

TVORBA DATABÁZE SPOJENÍ PRO ČESKOU REPUBLIKU

Ing.Monika Šeděnková¹, Doc.Dr.Ing. Jiří Horák¹, Ing.David Fojtík Ph.D.²

¹ Institut geoinformatiky, Hornicko-geologická fakulta, VŠB-TU Ostrava,
17.listopad u 15/2172, 708 33 Ostrava-Poruba, Česká Republika
monika.sedenkova@vsb.cz, jiri.horak@vsb.cz

² Katedra automatizační techniky a řízení, Fakulta strojní, VŠB-TU Ostrava,
17.listopad u 15/2172, 708 33 Ostrava-Poruba, Česká Republika
david.fojtik@vsb.cz

Abstrakt. Pro potřeby Integrovaného informačního portálu Ministerstva práce a sociálních věcí České republiky a další činnosti především pro úřady práce realizovala Vysoká škola báňská – Technická univerzita Ostrava přípravu databáze dopravních spojení, která obsahuje záznamy o dopravních spojeních mezi jednotlivými obcemi České republiky pomocí veřejné hromadné dopravy vlakové a autobusové. K vyhledávání spojení bylo využito aplikace IDOS Jízdní řády společnosti CHAPS s.r.o. a vyvinutého programu NEWDOK. Jednotlivá spojení musí odpovídat dojíždění do zaměstnání na ranní, odpolední a noční směnu a doba cestování nesmí překročit 90 minut. První vyhledávání spojení proběhlo v červnu 2006 a jeho aktualizace proběhla o rok později. Databáze je využita k vyhledávání volných pracovních míst dostupných veřejnou dopravou z místa bydliště na zmíněném informačním portálu i u zprostředkovatelů na úřadech práce. V příspěvku se zaměřujeme na problémy spojené s vytvořením takové databáze a prezentujeme výsledky hodnocení situace obcí v uvedených letech.

Keywords: dopravní dostupnost, GIS, databáze, veřejná hromadná doprava

Abstract. Development of transport connectivity database for the Czech Republic. VSB-Technical university of Ostrava prepared a traffic database which contains all connections between pairs of municipalities throughout the Czech Republic by using public transport (the train and the bus). The database is intended for Integrated Portal of the Ministry of Labour and Social Affairs of the Czech Republic and labour offices. We utilised NEWDOK programme which is tailored for searching in time schedule IDOS (application for travel connection searching). The connections were tested for commuting for three work shifts and the time for commuting was 90 minutes. Searching connections was made in 2006 and its updating was made in 2007. The database is used for searching of free working place accessible by using public transport from home to work. In this paper we describe problems connected with preparing of such database and we discuss evaluation of transport connectivity results for municipalities in two years.

Keywords: tradic accessibility, GIS, database, public transport

1 Úvod

Relativní podíl veřejné linkové dopravy v České republice se postupně snižuje. V roce 2005 byl podíl veřejné dopravy na počtu přepravených osob 57,2% [1] a o v roce 2006 56,6% [2], i když významnou část tvoří podíl MHD. Podíl veřejné dopravy je tedy nezanedbatelný, navíc existuje silný veřejný zájem na udržování a rozšiřování použití veřejné linkové dopravy a to jak z ekologických tak i sociálních důvodů. Veřejná linková doprava je pro většinu občanů nepostradatelná veřejná služba, která zajišťuje dopravní obsluhu a rozvoj území a mnohdy na základě dopravní obslužnosti se občané rozhodují o místě svého trvalého bydliště.

Existence a parametry spojení veřejné linkové dopravy ovlivňují i situaci na trhu práce, i když zřejmě méně než se všeobecně soudí [3]. Při rozhodování o přijetí nabídky volného místa hraje otázka dopravního spojení nezanedbatelnou roli. Proto i Ministerstvo práce a sociálních věcí České republiky (MPSV ČR) na svém informačním portálu hodlá nabídku doplnit o informace o dopravním spojení.

Databáze dopravních spojení (tj. databáze pro vymezení spádových území dopravní obslužnosti) byly vytvořeny pro potřeby Integrovaného informačního portálu MPSV ČR a další činnosti MPSV a úřadů práce (ÚP). Základním cílem aplikace na informačním portálu je poskytnout možnost vyhledávání volných míst v dosahu veřejné linkové dopravy podle stanovených parametrů.

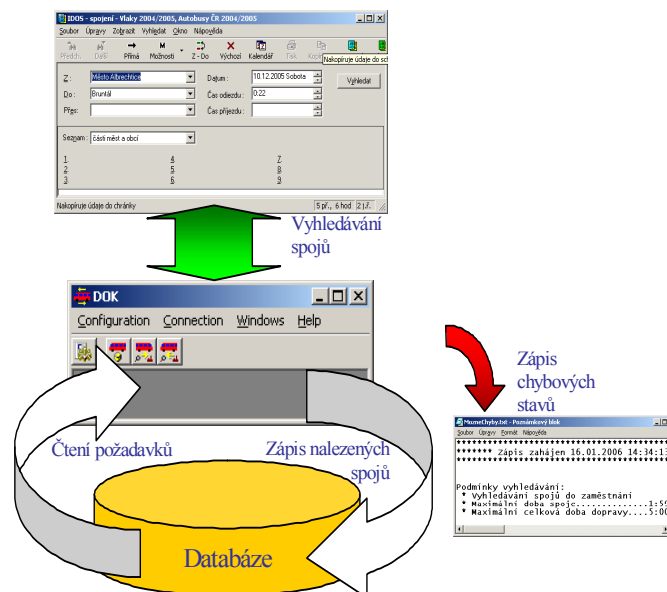
Pro vyhledávání spojení v programu Jízdní řády společnosti CHAPS s.r.o. byla vytvořena aplikace NEWDOK, která je na Institutu geoinformatiky ve spolupráci s Katedrou automatizační techniky a řízení vyvíjena od roku 2001.

2 Programy DOK a NEWDOK

Program NEW-DOK je určen pro automatizované zpracování Jízdních řádů. Jeho hlavní činností je vyhledat všechna platná dopravní spojení mezi sadou nástupních bodů spádové oblasti a cílovým bodem. Nástupními body jsou obce, části obcí, případně zastávky, které aplikace čte z databáze. Aplikace vyhledá všechna vyhovující spojení, a запиše je do databáze, kde jsou připravena k další analýze.

První verze aplikace DOK byla vytvořena ve Visual Basicu 5 v rámci diplomové práce studentem Ing. Tomášem Kettnerem. Později byla pro potřeby jiné diplomové práce upravena studentem Ing. Lukášem Růžičkou. Bohužel používání těchto verzí bylo těžkopádné a neintuitivní. Samostatnou kapitolou bylo navázání aplikace na novou verzi databáze jízdních řádů aplikace IDOS, které vyžadovalo provést řadu změn na mnoha místech přímo v nepřehledných zdrojových kódech aplikace. Proto bylo přistoupeno k přepracování s důrazem na snadné a intuitivní ovládání, konfigurovatelnou, vyšší spolehlivost a možné rozšíření aplikace o nové funkce. Realizaci nové verze zajistil Ing. David Fojtík, Ph. D. . Vznikla tak zcela nová verze aplikace NEWDOK, vytvořená ve Visual Basicu 6.0 s novým uživatelským rozhraním a přepracovaným jádrem zvyšující spolehlivost a rychlost aplikace.

Vlastní vyhledávání spojení aplikace NEWDOK nerealizuje přímo, ale prostřednictvím produktu IDOS pro Windows. Ten produkt firmy Chaps za úplatu poskytuje platné jízdní řády osobní veřejné linkové dopravy [CHAPS]. NEWDOK tento program ovládá mechanismem zasilání zpráv, tedy pomyslnou simulací klávesnice. Jednoduše vyplní všechna kritéria hledaného spojení a požádá o jeho nalezení. IDOS spojení vyhledá a po povelu je запиše do textového souboru, odkud je NEWDOK přečte a zpracuje.



Obr. 1 Princip práce aplikace NEWDOK

Jednotlivé fáze vyhledávání podkladových dat lze shrnout do několika bodů:

1. NEWDOK spustí aplikaci IDOS a provede její nastavení.
2. Z databáze přečte jeden požadavek (záznam) na nalezení spojení.
3. Metodou zasilání zpráv vyplní jednotlivé položky v rozhraní IDOSu (kritéria hledaného spojení).
4. Předá povel k vyhledání a čeká dokud IDOS hledání neukončí.
5. Po vyhledání předá povel k tisku nalezených spojů do souboru.
6. Otevře a načte soubor.
7. Nalezené spoje zpracuje a podle zadaných kritérií zkontroluje.
8. Vyhovující záznamy запиše do databáze, případně nevyhovující запиše do souboru chyb.
9. Přejde na další záznam požadavků a v cyklu opakuje činnosti od bodu 2.
10. Po zpracování celého seznamu požadavků IDOS uzavře.

Hlavním problémem uvedeného řešení je v rychlosti zpracování dat a v potencionální nestabilitě. Důvodem je onen způsob komunikace s aplikací IDOS (zasílání zpráv, zápis a čtení souboru), který je nejen výrazně pomalý, ale také velmi citlivý na jakýkoliv zásah uživatele. Přitom se aplikace používá k celostátní analýze dopravní obslužnosti, kdy počty hledaných spojů překračují velikost patnácti miliónu záznamů. Pochopitelně se tato rozsáhlá úloha řeší distribučně, paralelním zpracováním dílčích částí na co největším počtu počítačů (jedná se desítky počítačů). Přesto vlastní vyhledávání spojů trvá několik týdnů.

Z těchto důvodů se začalo hledat nové řešení, které by umožnilo vyhledávat spoje přímo, bez složité a pomalé komunikace s programem IDOS. Po dohodě se společností CHAPS byla zakoupena licence na přímé používání dynamicky linkované knihovny TT.DLL pro potřeby zpracování databáze dopravních spojení pro MPSV ČR. Byla připravena nová aplikace TRAM, která využívá přímé komunikace s knihovnou TT.DLL a pro uživatele osvědčené uživatelské rozhraní, takže nebylo potřeba měnit metodiku přípravy a následného zpracování dat.

3 Příprava databáze spojení

K tvorbě databáze dopravních spojení ČR byla testována spojení mezi všemi obcemi ČR do přímé vzdálenosti 70, resp. 100 km (toto omezení muselo být uplatněno kvůli geometricky rostoucím výpočetním nárokům). Vyhledávání spojení tedy probíhalo pouze na úrovni obcí (tedy mezi hlavními stanicemi v obcích startu a cíle), spojení jsou zjišťována pro dojížděku na 6., 7., 8., 14., a 22. hodinu, avšak bez ověřování cesty zpět. Vyhledání spojení je využito jízdních řádů autobusových a vlakových spojení, nepředpokládalo se využití místní dopravy. Vyhledávala se spojení s přestupy, v intervalu šesti hodin a maximální počet přestupů byl stanoven na 5. Seznam vyhledaných spojení byl následně omezen pouze na spojení kratší než 90 minut (požadavek MPSV ČR). Nastavení parametrů vyhledávání bylo limitováno rozsáhlostí úlohy a výpočetními nároky, rovněž některé parametry byly stanoveny odběratelem a nemohly být proto optimalizovány. Podrobnější popis k volbě parametrů a získané výsledky v území okresu Bruntál jsou dostupné v [3].

Veškerá dopravní spojení byla vyhledána k datu 20.6.2006 a následná aktualizace v roce 2007 k datu 19.6.2007. V době přípravy článku probíhá nové vyhledávání dle nových platných jízdních řádů (9.12.2007).

Pro zpracování se využívá databáze Microsoft Access (MSA). Bylo potřeba vyhledat 6,978,599 kombinací obcí na jednu směnu tedy celkově 20,935,797. Bohužel databáze MSA je omezena velikostí cca 2GB, což zapříčinilo rozdělení celku do dílčích databází dle krajů viz tabulka 1. Ovšem na konci vyvstává nutnost výsledky spojit a uložit do finální tabulky, jejíž struktura je uvedena v tabulce 2.

Tab.1 Rozdělení do dílčích databází dle krajů

Kraj	Počet záznamů na vstupu	Počet dílčích DBS na jeden vyhledávací čas
Moravskoslezský (MSK)	164 472	1
Středočeský (STK)	1 637 061	8
Olomoucký (OLK)	439 164	1
Zlínský (ZLK)	258 725	1
Liberecký (LLK)	195 649	1
Ústecký (USK)	307 518	1
Karlovarský (KVK)	55 995	1
Hlavní město Praha PRH	1 456	1
Jihomoravský (JMK)	781 306	3
Plzeňský (PLK)	436 842	2
Královehradecký (KHK)	514 414	2
Jihočeský (JCK)	563 765	5
Pardubický (PAK)	645 735	3
Vysočina (VYK)	976 510	5

V případě více spojení muselo být rozhodnuto o nejlepší variantě spojení, která byla použita pro určení požadovaných parametrů; jinak by byly uloženy údaje nekonzistentní, protože by jednotlivé atributy odpovídaly různým spojům (tedy např. doba cestování na 6.hodinu by byla určena z jiného spoje než cena) a to by mohlo vést k mylným interpretacím. Výběr varianty byl prováděn podle minimální vzdálenosti vyjma případů, kdy tato varianta vykazovala významně horší výsledky v dalších ukazatelech; tehdy se vybírala jiná optimální varianta.

První varianta vybírá spojení minimální **podle doby cestování**. Minimum je postupně stanovováno podle doby, počtu přestupů, plné ceny, vzdálenosti v kilometrech a zákaznické ceny. Nejprve se nalezne minimální doba cestování, kde v případě více případů se stejnou dobou cestování se vybere spoj s nejmenším počtem přestupů.

Existuje-li takových případů více, vybere se minimální dle plné ceny. Existuje-li takových případů více, vybere se podle minimální vzdálenosti a pokud i dále existuje takových případů více vybere se podle minimální zákaznické ceny.

Druhá varianta vybírá spojení minimální **podle plné ceny**. Minimum je postupně stanovováno podle plné ceny, doby cestování, počtu přestupů, vzdálenosti v kilometrech a zákaznické ceny. Nejprve se nalezne minimální plná cena při cestování. V případě více případů se stejnou cenou se vybere spoj s nejkratší dobou cestování. Dále pokračuje dle varianty první v jiném pořadí atributů.

Třetí varianta vybírá spojení minimální **podle počtu přestupů**. Minimum je postupně stanovováno podle počtu přestupů, doby cestování, plné ceny, vzdálenosti v kilometrech a zákaznické ceny. Poslední varianta vybírá spojení minimální **podle vzdálenosti**. Minimum je postupně stanovováno podle vzdálenosti v kilometrech, doby cestování, počtu přestupů, plné ceny a zákaznické ceny.

Rovněž se zjistí takzvaná nultá varianta, kdy jsou samostatně v jednotlivých polích určena minima. Může nastat, že pro danou kombinaci minimálních hodnot často neexistuje reálné spojení. Následně jsou posuzovány odchylky vybraných hodnot podle jednotlivých variant od nulté varianty. Je vybírána varianta č.4, pokud nedojde k hrubé odchylce, která je nastavena následovně:

- pokud je rozdíl ve vzdálenosti varianty 4 oproti nulové variantě větší než 20 km, vybere se varianta 4
- pokud je rozdíl v počtu přestupů varianty 3 oproti nulové variantě větší než 2, vybere se varianta 3
- pokud je rozdíl v plné ceně varianty 2 oproti nulové variantě větší než 20 Kč, vybere se varianta 2
- pokud je rozdíl v době cestování varianty 1 oproti nulové variantě větší než 20 minut, vybere se varianta 1

Uvedené limity je možné v případě potřeby upravit, stejně jako celou strategii výběru.

Tab. 2: Struktura tabulky DOJEZD

DOJEZD		
Název pole	Datový typ	Popis
KODOD	char	kód obce, ze které se spojení hledalo (výchozí obec)
KODDO	char	kód obce, do které se spojení hledalo (cílová obec)
PRIMEKM	integer	přímá (Euklidovská) vzdálenost v km mezi obcemi
VEREJNEKM	integer	vzdálenost v km mezi obcemi s využitím VHD
DOBA6	integer	doba dojížděky vybraného spojení na 6:00 ve formátu hodiny:minuty
PRESTUP6	integer	počet přestupů pro vybrané spojení na 6:00
CENA6	integer	cena (plná) vybraného spojení na 6:00
CENAZ6	integer	cena zlevněná (zákaznická) vybraného spojení na 6:00
POCET5_6	integer	počet spojení mezi 5. a 6. hodinou
DOBA7	integer	doba dojížděky vybraného spojení na 7:00 ve formátu hodiny:minuty
PRESTUP7	integer	počet přestupů pro vybrané spojení na 7:00
CENA7	integer	cena (plná) vybraného spojení na 7:00
CENAZ7	integer	cena zlevněná (zákaznická) vybraného spojení na 7:00
POCET6_7	integer	počet spojení mezi 6. a 7. hodinou
DOBA8	integer	doba dojížděky vybraného spojení na 8:00 ve formátu hodiny:minuty
PRESTUP8	integer	počet přestupů pro vybrané spojení na 8:00
CENA8	integer	cena (plná) vybraného spojení na 8:00
CENAZ8	integer	cena zlevněná (zákaznická) vybraného spojení na 8:00
POCET7_8	integer	počet spojení mezi 7. a 8. hodinou
DOBA14	integer	doba dojížděky vybraného spojení na 14:00 ve formátu hodiny:minuty
PRESTUP14	integer	počet přestupů pro vybrané spojení na 14:00
CENA14	integer	cena (plná) vybraného spojení na 14:00
CENAZ14	integer	cena zlevněná (zákaznická) vybraného spojení na 14:00
POCET13_14	integer	počet spojení mezi 13. a 14. hodinou
DOBA22	integer	doba dojížděky vybraného spojení na 22:00 ve formátu hodiny:minuty
PRESTUP22	integer	počet přestupů pro vybrané spojení na 22:00
CENA22	integer	cena (plná) vybraného spojení na 22:00
CENAZ22	integer	cena zlevněná (zákaznická) vybraného spojení na 22:00
POCET21_22	integer	počet spojení mezi 21. a 22. hodinou
DATUMHLED	char	datum, ke kterému vyhledání spojení provedeno
URL	char	URL řetězec potřebný pro vyhledání seznamu spojení mezi uvedenými 2 obcemi

Pro představu uvedeme příklad pro dva kraje. Databáze pro MSK je střední velikosti, databáze pro STK je s největším počtem hledaných spojení. Databáze pro MSK nebyla rozdělena na více dílčích databází, použila se pouze jedna databáze, která měla velikost 22MB. Po provedení vyhledávání se velikost zvýšila na cca 1GB v závislosti na počtu nalezených spojení. Databáze pro STK byla rozdělena na osm dílčích databází. Každá dílčí databáze měla velikost 28MB. Po provedení vyhledávání se velikost zvýšila až na 1,8GB a to pro jeden vyhledávací čas. Je tedy vidět, že při následném spojování do výsledné tabulky DOJEZD je nutno zpracovat např. Pro STK bylo použito 24 dílčích databází, každá o velikosti cca 1,8GB. Tedy náročnost na výpočetní techniku je vysoká.

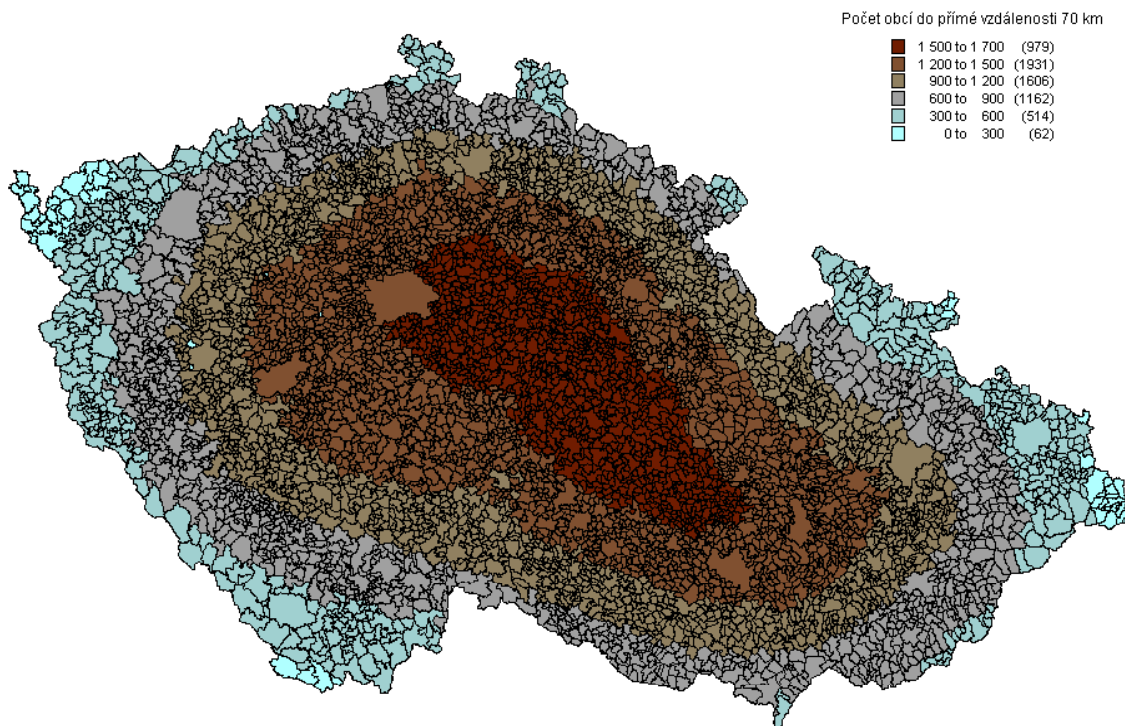
V roce 2006 byla vyhodnocena spojení pro celkem 6,978,599 kombinací obcí, představující obce vzdálené do 70 km přímou vzdáleností. Reálné spojení veřejnou hromadnou dopravou ve vymezených intervalech (celkově 4:00 – 8.00, 12:00 - 14:00, 20:00 - 22:00) však bylo nalezeno pro 880795 záznamů. Spojení pro dojížděku na 6. hodinu je evidováno v 407062 případech, na 7.hodinu v 621128 případech, na 8.hodinu v 505754 případech, na 14.hodinu v 338685 případech, na 22.hodinu v 83747 případech.

V roce 2007 bylo hodnoceno spojení pro 5,889,788 záznamů, představujících vazby mezi obcemi vzdálenými do 70 km přímou vzdáleností. V další etapě byly hodnoceny obce vzdálené 70-100 km. Reálné spojení veřejnou hromadnou dopravou ve vymezených intervalech (celkově 4:00 – 8.00, 12:00 - 14:00, 20:00 - 22:00) však bylo nalezeno pro 741668 záznamů. Počty nalezených spojení poklesly, což se projevilo i na jednotlivých charakteristikách.

4 Situace v ČR

Následné využití takovéto databáze je široké. Využití podobných databází spojení můžeme najít od lokálních aplikací určených např. pro studium dostupnosti potenciální investice a určení její spádové oblasti, až po úroveň republikovou, kdy je možné přehledně srovnat situaci v celém území, v univerzálním srovnání hledat rozdíly v dostupnosti.

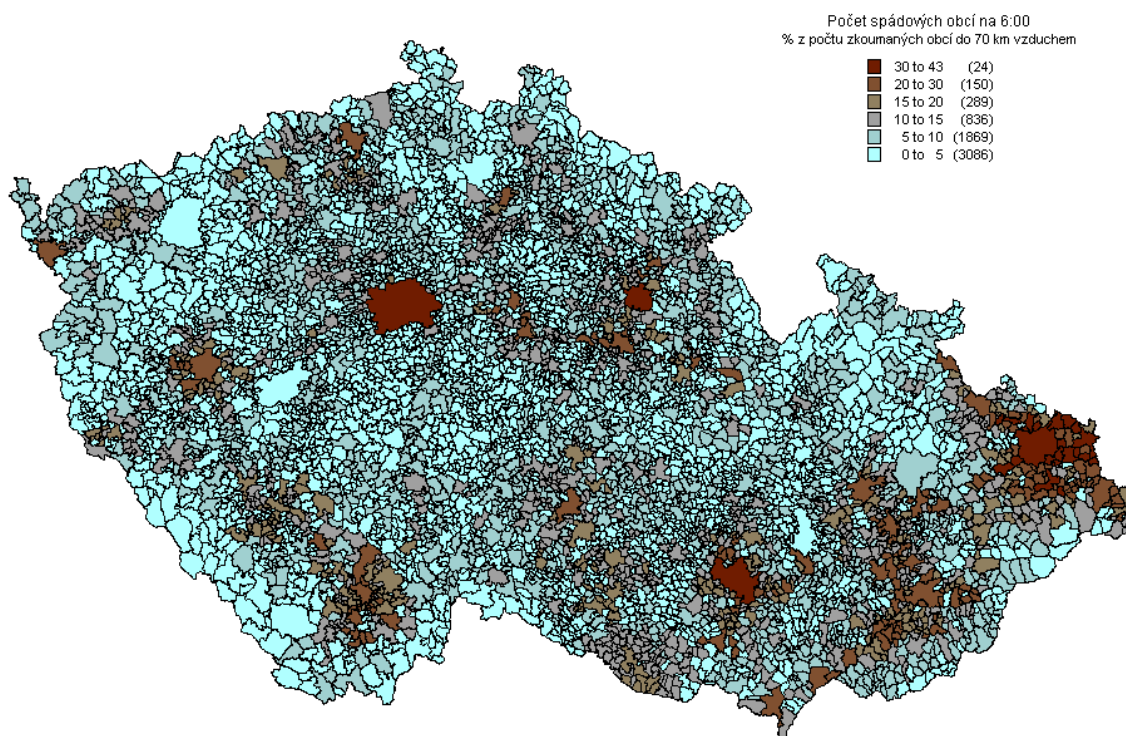
Na obr. 3 je zobrazen počet obcí do přímé (Euklidovské) vzdálenosti 70 km.



Obr.3. Počet obcí do přímé vzdálenosti 70km

Obr. 4 zobrazuje dostupnost obce počtem dostupných obcí na 6.hodinu vyjádřeným jako % z celkového počtu testovaných obcí. V ČR je samozřejmě nejlepší situace v Praze, do které je možné dojíždět ze 64 % všech obcí vzdálených do 70 km (přímá vzdálenost). Můžeme však pozorovat i další místa s výbornou dostupností. Do

nejvyšší kategorie nad 30% patří většina krajských měst, ovšem významnou koncentraci dobře dostupných obcí lze nalézt na střední Moravě. Velmi zajímavá je situace Ostravy a okolí, přičemž je to příhraniční oblast.



Obr.4 Podíl spádových obcí na 6:00

Z databáze spojení a volně dostupných dat je možné sledovat tyto základní ukazatele dopravní obslužnosti:

1. počet dopravních spojení pro dojížděku na všechny sledované časy ve vymezených oblastech
2. průměrný počet nalezených dopravních spojení v jednotlivých obcích
3. hustota dopravní sítě
4. počet přestupů

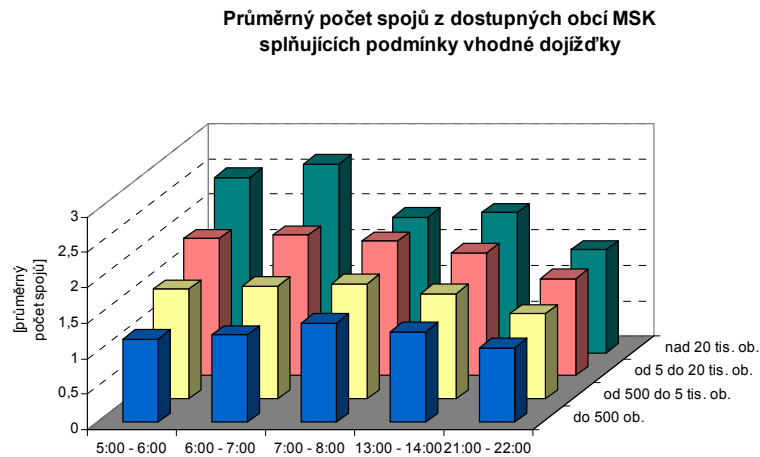
AD 2)

Pro srovnatelnost byly počty spojů veřejné hromadné dopravy vztaženy na celkový počet obyvatel obcí v uvedených krajích. Výsledné počty jsou uvedeny formou aritmetického průměru. Obce byly rozděleny podle následujících velikostních kategorií:

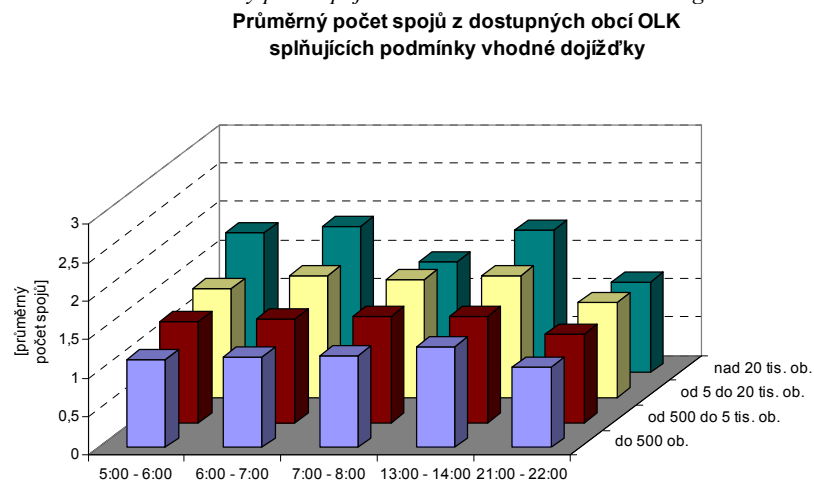
obce do 500 obyvatel.....“malé obce“
 obce od 500 do 5 000 obyvatel.....“velké obce“
 obce od 5 tisíc do 20 000 obyvatel...“malá a střední města“
 obce nad 20 000 obyvatel.....“velká města“

Velikostní rozdělení bylo převzato z [4].

Srovnání bylo provedeno pro kraje Moravskoslezský a Olomoucký. Cílem této analýzy je zjistit, do jaké míry jsou obce v jednotlivých velikostních kategoriích obsluhovány, jelikož ukazatel průměrného počtu spojů ze sledovaných obcí má výrazný vliv na dopravní obslužnost sledovaných území.



Obr. 5: Průměrný počet spojů z obcí MSK dle velikostních kategorií



obr. 6 Průměrný počet spojů z obcí OLK dle velikostních kategorií

Z výsledků jasně vyplývá, že u obcí obou zkoumaných území jsou nejméně obsluhovány obce v segmentech do 500 obyvatel a od 500 do 5 tis. obyvatel. V případě obcí do 500 obyvatel nejsou rozdíly mezi jednotlivými kraji tak výrazné, v ostatních kategoriích však dochází k výraznému vzrůstu hodnot u MSK, který nad OLK v této oblasti zcela dominuje. Pokud bychom sledovali průměrný počet spojů z jednotlivých obcí na základě sledovaných časů, pak u obou sledovaných území dochází k rozdílnému trendu. U obcí MSK, jejichž počet obyvatel nepřesáhl 5 tis., můžeme sledovat nejvyšší průměrný počet spojů na 8. hodinu, v ostatních kategoriích zcela viditelně převládá dojížd'ka na 6. a 7. hodinu. Je zřejmé, že u největších obcí je možné využít ve větší míře dojíždění na směnný provoz (počty spojů na 6, 14 a 22 hodin), výsledek je ovšem ovlivněn výrazně menším počtem obcí v této kategorii. Z hlediska počtu dopravních spojení lze konstatovat výrazně lepší situaci v MSK.

5 Závěr

Databáze dopravních spojení bude primárně sloužit pro vyhledání dostupných volných míst na úřadech práce. Předpokládáme, že do budoucna s růstem výkonnosti programu bude možné přejít na testování spojení mezi částmi obcí, což by lépe mapovalo situaci v území a dovolilo zpřesnit nabídku volných míst dostupných veřejnou dopravou. Studium výsledků v rámci celé republiky poskytuje zajímavé informace. Dovoluje mapovat situaci v území, sledovat vazby mezi sídly, ukazuje existenci jistých dopravních koridorů, které ne vždy korespondují s významem silniční a železniční sítě. Meziroční srovnání situace v krajích je popsáno v příspěvku "MODELLING PUBLIC TRANSPORT ACCESSIBILITY FOR MUNICIPALITIES IN THE CZECH REPUBLIC".

Poděkování

Data a přístup k programovému jádru byl získána od společnosti CHAPS s.r.o. Projekt je podporován Ministerstvem práce a sociálních věcí ČR.

Literatura

- [1] --- RD 2005. Ročenka dopravy České republiky 2005. Ministerstvo dopravy. Dostupný na WWW:<
<http://www.sydos.cz/cs/rocenka-2005/index.html>>
- [2] --- RD 2006 Ročenka dopravy České republiky 2006. Ministerstvo dopravy. Dostupný na WWW:<
<http://www.sydos.cz/cs/rocenka-2006/index.html>>
- [3] Horák, J., Horáková, B., Šeděnková, M., Šimek, M., Růžička, L., Peňáz, T.: Dostupnost zaměstnavatelů v okrese Bruntál (2006), 140 str. Dostupný na WWW:<
http://gis.vsb.cz/GACR_MTP/Clanky/dostupnostBR_X2.pdf>
- [4] Hnilová, L. : Vyhodnocení dostupnosti obcí na základě přímé a časové dostupnosti. (2007) Diplomová práce, VŠB-TU Ostrava, Institut geoinformatiky, 71.str.