Case Study on Geocoding Based Scheduling Optimization In Supply Chain Operations Management

Steffen S. Rasmussen, University of Copenhagen, steffen@audkus.dk
Malleswara Talla, Raul Valverde, JMSB, Dept. of SC&BTM, Concordia University, Montreal,
malleswara.talla@concordia.ca, raul.valverde@concordia.ca

Abstract: Traditional approach of ad-hoc geocoding for required addresses is quite time consuming and inefficient. In this paper, we propose and examine that the company-wide operations and addresses be geo-coded and stored in a database. As a case study, a geographical database is created for a company with primary focus to optimize operations management. This is done by storing the address information, the road network and a matrix containing time and distances to every address in the database. This approach prevents a lot of potential errors with ad hoc geocoding because every address is looked up in the database. The same goes for the time and distance matrix. Because the time and distances are stored for the distance between every address, no computations are needed at runtime. Results demonstrated that the proposed method improved the overall time taken starting from registration of an address to geocoding by 50%. When a route is prepared, only a TSP (Traveling Salesman Problem) algorithm is applied to the required addresses for the route optimization. The proposed method is proven to be very effective to a small and medium size company in supply chain logistics.

Keywords: Ad-hoc Geocoding, Geographical database, Time and Distance Matrix, Traveling Salesman Problem.

1. Introduction

The term geocoding refers to the process of finding an address location coordinates on a map that can be performed automatically or manually or a combination of the two. There is a limited amount of literature covering this process from a developer's point of view. Crosier presented the concepts, preparation, and process of geocoding in ArcGIS application [1]. Murray described the data type's requirements for geocoding setup and usage in Oracle [2]. Like in ArcGIS, different match degrees can be set to increase the number of geocoded objects. The disadvantage of using match degrees is that the location precision will be reduced. Studying user manuals from applications which supports geocoding, gives an insight into how other developers have attempted to create a geocoding solution. Arctur et. al. covered a broader aspect of geocoding focusing on the design and development of geodatabases from a whole range of different uses among which addresses, and locations is just one of them [3]. For a system designer with no prior knowledge of developing GIS systems this book gives a very good introduction into standards which is commonly used in the GIS industry. Zeiler's focus is on the map objects and how they can be used in a geographic database [4]. It gives a good overview of working with GIS but its contribution to design and development is limited. It was proposed that the positioning of an address is at least within in an address segment and not just e.g. in a postal district [5], pointing out the importance for the end user to know what kind of precision the data supports. Bakshi looked at geocoding precision and focused on increasing the street segment address precision [6]. Karimi evaluated uncertainties in geocoded data [7]. They look at the available data and use factors as completeness, correctness, consistency, currency and accuracy. In their research they test three different geocoding engines to compare the geocoding accuracy. The paper concludes that there are no major differences, in accuracy, between the geocoding engines. If the requirement is to improve the geocoding process, selecting another geocoding engine is not the solution, as such a new geocoding engine should be developed. The quality of addresses is important and the Customer Information Quality Committee (CIQ) has been formed to create an open XML for customer information and profile management. In this context, the Extensible Name and Address Language (xNAL) is interesting as well as the CIQ's description of their work, challenges and solutions to handle more than 200 address formats. They came up with a hierarchical address structure used to exchange this type of information between different information systems. Developing a hierarchical address structure for the database is quite appealing, but due to the address diversity this type of approach have proven to be inadequate when working with addresses in a geographical database, as pointed out in [3]. A legislation is covered in "Statutory Order on Road Names and Addresses" [8]. The address legislation is the corner stone for developing and setting up the rules to ensure the factors as completeness, correctness and consistency is built into the GIS extraction program and database. In short, this project aims to design and develop a geodatabase on which geocoded data is treated and stored for optimizing the supply chain operations management.

2. Problem Statement and Scope

The case study covered in this paper examines a windows cleaning company which manages a fleet of employees who visit approximately 10.000 customers each month. Current logistics is highly automated except for the part of operations concerning the fleet management which is manual task and highly inefficient. As a result, resources are used inefficiently making the company less competitive not only as a service provider but also as workplace. When the employed workforce is paid by piecework it is important that conditions for carrying out their assignment, provided by the company is optimized. In this case transportation time, although paid by the customer, is a factor which the employee can not alter to increase the hourly wages by working faster. If the company cannot provide short travelling times the employees will just find another employer who can. The positive impact route automation can have on the employees visiting the customers; can also have a negative effect on office personal due to the introduction of automatization which results in a changed work flow as well as less manual work. Software solutions for fleet management exists in the company but it is inefficient, when several thousand customer visits should be planned in a very short time span. To overcome these obstacles, this paper aims to design and develop a geodatabase on which geocoded data is treated and stored in such a way that it is optimized for operational management. From a theoretical view the optimization can be accomplished by:

1. Geocoding every address within a domain.
2. Predetermine shortest route between addresses.
3. Segment road database to reduce graph size.
4. Save all information in one database in which an algorithm can be applied for route optimization.

This effort requires several programs to be developed for combining and converting geographical data and its properties into a format which can be stored and manipulated in one database. For example, every address and every node segment should be extracted and converted into longitude and latitude coordinates. The extracted data should be imported into the developed database. A program can then be developed to find the shortest route between every geocoded address. This is static information in the form of distance and traveling time, which can be pre-calculated and stored in the database. In this way there will be no need to use computational resources for the task. Once all this information has been stored, an analysis on the data can be performed on the road segments to reduce the graph. If series of road segments can be clustered together the scheduling optimization can be performed on the clusters instead on every individual node which has to be visited. "To find the cheapest way of visiting all the cities and returning to the starting point". The [9] is known as the Traveling Salesman Problem (TSP), in this case the cities are the nodes. Optimizing the traveling route between the clusters and then within the clusters greatly reduces the problem and the calculation time required to solve it. Finally, a TSP algorithm will be selected and implemented to be used for the final operational management optimization. The original project proposal focuses much more on this part of selecting one or more TSP algorithms. Due to the extent of the project and the complexity of the TSP an extensive analysis will not be performed. Rather one algorithm will be selected customized and used on the developed database.

Computerized scheduling optimization has been one of the corner stones of computer science and mathematics research. Programs have been developed based on this research and are commercially available. Utilizing these programs in business can have a great beneficial affect which can give the business the extra edge in a competitive market. Even though these programs have been around for years they have never been able to deliver, in great extent, the promised scheduling optimization improvements. This is due to the extensively amount of manual work by expert users to correct data inconsistency between the different systems required for scheduling optimization. “It is not uncommon to have a geocoding hit rate of under 50%” [10]. This is especially a problem if there are hundreds if not thousands of dynamic entities which must be manually geocoded.

Once all the entities have been geocoded a scheduling optimization program can be applied to the data. Finding the shortest route between A and B can be done quite fast with the current algorithms but the challenge is that the problem is NP hard (nondeterministic polynomial-time hard) and the more destinations which are added will increase the calculation time considerably. For example, “the exact solution of a symmetric problem with 2392 cities was determined over a period of more than 27 hours on a powerful super computer” and just increasing the number to 7397 cities increased the total calculation time to 3-4 years [11]. Adding different kinds of constraints will just increase the computational time. Algorithms which generates an approximate solution to the optimization task has generally very good solutions which come close to the optimum but with a lot less computation power needed. Once the scheduling optimization has completed the data have to be passed on for further treatment, so it can be used in the business operation. As the scheduling optimization is not trivial task, there is no time for this kind of work. As a result, some businesses soon discover that they are unable to get the required return on their investments and should aboard their scheduling optimization project. In [3], several domain data models were presented:

- Streams and river networks
- Census units and boundaries
- Addresses and locations
- Parcels and the cadastre
- Federal land surveys

Here addresses and locations are relevant for this case study. In the case study, the authors point out areas the GIS database designer should be aware off, when designing a geographic database. Specifically, the use of a Zone entity for relating the address properties instead of building a hierarchical address structure was a major contribution of this source. Finding sources on how to store road networks, also known as graphs, in a relational database turned out to be quite a challenge. Traditionally graphs are used in an array. As such an array can directly be stored in a database but it is not very suitable for data retrieval. Most literature on the subject covers the problems with graph data retrieval, querying and how to improve it. It does not come with practical examples on how to implement graphs into a relational database, which is important for this case study when developing a geodatabase which should be optimized for creating and storing static data, minimizing the computational power needed at runtime. An example of storing Graphs in a relational database was found in an article about genes, proteins and metabolites by [12]. In the study of biological systems researchers are working with an increasing large amount of data in the form of graphs which representation is much like a graph representing a road network. The author’s practical solution for handling the data can be transferred to a road network and its implementation in the relational database.

3. Analysis and Design

The aim is to analyse the current system for improvements and then design and develop a geodatabase for storing the geocoded data for optimizing the supply chain operations. The process of working with the geographical data is done in the MapInfo Professional application. Since this application does not provide adequate geocoding tools for this project, a MapBasic program should be developed, which then can be executed within MapInfo Professional, to extract the required data into a table which then can be exported from MapInfo Professional. The exported data then should be imported into the developed TopSpeed database. For this an import program should be developed in Clarion. A program is developed to find the shortest route between every geocoded address. This is static information which can be pre-calculated and stored in the database using the Floyd-Warshall’s All-Pairs Shortest Path Algorithm (APSP) [13]. In this way there will be no need to use computational resources for the task. This algorithm will be developed in Clarion and applied once the data have been imported into the database. Once all this information has been stored, an analysis on the data can be performed on the road segments to reduce the graph. If series of road segments can be clustered together, the scheduling optimization can be performed on the clusters instead on every individual node which has to be visited. This type of calculation is known as the Traveling Salesman Problem (TSP) which is known to be NP hard. Optimizing the traveling route between the clusters and then within the clusters greatly reduces the problem and the calculation time required to solve it. For Clustering the addresses, a “Multi Level Clustering Approach” was used in [13]. To solve the TSP the “Nearest Neighbor Algorithm”, [14] is used for testing geodatabase operational management. Figure 1 presents a detailed architectural of this paper.

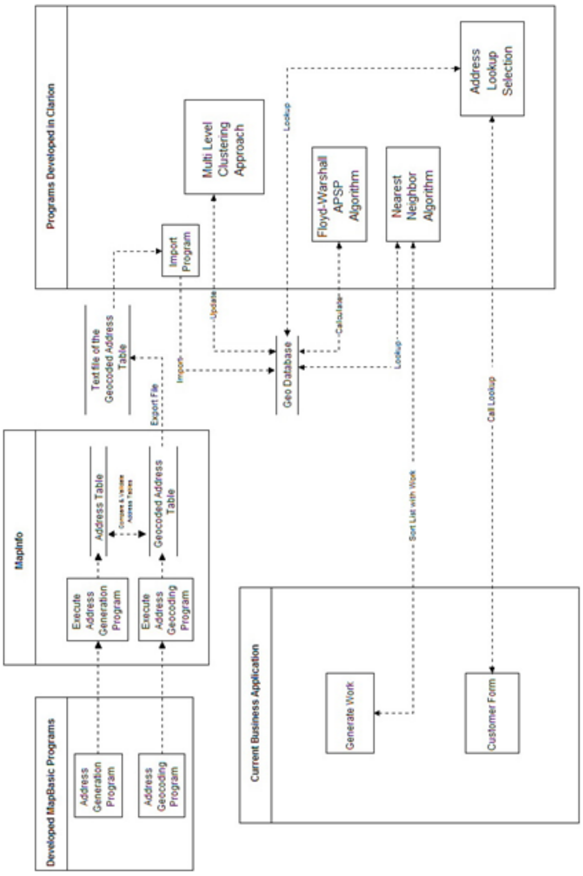

Figure 1. Architecture of the developed systems

Map Analysis

The primary focus of the analysis is on the available geographical data of the Danish road networks and how this data can be used in a geographical database for address lookup, coordinate location and route optimization. The use of multiple geographical thematic layers is considered to improve and validate the correctness of the data which will be extracted and then imported into the database. An algorithm examines every octant and all neighboring octants containing addresses are merged together. Once completed all clustered addresses are assigned a town name based on a file containing the town names and their respective coordinates which are assumed to be the centroid of the town. Any town that has not been assigned a cluster of address will be assigned to the nearest cluster of addresses without a town name within a radius of 200 meters. Figure 2 presents the merged octants.

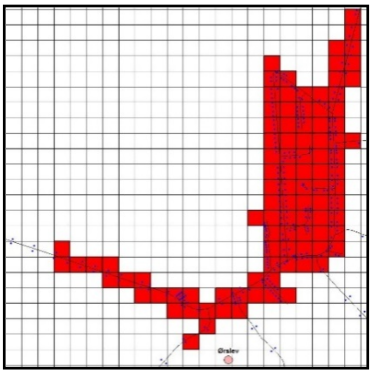

Figure 2. Octant clustering to locate the town Ørslev

4. Results and Discussion

Scheduling optimization in operations management is all about planning the most suitable order to visit several predefined locations so the total travel distance and/or time is minimized. To accomplish this, most systems rely on the ability to geocode every address on the route before an optimization algorithm can be applied and the shortest route can be calculated. Unless an error in the input, that prevents the geocoding process from being fully automatic. Any manual geocoding is quite time consuming and the probability that errors will occur is quite high. This is because the user, most likely, is not a domain expert on the address, and might make some erroneous assumptions which will place the address at a wrong location.

In operations management time is not always available so if the manual geocoding process is too time consuming it just can’t be used. The consequence of this is that route optimization cannot be performed on the address data and potential enhancements cannot be exploited.

In this case study, a different approach has been taken to overcome the known issues with geocoding. Instead of geocoding the individual addresses in the customer database a new geographical database has been developed which contains every address within the company's operational area. All the addresses stored in the geodatabase are pre-geocoded. In this way there is no manual geocoding and when the user has to enter a new customer address, it will be looked up in the geodatabase and the found address will contain all the information required for performing a route optimization. In this way all the possible errors with geocoding have been eliminated.

Getting to this stage has not been a trivial task, as this case study shows. Allot of obstacles had to be overcome before it was realized. Even the user interface had its complications which resulted in the first developed program had to be scraped and a new approach to the design and development had to be taken due to the sheer amount of data which had to be handled.

Address Registration and Geocoding

The final user test was conducted with three users of different experience and their average times to complete a customer address registration. The test was first conducted in the existing system which they were familiar with. The next test was conducted in the new system which was integrated with the geodatabase. The results of each test can be seen in Figure 3.

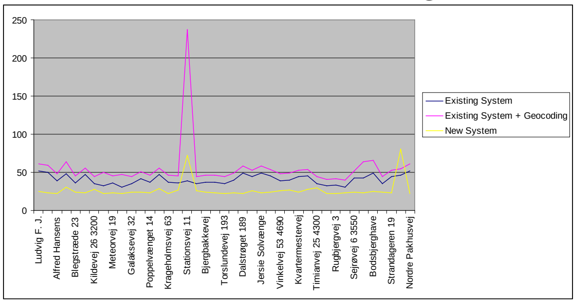

Figure 3. Average time for address completion in the old and in the new system

In average there is a significant speed improvement of 33,71% so looking up an address is much faster than typing it. As of Figure 3, it isn't a complete improvement, but some deviations do exist where the users in average used much more time to find the address then they would have used if it was just typed into the customer database (comparing the "Existing System" and "New System").

The increased time used for selecting an address occurred when the address existed in more than 15 postal districts, so the user had to scroll the screen in order to locate the specific address. In the other instance the address did not exist in the geodatabase because it was outside the operational area of the company and it took the users in average 81 seconds before they realized this.

Adding the geocoding process to the existing system increases the overall time for completion. 34 of the 40 test cases were geocoded automatically and the remaining 6 had to be manually geocoded. The geocoding process was only done by one super user. The test case did not contain any misspelled addresses which would just increase the overall time for geocoding the addresses. Stationsvej 11 was especially difficult to geocode, as can be seen in the high peek of Figure 3. It required additional research in existing maps before the user could select the correct address to be geocoded. Isolated this example might not seem to be too much work since in average only one minute is used per customer for address registration and geocoding. It might not be if the process only contains 40 addresses, but if the process requires hundreds if not thousands of addresses the improved time performance can be considerable. When comparing the whole process of address registration and geocoding with the new system then there is an overall improvement of 51,33% in this test case. Even though there was a consistent time improvement in the new system, the lookup did have some latency because the users increased their entry speed with the selections.

Route Optimization

Once all the addresses in the customer database have been assigned a geocoded address, route optimization can be performed. In the current system a global route has been created and adding a new address typically takes between 3 and 5 minutes. The manual algorithm used by the users complies very closely with the "Nearest Neighbor Algorithm" is used in the new system [14]. Here the automated route process takes less than a second and since the route is not part of a global scheme the optimization is much better when it only uses the tour in its calculation. Another advantage is that user errors, which could increase the tour considerably, are prevented as Figure 4 illustrates. Here the tour can be improved by 20,5%

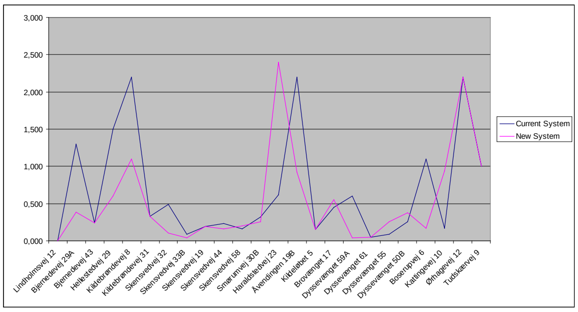

Figure 4. Example of distance improvements from the current system to the new

Generally, the tour improvements are expected to be much less although from the three test cases two of them could be improved considerably. The average improvement for these three test cases is 14,78%. Time savings in address registration and geocoding as well as automated route optimization is not the only advantage of using a geodatabase for operational management. Using an integrated system reduces the necessary expertise which otherwise was required for each system which had to handle the data. Here just one system is used, and no additional expertise is required for preparing the data for farther treatment in the operational management process. This makes the system less vulnerable to expert absence which otherwise could stop the whole process at a critical time period of the operational management.

5. Conclusions

The process of geocoding is time consuming and in an operational environment where several thousand customer visits should be planned in a short time span, the traditional geocoding process is inefficient. As an alternative to ad-hoc geocoding, a geographical database is created with every address and its coordinates for an operational area. Instead of creating an address and then geocoding it, the users will look-up the pre-geocoded address. This way, the geocoding obstacles are eliminated for the end user. To reduce the amount of calculation needed a matrix is created and stored in the geographical database containing the distance and time to every address in the system. During the test period with the users the solution proved to be efficient, easy to use and the whole complexity of the system is totally transparent to the user. The initial focus was the route optimization and how this could be improved and become more efficient by creating a geographical database which is optimized to handle the task. The program was tested on a small subsection of the database and it used a substantial number of filters. Filters are excellent for creating complicated data views but with a large amount of data they add a great performance penalty. Testing reviled that addresses which are not included in the geodatabase caused considerable delays in address registration. To prevent this, the geodatabase should also contain a complete list of potential address outside the operational domain, so that the address can be found by the user and the system can notify the user that the address is not in the company's operational area. The geocoding program still has some areas where it could be improved. For example, when roads run close to each other the geocode addresses from two different roads can overlap. To speed up the program development the import program is only able to import a complete road network file with addresses. Updating the geographical database will probably best be accomplished by linking the system up to a GIS system where the updates can be done interactively by applying new layers and manipulating existing objects. Until this kind of link has been established, any update requires that the whole geodatabase is deleted prior to importing the new updated road network with addresses. It is a common practice that the customer can complain about the performed work within 3 days, therefore the cleaner can be assigned to nearby work to react to such calls. The method is proven to be very effective in improving operations of a small to medium size company. Future research can cover the geodatabase for an international region with an extensive amount of data from each country.

REFERENCES

[1] Crosier, Scott (2004) Geocoding in ArcGis: ArcGis 9. Redlands: ESRI Press. ISBN 9781589481121.

[2] Murray, Chuck et al. (2005) Oracle Spatial User's Guide and Reference, 10g Release 2 (10.2). Oracle Corporation. Part No. B14255-01.

[3] Arctur, David & Zeiler Michael (2004) Designing Geodatabases: Case Studies in GIS Data Modeling. Redlands: ESRI Press. ISBN 1-58948-021-X.

[4] Zeiler, Michael (1999) Modeling our World: The ESRI Guide to Geodatabase Design. Redlands: ESRI Press. ISBN 1-879102-62-5.

[5] Cayo, Michael R. and Talbot, Thomas O. (2003) Positional error in automated geocoding of residential addresses. International Journal of Health Geographics. Available from: http://www.ij-healthgeographics.com/content/2/1/10.

[6] Bakshi, Rahul; Knoblopck, Craig A. and Thakkar, Snehal (2004) Exploiting Online Sources to Accurately Geocode Addresses. Proceeedings of the 12th annual ACM international workshop on Geographic information systems. Avalible from: http://portal.acm.org/citation.cfm?id=1032222.1032251

[7] Karimi, Hassan A; Curicik, Matej and Rasdorf, William (2004) Evaluation of Uncertainties Associated with Geocoding Techniques. Computer-Aided Civil and Infrastructure Engineering 19 (3), 170–185. doi:10.1111/j.1467-8667.2004.00346.x. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8667.2004.00346.x

[8] Bek. Nr. 1398 12 (2006) Statutory order on Road Names and Addresses [Online] Available: http://danmarksadresser.dk/file/389699/0_CountryReport_2010_DK_EN.pdf

[9] Applegate, David L. et al. (2006) The Traveling Salesman Problem: A Computational Study. Printceton: Princeton University Press. ISBN 978-0-691-12993-8

[10] Anon (2006) MapInfo Professional User Guide [Online]: http://reference.mapinfo.com/software/mapinfo_pro/english/8.5/MI_UG.pdf (Accessed: 12. April 2007).

[11] Helsgaun, Keld (1998) An Effective Implementation of the Lin-Kernighan Traveling Salesman Heuristic. Roskilde: Department of Computer Science. Available: http://www.akira.ruc.dk/~keld/research/LKH/LKH-1.3/DOC/LKH_REPORT.pdf (Accessed: 15 July 2007)

[12] Stephens, Susie M.; Rung, Johan and Lopex, Xavier (2004) Graph Data Representation in Oracle Database 10g: Case Studies in Life Sciences. Oracle [Online]: http://www.oracle.com/technology/industries/life_sciences/pdf/ls_ieee_graph.pdf

[13] Brandes, Ulrik & Erlebach, Thomas (2005) Network Analysis: Methodological foundations. Berlin: Springer-Verlag. ISBN 3-540-24979-6.

[14] Gross, Jonathan L. & Jay Yellen (2006) Graph Theory and its applications. Boca Raton: Chapman &Hall. ISBN 978 1-58488-505-4