Transactions of the VŠB – Technical University of Ostrava, Mechanical Series
No. 2, 2009, vol. LV,
article No. 1686

David FOJTÍK[*], Jiří HORÁK[**], Igor IVAN[***]

AUTOMATIC CREATING DATABASE OF PUBLIC TRANSPORT CONNECTIONS

AUTOMATIZACE TVORBY DATABÁZE DOPRAVNÍ OBSLUŽNOSTI

Abstract

The Integrated Portal of the Ministry of Labour and Social Affairs of the Czech Republic provides the unemployed the possibility to find appropriate vacancy according to various criteria. The possibility of public transport connection existence between municipality of residence and municipality of potential vacancy is one of many various criteria of searching. Data for this option are guaranteed by VSB-TU of Ostrava. Valid time tables are used as information source of public transport connection existence and manual or automatical processing can be used for data mining. Due to the fact, that about 13.5 million of connections are calculated, it has to be used automatical processing. This paper describes the software solution, where the whole process is automatically distributed to many powerful workstations for parallel processing of all public transport connection combinations.

Abstrakt

Integrovaný portál Ministerstva práce a sociálních věcí umožňuje nezaměstnaným vyhledávat volná pracovní místa dle různých kritérií. Jedním z nich je také vyhledávání volných pracovních míst dle existence spoje veřejnou linkovou dopravou mezi obcí bydliště a obcí s potencionálním zaměstnavatelem, respektive dle vlastností tohoto spoje. Právě tuto možnost datově zabezpečuje VŠB – TU Ostrava. Informace o existenci dopravního spojení je získávána z platných jízdních řádů buďto manuálně nebo automatizovaným procesem. Vzhledem k faktu, že se celkem pracuje s 13,5 milióny kombinacemi obcí, je třeba využít proces automatizovaný. Tento článek popisuje programové řešení, které automatizovaně distribuuje celý tento výpočetní proces mezi mnoho výkonných pracovních stanic pro paralelní zpracování.

1 INTRODUCTION

The role of public transport services differs among countries. The countries with the highest extent of public transport are located in the middle and east Europe, in spite of continuous increases of the individual transport share. The relatively high level of bus transport density in countries of previous communist block was documented by e.g. Jordan (1995). In the Czech Republic, the ratio of people transported by public means in 2001 was 60.2% and since that time it slowly declines to 56.0% in 2007 (RD 2007), city public transport is included.

The situation of public transport usage is also depicted by results of the European survey from 2007 (Attitudes, 2007), which covered all 27 EU Member States on a randomly selected sample of over 25767 individuals of at least 15 years of age. The graph (fig. 1) shows what main mode of

[*] Ing., Ph.D., Department of Control Systems & Instrumentation, VŠB -Technical University of Ostrava, av. 17. listopadu 15, CZ-708 33 OSTRAVA-Poruba, 59 732 4193, david.fojtik@vsb.cz
[**] doc., Dr., Ing. Institute of Geoinformatics, VŠB -Technical University of Ostrava, av. 17. listopadu 15, CZ-708 33 OSTRAVA-Poruba, +420 597 32 5457, jiri.horak@vsb.cz
[***] Ing. Institute of Geoinformatics, VŠB -Technical University of Ostrava, av. 17. listopadu 15, CZ-708 33 OSTRAVA-Poruba, +420 597 32 5457, igor.ivan@vsb.cz

transport do they use for their daily activities. Using public transport is most popular in Latvia (45%) and Slovakia (37%). In the Czech Republic, Estonia and Bulgaria, there are very similar results around 35%. The public transport is at least popular in the Netherlands and Slovenia (both only 11%).

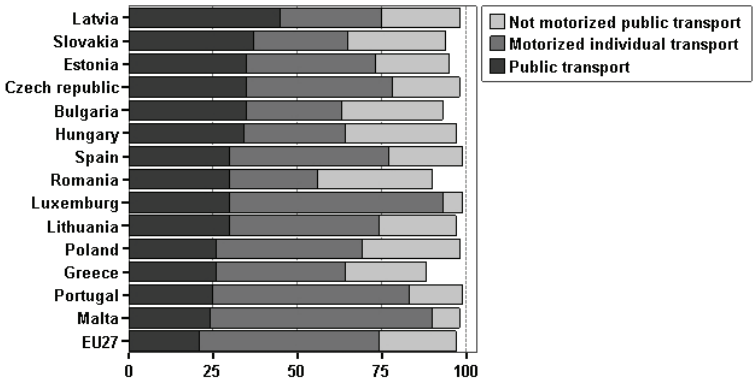

Fig. 1 Main mode of transport for daily activities in countries of EU (selection above average)

The Ministry of Social Affairs of the Czech Republic (MSA hereafter) is conscious that low-income groups of employees and unemployed at all cannot mostly afford to commute via private cars and that is why the role of the public transport for commuting is very important. The Integrated Portal of MSA represents a central (and official) site for searching vacant job places. People, who cannot find any appropriate job in their residence municipality, can try to find some other appropriate job in surrounding municipalities. For this reason the searching interface has been enlarged by special option for commuting conditions since 2007. Job applicant can specify the distance or time interval of the commute via public transport. Based on applicant's requirements the output list of appropriate job offers is created and it contains prices, connection duration, number of changes and a detail description of one recommended connection for five time intervals.

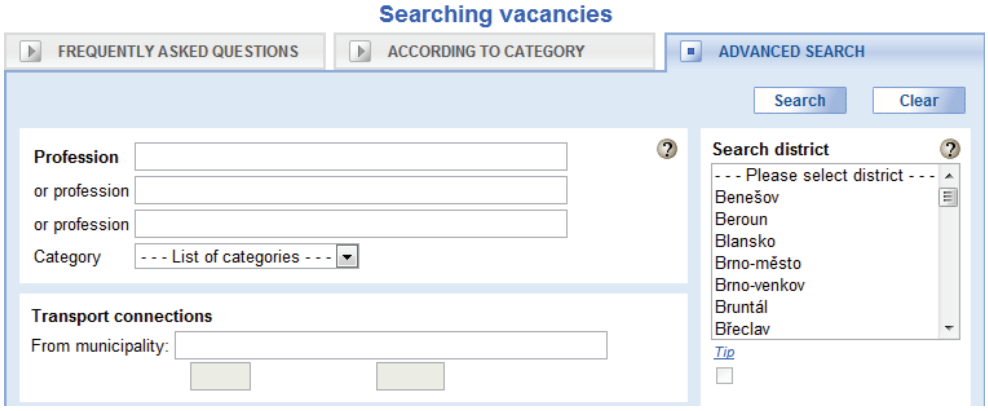

Fig. 2 Basic searching options of the Portal

The best solution would be the online searching of public transport connection as well as the analysing and listing of the best connection parameters but this is impossible. The only solution is the preprocessing of all possible commuting options and all results have to be stored in a large database.

There are stated some limitation for the database of public transport accessibility (Database hereafter). Not all municipality combinations are searched but public transport connectivity is analysed between municipalities within the Euclidean distance less than 100 kilometres and the duration of connection has to be less than 90 minutes. Commuting time is analysed for selected time periods according to usual working time intervals (shifts). The potential employee must have

possibility to return back from work to the residence municipality and that is why the appropriate return way is searched as well and its parameters are stored also to the Database.

To approximate to the online status of the Database and to provide true information the portal has to be updated regularly. In the Czech Republic the time tables are generally updated by providers twice a year. Therefore the completely new database has to be created right after publication of this new time table. In the meantime during the year there are some smaller changes of the time tables but only at concrete local links. Basically the database could be updated whenever by every change of time table. But practically this process of updating means to find a public transport connection between almost all municipalities in the Czech Republic (see conditions above) and make sophisticated analysis. The whole process is very time-consuming and demands the parallel way of processing. The processing distribution represents a core of this paper.

2 THE DATABASE BUILDING

The building process of the database consists of two phases:

1. searching phase, including massive searching of various possible public transport connections,
2. analyses of found data and recording of parameters of the optimal commuting variant for the given working time interval.

The first phase consumes the most time and consequently it is the main problem of database of transport services creating.

3 THE SEARCHING PRINCIPLE

During the first phase of database creation it is necessary to find all public transport connections for all requested pairs of municipalities. Each pair respecting defined time criteria has to be individually delivered to a searching module for processing. It means to process almost 13 million of municipal combinations within 100 kilometers. This part of processing is very demanding due to a large number of combinations.

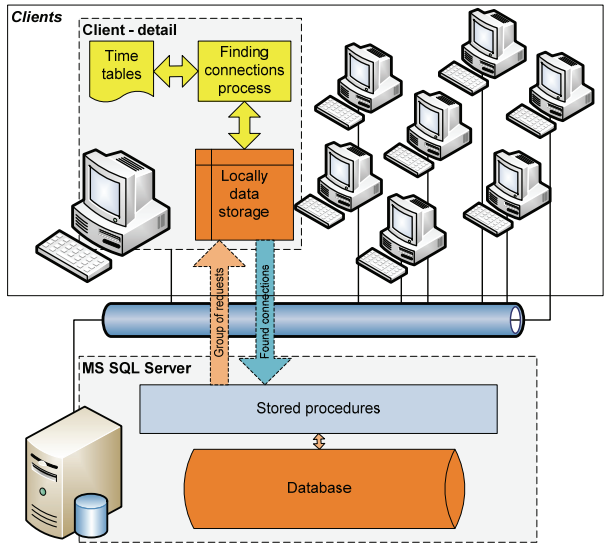

Fig. 3 Distributing method principle

To reduce the process of connection searching it is necessary to distribute the whole process into a set of workstations. Particular data segments are processed here parallelly.

A previous application distributed segments manually (it means cutting the database to pieces and copying them to individual stations for processing by DOK and NEWDOK utilities). Then the results were collected manually together to some group for next processing. Naturally such distribution is very demanding, time-consuming and error prone.

The new solution automates the whole process. The new method is based on client-server technology (Fig. 3), where the server part provides database of MS SQL Server 2008. This database contains all searched combinations of municipalities and it saves all found connections as well. The client part of this solution is created of circa 30 computers with special software for connection searching.

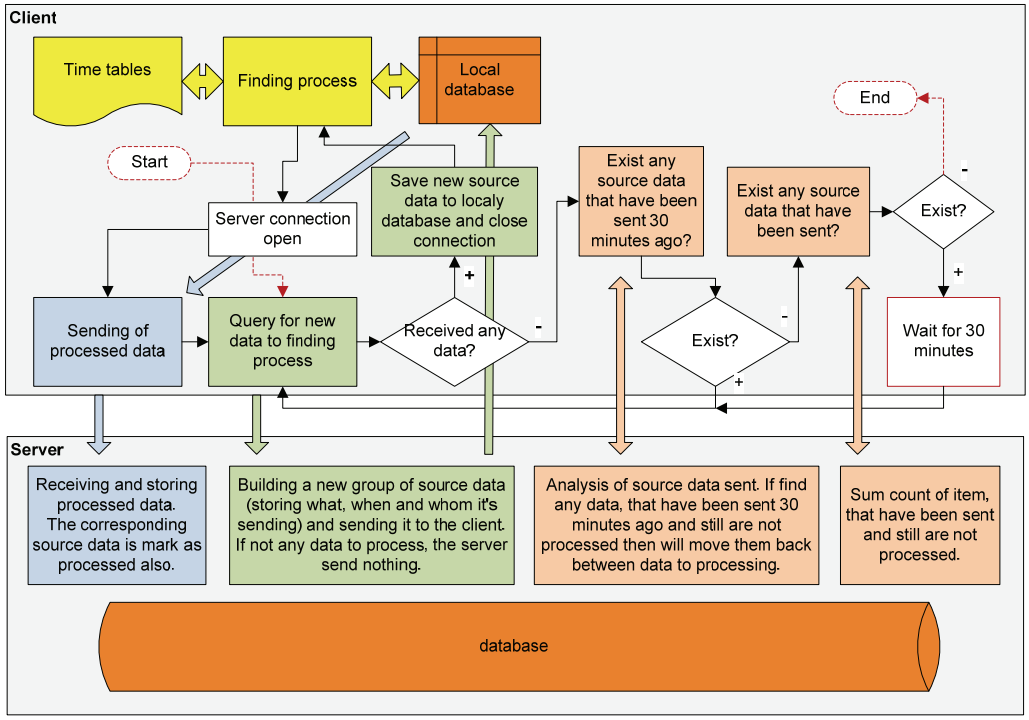

Fig. 4 Searching process

The client starts the proper requests distribution with a data request (Fig. 4). Concrete client calls the saved procedure on the server. Then the server answers this request with sending group of municipalities' combinations. After that a link between server and client is closed and the client starts the searching process. After finishing the link with server is created again and data (found connections) are transferred to a server and a client requests a new group of municipalities' combinations. This starts a new searching cycle.

The client asked for data via on the server saved procedure. This procedure separates a new group from non searched requests (by default 1200 records) and it sends this group to a client. This procedure notes information about which data, whom and when they were sent. When results are retrieved, the procedure marks the records as completed. The server keeps information about to whom and when requests have been sent and by whom were searched. This information is necessary in case of some client collision. The server recognizes such situation so that it does not get a request till 30 minutes later after sending (five times more than the average processing time of one group).

This evaluation is not made continuously but as a final step of the whole processing. Practically it is evaluated when the client requests a new group of data but all requests have been already sent to clients. If the client gets an empty set, it asks the server for selecting some already sent

but non-processed requests, which are older than 30 minutes. If there are some of them, they are marked as non-sent. The procedure sends the number of these records to the client. The client will either repeat the request for a new group of data or (if it is null) it will ask for a number of requests which were sent and it detects if there are any of them at all, waiting for the processing. If the server returns zero, it means, that all requests were processed and client will end the work. Otherwise the client waits 30 minutes and then it repeats the whole process again. The whole process is very robust and it counts with every possibility, like some client crash by saving found data, late reaction by saving found data, late reaction by that limit etc.

Client software is based on .NET platform in MS Visual Basic 2008. As it was written above, a client will disconnect immediately when it gets the data from a server. This fact saves system sources of the server. A client is searching connection between municipalities during the time in disconnected status. After it finishes the searching process, it connects to the server to send the results. So the potential of .NET platform is fully exploited. For physical connection searching process a module from CHAPS Company is used there. The client part is approaching the valid time table through this module, as well as searching for connections, computing prices etc.

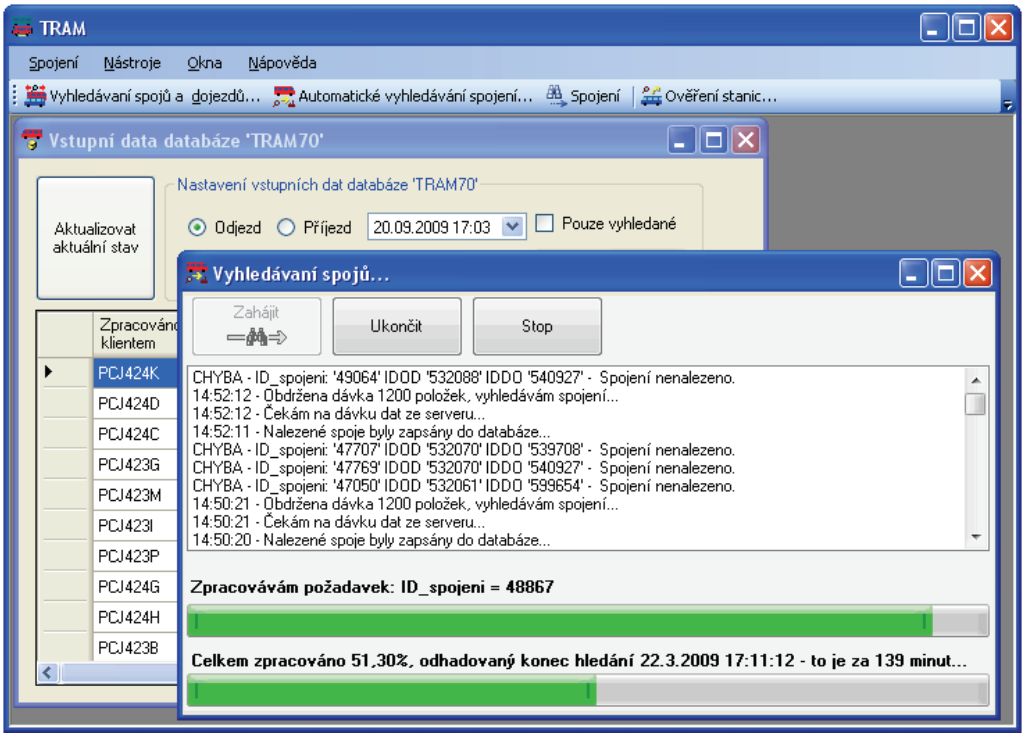

Fig. 5 Application during searching process

4 CONCLUSIONS

Building the database of public transport services is very time-consuming task. Most of the time is taken by the initial search of all the transport connections between communities. This is illustrated by the fact that the output database file reaches the size about 210 gigabytes, containing about 120 million transport connections.

Since the beginning the task has been solved by parallel processing on a group of computers. In the past the data distribution on computers was carried out manually. Currently the new solution already carries out the distribution completely silently. Thereby it is reducing the total time of the creation of database from the original two weeks to a mere 4 days. The great advantage of the new

solution is its independence on the number of parallel running computers. So the time duration of the processing can be reduced by adding other computers. Possible increase in performance is very important not only from the perspective of the needs to carry out more frequent updates, but also allows the future database building with more details. For example, creation of a database transport services is planned, for the transport connection of parts of towns and municipalities. This will significantly increase the number of combinations for processing and thus the intensity of the whole task.

The database can be utilized not only for searching parameters of a commuting request, but also for an evaluation of transport accessibility. The concept of accessibility of geographical objects has been studied since the 1950's. The practical provision of accessibility analysis is demonstrated e.g. in Bracken (1994) or Burrough et al. (1998). The next development of the database will include an automation of this step. Comparison and aggregation of transport commuting conditions for a municipal level will be followed by an evaluation of conditions for district and regional levels. The public transport accessibility progress in Czech Republic between 2006 and 2008 is depicted in Horák et al. (2008). Results serve regional authorities for monitoring and taking appropriate measures in their transport policy.

Acknowledgement: The data and the transport searching engine are provided by courtesy of CHAPS L.t.d. The project is supported by the Ministry of Social Affairs of the Czech Republic.

REFERENCES

[1] Attitudes on issues related to EU Transport Policy (2007) Directorate General Communication. [online] Available from WWW: <URL: http://ec.europa.eu/public_opinion/flash/fl_206b_en.pdf>

[2] BRACKEN, I. (1994) Towards improved visualization of socio-economic data. In Hearnshaw H., Unwin D. (ed.): Visualization in Geographical Information Systems. UK, John Wiley & Sons Ltd., 1994, p.76-84, ISBN 0-471-94435-1.

[3] BURROUGH, P., MCDONELL, R., MCDONELL, R. 1998. *Principles of Geographical Information Systems*. Oxford, Oxford University Press, 336 pp., ISBN 0-19-823365-5.

[4] FARANA, R. *Metodika návrhu informačních systémů s důrazem na použití v průmyslové praxi* Vědecké spisy FS, VŠB-TU Ostrava, Edice Habilitační a inaugurační spisy, sv. 37, 2007, 36 s. ISBN 978-80-248-1395-0

[5] HORÁK, J., ŠEDĚNKOVÁ M., HORÁKOVÁ B.: Public Transport Accessibility based on Time Schedule Analyses. In Hana Svatoňová et al.: Geography in Czechia and Slovakia. Theory and Practice at the Onset of 21st Century. Brno, 2008. ISBN: 978-80-210-4600-9.

[6] HORÁK, J., ŠEDĚNKOVÁ M., IVAN, I.: Modelling of transport accessibility for municipalities of the Czech republic. In Proceedings of GIS Ostrava 2008, 27. - 30. 1. 2008, VŠB-TU Ostrava, ISBN 978-80-254-1340-1.

[7] CHAPS 2009. *Website of Chaps Company*. [online]. Available from WWW: <URL: http://www.chaps.cz/ >

[8] JORDAN, P. 1995. *Functional regions in East-Central Europe defined on the basis of the frequency of public bus traffic*. Geografický časopis, volume 47, 1995, issue 1, pp 9-15.

[9] RD 2007. Transport Yearbook of the Czech Republic. Ministry of Transport of the Czech Republic [online] Available from WWW: <URL: http://www.sydos.cz/cs/rocenka-2007/index.html>