

BÚSQUEDA ESPACIAL

SPATIAL SEARCH

Luzmila Pró, Carlos Yañez, Juan Carlos Gonzales *

RESUMEN

La búsqueda espacial es un tipo de búsqueda local mediante el internet, por el cual se puede ubicar dentro de un área de espacio o distancia a un centro específico de referencia, a este tipo de búsqueda con respecto a un punto de referencia se le denomina búsqueda aproximada geoespacial (1).

La base de datos necesaria para esta búsqueda por aproximación requiere una preparación de los datos en varios pasos:

- El primer paso consiste en localizar el texto que la aplicación geográfica puede mapear por medio de una localización física, este paso es fácilmente para aplicaciones de directorio como los de las páginas amarillas por internet, en este caso dado que está conformado de una simple estructura de base de datos que definen significativamente para cada campo.

Para derivar el contenido por internet el problema es complejo porque el texto con significado geográfico puede ser cualquiera sobre una página.

- El segundo paso es referente a la transformación de la designación textual dentro de las coordenadas físicas como el espacio geográfico como es la del espacio tierra, luego para la detección geográfica del contenido en el texto de los documentos, el lado de la barra describe el tópico de detección del contenido geográfico entre el texto de los documentos, el cual es de interés comercial e investigación (7).

Palabras clave: Geobúsqueda, Búsqueda Espacial, Mapear, Coordenadas Geográficas, Geocódigos, Algoritmos de Búsqueda, Query Aproximado, Query Geográfico.

ABSTRACT

The spatial search, is an type of search local with internet, which local search to locate into an area of space o distance an specific center of reference, at this type search with point reference known geospatial proximity searching (1).

The data base requires for this search preparing of data in several steps:

The first step is to locate text that the application geographic can be map for the phisical location, this step, is easily for aplicacions of the directory as equal the yellow pages for internet, in this caseas this conformed of a structure simple of data base with defined maenings for each field.

For Intenet derived the content the problem is complex because the text wirth geographic significance can be anywhere on a page.

The second step is referent at the transform location of the textual designation into the physical coordinates on the earth´s surface, in order for the detection geographicof the content in the text of the documents, sidebar describes, topic of the detectinggeographic into the text documents has generated interest of business and the investigation (7).

Key words: Geosearch, Spatial Search, Map, Geographics Coordinates, Searching Algorithms, Aproximatly Query, Geographic Query.

* Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos, Lima - Perú.
E-mail: lproc2003@hotmail.com, cenyanes@yahoo.es, jgonzale@aasa.com.pe

1. INTRODUCCIÓN

La búsqueda espacial es un tipo de búsqueda local mediante el internet, por la cual se puede ubicar dentro de un área de espacio o distancia a un centro específico de referencia. A este tipo de búsqueda con respecto a un punto de referencia se le denomina búsqueda aproximada geoespacial.

La base de datos necesaria para esta búsqueda por aproximación requiere una preparación de los datos en varios pasos:

- Primer paso consiste en localizar el texto que la aplicación geográfica puede mapear por medio de una localización física, este paso es útil y fácil para aplicaciones de directorio como los de las páginas amarillas por internet. En este caso dado que está conformado de una simple estructura de base de datos que definen significativamente para cada campo.

Para derivar el contenido por internet el problema es complejo porque el texto con significado geográfico puede ser cualquiera sobre una página.

- El segundo paso es referente a la transformación de la designación textual dentro de las coordenadas físicas como el espacio geográfico como es la del espacio tierra, luego para la detección geográfica del contenido en el texto de los documentos, el lado de la barra describe el tópico de detección del contenido geográfico entre el texto de los documentos, el cual es de interés comercial y de interés por el lado de la investigación.

En el mundo desarrollado una dirección de una calle o postal es el camino más común para referir a la ubicación y particularmente a la búsqueda local. Las aplicaciones con geocódigos intentan resolver un grupo de señales dentro de un par de coordenadas, un geocódigo retornado.

Los mejores geocódigos son exactos entre unos pocos metros, usualmente las coordenadas específicamente menores se refieren al centro de una gran región.

La base de datos de los geocódigos son grandes y dinámicas, en este caso las calles dentro de la red que representan en el trabajo.

Eficientemente procesando la proximidad de las consultas nuevas dentro de un conjunto de datos grandes, tales como el ancho y largo de un país o ciudad, directorios comerciales de unos 14 millones de negocios o profesionales o que el internet requiere un método de acceso espacial.

La idea básica de SAMs es para realizar mapas bidimensionales (o n-dimensionales), mediante el

sistema de coordenadas, en este caso la latitud y la longitud sobre un sistema de coordenadas de dimensión simple.

Por ejemplo si tomamos una región de la superficie de la tierra puede ser demostrado con un atributo simple a llave espacial referida, de los cuatro atributos que son necesarios para describir el área de un rectángulo: las coordenadas x e y del rectángulo, en la parte izquierda y los de la parte de la derecha.

Las llaves espaciales son calculadas como el conjunto de datos para ser geonavegables al construir. En el caso de un directorio comercial, la llave espacial para cada uno de los negocios es guardada o almacenada como un campo adicional a lo largo con los otros campos para los negocios.

Si se aplica una búsqueda indexada sobre un texto no estructurado, éste adiciona las llaves espaciales a un término adicional para indexar y construir la página.

Un texto estructurado, tal como el de las páginas web, puede requerir de varias llaves espaciales porque ellos pueden contener diferentes direcciones.

El tiempo de búsqueda para determinar el conjunto de negocios sobre páginas web, que satisface un query aproximado, la aplicación de búsqueda en los mapas de los usuarios que no hacen queries geográficos, en términos de que ellos pueden ser comparados para el precálculo de llaves espaciales con el conjunto de datos que están siendo buscados.

La aplicación de búsqueda se refiere a los usuarios no geográficos en términos para determinar en cuáles de los datos está el radio, y para ello requiere hacer un emparejamiento o matching con el usuario principal.

Los algoritmos de búsqueda aproximados pueden ordenar resultados por distancia, así como resultado cerrado para la búsqueda central como listado de lugares remotos.

Ordenando la rutina IYP (Aplicaciones en internet y páginas amarillas), esta ordenación puede ser compleja para las páginas web porque los números de estas páginas son potencialmente grandes, y es necesario clasificarlos para su utilización.

2. ANTECEDENTES

En 1998, un grupo de investigadores de Vicinaty desarrolló un sistema prototipo para asuntos geográficos.

El contenido de la web Vicinaty modificando los métodos de acceso espacial, ha desarrollado para estos IYP

y negocios, que permite la localización de productos, trabajos, esto como un componente de software en conjunción con aplicaciones de búsqueda.

En 1999, el mismo equipo de investigadores de Vicinaty llevó a cabo un experimento para incluir los cuerpos de la web general.

En el 2002 Microsoft, con respecto a la búsqueda geoespacial, compró a Vicinaty. Y se publicó desde abril de 2000 hasta marzo de 2002.

En este tiempo la búsqueda ya procesaba hasta 300 millones distintos de páginas web (3).

Las experiencias con la geobúsqueda proveen a las bases de datos para manejar dos tipos de observaciones:

- La Internet es una fuente rica de contenido local.
- La información local sobre la Internet posee ciertas características que simplifican, el trabajo de agregación.

La idea básica de geobúsqueda es que este transformador local de información en el documento de texto dentro de una forma de búsqueda de ingeniería puede ser eficientemente de búsqueda próxima. Estos primeros pasos como para escanear documentos utilizados en el reconocimiento de patrones que representan entidades geocodificables.

La geobúsqueda aborda semánticamente análisis de textos preferentemente para dejar libre la determinación de un documento materia de estudio para el análisis de información de algoritmos de aplicación de búsqueda referente sobre el hecho que una significativa porción del contenido de ese valor para la búsqueda local, contiene direcciones bien formadas, direcciones postales, números telefónicos o ambos.

La presencia de una o más direcciones es una indicación acerca de un documento presentado por los diseñadores de aplicación de búsqueda con algoritmos de ranking relevante.

Una ventaja para esta aproximación es que esa geobúsqueda es portable. Este es un software componente que es insertado a un punto conveniente dentro de una aplicación de búsqueda o workflow.

3. FUNDAMENTOS TEÓRICOS

3.1. La Geobúsqueda

La Geobúsqueda es un tipo de búsqueda local mediante el internet, por el cual se puede ubicar dentro de un área de espacio o distancia a un centro espe-

cífico de referencia. A este tipo de búsqueda con respecto a un punto de referencia se le denomina búsqueda aproximada geoespacial.

3.2. Reconocimiento de direcciones

El reconocimiento de direcciones en la geobúsqueda detecta las direcciones que la conforman consistiendo de un código, una porción significativa del contenido de ese valor para la búsqueda local y contiene direcciones bien formadas como las direcciones postales, números telefónicos o ambos.

3.3. Observaciones de Geobúsqueda

Por dos años la Geobúsqueda fue publicada y disponible y por los años siguientes, Vicinaty que investigó las técnicas usadas para realizar la observación del internet y la identidad de derivada de la búsqueda local, así como sus fortalezas y debilidades y las perspectivas futuras. Sin embargo a gran escala, el concepto de geobúsqueda está excediendo las expectativas (5).

3.4. Los datos locales difundidos por la web

Cuando Vicinaty investiga sobre el desarrollo de Geobúsqueda en el año 1998, evaluó el conjunto de páginas web que constaba de varias investigaciones de ingeniería y portales populares. Sobre una base consistente de más del 20 por ciento, esas páginas contenían direcciones tanto de Canadá como de Estados Unidos, y números telefónicos.

El porcentaje restante fue casi constante por los dos años de Geobúsqueda disponible sobre el internet, sin embargo la geobúsqueda solamente observa las direcciones de Norteamérica, las páginas examinadas fueron no reestrictas para Norteamérica. Además el porcentaje de páginas con las direcciones de las ciudades están por sobre el 20 por ciento de geobúsqueda encontrada (2).

3.5. Las Direcciones bien formadas son la ruta, no la excepción

La eficiencia del reconocimiento de las direcciones procesadas fue referente a todos, concerniente para todos los grupos de ingeniería, las investigaciones de Vicinaty.

En muchos casos las direcciones sobre las páginas web conforman al código postal estándar usado para deliberar y enviar e-mails.

Ocasionalmente, un código postal que un grupo de direcciones que comparte, fue agrupado o individualizado y el lugar del inicio de la tabla.

Los números telefónicos excluidos de la Geobúsqueda, se encontraban al final de las direcciones haciendo un aproximado de 15 por ciento de páginas.

Algunas empresas usan el número de teléfono como el primer punto de contacto. Por ejemplo servicio de un área geográfica y que confían sobre un número de teléfono, además que se guardan para almacenar la presencia local estable.

Alguna de las direcciones reconocidas encontraron un número de teléfono, pero no está acompañado de dirección.

En esos casos la presencia de un número de teléfono puede ocasionar el resaltar la forma más intensiva, escaneando alrededor de todo el texto para una dirección.

El resultado básico fue favorable, sin embargo puede no ser considerado adicional de trabajo sobre las direcciones reconocidas para una alta prioridad.

3.6. Direcciones son las llaves para un contenido exógeno y valioso

La mayoría de negocios a través de la dirección proveen bastante información para construir una imagen de un negocio local y familiar, de una vecindad, o usar como un índice para encontrar la localización de un mapa.

Una dirección no es directamente usada por la distancia computacional y el mapeo o el ruteo de las aplicaciones basados en computación localizada que requiera.

Este es el trabajo de las aplicaciones de geocódigos que asocia una dirección y una ubicación física de la superficie de la Tierra.

Las bases de datos para esas aplicaciones usan representación significativa del capital intelectual.

Ejemplo: La Red de productos de trabajo de datos geográficos de tecnología y un principal proveedor de base de datos de geocódigos, contiene un aproximado de más de 14 millones de direcciones de calles segmentadas postales y población de censos y tiendas comerciales importantes.

La compañía procesa más de un millón de cambios para esta base de datos, en forma mensual. (www.geographic.com/home/productstandservices.cfm).

Una dirección es la llave que asocia valiosa información exógena con el contenido de la web.

Las direcciones son proporcionalmente evaluadas para la búsqueda local porque son fácilmente computables para detectar (7).

3.7. Las direcciones son metadatos

La www es una asociación de meta datos para una máquina de baja estandarización de información para la web (www.w3.org/metadata).

Se incorporan meta datos por fechas dentro de la ingeniería de búsqueda y la métrica relevante. Los Meta tags HTML son ignorados porque ambos son mal usados o los meta datos estándar no tienen valor mas si son desagregados.

3.8. Las páginas tienden hacia la buena calidad

La búsqueda local es más importante en el aspecto de negocios o comercial, es decir a través del internet mucha gente puede realizar compras o transacciones.

Algunas personas realizan sus compras por necesidad o a veces por ser de preferencia, que generalmente los productos son proporcionados mediante catálogos, o por televisión o radio u otro tipo de distribución de información en el medio.

Las características de los datos y sus contenidos en el internet es que éstos sean claros, uniformes, y a veces son escasos y lentos para cambiar.

Estos datos dirigidos o rutas de circuitos que son compilados al iniciar y finalizar en listados como las páginas amarillas por internet IYP.

La lista de compilación de los vendedores son tradicionalmente customizados para esos productos por negocios, correos, listas, venta que suministran a una IYP de datos.

La compilación de datos fuente es impresa, los directorios telefónicos, los cuales son convertidos para información digital con OCR dispositivos.

En Estados Unidos, la lista incluye un aproximado de 14 millones de negocios de 5200 directorios telefónicos (www.infousa.com) y registra alrededor de 17 millones de llamadas.

Los IDYP son directorios locales de negocios similar a los IYP, pero ricos en contenido. La esencial diferen-

cia es que la información en el IDYP es derivada directamente desde el internet no desde la fuente fuera de línea.

La presencia de Web. Organizaciones que tienen relaciones, navegación de confianza en ambos sean públicos y privados los cuales proveen información y son certificados o creados para ejecutar la regla establecida.

Dos ejemplos de organización que pueden servir como barreras y basados sobre una geografía basado sobre un mercado, tales como un lugar de comercio y basado sobre un segmento de mercado, tales como el funcionamiento de la organización.

Una función primaria de ambos tipos de representación de organizaciones es para comparar y distribuir información exacta sobre el interés de sus miembros. Ambos tipos de organización están a menudo acordes a la tecnología web, y no tienen uno propio sobre la presencia web.

Mientras que son 14 millones de negocios en Estados Unidos, la mayoría de ellos pequeños negocios y el funcionamiento de organizaciones en números en cientos (4).

Por otra parte se observa la presencia de información fraudulenta y con compromisos atrasados que comprometen la integridad de lo que constituye la información.

La principal función de una licencia autorizada es para proponer sobre parte de información para IDYP de un miembro sobre cierta información o miembros que ésta tiene directamente propuestos para los IDYP. La segunda función la seguridad es orientada como minimizar la cantidad de fraudes o errores de datos que tiene este camino sobre IDYP.

3.9. Seguridad

A diferencia de una búsqueda virtual pura, el asunto de que trata la seguridad es de una búsqueda local que tiene una exigencia física que puede ser confirmada. Es una búsqueda local más resistente a fraude, son las búsquedas virtuales.

El modelo IDYP si no es usuario autorizado y garantizado para negocios, pero no serán incluidos en el IDYP, este es un requerimiento o las reglas de juego del sistema que lo establece para los negocios que hacen uso de ella.

El potencial de Internet para proveer seguridad acerca de empresas locales que puedan exceder de los actuales directorios service.

No es posible garantizar sobre las páginas amarillas, proveer confiabilidad acerca de la reputación de los negocios.

Los IDYP, sin embargo, pueden crecer esta información con varios datos fuente, tales como: cuáles son buenos negocios, y qué son independientemente analizados los datos públicos. En adición el IDYP puede usar prácticas que tienen acogida popular sobre la internet para clasificar productos, servicios y ventas.

4. CONCLUSIONES

La Búsqueda Geoespacial permite realizar búsquedas locales en espacios aproximados de interés dado los puntos o centros de referencia, además ofrece ventajas para realizar aplicaciones, hay muchas expectativas futuras, como los sistemas de localización geográfica (GPS o Sistema de Posicionamiento Global, denominado originalmente NAVSTAR, que es un sistema de navegación por satélite) (GNSS) (6).

La Búsqueda Local, en fusión con otras tecnologías como Web Services, Sistema Inteligentes, etc., puede brindar aportes muy importantes a la sociedad en aplicaciones como servicios médicos de emergencia, servicios policiales, en casos de desastres de la naturaleza para brindar ayuda oportuna.

REFERENCIAS BIBLIOGRÁFICAS

1. HERRIGAN, J., Pew Internet Project, *pew internet proyect data memo*, 2004, www.pewinternet.org/pdfs/PIP_Broadband04.Data Memo.pdf
2. HERRIGAN, J., Pew Internet Project, *The broadband difference*, 2002, www.pewinternet.org/pdfs/PIP_Broadband_Report.pdf.
3. KELSEY GROUP & BIZRATE.com, *Local search now 25% of internet commercial activity*, 2004, www.kelseygroup.com/press/pr04211.htm.
4. KERNER S., *Majority of US consumers research online, buy online, buy offline*, USA, 2004, www.clicz.com/stars/markets/retailing/article.php/3418001.
5. STERLING, G., *Is 2004 the year of local search?*, 2003, www.imediacion.com/content/2343.asp.
6. http://es.wikipedia.org/wiki/sistema_de_posicionamiento_global, 2004.
7. MARTY HIMMELSTEIN, *Local search: The internet is yellow pages*, Feb 2005, Computer, Innovative Technology for Computer Professionals.