

Geocoder compuesto – solución híbrida en el mundo de ruido informativo

A. J. Florczyk¹, F. J. López-Pellicer¹,
P. Rodrigo¹, R. Rioja², P. R. Muro-Medrano¹

¹ Depto de Informática e Ing. de Sistemas
Universidad de Zaragoza

C/ María de Luna, 1, 50.018, Zaragoza, España
{florczyk, fjlopez, prodrig, prmuro}@unizar.es

² GeoSpatiumLab

C/ Carlos Marx, 4, local der., 50.015 Zaragoza, España
rodolfo@geoslab.es

Resumen

La creciente demanda del acceso a la información geográfica en cada tipo de actividad humana aumenta el número de servicios de georreferenciación accesibles vía Web. Los sistemas de información geográfica están presentes en nuestra vida cotidiana (traslado dentro de la ciudad, información meteorológica), ocio (viajes, información turística, redes sociales) o trabajo (viajes de negocios). Para ello utilizamos servicios Web de georreferenciación. Pero los servicios Web de georreferenciación se diferencian entre ellos en su nivel de detalle, cobertura, fiabilidad o precisión. En un escenario con tanta diversidad, son problemas importantes cómo buscar los servicios adecuados y, sobre todo, cómo evaluarlos para elegir el más apropiado para satisfacer a un usuario que no suele estar interesado en la procedencia de la información pero sí en su calidad.

Palabras clave: geocoder, infraestructuras de datos espaciales

1. Introducción

Hoy en día existen muchos proveedores de servicios de georreferenciación accesibles vía Web debido a la creciente demanda de acceso a información geográfica en contextos como la vida cotidiana (p.e. callejeros), el ocio (p.e. redes sociales), el trabajo (p.e. portales de viajes) y los servicios públicos (p.e. sistemas de emergencia). Independiente del caso de uso se necesita información geográfica ajustada a los requisitos. Las características básicas de cada uno de los servicios están formadas por el tipo de datos proporcionados (direcciones, puntos de interés¹, nombres históricos², etc.), la cobertura (España, municipio

¹http://www.nearby.org.uk/location_search.cgi

²<http://www.port.ac.uk/research/gbhgis/about/historicalgis/historicalgazetteers/>

de Zaragoza o el mundo) y el nivel de detalle, llamado granularidad (ej. calle ³, dirección ⁴, código postal, etc.). La calidad del servicio de georreferenciación está condicionada por factores dependientes de los requisitos de calidad en el servicio (QoS) (tiempo de respuesta, fiabilidad) y de los datos (variaciones en nivel de granularidad, precisión). Además, los servicios se diferencian por el modo de acceso (libre, restringido, de pago) muy influenciado por el dominio de los servicios de georreferenciación por las empresas privadas. Los servicios dedicados de pago garantizan la mejor calidad y pueden servir como soporte a los sistemas que requieren información muy precisa como los servicios de emergencia. En la oferta de los proveedores más grandes hay acceso libre con limitaciones a sus datos vía su propia página Web, generalmente en la forma de un servicio de callejero de buena calidad. Estas tácticas son aplicadas, por ejemplo, por Google⁵. El servicio de GoogleMaps⁶ o la aplicación GoogleEarth⁷ son conocidos por la mayoría de los usuarios de la red. Otros conocidos proveedores son ViaMichelin⁸, Yahoo⁹ y Geonames¹⁰. Además de un acceso vía un portal Web, estos proveedores ofrecen un acceso restringido vía servicios a sus datos aportando un API. Las restricciones suelen condicionar la visualización (la licencia exige uso de los visualizadores propios del proveedor) y prohíben la reutilización de los datos. Además, las condiciones del uso del servicio libre perjudican la calidad de las aplicaciones construidas sobre el dicho servicio al establecer un tiempo mínimo entre siguientes preguntas y un límite en el número de preguntas por día, entre otras restricciones.

El cuadro 1 presenta las condiciones del acceso restringido de los principales proveedores de los servicios de georreferenciación.

³<http://www.thensg.org.uk/iansg/welcome.htm>

⁴<http://www.nlpg.org.uk/nlpg/welcome.htm>

⁵<http://code.google.com/apis/maps/>

⁶<http://maps.google.es/>

⁷<http://earth.google.es/>

⁸<http://dev.viamichelin.co.uk/wswebsite/gbr/jsp/vmdn/VMDNHomePage.jsp>

⁹<http://developer.yahoo.com/maps/rest/V1/geocode.html>

¹⁰<http://www.geonames.org/maps/reverse-geocoder.html>

Cuadro 1: Comparación de condiciones del acceso restringido entre principales proveedores de los servicios de georreferenciación

	Control del acceso	Límite de número de peticiones por día	Otros límites	Restricción de uso
GoogleMaps	código del acceso	15.000	Tiempo mínimo entre peticiones consecutivas	Solo en visualizador de GoogleMaps
YahooMaps	IP	5.000		Solo en visualizador de YahooMaps
ViaMichelin	código del acceso	1.000		Solo en visualizador de ViaMichelin
Geonames	IP	50.000		

El uso típico de los servicios de georreferenciación por los usuarios de la banda ancha (aplicaciones del escritorio, aplicaciones Web 2.0) es el llamado “geotagging”, la asignación de la localización a cualquier tipo de información (p.e. a las fotos, noticias, publicidad, blogs, etc.). La tendencia más reciente del mercado de servicios de georreferenciación es ofrecer a los usuarios la posibilidad de añadir su propia información geográfica (puntos de interés, fotos, etc.) en el sistema, lo que enriquece el servicio proporcionado. Esto ha propiciado la aparición de aplicaciones basadas en información georreferenciada creada y gestionada por una comunidad de usuarios (p.e. Geonames, Megalithic Portal¹¹, Geograph British Isles¹², flickr¹³ o OpenStreetMap¹⁴).

Actualmente, los proveedores de la información geográfica han dirigido sus esfuerzos al mundo de bajo ancho de banda. El mundo de telefonía móvil está dominado por las aplicaciones basadas en la localización o LBS que exigen un soporte de los servicios de georreferenciación. Tanto el seguimiento de localización de usuario como la geocodificación inversa juegan un papel significativo en aplicaciones Web destinadas a este mercado. Esto abre la puerta a nuevas necesidades. Por ejemplo, conocer la calidad una fuente de datos, su coste y su disponibilidad. Algunas compañías como Google están apostando fuerte para cubrir estas nuevas exigencias. Por ejemplo Google Gears¹⁵ permite a las aplicaciones Web obtener la localización del usuario y seguirlo (guardar en su sistema su localización, lo que puede servir para calcular su velocidad y predecir su ruta, para aportar mejor servicio). Otro proyecto del Google es

¹¹<http://www.megalithic.co.uk/>

¹²<http://www.geograph.org.uk/>

¹³<http://www.flickr.com/>

¹⁴<http://www.openstreetmap.org/>

¹⁵<http://gears.google.com/>

Android¹⁶ en el que participan 30 compañías de tecnología y del móvil, para desarrollar una completa, abierta y libre plataforma móvil. En su modelo permite elección de proveedor a base de sus características, como exactitud, soporte a los datos adicionales (altura, entidades relacionadas), pero también condicionadas por el entorno de trabajo (necesidad de engría, red celular, acceso al Web, GPS o precio). Estos criterios están definidos por la aplicación final y sirven como base para elección del servicio. También desde la comunidad Open Source hay iniciativas en este ámbito. Un buen ejemplo es GeoClue¹⁷, una utilidad de localización del usuario por aplicaciones del entorno móvil. En este caso se utiliza mecanismo de comunicación D-Bus habitual en distribuciones Linux para simplificar el desarrollo de las aplicaciones proporcionando un interfaz único para el acceso de servicios de geocodificación (Yahoo, Geonames) y geocodificación inversa (Geonames).

Las características específicas de los aparatos de bajo ancho de banda exigen nuevas características de los sistemas de georreferenciación, pero el crucial problema de elección de proveedor adecuado sigue estando abierto.

De los sistemas de información geográfica también depende nuestra salud o seguridad. Un ejemplo conocido son los sistemas de emergencia que requieren servicios de georreferenciación de alta calidad. Esta necesidad crece también en mundo de investigaciones científicas, donde cada vez más los resultados de los trabajos de investigación dependen estrictamente de la calidad de los datos de georreferenciación, por ejemplo, medicina[2, 3, 7], criminología[4, 6, 1]o estudios del medioambiente[5, 4].

La elección del servicio depende de los requisitos del caso de uso, y, aunque hay mucha oferta de servicios de georreferenciación de acceso libre, la calidad de estos no siempre satisface las necesidades de los usuarios. Hay que añadir que las autoridades locales siguen en posesión de los datos geográficos de la mejor calidad, no obstante no tienen todos los datos que requieren. Con tanta diversificación aparece la pregunta: cómo buscar a los servicios adecuados en función de nuestra necesidad y cómo evaluarlos para elegir lo más apropiado. En muchos casos el usuario necesita una respuesta y no está interesado en saber dónde buscar la información adecuada.

Nuestra propuesta de servicio de geocoder compuesto intenta responder a esta nueva necesidad, ya que actualmente el problema no es el acceso a la información sino la elección del mejor disponible. Este servicio está orientado a los usuarios "cotidianos" y es una solución que hibrida varios servicios existentes, componiendo una respuesta al usuario gracias al acceso a varios proveedores de información geográfica reconocidos. Por ejemplo, en el caso de un usuario de la administración pública nacional, esta estrategia permite acceder de forma transparente a diferentes servicios de direcciones, nacionales y locales, garantizando que siempre se tiene la información más reciente.

El artículo está organizado en siguiente manera. En la parte 2 presentamos el caso del uso sobre el cual se basa el trabajo presentado. Después presentaremos

¹⁶<http://code.google.com/android/>

¹⁷<http://www.freedesktop.org/wiki/Software/GeoClue>

el estado del arte entorno el tema del trabajo. Las secciones 3 y 4 describen la arquitectura del servicio propuesto y su implementación con resultados. El artículo finaliza con conclusiones y descripción del trabajo futuro.

2. Caso de uso: incidencias en la administración local

Un municipio necesita de sistemas de gestión de incidencias de tráfico que le permitan manejar tanto las incidencias que ocurran en su término municipal como tener una visión de las incidencias en términos cercanos que puedan afectarle. Además esta información es una importante fuente de datos para las aplicaciones de GPS del sector privado (ej. taxis, sistemas de la logística de las empresas). Otro sistema de gestión similar es el sistema de gestión de emergencias, como las inundaciones o incendios, donde las zonas afectadas suelen traspasar las fronteras administrativas. La efectiva gestión de recursos del plan de emergencia tiene que estar centralizada y coordinada. Tal sistema necesita un servicio de georreferenciación de alta calidad (la cobertura al nivel nacional, alta precisión y granularidad de datos, fiabilidad) y disponer de los datos actualizados. Los datos de la mejor calidad están en posesión de diferentes administraciones públicas locales, lo que dificulta la centralización de la información, teniendo en cuenta el cambio continuo que presencian los datos geográficos, especialmente en zonas urbanas.

La aplicación de la solución propuesta permite complementar datos de una fuente con datos de otra, lo que asegura acceso a los datos recientes. También permite al usuario mas libertad en elección de manera de búsqueda. Teniendo el acceso a varios servicios permite obtener lo mejor respuesta de todo el sistema (a base de evaluación inicial de cada de los servicios y respuestas obtenidas de cada fuente), la mejor respuesta de cada una de las fuentes o las mejores respuestas de una fuente elegida. Además, la facilidad de adición de diferentes fuentes de datos permite incorporar al sistema cualquier tipo de datos georreferenciados necesarios para el funcionamiento correcto del sistema de gestión.

3. Estado del Arte

En España existen varias alternativas de servicios de georreferenciación en el ámbito de sector público, como Cartociudad¹⁸, Geopista¹⁹, Servicios Web Catastro²⁰ o servicios de callejeros propios de los Ayuntamientos como el del Ayuntamiento de Zaragoza²¹.

Los Servicios Web Catastro proporcionan un servicio de datos catastrales al nivel nacional, el servidor de mapas, conversor de coordenadas y Callejero

¹⁸<http://www.cartociudad.es/ignnomenclator/>

¹⁹<http://www.geopista.com/>

²⁰<http://ovc.catastro.minhac.es/>

²¹<http://idezar.unizar.es/>

catastral. El callejero catastral se caracteriza por alta fiabilidad y granularidad. La precisión de los datos de georreferenciación de las direcciones esta restringida por el tipo del servicio (alta precisión de georreferenciación de la parcela pero baja de la dirección). Además, la búsqueda no siempre es cómoda, porque exige definición de la zona de búsqueda (provincia y municipio).

Otra alternativa es Cartociudad, que proporciona varios servicios Web, entre ellos el Servicio de Nomenclátor (WFS-NEM) para búsqueda de topónimos y Web Feature Service (WFS) para descargas selectivas de elementos. El corazón del sistema es una base de datos cartográficos oficial de las poblaciones de España (redes de calles y carreteras topológicamente estructuradas y continuas) creada por la armonización e integración de los datos digitales oficiales producidos por varios proveedores (Dirección General del Catastro, Instituto Nacional de Estadística, Oficina de Correos, Dirección General del Instituto Geográfico Nacional, organismos estadísticos y de catastro de La Comunidad Foral de Navarra y el País Vasco). Cada uno de los proveedores es responsable de un tipo de los datos originales, y de su mantenimiento. Los datos de dirección georreferenciados proceden principalmente del Catastro y son contrastados con los provenientes de las otras fuentes. La filosofía del sistema es ofrecer una representación del estado actual, lo que condiciona el modelo de datos, donde falta el soporte al historial de los datos que aporta visión histórica de los cambios que perciben continuamente los entornos urbanos. Además, falta el soporte a la información de la procedencia de los datos y su tratamiento – la información vital para los usuarios de los datos. Esto perjudica el proceso de mantenimiento pero también el conocimiento aportado por el sistema. A lo largo del año 2008 se espera implementar una metodología de actualización para los primeros municipios. La georreferenciación de una dirección es al estilo del catastro (indicación la provincia y municipio). Además, el servicio aporta poca cobertura en termino nacional – 50 % de España.

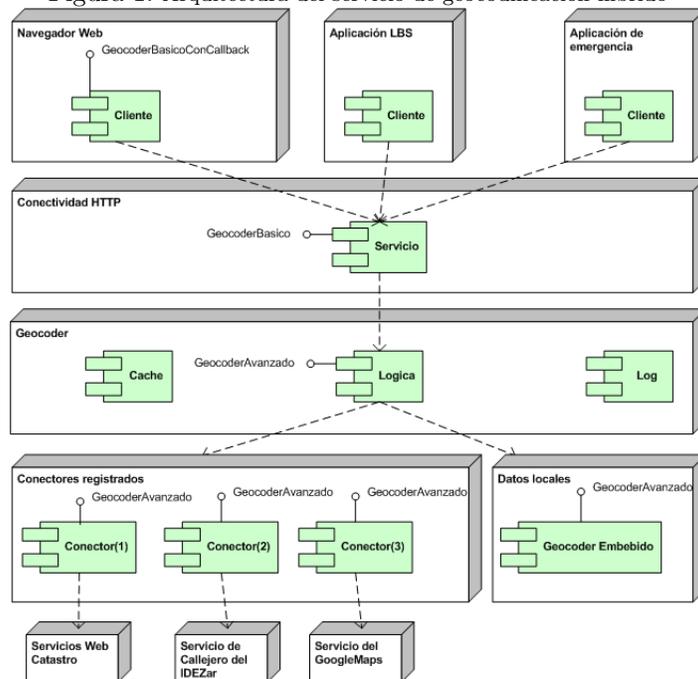
Otra propuesta es el Sistema de Información Territorial para los Ayuntamientos proporcionado, en su momento, por Geopista. Ofrece un sistema de gestión para los municipios de pequeño o mediano tamaño, lo que condiciona su ámbito solo a nivel local. La georreferenciación proporcionada se basa en una base de datos local dedicada. Los datos proceden del catastro, lo que supone necesidad de actualización cada cierto tiempo. Los usuarios pueden también añadir sus propios datos de georreferenciación. El principal problema es la falta de acceso al los datos entre municipios.

Los servicios de Callejeros propios de los Ayuntamientos, como del IDEZar, se caracterizan sobre todo por alta precisión. El nivel de granularidad es diferente dependiendo de la zona (centros urbanos, pequeños pueblos) y suelen existir las zonas sin cobertura, como las autopistas o nuevas urbanizaciones.

4. Arquitectura

El servicio del geocoder propuesto es un servicio de callejero que permite al usuario común de la Web obtener información procedente de varias fuentes a la

Figura 1: Arquitectura del servicio de geocodificación híbrido



hora de añadir información geoespacial a diferentes contenidos, como paginas Web, RSS feeds, imágenes. Su principal ventaja es el uso de varios proveedores de información georreferenciada lo que proporciona un servicio de mayor calidad basado en las necesidades del usuario y capaz de complementar la información obtenida. Sus características permiten solventar los problemas indicados en las alternativas anteriormente presentadas, como la necesidad de la actualización de los datos, necesidad de definición de la zona búsqueda o faltas en cobertura o granularidad.

La arquitectura general del servicio propuesto se detalla en la figura 1. Lo elementos mas importantes de la arquitectura son el componente de lógica y los conectores.

El componente de lógica es el responsable de la selección del servicio de georreferenciación y de la evaluación de la respuesta. La estrategia de elección del proveedor se basa en las características de la fuente mientras que la estimación de la respuesta se realiza en base a las necesidades del usuario.

Los conectores sirven para realizar las peticiones a cada uno de los servicios de georeferenciación registrados en el sistema. Esta estrategia permite utilizar diversos servicios independientemente del protocolo de comunicación, del estilo de invocación (REST, SOAP), del interfaz (propietario, abierto), etc. El conector es el responsable también de la integración de los datos mediante el mapeo entre el modelo de datos de la fuente y el modelo interno del servicio compuesto.

Adicionalmente, si es necesario, también se realiza el cambio del sistema de coordenadas que utiliza el servicio ajeno para homogenizar las respuestas.

5. Implementación y resultados

El servicio propuesto utiliza los siguientes servicios de geocodificación: Servicio del callejero del Ayuntamiento de Zaragoza, Servicio del GoogleMaps, Servicio de Catastro de España, y Servicio del Nomenclátor Conciso (IGN). Las características de estos servicios presenta el cuadro 2, donde:

- SOAP (Simple Object Access Protocol) es un protocolo RPC (Remote Procedure Call) que permite invocar remotamente los métodos de un objeto,
- GET/ POST: Los típicos métodos de HTTP (petición por un recurso / inserción de datos),
- SRW (Search/Retrieve Web service) – servicio Web para búsqueda y recuperación de la información (ahora conocido como SRU VIA http SOAP).

Cuadro 2: Los servicios de georreferenciación u sus diferencias.

Servicio	Protocolo	Estilo	Interfaz	Estándar	CRS
GoogleMaps	HTTP	GET	propietario	mixto (xAI)	EPSG:4326
Callejero de IDEZar (SRW IDEZar)	HTTP	SOAP	abierto (SRW)	no (exclusivo del ayuntamiento de Zaragoza)	EPSG:23030
Catastro	HTTP	SOAP	abierto	sí (catastro)	varios
Servicio del Nomenclátor IGN Conciso	HTTP	GET/ POST	abierto (WFS)	sí (MNE)	EPSG:4230

La implementación de la lógica del servicio avanzado de geocodificación exige la evaluación inicial de los servicios ajenos, lo que es crucial para comportamiento del servicio compuesto. La evaluación de cada uno de los proveedores sirve para obtener los calificadores de la elección – los valores en los cuales se basa la selección de la fuente. Para obtenerlos hay que tener en cuenta cuatro características principales de las fuentes:

- cobertura - define la área donde se encuentran los datos espaciales proporcionados y suele corresponder a una entidad administrativa (servicio de IDEZar: municipio de ZARAGOZA; Catastro y Nomenclátor IGN: España; GoogleMaps: el mundo);

- fiabilidad - informa sobre su correspondencia al mundo real. Una fuente fiable de datos espaciales para una zona contiene todos los elementos del mundo físico. Para poder definir el nivel de fiabilidad de un servicio desconocido es necesario disponer de una fuente fiable (la fuente de referencia, los datos oficiales de las autoridades);
- nivel de granularidad - esta característica esta vinculada con la variación de la fiabilidad en diferentes zonas (grandes ciudades, pequeños pueblos);
- precisión - describe el error medio de los datos espaciales de una fuente en comparación con datos de la fuente de referencia (IDEZar: muy alta; GoogleMaps: aceptable; Catastro: alta - datos catastrales y aceptable - direcciones).

A partir de una dirección introducido en el formato: nombre de calle, numero (“C/ mayor, 20”, “coso”, “plz España”), el servicio transforma la cadena en un objeto que representa una pregunta de una dirección compleja en un modelo general, particular al geocoder compuesto. Dependiendo de la configuración elige uno o varios servicios a los que preguntar la dirección. Cada uno de los conectores a cada servicio conoce como traducir la pregunta al modelo propio del servicio, ejecutarla, y devolverla transformada a un modelo unificado de respuesta al geocoder compuesto. La evaluación de los resultados se realiza primero en el conector (por sus características específicas) y después en la lógica del geocoder compuesto a base de criterios predefinidos por fichero de configuración o por el usuario.

Actualmente existe una instancia de este servicio que está siendo aplicada en el callejero Web del Ayuntamiento del Zaragoza²² (figura 2). La principal fuente de datos es el servicio del callejero del ayuntamiento de Zaragoza (SRW IDEZar) pero complementando sus resultados con otras fuentes, lo que mejora la fiabilidad del servicio de callejero de cara al usuario.

²²<http://idezar.zaragoza.es/callejero/index.jsp>

Figura 2: El servicio de Callejero de Ayuntamiento de Zaragoza a base de geocoder compuesto



6. Conclusiones y trabajo futuro

Los resultados de búsqueda del servicio propuesto están enriquecidos por las informaciones procedentes de diferentes servicios ajenos, lo que permite complementar los resultados de un servicio con datos del otro y aportar una respuesta mejor al usuario. Los datos procedentes de las fuentes aportados por las autoridades locales suelen ser bastante completos y de alta precisión lo que permite suponer que teniendo un conjunto de servicios utilizando estos datos, puede permitir desarrollar aplicaciones fiables y de alta calidad necesarias para los servicios públicos.

En nuestro trabajo hemos detectado que el principal problema es la selección y elaboración de los indicadores requeridos para la selección del proveedor. Como trabajo futuro nos enfocaremos en el desarrollo de la metodología para la evaluación de los servicios de georreferenciación. Para ello necesitamos describir mejor cada uno de los servicios, lo que exige el soporte de metainformación acerca de los servicios. Como siguiente paso, utilizaremos métodos estadísticos para evaluar la respuesta y para mejorar los indicadores de la precisión y granularidad. Con esta información será factible desarrollar técnicas adecuadas de comparación de diferentes servicios de georreferenciación.

Agradecimiento. Este trabajo ha sido parcialmente financiado por GeoSpatium-Lab, el Instituto Geográfico Nacional (IGN) y el Ministerio de Ciencia e Innovación (ref. TIN2007-65341 y PET2006.0245). El trabajo de Rodolfo Rioja ha sido cofinanciado por el Ministerio de Ciencia y Tecnología (programa Torres Quevedo ref. PTQ06-2.0790).

Referencias

- [1] C. R. Block, M. Dabdoub, and S. Fregly. Geocoding of crime incidents using the 1990 tiger file: the chicago example. *Crime Analysis Through Computer Mapping* (189-193), Police Executive Research Forum, Washington, DC, 1995.
- [2] M. R. Bonner, D. Han, J. Nie, P. Rogerson, J. E. Vena, and J. L. Freudenheim. Positional accuracy of geocoded addresses in epidemiologic research. *Annals of Epidemiology*, 17:464 – 470, 2003.
- [3] Yang Duck-Hye, Lucy Mackey Bilaver, O. Hayes, and R. Goerge. Improving geocoding practices: Evaluation of geocoding tools. *Journal of Medical Systems*, 28(4):361 – 370, 2004.
- [4] S. M. Gilboa, P. Mendolab, A. F. Olshana, C. Harnessc, D. Loomisa, P. H. Langloisd, D. A. Savitza, and A. H. Herring. Comparison of residential geocoding methods in population-based study of air quality and birth defects. *Environmental Research*, 101:256–262, 2006.
- [5] J. H. Ratcliffe. On the accuracy of tiger type geocoded address data in relation to cadastral and census areal units. *International Journal of Geographical Information Science*, 15:473–485, 2001.
- [6] J. H. Ratcliffe. Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science*, 18:61–72, 2004.
- [7] F. B. Zhan, J. D. Brender, I. De Lima, L. Suarez, and P. H. Langlois. Match rate and positional accuracy of two geocoding methods for epidemiologic research. *Annals of Epidemiology*, 16:842 – 849, 2006.