

Posudek oponenta magisterské práce

Pavel Břichnáč: Bezpečnost publikování prostorových dat na internetu

Předložená magisterská práce Pavla Břichnáče přináší velmi zajímavé téma ochrany prostorových dat na internetu. Jedná se především o popis možností zabezpečení dat, která jsou publikována pomocí nejrozšířenějších technologií současnosti - rastrové dlaždice, služeb WMS a WFS. Autor si vytyčil cíl popsat technologie publikace prostorových dat, možnosti zabezpečení na datové úrovni (degradace dat, vodoznaky apod.) a rámcově představit i zabezpečení dané vnějšími prostředky (možnosti sítě internet, autorizace apod.). Autor předkládá logické dělení jednotlivých metod zabezpečení dat. Na základě tohoto dělení si autor volí příklady jednotlivých mapových serverů a pro každý příklad testuje možnosti stažení prostorových dat. Díky tomuto praktickému postupu velice dobře ukazuje a také dokazuje existenci slabých míst jednotlivých mapových serverů. Přesnost a kvalita stažených dat je definována ke vzorku originálních dat.

Magisterská práce má rozsah 70 stran včetně seznamu zdrojů, obrázků, tabulek, zkratk a skriptů. Přílohou práce je CD, které obsahuje digitální verzi práce, stažená testovací data a vlastní autorovo skripty. Text je psán velmi srozumitelně a čitelně a podává ucelený průřez problematikou a vhodně dokumentuje autorův přínos. Přesto je na některých místech vidět nepřesnost ve formulacích a lehká nesrozumitelnost. To ukazuje na pravděpodobně nedostatek času na finální dokončení práce. Jako příklad může sloužit str. 53 a poslední nedokončená věta („Hodnoty jsou vyšší než v předchozím případě a tedy výsledek vektorizace dopadl hůře zejména z důvodu náročnějšího zpracování. Hodnoty jsou však stále “) nebo na str. 45 věta: “Tím byly topologické rozdělení objektů.” Po formální stránce je práce bez výraznějších chyb, ačkoli seznamy zkratk, obrázků apod. se v současnosti většinou řadí na začátek práce. Zde bych měl připomínku k popisu obrázků a schémat - je nutné vysvětlit a popsat použité prvky a značky. Autor to v práci dodržuje, nicméně např. na obr. 2 není zřejmé použití čísel ve schématu (vysvětlení není možné dohledat ani v textu). Autor vychází z řady literárních zdrojů, které v práci vhodně cituje (formou číselného odkazu, možná vhodnější použít formu - název autora, rok).

Magisterská práce se skládá ze 7 kapitol a obsahově se dělí na 4 logicky uspořádaná témata - vymezení a typologie prostorových dat, technologie publikace prostorových dat na internetu pomocí mapových serverů, praktická ukázka možné realizace útoku na data a návrhy opatření ke ztížení automatického stahování dat. Jednotlivé kapitoly na sebe vhodně navazují a dovolují čtenáři dobrou orientaci v problematice. V definici a popisu typů prostorových dat autor správně vychází z používané terminologie a informačních zdrojů. Oceňuji, že autor svůj popis příliš nezobecňuje, ale uvádí především vlastnosti, které mají vztah k publikaci prostorových dat na internetu. Tím zbytečně nerozšiřuje práci mimo vlastní téma, ale také se tím občas dopouští jistých nepřesností. Jako příklad můžeme vidět poznámku, která se věnuje topologii v geoinformatice. Např. na str. 11 „... ale vzájemný vztah objektů (sousednost)” (tím může docházet k zjednodušení, že topologie je v geoinformatice jen záležitostí vztahu „sousednost,” ačkoli může existovat i jiný topologický vztah dvou objektů). Kromě technické popisu jednotlivých typů data je velice správné uvedení aktuální státní datové politiky (INSPIRE) a také argumentů a motivací, které vedou uživatele ke snaze nelegálně získávat data. Autor přináší přehled vybraných mapových portálů v ČR i v Evropě (tabulka 1 a 2). Zde bych u národních mapových portálů viděl jako zajímavé doplnění informace o vztahu portálu k INSPIRE.

Klasifikaci mapových portálů podle způsobu zobrazení výsledků uživateli považuji za správně zvolenou. Čitelnosti práce prospívá, že autor v popisu standardů mapových služeb nejde do velkých detailů (což není cílem práce), ale ukazuje dílčí způsob komunikace se serverem a rozpoznání použité technologie. Použitými technologiemi jsou WMS, WFS, publikace rastrových dlaždic při vytvoření cache na serveru, vytváření dočasných souborů dle požadavků klienta. Jedinou výtku bych viděl v absenci alespoň obecné informace o některých proprietárních řešení, které nevyužívají standardy (např. ESRI řešení). Z hlediska metodologie je použití standardní mapové služby správný postup, ale pokud jde o získání dat, tak lze poměrně často narazit na zdroje, které používají vlastní implementaci.

Stěžejní části práce vidím v kapitolách 5 a 6. Jedná se o praktickou realizaci „útoku na data” a návrhy způsobů zabezpečení dat. Pro praktickou realizaci stažení dat jsou zvolené vrstvy, které jsou velmi malé svým rozsahem a významem. To považuji za správné z etického hlediska. Práce si klade za cíl popsat řešení zabezpečení dat a ne stažení velkého rozsahu atraktivních dat. Výběr mapových serverů pro praktickou realizaci je velmi zdařilý. Důležitým aspektem je zvolení serverů, které využívají rozdílné technologie, ale zároveň je možné stažená data kombinovat dohromady (přehled specifik vybraných serverů na str.24). Je zřejmé, že postup získání dat (odhalení slabého zabezpečení) je rozdílný pro rastrová a vektorová data. Autor po prozkoumání konkrétních technologií navrhuje postup a pomocí skriptů v jazyce php i realizuje postupné stažení rastrových souborů, které představují dané území či vrstvu. Dílčí rastrové soubory jsou ukládány i se svým souřadnicovým vymezením (world file). Autor pak může v dříve uvedené soubory v desktop GIS aplikaci georeferencovat a případně vytvářet mozaiku se snímků. V případě vektorových dat je navržena ještě metoda vektorizace rastrových obrazů a získání atributových dat (WMS služba). K použité technologii nemám výhrady - je samozřejmě možné některé kroky (skripty) více optimalizovat, ale to

nebylo cílem práce. Velmi důležitou a nutnou částí práce je verifikace výsledků a porovnání jejich přesnosti k originálním datům. Dosažené výsledky, kdy odchylka po vektorizaci je menší než 5m, jsou vzhledem procesu vektorizace staženého rastrového obrazu velmi dobré. Otázkou zůstává porovnání rychlosti získání dat (doba stažení rastrových obrázků, georeferencování, vektorizace) a polohové věrnosti vůči originálu. Bylo by zajímavé porovnat závislost polohové přesnosti k velikosti prostorovému rozlišení stahovaných rastrů (v práci většinou 1m/1pixel). Další otázkou je tvarová podoba originálu a získaných vektorových dat (nejenom rozdílnost jednotlivých lomových bodů, případně velikost plochy polygonu).

Navržená opatření ke stížení automatického stahování dat vychází z předchozích praktických postupů a metod. Popis jednotlivé prvky zabezpečení systematizuje dle datové závislosti konkrétního opatření. Domnívám se, že tato kapitola měla být vzhledem k cílům práce obsáhlejší. Práci by prospělo, pokud by se autor pokusil jednotlivé prvky zhodnotit podle míry zabezpečení. Zároveň mi zde chybí definování obecného postupu, který by vedl k maximálnímu možnému znemožnění nelegálního stahování dat. Dalším prvkem této části mohlo být i autorovo zhodnocení jednotlivých technologií publikace prostorových dat na internetu a např. zhodnocení vybraných mapových serverů podle míry zabezpečení. S uvedenými prvky zabezpečení lze nicméně s autorem souhlasit a jejich oprávněnost při použití je velmi dobře dokumentována v kapitole 5. Tím lze ve výsledku práci a uvedené možnosti zabezpečení hodnotit pozitivně. Závěrečné zhodnocení dosažených cílů je věcné a v pořádku.

Magisterská práce je svým tématem velmi aktuální a určitě přináší zajímavý pohled na převládající způsob distribuce prostorových dat. Je zřejmé, že se autor musel seznámit s řadou technologií a detailním principem fungování mapových serverů. Uvedenou práci přes řadu výše zmíněných připomínek či nápadů hodnotím velmi dobře a doporučuji k obhajobě a navrhuji hodnocení výborně.

V Prachaticích dne 18. 9. 2010

Mgr. Stanislav Grill

Katedra aplikované geoinformatiky a kartografie PŘF UK Praha