de un conjunto de datos se llama aprendizaje de árboles de decisión.

Si se hace referencia a los árboles de decisión como una técnica, se dice que son un método para aproximar funciones objetivo de valor discreto.

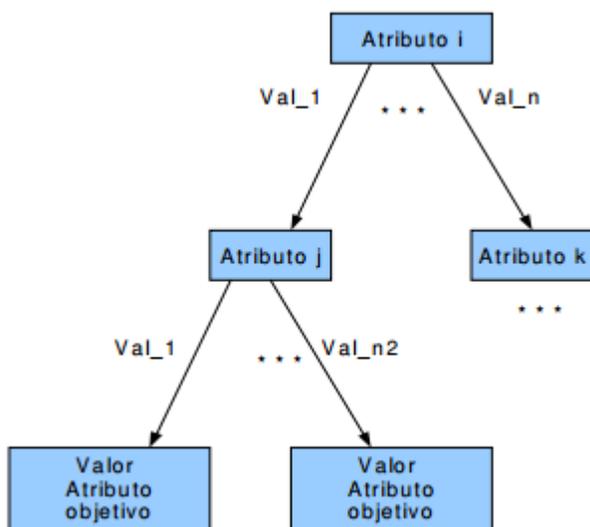


Fig 5. Estructura de un árbol de decisión. [8]

III. ALGORITMOS UTILIZADOS EN ESTE CONTEXTO

Algoritmo ID3: es utilizado dentro del ámbito de la inteligencia artificial. Su uso se engloba en la búsqueda de hipótesis o reglas en él, dado un conjunto de ejemplos.

El algoritmo id3 realiza un proceso recursivo sobre todas las ramas del árbol generado, tal y como se refleja en la fig 3. Inicialmente, se le llama con el conjunto de ejemplos, el de atributos, y el nodo raíz del árbol de decisión que estará vacío. Como ya se ha comentado, el proceso se realiza hasta que todos los ejemplos de la rama en cuestión pertenecen a una única clase. Sin embargo, esto no siempre es posible. En primer lugar, puede ocurrir que se hayan agotado todos los atributos y, sin embargo, sigan existiendo ejemplos con distintos valores de clase. De otra parte, también puede suceder que, una vez elegido un atributo para un nodo de decisión, no exista ningún ejemplo para una de las ramas generadas por dicho atributo. En esos casos, se etiqueta el nodo hoja con la clase mayoritaria. En los otros casos, se selecciona un atributo de acuerdo a la heurística definida anteriormente. Por cada valor del atributo escogido se crea un nodo sucesor de N etiquetado con el valor correspondiente del atributo, a través de la función crear-nodo. Ese nodo se añade al conjunto de nodos sucesores del nodo raíz N y finalmente, se llama recursivamente a id3. La llamada recursiva se hace con el conjunto de ejemplos que tienen ese valor concreto para el atributo escogido A_i , el conjunto de atributos menos el que acabamos de utilizar, y el nodo que acabamos de crear[9].

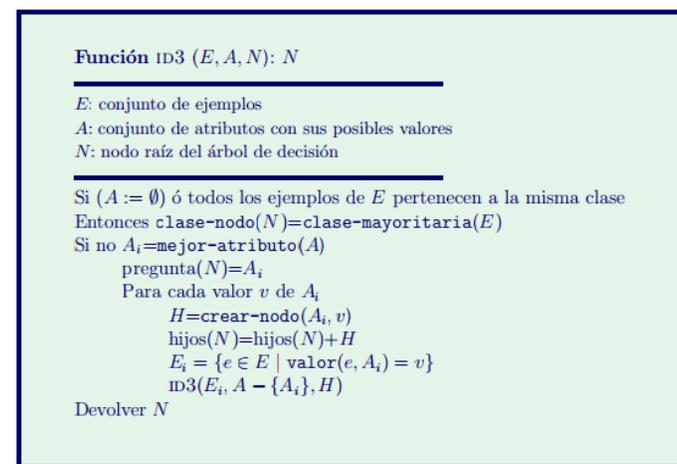


Fig 6. Procedimiento ID3.

Algoritmo C4.5: es un algoritmo usado para generar un árbol de decisión desarrollado por Ross Quinlan.¹ C4.5 es una extensión del algoritmo ID3 desarrollado anteriormente por Quinlan. En esta sección, se presenta el algoritmo de conglomerados en la búsqueda de C4. Detalles y pruebas de seguimiento. Muchos lectores sólo

se quiere una breve explicación del algoritmo y le sugerimos que examinen el diagrama de flujo del algoritmo dado en la Figura 2 y lea esta sección Información general. Para aquellos que desean más detalles, cada paso se describe con más detalle a lo largo del resto de esta sección. Se discute la aplicación del algoritmo de C4 a los datos de SDSS en la Sección 7.

El algoritmo C4 comienza colocando cada galaxia en un siete dimensiones del espacio de ascensión recta (AR), la declinación (DEC), el corrimiento al rojo y cuatro dimensiones de color (u-g, g-r, r-i, i-z). En cada una de estas galaxias objetivo, entonces realizamos los siguientes pasos:

1. Ponemos una abertura alrededor de cada objetivo que sólo incluir las galaxias en un rango específico y desplazamiento hacia el rojo. Luego, medimos la probabilidad de que cada galaxia dentro de esta abertura espacial tiene colores igual a la galaxia de destino. Las probabilidades son sumadas para obtener una "cuenta de número". [9].

Modelos estadísticos: Los modelos estáticos se ocupan de determinar una respuesta para una serie especial de condiciones fijas que probablemente no cambiarán significativamente a corto plazo es decir, la solución esta basada en una condición estática.[10]

Agrupamiento o Clustering: El problema de formar grupos en un conjunto de datos es muy importante para el conocimiento del comportamiento de una población de la cual solo se tiene una cantidad n de sus elementos. La solución de estos problemas se realiza mediante la creación de algoritmos de agrupamiento.

Entre los métodos de agrupamiento paramétricos se encuentran las mixturas finitas, éstas son una poderosa herramienta para modelar densidades de probabilidad de conjuntos de datos univariados y multivariados, modelan observaciones las cuales se asume que han sido producidas por un conjunto de fuentes aleatorias alternativas e infieren los parámetros de estas fuentes para identificar qué fuente produjo cada observación, lo que lleva a un agrupamiento del conjunto de observaciones.

Los métodos de agrupamiento no paramétricos pueden dividirse en tres grupos fundamentales: jerárquicos, particionales y basados en densidad.

Los algoritmos jerárquicos son aquellos en los que se va particionando el conjunto de datos por niveles, de modo

tal que en cada nivel generalmente , se unen o se dividen dos grupos del nivel anterior, según si es un algoritmo aglomerativo o divisivo.

Los algoritmos particionales son los que realizan una división inicial de los datos en grupos y luego mueven los objetos de un grupo a otro según se optimice alguna función objetivo.

Los algoritmos basados en densidad enfocan el problema de la división de una base de datos en grupos teniendo en cuenta la distribución de densidad de los puntos, de modo tal que los grupos que se forman tienen una alta densidad de puntos en su interior mientras que entre ellos aparecen zonas de baja densidad[11]

Algoritmo K-means: Es uno de los más simples y conocidos algoritmos de agrupamiento, sigue una forma fácil y simple para dividir una base de datos dada en k grupos (fijados a priori). La idea principal es definir k centroides (uno para cada grupo) y luego tomar cada punto de la base de datos y situarlo en la clase de su centroide más cercano. El próximo paso es recalcular el centroide de cada grupo y volver a distribuir todos los objetos según el centroide más cercano. El proceso se repite hasta que ya no hay cambio en los grupos de un paso al siguiente.

El problema del empleo de estos esquemas es que fallan cuando los puntos de un grupo están muy cerca del centroide de otro grupo ver ejemplo en, también cuando los grupos tienen diferentes tamaños y formas [11].

Algoritmo K-medoids: El algoritmo K -means es sensible a los valores atípicos desde un objeto con un extremadamente valor grande puede distorsionar sustancialmente la distribución de datos. ¿Cómo podría modificarse el algoritmo para disminuir tal sensibilidad? En lugar de tomar el valor medio de los objetos en un grupo como punto de referencia, un medoide puede ser utilizado, que es el objeto más céntrico de un cluster. Así, el método de partición puede ser todavía realizado basado en el principio de reducir al mínimo la suma de las diferencias entre cada objeto y su correspondiente punto de referencia. Esto forma la base del método K -medoids. La estrategia básica de algoritmos de agrupamiento k - Medoids es encontrar grupos de k en n objetos por primera búsqueda arbitrariamente un representante objeto (los medoids) para cada grupo. Cada restante objeto está agrupado con la medoide a la que es la más similar .Método K- medoids utiliza representante objetos como puntos de referencia en

lugar de tomar la media valor de los objetos en cada grupo. El algoritmo toma el parámetro de entrada k, el número de grupos sea dividido entre un conjunto de n objetos. Un algoritmo de k - Medoids típico para una partición basada en medoids u objetos centrales es el siguiente:

k- medoids es una técnica de partición clásica de la agrupación que agrupa el conjunto de datos de n objetos en k grupos conocidos a priori. Una herramienta útil para determinar k es la silueta.

Es más robusto al ruido y los valores atípicos en comparación con k-medias, ya que minimiza una suma de diferencias por pares en lugar de una suma de las distancias euclidianas al cuadrado.

k- medoids se puede definir como el objeto de un clúster, cuyo promedio de disimilitud a todos los objetos del clúster es mínima, es decir, es un punto más céntrico de la agrupación [12].

Reglas de asociación: Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.

Según el objetivo del análisis de los datos, los algoritmos utilizados se clasifican en supervisados y no supervisados. Algoritmos supervisados (o predictivos): predicen un dato (o un conjunto de ellos) desconocido a priori, a partir de otros conocidos.

Algoritmos no supervisados (o del descubrimiento del conocimiento): se descubren patrones y tendencias en los datos.

Tenemos planteado utilizar el demo del programa MaltegoCE v3.4.0 el cual sirve para realizar una minería de datos empleando un motor de búsqueda que permite ubicar los correos y datos de personas y empresas.

IV. CONCLUSIONES

Los autores consultados coinciden en la importancia de la tecnología para el desarrollo de herramientas o aplicaciones que localicen datos de personas con características específicas de la sociedad. Debido a la amplia gama de aplicaciones que ofrece las TIC, esto se hace aún más atractivo para entidades cuyo se ve reducido debido a diferentes motivos, como la falta de conocimiento de las mismas.

Teniendo en cuenta la constante negativa que gran porcentaje de pequeños empresarios pueden tener con este tipo de tecnologías, algunas veces por el simple hecho

de desconocerlas, por lo que creemos que además de la creación de una herramienta, es necesaria la creación de metodologías pedagógicas que logren instruir a las personas o entidades interesadas sobre lo que son la minería de datos y las grandes ventajas a las que puede acceder.

Algunos autores hacen referencia de manera especial al uso de páginas web para localizar datos de personas, por ser esta una herramienta que bien estructurada, puede lograr ser un puente directo con los clientes. Además de que dan una serie de recomendaciones que podrían llevar al éxito de su aplicación.

Dado que el objetivo de este proyecto es innovar y buscar nuevas formas ubicar o localizar datos de personas utilizando las redes sociales, se propone el desarrollo de una aplicación en el que se localice datos de personas, esto con el fin de hacer que el usuario final del producto se vea más interesado por el producto desde el momento en el que lo vea presentado en el dicha aplicación.

REFERENCIAS

[1] *Cómo buscar a una persona por Internet: Pasos para mejorar la búsqueda*, <http://www.respuestario.com/como/como-buscar-a-una-persona-por-internet-pasos-para-mejorar-la-busqueda>, [Consulta: martes, 03 de junio de 2014]

[2] J. Rivero, *Crean un buscador de redes sociales inspirados en hormigas*, [Http://www.muyinteresante.es/tecnologia/articulo/crean-un-buscador-de-redes-sociales-inspirados-en-hormigas](http://www.muyinteresante.es/tecnologia/articulo/crean-un-buscador-de-redes-sociales-inspirados-en-hormigas), [Consulta: martes, 03 de junio de 2014]

[3] J. Mot, *Cómo serán las redes sociales en el futuro*. <http://demuchounpoco.com/como-seran-las-redes-sociales-en-el-futuro/>. [Consulta: martes, 03 de junio de 2014]

[4] J. Aguilar, *Sistemas semánticos para la búsqueda inteligente de información por contexto para la web* <http://erevistas.saber.ula.ve/index.php/cienciaingenieria/article/view/3536>. [Consulta: martes, 03 de junio de 2014]

- [5] *El Perceptrón*
<http://148.204.64.201/paginas%20anexas/POO/EL%20PERCEPTRON.htm> [Consulta: martes, 03 de junio de 2014]
- [6] J. Rodríguez Blázquez, *Redes Multicapas*,
<ftp://decsai.ugr.es/pub/usuarios/castro/Actividades/Redes-Neuronales/Apuntes/Apuntes%20Javier%20Rodriguez%20Blazquez/Redes%20Multicapa%20-%20Algoritmo%20Backpropagati> [Consulta: martes, 03 de junio de 2014]
- [7] I. Valdez, Alfaro, *Regresión lineal*.
<http://www.dcb.unam.mx/profesores/irene/Notas/Regresion.pdf>
- [8] L. Jimenez Moscovitz, “Un Modelo Conceptual para el Desarrollo de Arboles de Decisión” Fundación Universitaria Konrad Lorenz. Dr. Nelson Obregon Neira
- [9] I. Kononenko and E. Simec. “Induction of decision trees using relief”, 1995. [Kononenko and Simec, 1995]
- [10] G. Hernandez Coca, “Tipos de Modelos en Investigación de Operaciones” [Julio – Diciembre 2011] Universidad Autónoma del Estado de Hidalgo.
- [11] D. Pascual 1, F. Pla 2, S. Sánchez 2, “Algoritmos de agrupamiento” Departamento de Computación .Universidad de Oriente Avda. Patricio Lumumba S/N. 90500 Santiago de Cuba, Cuba. dpascual@csd.uo.edu.cu
- [12] T. Velmurugan and T. Santhanam “Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points”, Department of Computer Science, DG Vaishnav College, Chennai, India