

Slovenská technická univerzita v Bratislave, Stavebná fakulta

Ing. Tomáš Kliment

Autoreferát dizertačnej práce

**VYHLADÁVANIE ZDROJOV GEOPRIESTOROVÝCH
INFORMÁCIÍ PROSTREDNÍCTVOM INTERNETU**

na získanie akademického titulu doktor (philosophiae doctor, PhD.)

v doktorandskom študijnom programe: 3636 geodézia a kartografia

Bratislava 2012

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Katedre geodetických základov Stavebnej fakulty Slovenskej technickej univerzity v Bratislave.

Predkladateľ: Ing. Tomáš Kliment
Katedra geodetických základov
Stavebná fakulta, STU v Bratislave
Radlinského 11, 813 68 Bratislava

Školiteľ: doc. Ing. Ernest Bučko, PhD.
Katedra geodetických základov
Stavebná fakulta, STU v Bratislave
Radlinského 11, 813 68 Bratislava

Oponenti: Prof. Ing. Ján Tuček, CSc.
Katedra hospodárskej úpravy lesov a geodézie
Lesnícka fakulta, Technická univerzita vo Zvolene
Ul. T. G. Masaryka 2117/24, 960 53 Zvolen

doc. Ing. Petr Rapant, CSc.
Institút geoinformatiky
Hornicko-geologická fakulta
Vysoká škola báňská - Technická univerzita Ostrava
17. listopadu 15, 708 33 Ostrava – Poruba, Česká republika

Ing. Martin Tuchyňa, PhD.
Centrum environmentalistiky a informatiky
Odbor informačných technológií
Slovenská agentúra životného prostredia
Tajovského 28, 975 90 Banská Bystrica

Autoreferát bol rozoslaný:

Obhajoba dizertačnej práce sa koná o h
na Katedre geodetických základov SvF, STU v Bratislave, Radlinského 11, 813 68 Bratislava

Prof. Ing. Alojz Kopáčik, PhD.
dekan Stavebnej fakulty STU
Radlinského 11, 813 68 Bratislava

Obsah

1	Úvod	3
2	Súčasný stav v riešenej problematike	3
2.1	Koncept vyhľadávania v registroch	4
2.2	Koncept vyhľadávania v indexoch.....	5
3	Cieľ dizertačnej práce	6
4	Metodika práce a metódy skúmania	7
4.1	Materiál použitý pre praktické experimenty	7
4.1.1	Geopriestorové dáta a metadáta z domény geodézie a kartografie v SR.....	7
4.1.2	Existujúce programy a technológie použité pri praktických experimentoch.....	7
4.1.3	Programy vyvinuté pre realizáciu praktických experimentov	8
4.2	Metodické postupy a ich aplikácia v praktických experimentoch	10
4.2.1	Metodika tvorby a overenia geopriestorových metadát	10
4.2.2	Aplikácia metodiky na zdrojoch domény geodézie a kartografie – experiment č.1	11
4.2.3	Metodika na vyhľadávanie OGC služieb v internetovom vyhľadávači Google a extrahovanie metadát k zdrojom geopriestorových informácií	13
4.2.4	Aplikácia metodiky pre súčasné OGC služby – experiment č.2	13
5	Výsledky a diskusia	16
5.1	Experiment č.1 - kvantitatívne a kvalitatívne hodnotenie metadát vytvorených pre dáta ZB GIS	16
5.2	Diskusia k výsledkom experimentu č.1	16
5.3	Experiment č.2 – OGC služby nájdené v Google a poskytované metadáta ku zdrojom GI.....	17
5.3.1	Nájdené URL na GetCapabilities požiadavky pre jednotlivé typy OGC služieb	17
5.3.2	Výsledky vyhľadávania podľa verzií jednotlivých typov OGC služieb.....	19
5.3.3	Výsledky po ukončení procesu harvesting.php	20
5.3.4	Kvantitatívne a obsahové vyhodnotenie vytvorených metadát	20
5.4	Hodnotenie výsledkov experimentu a vyplývajúca diskusia	24
6	Záver	26
7	Prínos pre vedu a prax v oblasti geodézie a geoinformatiky	26
8	Summary	27
9	Zoznam použitej literatúry	28
10	Zoznam publikovaných prác doktoranda	30

1 Úvod

World Wide Web (WWW) alebo celosvetová pavučinová sieť má v súčasnosti geopriestorový rozmer (Pellicer, 2011). Webové stránky využívajú geopriestorové informácie (GI) ako polohovú referenciu ich obsahu k Zemi. Webové portály ako Google Maps využívajú geopriestorové dáta rôznej tematiky (cestná sieť, lesy, hydrografia a iné) produkované rôznymi autoritatívnymi poskytovateľmi vo svete (Ministerstvo dopravy v USA, Európska vesmírna agentúra, Environmentálne informačné centrum v Indii a iné). Zručnosť používateľov a poskytovateľov, súvisiaca s manažmentom GI prostredníctvom internetu, tvaruje súčasnú podobu tzv. geopriestorového webu. Geopriestorový web môžeme zdefinovať ako kolekciu webových služieb, geopriestorových dát a metadát, ktoré napomáhajú pri používaní geopriestorových dát v rozsahu ich aplikácie do praxe v rámci konkrétnej domény (biodiverzita, povodne, krajinné plánovanie, doprava a iné).

V dizertačnej práci používame termín geopriestorové webové služby ako webové informačné služby spojené s geopriestorovými dátami a metadátami. Geopriestorové webové služby, geopriestorové dáta a metadáta o geopriestorových zdrojoch (službách a dátach) sú v práci spoločne zdefinované ako geopriestorové webové zdroje, alebo zdroje GI na webe.

Vyhľadávanie špecifickej informácie je dôležitá funkcia, ktorú musí akýkoľvek informačný systém podporovať (Florczyk, 2012). V doméne GI sa v súčasnej dobe pre vyhľadávanie jej zdrojov používajú najmä aplikácie geokatalógov v rámci realizácií infraštruktúr pre priestorové informácie (IPI). Okrem toho, že správcovia vytvárajú zdroje GI v rámci svojej organizácie, potrebujú ich zaregistrovať v rámci geokatalógu IPI vytvorením metadát, ktoré majú preddefinovanú štruktúru. Takto dochádza často k duplikáciám informácií, ktoré sú z časti už obsiahnuté v pôvodných zdrojoch GI (geopriestorových dátach a službách). Na druhej strane internetové vyhľadávače nevyžadujú žiadnu prácu na viac a ako ukazujú výskumy, v súčasnej dobe už poskytujú možnosti vyhľadania zdrojov GI vo významnej miere.

Dizertačná práca sa zaoberá dvomi okruhmi problémov, prvým je automatizácia tvorby metadát potrebných pre vyhľadávanie zdrojov GI v rámci IPI extrahovaním informácií obsiahnutých v pôvodných zdrojoch GI vo forme OGC (Open Geospatial Consortium) služieb. Druhým je stanovenie rozsahu GI poskytovaných internetovými vyhľadávačmi pomocou vyhľadávania URL adries OGC služieb a následnej extrakcie metadát k týmto službám a zdrojom GI, ktoré tieto služby poskytujú.

2 Súčasný stav v riešenej problematike

Geopriestorový web poskytuje v súčasnosti rôzne metódy ako vyhľadávať a pristupovať ku geopriestorovým dátam. Zdefinované sú tri hlavné metódy a to online GIS, infraštruktúra priestorových informácií (IPI) a systémy priestorového prezerania (Pellicer, 2011).

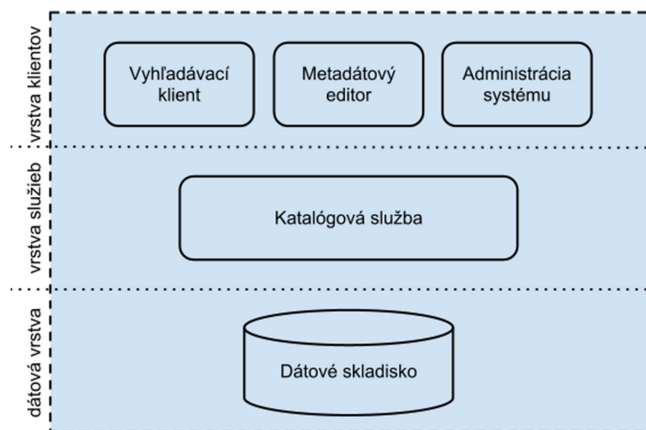
Online GIS systémy sú jedny z prvotných komponentov geopriestorového webu. Tieto systémy by sa nemali mýliť so systémami GIS, ktoré sú nainštalované na jednom počítači a aj používateľ pristupuje a pracuje s nimi z toho istého počítača. Takéto GIS sa nazývajú tzv. stand-alone GIS. GIS komunita bola jednou z prvých, ktorá prijala web ako komunikačnú platformu. Dôsledkom toho bolo, že hlavný predajcovia GIS softvérov ako napr. ESRI začali poskytovať nadstavby nad svoje stand-alone GIS pre potreby publikovania geopriestorových dát na webe. Také systémy komunikujú na webe, ale problémom je, že nepoužívajú "spoločný jazyk", pretože každý z nich je často vyvinutý pre rôzne špecifické účely definované požiadavkami ich používateľov. Na vyriešenie tohto problému s interoperabilitou sa posledných približne 20 rokov vyvíjajú iniciatívy, ktoré sa orientujú na vývoj technologických a organizačných štruktúr, politík a štandardov, ktoré by spolu mali umožniť efektívne vyhľadávanie, výmenu a používanie geopriestorových dát na webe. Tieto iniciatívy boli pomenované ako infraštruktúry pre priestorové informácie (angl. Spatial Data Infrastructure – SDI, Nebert, 2009).

Napriek vyššie popísaným skutočnostiam, keď bežný používateľ prehliada web a hľadá informácie, ktoré majú mať priestorovú lokalizáciu, ako napr. určitá adresa v Bratislave, alebo obrázky o hrade Krásna hôrka a jeho okolí, tak nepoužije ani online GIS aplikáciu, ani aplikáciu v rámci IPI. Naopak použije napr. aplikáciu Mapy Google, alebo slobodnú wiki-mapu sveta OpenStreetMap, či webovú stránku, ktorá má vyššie aplikácie včlenené. Takéto aplikácie poskytujú práve funkcionálnu na priestorové prezeranie (angl. spatial browsing). Priestorové prezeranie kombinuje priestorové dopyty s interaktívnym zobrazovaním digitálnych máp takým spôsobom, ktorý poskytuje používateľovi preskúmať geopriestorovú dimenziu jedného, alebo niekoľkých informačných systémov (Larson, 1996). Takýto systém môže používať štandardné geopriestorové webové služby, ako napr. OGC webovú mapovú službu (WMS) a webovú objektovú službu (WFS) na zobrazovanie a prístup ku geopriestorovým dátam uloženým vo vzdialených informačných systémoch. Tu je dôležité zdôrazniť, že najviac používané systémy priestorového prezerania na webe (napr. Mapy Google) používajú práve proprietárne špecifikácie, čo pramení z ich pôvodnej koncepcie. Tieto systémy sú vyvíjané pre masívne používanie a berú do úvahy obmedzenia a dynamiku, ktoré sú špecifické iba v kontexte webu. Možnosť ich použitia v rámci externých aplikácií je umožnená s využitím špecifických aplikačných programových rozhraní,

ktoré poskytujú za určitých licenčných podmienok. Táto skutočnosť umožňuje vývoj nových aplikácií na priestorové prezeranie, ktoré sú postavené na týchto existujúcich systémoch. Takáto aplikácia sa nazýva geomashup.

2.1 Koncept vyhľadávania v registroch

Pojem register predstavuje katalóg alebo zoznam, pomocou ktorého poskytovateľa informačných zdrojov publikujú popisné informácie vo forme metadát. Používatelia potom môžu definovať dopyty na tieto informácie pre potreby vyhľadávania, hodnotenia a prístupu k informačným zdrojom. Register plní v tomto prípade úlohu zoznamu (Zhao et al., 2004), ktorý ukladá všetky informácie do úložiska, ktoré môže byť buď relačný systém riadenia bázy dát alebo XML databáza. V distribuovaných infraštruktúrach je register prezentovaný ako webová služba, s ktorou spolupracujú klientske aplikácie alebo ďalšie webové služby (Alameh, 2001). V takomto prípade sú používané štandardizované rozhrania pre register (UDDI), výmenu správ (SOAP) a popis jednotlivých webových služieb (WSDL), ktoré tvoria základné piliere servisne orientovanej architektúry (SOA). V kontexte IPI sa na vyhľadanie, hodnotenie a prístup ku geopriestorovým dátam používajú systémy registrov nazývané katalógy (Nebert, 2009), alebo aj geokatalógy (Shvaiko et al., 2010). Riešenie geokatalógu pozostáva z webového klienta na vyhľadanie a katalógovej služby. Niektoré komplexné riešenia geokatalógov poskytujú aj webové rozhranie na tvorbu metadát v preddefinovanej štruktúre. Všetky informácie sú ukladané v rámci databázy geokatalógu. Schematické znázornenie architektúry geokatalógu je na obrázku 2.1.



Obr. 2.1 Hlavné komponenty geokatalógu

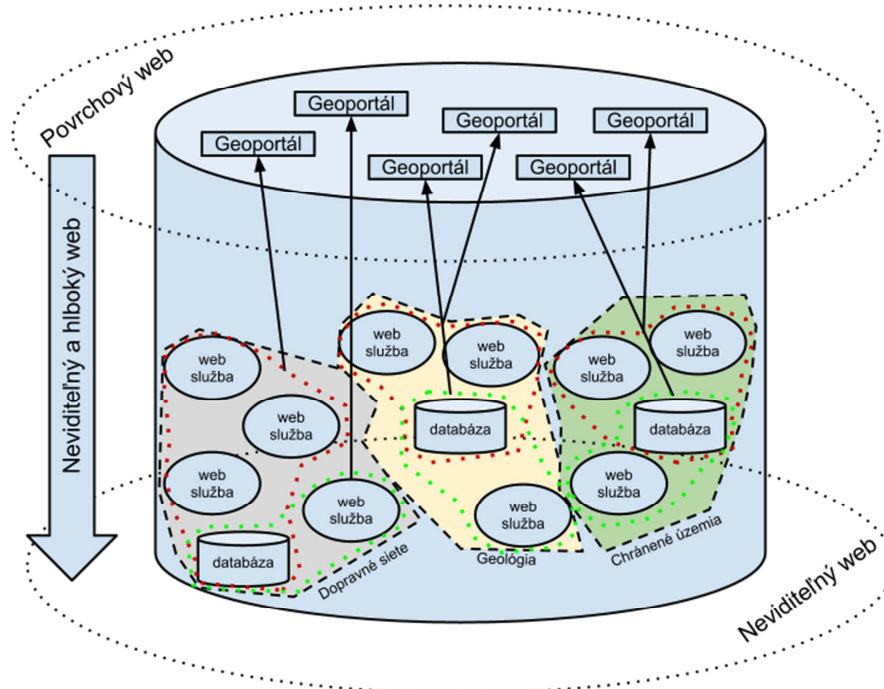
Geokatalógy poskytujú grafické a aplikačné programové rozhrania na definovanie požiadaviek na vyhľadanie, ktoré sa následne porovnávajú s informáciami uloženými v rámci databázy metadát (Nebert, 2009). Metadáta, ktoré sa v súčasnosti používajú v rámci implementácií IPI používajú rôzne metadátové modely definované v rámci existujúcich štandardov. Napríklad metadáta implementované v rámci Európskej IPI (INSPIRE, 2007) sú definované implementačnými pravidlami, ktoré vychádzajú z metadátových modelov definovaných v štandardoch EN ISO 19115 a EN ISO 19119 (INSPIRE, 2010). V rámci národnej IPI v USA sa zdefinoval metadátový model štandardu CSDGM (Content for Digital Geospatial Metadata) a profil NAP (North American Profile) vyvíjaný ako komunitný profil vychádzajúci z medzinárodného štandardu ISO 19115 (ISO, 2003). Iným príkladom metadátového štandardu je EML (Ecological Metadata Language), ktorý bol vyvinutý a používa sa na popis dát v rámci ekologickej domény naprieč celým svetom (Michener et al., 1997). Hoci sa metadátové štandardy líšia medzi jednotlivými IPI, je možné transformovať informácie medzi ich modelmi pomocou tzv. metadátových transformácií (angl. metadata crosswalks) (Nogueras-Iso et al., 2005; Batcheller, 2008; Kliment et al. 2012). Nougeras-Iso et al. (2005) popisujú transformáciu medzi metadátovými modelmi definovanými štandardmi ISO 19115 a Dublin Core. Transformácia má veľký význam, pretože spája geopriestorovú doménu s akoukoľvek inou doménou, keďže Dublin Core sa používa na všeobecný popis rôznych typov informačných zdrojov (Powell et al., 2007). Dublin Core je od roku 2009 medzinárodným štandardom ISO 15836:2009 Information and documentation - The Dublin Core metadata element set (ISO, 2009). Ďalším príkladom je transformácia, ktorú popisujú Kliment a Oggioni (2011) medzi štandardom EML a štandardom ISO 19115, ktorá spája ekologickú doménu s geoenvironmentálnou definovanou v rámci smernice INSPIRE, alebo aj s akoukoľvek inou geografickou doménou. Implementácia metadátových transformácií v rámci geokatalógu umožňuje používateľom realizovať rôzne stratégie vyhľadávania medzi rôznymi úrovňami IPI, a to napr. centralizované, alebo distribuované vyhľadanie.

2.2 Koncept vyhľadávania v indexoch

Druhým prístupom k vyhľadávaniu zdrojov GI na internete je s využitím populárnych internetových vyhľadávačov, akým je napr. Google (Li et al., 2010). Takýto prístup sa nazýva aj vyhľadávanie s využitím indexu (Zhao et al., 2007). Jedná sa o zhromažďovanie zdrojov informácií pomocou tzv. webového robota, ktorý prechádza cez všetky webové stránky a zhromažďuje metadáta týchto stránok, ktoré potom indexuje vo svojej databáze. Internetové vyhľadávače, ktoré poskytujú používateľské rozhranie na vyhľadávanie v rámci týchto databáz predstavujú v súčasnosti najviac používané webové stránky. Takýto vyhľadávač v odpovedi na požiadavku používateľa poskytne zoradený list informačných zdrojov pre potenciálne relevantné webové zdroje ako napr. webové stránky, videá, snímky, mapy a iné. Podľa (Inthiran et al., 2010) môžeme rozlíšiť tri typy internetových vyhľadávačov: založené na adresároch, na webových robotoch a hybridné vyhľadávače. Prvú skupinu tvoria vyhľadávače založené na vyhľadávaní vo webových adresároch (napr. Yahoo). Tieto vyhľadávače obyčajne zabezpečujú vyššiu presnosť vo vyhľadávaní, ale pokrytie webu je menšie a často doménovo orientované (vzdelanie, zdravie, veda, správy atď.). Geokatalógy popísané v predchádzajúcej kapitole môžeme klasifikovať ako internetové vyhľadávače založené na adresároch (Florczyk, 2012). Na rozdiel od tohto typu vyhľadávačov je databáza informácií vyhľadávačov založených na webových robotoch vytvorená a spravovaná automaticky. Automatické webové roboty majú za úlohu vyhľadávať, analyzovať a indexovať webové zdroje. Z pohľadu dostupnosti zdrojov pre webové roboty môžeme web rozdeliť do troch vrstiev:

- Povrchový web (Surface Web) – je to vrstva webu, v ktorej sú zdroje dosiahnuteľné webovým robotom pomocou hypertextových odkazov (Bergman, 2001).
- Hlboký web (Deep Web) – túto vrstvu tvoria databázy, ktorých dáta môžu byť prístupné cez online formuláre, aplikácie alebo webové služby (Florczyk, 2012).
- Neviditeľný web (Invisible Web) – ostatná časť webu, ktorá je veľmi ťažko dosiahnuteľná webovým robotom (Price a Sherman, 2001).

Doménové portály, nájdené v internetových vyhľadávačoch, vyžadujú zvyčajne ďalší prieskum používateľa, aby sa dostal priamo k zdrojom, ktoré vyhľadáva (Green, 2000). Príkladom takého portálu z domény GI je geoportál, ktorý si vyžaduje ďalšie používateľské interakcie, kým používateľ začne pristupovať priamo k zdrojom z hlbokého webu.



Obr. 2.2 Geoportál ako prístupový bod k multidisciplinárnym geopriestorovým zdrojom (podľa Florczyk, 2012)

Obrázok 2.2 zobrazuje umiestnenie zdrojov GI v rámci vyššie uvedených vrstiev webu. Databázy a webové služby, ktoré operujú nad nimi v rôznych oblastiach (dopravné siete, geológia, chránené územia a pod.) sa nachádzajú v tých vrstvách webu, ktoré webové roboty nevidia nájsť, indexovať ich a poskytovať koncovému

používateľovi internetového vyhľadávača. Na druhej strane geoportály, ako aplikácie, ktoré poskytujú rozhranie medzi používateľom a webovými službami, sa nachádzajú vo vrstve povrchového webu, čiže webové roboty ich nachádzajú.

Zdroje poskytované GIS patria v súčasnosti do skupiny, o ktorú majú všeobecné internetové vyhľadávače ako Google, Yahoo, Bing atď. nízky záujem. (Florczyk, 2012). Nevyhnutné úsilie potrebné pre vyhľadanie, indexovanie a podporu vyhľadávacej funkcionality pre tieto zdroje nie je kompenzované záujmom bežných používateľov webu. Znamená to, že hlavní hráči v doméne internetových vyhľadávačov nevyvíjajú vo veľkom rozsahu softvéry, ktoré by boli takto špecificky orientované. Na základe špecializácie internetového vyhľadávača ich môžeme rozdeliť na:

- Všeobecné (napr. vyhľadávač Google) - poskytujú vyhľadávanie v horizontálnej rovine (v rámci povrchového webu).
- Doménové vyhľadávače (napr. Google Books, Google Maps, alebo Google Scholar) - poskytujú vertikálne vyhľadávanie (Rajashree, 2008).

Doménové vyhľadávače sú optimalizované tak, aby podporovali špecifiká domény ako tematické vyhľadávanie, typ webového zdroja (napr. webová služba (Al-Masri et al., 2008), obrázok, video a pod.), pokrytie webu (riadený zoznam webových adries), alebo kombináciu týchto charakteristík.

3 Ciel' dizertačnej práce

Súčasný internet umožňuje pristupovať ku geopriestorovým zdrojom na rozličných úrovniach, ktoré medzi sebou nie sú vždy kompatibilné, pričom táto práca hľadá riešenia ako úrovne prepojiť na dosiahnutie lepšej viditeľnosti a tým aj použiteľnosti geopriestorových zdrojov. Webové geopriestorové zdroje sa nachádzajú v hlbokaj až neviditeľnej vrstve webu, ktorá nie je vôbec, alebo ak aj je, tak je ťažko prehliadateľná pomocou webových robotov internetových vyhľadávačov, akým je napr. Googlebot od spoločnosti Google. Väčšina laických používateľov internetu však vyhľadáva informácie práve pomocou týchto vyhľadávačov a geopriestorové zdroje sú pre nich často nedostupné.

Cieľom dizertačnej práce je analyzovať formy vyhľadávania publikovaných, on-line dostupných geopriestorových zdrojov, vrátane overenia možností ich publikovania spôsobom umožňujúcim ich vyhľadanie v heterogénnom prostredí internetu.

Na splnenie cieľa práce sú stanovené tieto čiastkové ciele:

- Analyzovať možnosti vyhľadávania geopriestorových zdrojov v rámci infraštruktúry pre priestorové informácie.
- Analyzovať možnosti vyhľadávania geopriestorových zdrojov v najpoužívanejšom internetovom vyhľadávači Google.
- Navrhnuť a implementovať komplexné riešenie na tvorbu, overenie, publikovanie a distribúciu metadát pre potreby vyhľadávania geopriestorových zdrojov z domény geodézie a kartografie v rámci infraštruktúry pre priestorové informácie s dôrazom na automatizáciu procesov a využitia existujúcich zdrojov geopriestorových informácií.
- Navrhnuť a implementovať model vyhľadávania OGC služieb v internetovom vyhľadávači Google a ich následného vyhodnotenia a aktualizácie.
- Navrhnuť a implementovať model na naplnenie databázy geokatalógu zdrojmi geopriestorových informácií nájdených v rámci internetového vyhľadávača Google.
- Definovať rozsah geopriestorových informácií poskytovaných internetovým vyhľadávačom Google pomocou analýz realizovaných na zozbieraných metadátach.

4 Metodika práce a metódy skúmania

4.1 Materiál použitý pre praktické experimenty

Cieľom tejto kapitoly je popísať materiál, ktorý bol použitý na realizáciu praktických experimentov podľa navrhnutých metodík, ktoré sú popísané v ďalšej kapitole 4.2. Pod termínom materiál rozumieme štandardy, softvéry a technológie použité na realizáciu experimentov.

Realizovali sme dva nezávislé experimenty. V experimente č.1 sme navrhli a realizovali proces automatickej extrakcie metadát z existujúcich rozhraní webových služieb OGC a ich následnej úpravy podľa požiadaviek vyplývajúcich z INSPIRE vykonávacích pravidiel pre metadáta a rezortných metadátových profilov. V rámci experimentu č.2 sme vyhľadávali geopriestorové webové služby definované špecifikáciami OGC v internetovom vyhľadávači Google. Po spracovaní výsledkov vyhľadávania bola databáza geokatalógu naplnená metadátami, ktoré popisujú nájdené OGC služby ako aj ich zdroje GI, a to extrakciou metadát z týchto zdrojov spôsobom použitým aj v experimente č.1.

4.1.1 Geopriestorové dáta a metadáta z domény geodézie a kartografie v SR

Geodetický a kartografický ústav (GKÚ) poskytol pre potreby praktického experimentu č.1 nasledovné zdroje GI:

- výrez z personálnej geodatabázy ZB GIS vo formáte *.mdb* (*ZBGIS_vyber_NR_JTSK.mdb*), ktorý obsahoval 126 tried objektov definovaných v KO (ÚGKK, 2008),
- metadátové profily navrhnuté pre popis dát katastra nehnuteľností (*profil_CSKN_v10.xml*), ZBGIS (*profil_ZBGIS_v21.xml*) a pre popis webových služieb (*profil_UGKK_rev6.xml*) vo formáte XML,
- vzorové metadátové záznamy pre INSPIRE témy priestorových dát z príloh I a II smernice: I.3 Zemepisné názvy (*nazvoslovie_111327100150_1913836747.xml*), I.4 Správne jednotky (*uzemnospravne_clenenie_112027100145_1247679266.xml*), I.6 Katastrálne parcely (*VKM_114127100158_1562284236.xml*), I.7 Dopravné siete (*doprava_111827100122_1198169862.xml*), I.8 Hydrografia (*hydrografia_112427100140_1050797094.xml*) a II.1 Výška (*dmr_113827100106_2004896331.xml*).

4.1.2 Existujúce programy a technológie použité pri praktických experimentoch

Na realizáciu experimentu č.1 sme použili nasledovné programy a súvisiace technológie:

- Geoserver – open source program, ktorý umožňuje distribuovať geopriestorové dáta z mnohých súčasných priestorových dátových typov. Zároveň je referenčnou implementáciou pre najviac používané OGC webové služby ako WMS, WFS a WCS (Web Coverage Service),
- GeoNetwork opensource – open source program, ktorý poskytuje komplexné riešenie pre správu metadát k zdrojom GI. Okrem základných komponentov ako metadátový editor, vyhľadávací klient, či katalógová služba, poskytuje množstvo ďalších užitočných prvkov, potrebných pre administráciu a konfiguráciu systému, ako aj rozhranie pre programovanie aplikácií (API),
- Altova MapForce – proprietárny program typu ETL (Extract, Transform and Load), ktorý poskytuje grafické rozhranie na mapovanie vstupných dát (databázy, XML/TXT/CSV, EXCEL súbory a webové služby) na výstupné modely (napr. XSD aplikačné schémy) s využitím mnohých typov knižníc pre transformácie. Po vytvorení a otestovaní mapovania je možné ho exportovať do skriptov ako napr. XSLT 1.0/2.0, alebo kódov programov v jazykoch Java/C++/C#,
- Altova XMLSpy – proprietárny XML editor, ktorý patrí do toho istého balíka programov ako MapForce. Poskytuje programové prostredie pre modelovanie, editáciu, transformovanie a overovanie dát založených na XML technológiách (XSL/XSLT, DTD/XSD a JSON). Umožňuje aplikovať operácie hromadne na dáta uložené v rámci projektu a generovať kódy programovacích jazykov Java/C++/C#. Pre potreby vývoja webových stránok poskytuje aj prostredie na editovanie HTML a CSS.

Na realizáciu experimentu č.2 sme použili nasledovné programy, ich rozhrania pre programovanie aplikácií a súvisiace technológie:

- Internetový vyhľadávač Google - najviac využívaný vyhľadávač v rámci celosvetovej internetovej siete, ktorý denne spracuje niekoľko stoviek miliónov požiadaviek. Vyhľadávač bežne pracuje s požiadavkami vo formáte voľného textu, ktorý používatelia definujú v rámci jeho grafického

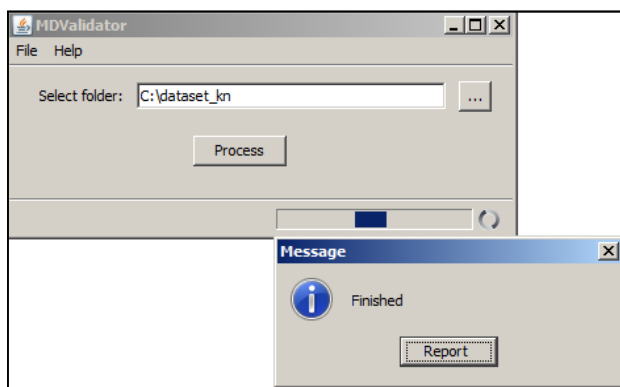
rozhrania a ktoré následne rozdeľuje na sekvencie termínov, ktoré by sa mali vyskytnúť vo výsledkoch. Okrem vyhľadávania voľného textu umožňuje definovať požiadavky aj s použitím tzv. pokročilých operátorov na obmedzenie rozsahu vyhľadávania a tým zvyšuje relevantnosť výsledkov. Viac ako 15 operátorov je možné použiť pri definícii požiadavky na vyhľadávanie, napr. operátor *site* (obmedzuje vyhľadávanie na doménu, ktorá je zadaná v požiadavke, napr. *site:sk* bude vyhľadávať v rámci stránok registrovaných pod doménou sk), *intitle* (obmedzuje vyhľadávanie na názvy stránok), *inurl* (obmedzuje vyhľadávanie v URL adresách stránok, napr. *inurl:geoportal* bude vyhľadávať webové stránky, ktoré majú text geoportál v rámci URL adresy), *filetype* (vyhľadáva len dáta zadaného typu ako napr. *filetype:pdf* vráti výsledky typu PDF súborov) a ďalšie. Používateľ môže na vyhľadávanie v rámci Google použiť buď priamo webového klienta, alebo externú aplikáciu s využitím Google XML API,

- OutWit Hub – voľná verzia tohto programu umožňuje vyextrahovať dáta z webových stránok podľa preddefinovaných pravidiel. Dáta z hypertextových odkazov, obrázky, emailové adresy, správy vo formáte RSS. Dátové tabuľky môžu byť extrahované a následne exportované do textových súborov typu CSV/HTML, alebo do súborov Excel či SQL skriptov pre potreby ich ďalšieho spracovania,
- XML služby v rámci GeoNetwork API – program GeoNetwork poskytuje v rámci svojho API tzv. XML služby pre ovládanie logickej vrstvy pomocou vzdialených aplikácií s použitím komunikácie založenej na výmene POST XML správ pomocou HTTP protokolu. Pomocou XML služieb je možné prihlasovať/odhlasovať sa do/zo systému, spravovať skupiny používateľov, manažovať metadáta v katalógu, konfigurovať samotný systém, ukladať súbory z databázy programu na lokálny disk, konfigurovať harvestovacie úlohy (získať/aktualizovať informácie o existujúcich úlohách, pridať nové úlohy, vymazať/aktivovať/spúšťať/zastavovať úlohy) a ďalšie. Uvedené umožňuje využiť logiku naprogramovanú v programe GeoNetwork ľubovoľnými externými klientskymi aplikáciami.
- Kategorizácia informácií v geokatalógu GeoNetwork – program umožňuje v rámci svojej administrátorskej časti vytvárať kategórie zdrojov a tie následne priradovať k novo vytvoreným, importovaným, alebo prebratým (harvestovaným) metadátam. Význam kategorizácie spočíva v tom, že používateľ systému následne pri vyhľadávaní môže prehliadať výsledky na základe kategórií a tak sa dostať bližšie k tomu, čo hľadá.

4.1.3 Programy vyvinuté pre realizáciu praktických experimentov

MDValidator - program pre hromadné overovanie metadát voči INSPIRE požiadavkám

Jeden z krokov v navrhutej metodike je overenie vytvorených metadát voči štandardu ISO 19139 a INSPIRE špecifickým požiadavkám, ktoré sú zadané v implementačných pravidlách pre metadáta (INSPIRE, 2008). Dostupné nástroje na overovanie metadát, napr. vstavaný validátor v geokatalógu GeoNetwork, alebo on-line dostupný INSPIRE validátor neumožňujú realizovať hromadné overenie viacerých metadátových záznamov. Preto sme takýto program vytvorili (Kliment et al., 2012). Využili sme zdrojový Java kód dostupný na stránke INSPIRE Geoportálu, pomocou ktorého je možné spúšťať INSPIRE RESTful webovú službu na overenie metadát pomocou HTTP POST požiadavky. Služba následne vracia späť odpoveď (validačný report) vo formáte XML, alebo HTML. Naprogramovaná aplikácia MDValidator (obr. 4.1) je jednoduchý desktopový program, ktorý inkrementálne načítava metadátové záznamy vo formáte XML (ISO, 2007a) z adresára na lokálnom počítači a následne ich zasiela na INSPIRE validačnú službu, ktorá ich overuje a posiela validačný report.



Obr. 4.1 Grafické rozhranie aplikácie MDValidator s možnosťou otvorenia validačného reportu

Po ukončení procesu overenia je možné prezerat' validačný report vo formáte HTML s využitím internetového prehliadača. V reporte sú uvedené informácie o správnych a nesprávnych, alebo chýbajúcich metadátach podľa INSPIRE implementačných pravidiel pre metadáta (INSPIRE, 2008).

Je dôležité uviesť, že MDValidator poskytuje v súčasnosti overenie metadátových záznamov len voči INSPIRE špecifickým požiadavkám, pretože používa INSPIRE validačnú službu. V praxi to znamená, že ak metadátový záznam obsahuje všetky povinné elementy podľa INSPIRE, výsledok validácie v aplikácii MDValidator bude bez chýb aj vtedy, ak tento záznam nebude vytvorený celkom presne podľa pravidiel ISO schémy. Preto je potrebné pre zabezpečenie interoperability metadátových záznamov aj s inými aplikáciami realizovať overenie voči tejto schéme. Realizácia je možná pomocou akéhokoľvek XML editora s využitím XSD schémy, ktorá je on-line dostupná.

Naprogramované PHP Skripty pre realizáciu experimentu č.2

Pre potreby praktickej realizácie experimentu č.2, ktorého metodika je popísaná v kapitole 4.2.3 sme naprogramovali tri skripty pre aplikáciu navrhnutých krokov. Použili sme skriptovací programovací jazyk PHP na naprogramovanie jednotlivých krokov, databázový systém MySQL na ukladanie nájdených a spracovaných dát a webový server Apache HTTP na spúšťanie skriptov.

Prvé dva sú skripty *import.php* a *crawl.php*, ktoré spracovávajú informácie, ktoré obsahujú vstupné súbory typu CSV a importujú ich obsah a výsledky ďalšieho spracovania do databázy MySQL. Detailný popis jednotlivých krokov je nasledovný:

- načítať informácie zo vstupných súborov typu CSV, prezrieť správnosť zápisu jednotlivých URL adries a import/aktualizácia dát do/v databáze (*import.php*),
- zistiť verziu OGC služby – tento krok posieľa požiadavky na jednotlivé služby, ktoré boli importované do databázy, zisťuje kód odpovede servera (či server beží, či použitá URL adresa je presmerovaná, atď.) a z odpovede extrahuje verziu služby, ktorej hodnotu ukladá do databázy. (*crawl.php*),
- zdefinovať zistené hodnoty v databáze do stĺpcov: *type* (typ OGC služby), *version* (verzia daného typu OGC služby) a *status* (stav po poslednom zisťovaní). Hodnoty v stĺpci *status* sú dôležité pre ďalšie spracovanie a môžu obsahovať nasledovné hodnoty:
 - status=0 definuje záznamy, ktoré boli importované do databázy (*import.php*) a ktorých verzia sa bude zisťovať (*crawl.php*),
 - status=1 definuje záznamy, ktorých verzia sa úspešne zistila a uložila do databázy (*crawl.php*),
 - status=2 definuje záznamy, ktorých zistenie verzie zlyhalo (odpoveď servera neposkytla informácie o verzii v preddefinovanej adrese v rámci hierarchickej XML štruktúry) a po aplikácii skriptu *harvesting.php* sa ich status prepíše na kód 0, aby bolo možné pri ďalšom opakovaní zisťovať ich verziu (server mohol byť dočasne nedostupný v čase spustenia skriptu *crawl.php*),
 - status=3 definuje vždy jeden záznam, pre ktorý sa úspešne vytvorila harvestovacia úloha (zapísala sa hodnota do stĺpca *id_gn* vyextrahovaná z odpovede GeoNetwork), úloha je spustená, ale ešte neskončila,
 - status=4 definuje záznamy, ktorých zlyhalo buď vytvorenie úlohy, alebo proces harvestovania metadát v prostredí GeoNetwork. Zlyhanie vytvorenia úlohy môže byť zapríčinené tým, že určitá verzia, alebo typ OGC služby nie sú podporované v súčasnej verzii GeoNetwork (napr. WFS 2.0.0, alebo WMTS). Samotné zlyhanie procesu harvestovania programom GeoNetwork, môže mať viacej príčin, ktoré neboli detailne analyzované v rámci práce,
 - status=5 definuje záznamy, pre ktoré bola úspešne vytvorená úloha, čiže GeoNetwork odpovedal na požiadavku s uvedením hodnoty ID úlohy vytvorenej v rámci databázy *metadata*, ktorá bola vytvorená ako úložisko informácií pre GeoNetwork. Hodnota ID bola uložená v tabuľke *services* v stĺpci *id_gn*. Po spustení jednotlivých úloh sa predpokladalo, že úspešne všetky prebehnú. V prípade akýchkoľvek problémov v procese harvestovania programom GeoNetwork neboli ich detaily dokumentované v databáze *geoserver*, ale priamo v databáze *metadata*,
 - status=6 definuje záznamy, ktorých priebeh harvestovacej úlohy bol prerušený manuálne.

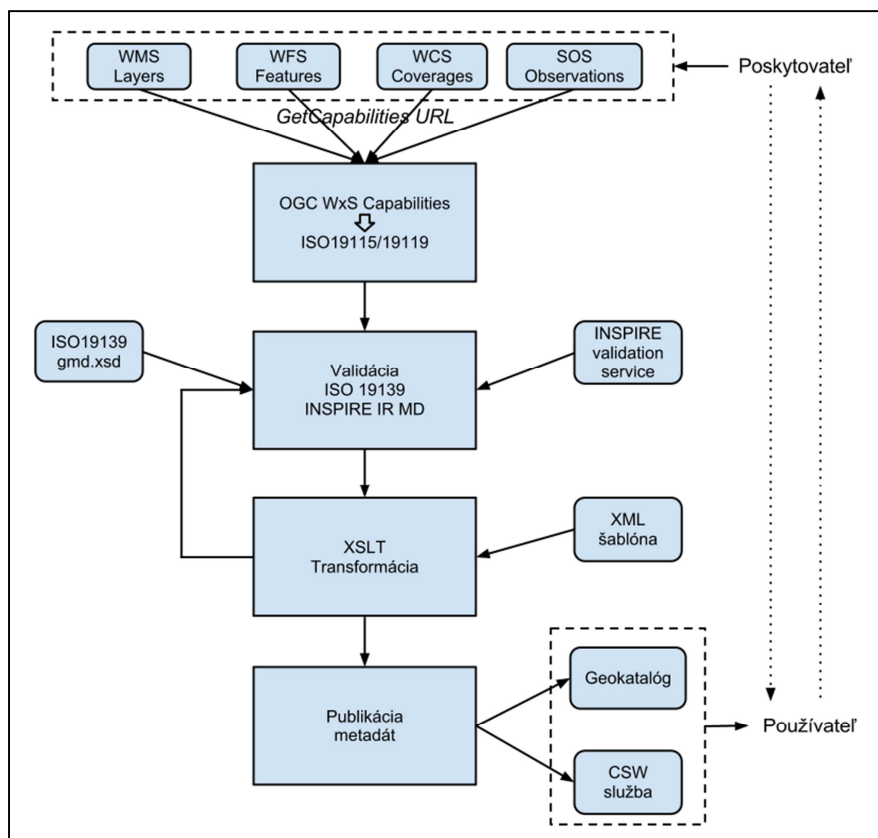
Tretí skript s názvom *harvesting.php* realizuje nasledovné kroky:

- vytvoriť pre záznamy v databáze s hodnotou 1 v stĺpci *status* harvestovacie úlohy pomocou XML služieb v rámci GeoNetwork API. Definovať obsah požiadavky a poslať na preddefinovanú URL adresu pomocou XML POST, prijať odpoveď, z ktorej vyselektovať identifikátor (ID) harvestovacej úlohy uložený v rámci databázy *metadata* v GeoNetwork a ten následne uložiť do databázy *ogcservices* v MySQL do stĺpca *gn_id*,
- spúšťať harvestovacie úlohy pre jednotlivé záznamy v tabuľke *services*, ktoré obsahujú hodnoty ID v stĺpci *gn_id* získané v predchádzajúcom kroku a hodnotu 3 v stĺpci *status*,
- ukladať výsledky procesu do tabuľky *services* podľa toho, ako prebehol. Ak prebehol v poriadku, zapísať hodnotu 5 do stĺpca *status*, ak zlyhal hodnotu 4.

4.2 Metodické postupy a ich aplikácia v praktických experimentoch

4.2.1 Metodika tvorby a overenia geopriestorových metadát

Navrhovaná metodika vychádza zo scenára, kde poskytovateľ informácií z určitej domény (napr. životné prostredie, geodézia a kartografia, doprava a iné) už má implementované OGC webové služby (napr. WMS, WFS, WCS, SOS (Sensor Observations Service)), ktoré pracujú nad jeho geopriestorovými zdrojmi a potrebuje ich zaregistrovať a následne poskytovať v rámci INSPIRE, resp. národnej IPI. Inak povedané, potrebuje vytvoriť metadátové záznamy pre jednotlivé dátové súbory, série dátových súborov a služby (typy zdrojov definované v INSPIRE (INSPIRE, 2008)) pre potreby ich vyhľadávania koncovým používateľom.



Obr. 4.2 Schematické znázornenie navrhutej metodiky

V rámci metodiky, ktorá je schematicky zobrazená na obrázku 4.2 boli navrhnuté nasledovné kroky pre naplnenie používateľského scenára:

1. Vytvorenie metadát pre služby, ktoré sú definované štandardom ISO 19119/19139 transformáciou odpovede na požiadavku *GetCapabilities* OGC webovej služby (WMS, WFS, WCS a SOS),
2. Vytvorenie metadát pre dátové súbory, ktoré sú definované štandardmi ISO 19115/19139 transformáciou odpovede na požiadavku *GetCapabilities* OGC webovej služby nasledovne:

- pre služby typu WMS extrahovaním informácií uvedených pre vrstvy poskytované službou v elemente <Layer>,
 - pre služby typu WFS extrahovaním informácií uvedených pre objekty poskytované službou v elemente <FeatureType>,
 - pre služby typu WCS extrahovaním informácií uvedených pre pokrytia poskytované službou v elemente <CoverageSummary>,
 - pre služby typu SOS extrahovaním informácií uvedených pre pozorovania poskytované službou v elemente <ObservationOffering>.
3. Overenie vytvorených metadátových záznamov voči ISO schéme (ISO, 2007a) a INSPIRE špecifickým požiadavkám pre metadáta (INSPIRE, 2008).
 4. Doplnenie chýbajúcich informácií transformáciou na preddefinované šablóny a opätovné overenie.
 5. Publikovanie metadátových záznamov v geokatalógu.

4.2.2 Aplikácia metodiky na zdrojoch domény geodézie a kartografie – experiment č.1

Popísanú metodiku sme aplikovali na zdroje domény geodézie a kartografie popísané v kapitole 4.1.1. GKÚ v súčasnej dobe neposkytuje verejne dostupné OGC webové služby, bolo ich potrebné na začiatku experimentu vytvoriť. Na vytvorenie OGC webových služieb sme použili aplikáciu Geoserver, ktorá umožňuje dáta vo formáte ESRI shapefile importovať a publikovať ako WMS a WFS služby. Nie všetkých 126 tried objektov zo vstupnej geodatabázy obsahovalo reálne objekty pre záujmové územie, preto sme pred samotným importovaním do Geoservera vykonali redukovanie jej obsahu o prázdne triedy objektov. Po zredukovaní obsahovala databáza 46 tried objektov, ktoré boli importované do Geoservera. Pre každú importovanú triedu objektov sme zadefinovali názov, abstrakt, kľúčové slová a súradnicový systém. Po dokončení bolo všetkých 46 tried objektov dostupných ako vrstvy služby WMS, alebo aj ako objekty služby WFS.

Implementovanú WMS službu sme ďalej použili v rámci realizácie prvého bodu v navrhutej metodike.

Vytvorenie ISO metadát pre WMS službu a jej vrstvy

Na vytvorenie ISO metadát pre WMS službu ako aj vrstvy (dátové súbory), ktoré táto služba zobrazuje, sme použili funkcionality geokatalógu GeoNetwork, ktorá sa nazýva harvesting (preberanie metadát). V princípe ide o transformáciu informácií kódovaných v OGC XSD aplikačnej schéme na ISO XSD aplikačné schémy pre služby a dátové súbory, ktorá je definovaná štandardom ISO 19139. Ide o transformáciu medzi súbormi typu XML, preto použitou technológiou je XSLT (Extensible Stylesheet Language Transformations). Geokatalóg GeoNetwork používa túto technológiu s XSLT procesorom Saxon implementovaným ako Java knižnica. Pre samotnú transformáciu sa používajú pre vytvorenie ISO metadát OGC webových služieb a dátových súborov z vrstiev transformačné súbory *OGCWxSGetCapabilities-to-19119.xsl* a *OGCWMSGetCapabilitiesLayer-to-19139.xsl*. Je možné vytvoriť ISO metadáta aj pre objekty WFS, pokrytia WCS a pozorovania zo služby SOS.

Pre vytvorenú WMS službu sme v prostredí geokatalógu GeoNetwork vytvorili tzv. harvestovaciu úlohu so zadefinovaním potrebných parametrov ako názov, typ a verzia OGC služby, koncový bod služby, spôsob extrahovania (vytvoriť metadátový záznam iba pre službu alebo aj pre dáta), kategóriu pre službu a pre dáta, privilégia a plánovač, v ktorom sme nastavili spustenie úlohy iba jedenkrát. Takto nakonfigurovanú úlohu sme spustili a následne získali metadáta pre WMS službu (1 záznam) a dátové súbory (46 záznamov) už v štruktúre ISO. Transformácia mohla byť použitá iba informácie obsiahnuté v odpovedi WMS služby na požiadavku *GetCapabilities*, čiže informácie, ktoré sme zadali v prostredí Geoserver pre službu a jednotlivé vrstvy. Na zistenie chýbajúcich informácií podľa INSPIRE požiadaviek sme realizovali overenie vytvorených metadátových záznamov.

Overenie vytvorených ISO metadátových záznamov

Na overenie vytvorených metadátových záznamov sme použili nástroj MDValidator. Tak, ako sme predpokladali, podstatná časť informácií chýbala v rovnakom rozsahu pre všetky záznamy dátových súborov (17 správnych a 8 chýbajúcich elementov) a pre službu (15 správnych a 8 chýbajúcich). Zoznam validačných chýb je uvedený v tabuľke 4.1.

Tabuľka 4.1 Zoznam validačných chýb pre metadáta služby a vrstiev (dátových súborov)

Metadátový záznam služby	Metadátové záznamy vrstiev
(2.3.2) Spatial data service type is missing or has a wrong value	(2.2.5) Unique resource identifier is missing.
(2.4) For spatial data services, at least the category or subcategory of the service as defined in Part D 4 shall be documented	(2.3.1) Topic category is missing.
SC12. There shall be at least one instance of MD_Metadatas.identificationInfo [1].MD_Identification.resourceConstraints.	(2.4) For datasets and series at least one keyword of GEMET thesaurus shall be documented
(2.10.1) Organisation name (Resource) is missing.	(2.7.1) Lineage is missing.
(2.10.1) Electronic mail address(resource) is missing	SC12. There shall be at least one instance of MD_Metadatas.identificationInfo[1].MD_Identification.resourceConstraints.
(2.10.1) Role (resource) is missing or has a wrong value	(2.10.1) Electronic mail address(resource) is missing .
(2.11.3) Metadata language is missing or has a wrong value	(2.11.3) Metadata language is missing or has a wrong value
(2.11.1) Metadata responsible party (electronic mail address) is missing.	(2.11.1) Metadata responsible party (electronic mail address) is missing.

Vytvorenie transformačného súboru podľa šablóny ZB GIS

Na základe výsledkov overenia sme upravili metadátové záznamy tak, aby spĺňali požiadavky ISO/INSPIRE, ale aj požiadavky zadefinované v rezortných MD profiloch pre dáta ZB GIS a pre sieťové služby. Metadátový záznam pre WMS službu sme upravili manuálne s využitím metadátového editora v geokatalógu GeoNetwork. Na definovanie správnych informácií sme použili metadátový profil pre služby (UGKK_rev6.xml). Pre metadáta dátových súborov nebolo vhodné aplikovať manuálnu opravu, pretože pri väčšom množstve je to neefektívne. Vytvorili sme mapovanie jedného metadátového záznamu pre dátový súbor na metadátový profil ZB GIS (ZBGIS_v21.xml). Použili sme program Altova MapForce a vygenerovaný transformačný súbor *WMSLayer2ZBGIS.xslt* sme použili na hromadnú transformáciu metadátových záznamov v prostredí XML editora Altova XML Spy.

Mapovanie pozostávalo z nasledovných blokov:

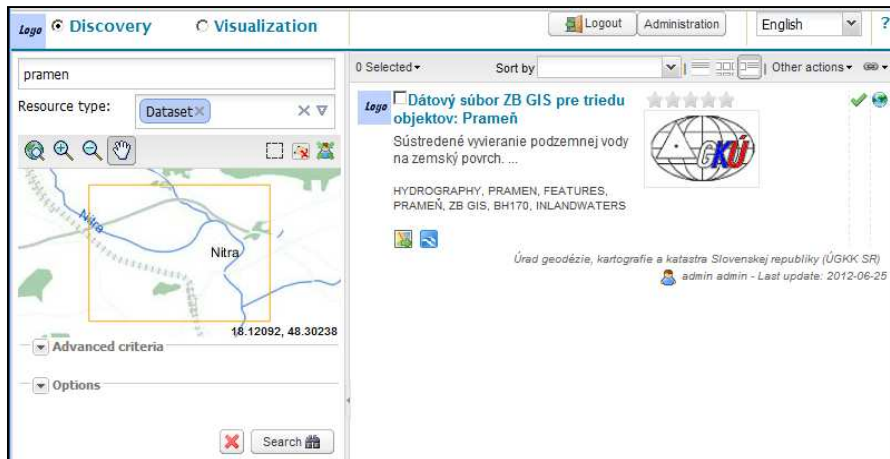
- priame prebratie informácií obsiahnutých vo vstupnom metadátovom zázname (vytvorenom transformáciou OGC na ISO) do XML šablóny pre dátové súbory ZB GIS,
- zadefinovanie konštánt pre výstupný metadátový záznam podľa ZB GIS šablóny. Išlo o informácie, ktoré sa opakujú v rámci každého záznamu ako napr. informácie o zodpovednej organizácii, kontaktnom mieste, identifikácii katalógu objektov, distribúcii, údržbe atď.,
- zadefinovanie podmienok transformácie pre niektoré triedy objektov ZB GIS, ktoré boli relevantné pre dátové témy INSPIRE, a preto musia poskytovať informácie navyše, ako napr. súlad so špecifikáciami INSPIRE, kľúčové slová (názov INSPIRE témy priestorových dát z európskeho environmentálneho tezauru GEMET) a ďalšie.

Podľa informácií uvedených vo vzorových metadátových záznamoch, ktoré poskytol GKÚ, sme pre potreby praktickej ukážky zadefinovali podmienky pre nasledovné triedy objektov (ÚGKK, 2008), ktoré spadajú do INSPIRE témy Hydrografia: Vodný tok; Most; Prieput; Hrádza; Hať, stavidlo; Násyp, ochranná hrádza; Koruna hrádze; Zárez ;Vodopád; Brod; Prameň; Nádrž plavebnej komory; Plavebná brána; Plocha vodného toku; Močiar, slanisko.

Po aplikácii hromadnej transformácie sme pre overenie jej správnosti vykonali validáciu metadátových záznamov voči ISO schéme priamo v XML editore Altova XML Spy a voči INSPIRE požiadavkám pomocou aplikácie MDValidator. Všetkých 46 metadátových záznamov bolo validných voči ISO schéme a ako sme predpokladali, nie všetky boli bez chýb voči INSPIRE požiadavkám (bez chýb boli 4 záznamy , ktoré popisovali dátové súbory vytvorené zo skupiny tried objektov definovaných pre INSPIRE tému Hydrografia a ostatných 42 malo dve chyby, ktoré súviseli s tým, že neboli relevantné pre túto tému).

Publikácia metadátových záznamov v geokatalógu GeoNetwork

Otestované metadátové záznamy sme importovali do geokatalógu GeoNetwork. Tieto záznamy boli následne prístupné pre vyhľadávanie pomocou vyhľadávacieho klienta vstavaného v geokatalógu (obr. 4.3), ako aj prostredníctvom katalógovej služby pre potreby vyhľadávania vzdialenými klientmi (napr. Národný geoportál SR, alebo INSPIRE Geoportál).



Obr. 4.3 Vyhľadanie metadátového záznamu pre triedu objektov Prameň v programe GeoNetwork

4.2.3 Metodika na vyhľadanie OGC služieb v internetovom vyhľadávači Google a extrahovanie metadát k zdrojom geopriestorových informácií

V rámci experimentu č.2 sme navrhli metodiku, pomocou ktorej je možné vyhľadať OGC služby v internetovom vyhľadávači Google a následne vyextrahovať GI z metadát popisujúcich tieto služby, ale aj súvisiace geopriestorové zdroje. Boli navrhnuté nasledovné kroky pre aplikáciu v praktickom experimente č.2:

- vyhľadať URL adresy na požiadavky *GetCapabilities* jednotlivých typov OGC webových služieb (WMS, WMTS (Web Map Tile Service), WFS, WCS, WPS (Web Processing Service)), SOS a CSW (Catalogue Service for the Web)) v rámci webového vyhľadávacieho klienta Google s použitím pokročilého operátora *inurl* na filtrovanie výsledkov,
- exportovať výsledky vyhľadávania v Google do dátových súborov CSV v preddefinovanej štruktúre pre všetky typy OGC služieb,
- prezerat' nájdené URL adresy na zistenie stavu serverov (kód HTTP odpovede) a verzie OGC služby z obsahu odpovede (odpoveď na *GetCapabilities* by mala byť v XML štruktúre definovanej XSD schémami podľa typu služby),
- vytvoriť harvestovacie úlohy v geokatalógu GeoNetwork s využitím XML služieb v rámci jeho API,
- načasovať a spustiť harvestovacie úlohy a priebežne ich aktualizovať,
- spracovať informácie poskytované v rámci metadát, ktoré vznikli ako výsledok harvestovacích úloh.

4.2.4 Aplikácia metodiky pre súčasné OGC služby – experiment č.2

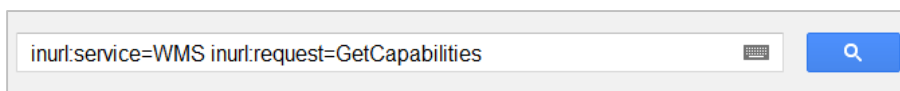
Metodiku sme aplikovali na vyhľadanie súčasných typov OGC služieb v rámci webového rozhrania internetového vyhľadávača Google. Jednotlivé kroky sú popísané detailne v nasledujúcej časti.

Vyhľadanie URL adries na OGC WxS *GetCapabilities* vo vyhľadávači Google

Pre vyhľadanie URL adries, ktoré obsahujú definíciu požiadavky *GetCapabilities* na OGC služby sme použili operátor *inurl* nasledovným spôsobom:

- filter *inurl:service=WxS* sme použili na definíciu typu vyhľadávanej služby, kde boli pre hodnotu WxS použité štandardizované skratky: WMS, WFS, WCS, WPS, SOS, CSW a WMTS,
- filter *inurl:request=GetCapabilities* sme použili na definíciu typu požiadavky, pre ktorú sa mali vyhľadať URL adresy.

Na obr. 4.4 je zobrazená výsledná štruktúra pre definovanie filtra na vyhľadanie URL adries na požiadavku *GetCapabilities* pre OGC WMS služby.



Obr. 4.4 Definícia filtra vyhľadávania v Google pre OGC WMS služby

Takým spôsobom sme realizovali vyhľadávanie pre všetky súčasné typy OGC služieb v troch etapách, medzi ktorými bol časový posun.

Extrahovanie informácií z výsledkov vyhľadávania a ich export do CSV

Po zadefinovaní uvedených požiadaviek na vyhľadávanie sme získali zoznamy výsledkov pre jednotlivé typy OGC služieb v rámci internetového vyhľadávača Google. Pre potreby ďalšieho spracovania sme ich extrahovali do preddefinovanej štruktúry a exportovali do súborov CSV. Na extrahovanie a exportovanie výsledkov sme použili program OutWit Hub, v ktorom sme najprv zadefinovali štruktúru informácií, ktoré chceme exportovať z výsledkov a ich lokalizáciu v rámci HTML kódu odpovede vyhľadávania v Google. V tabuľke 4.2 sú uvedené informácie, ktoré sme extrahovali z výsledkov a ich vymedzenie HTML značkami v rámci kódu odpovede.

Tabuľka 4.2 Definícia dát na exportovania z výsledkov vyhľadávania

Exportované dáta	Začiatočná značka	Koncová značka
Title	<h3 class="r">	</h3>
Description		
URL	<h3 class="r"><a href="	"

V rámci extrahovaných dát sme zadefinovali položku *Title*, do ktorej sa exportovali informácie, ktoré Google poskytuje ako názov v rámci výsledkov vyhľadávania, *Description* pre informácie, ktoré popisujú zdroj a *URL* adresa, ktorá mala byť HTTP GET požiadavkou na metadáta služby (*GetCapabilities*).

Takto formátované výsledky sme potom exportovali jednotlivo pre všetky typy OGC služieb do CSV súborov z názvami: *wms.csv*, *wfs.csv*, *wcs.csv*, *wps.csv*, *sos.csv*, *csw.csv* a *wmts.csv*. Tieto súbory boli použité ako vstupné dáta pre ďalšie kroky v rámci experimentu.

Testovanie nájdených URL adries

Na testovanie URL adries nájdených v rámci Google indexu sme použili skripty *import.php* a *crawl.php* popísané detailne v kapitole 4.1.3. Skripty boli aplikované na vstupné dáta typu CSV, ktoré importovali výsledky do vytvorenej databázy *geoserver* a zisťovali, či je uvedená URL adresa správna a aké verzie OGC služieb sa nachádzajú pod nájdenými *GetCapabilities* URL adresami.

Vytvorenie a spustenie harvestovacích úloh v GeoNetwork

Na aplikáciu tohto kroku experimentu sme použili PHP skript *harvesting.php*. Tento skript načítaval postupne informácie z databázy *geoserver*, ktorá bola naplnená dátami v predchádzajúcom kroku a používal ich ako premenné v XML POST správach typu *harvesting.add*, ktoré potom zasielal programu GeoNetwork na preddefinovanú URL adresu v štruktúre definovanej nasledovne:

```
<node type="ogcwxss">
  <site>
    <name>IGME SGN EN GEOLOGY</name>
    <url>http://mapas.igme.es/gis/services/PSysmin/IGME_SGN_EN_Geology/MapServer/WMServer</url>
    <ogctype>WMS1.1.1</ogctype>
    <icon>wms.gif</icon>
  </site>
  <options>
    <every>10080</every>
    <oneRunOnly>>true</oneRunOnly>
    <useLayer>>true</useLayer>
    <useLayerMd>>false</useLayerMd>
    <datasetCategory>2</datasetCategory>
  </options>
  <privileges>
    <group id="1">
      <operation name="view"></operation>
      <operation name="dynamic"></operation>
      <operation name="featured"></operation>
    </group>
  </privileges>
</node>
```

```

</privileges>
<categories>
  <category id="11"></category>
</categories>
</node>

```

Štruktúra požiadavky typu *harvesting.add* pozostávala z týchto častí:

- Koreňový element <node> - definoval v atribúte *type* typ úlohy, ktorá sa má uložiť v databáze GeoNetwork ako pripravená na prvé spustenie (waiting for first run). GeoNetwork definuje mnoho typov týchto úloh v závislosti od protokolu vzdialeného systému, z ktorého preberá metadáta. Použili sme dva typy: *type="csw"* a *type="ogcwxss"*, kde *csw* definuje úlohu na prevzatie metadát zo vzdialenej katalógovej služby CSW a *ogcwxss* slúži pre všetky ostatné typy OGC služieb.
- Element <site> – definuje elementy <name> a <url>, do ktorých skript vkladá hodnoty z databázy pre všetky služby postupne zo stĺpcov *title* a *url*. Ďalší element <ogctype> definuje typ a verziu OGC služby, ktorá sa bude harvestovať. Tieto hodnoty skript prevezme zo stĺpca *version* pre všetky typy OGC služieb.
- Element <options> – obsahuje elementy, ktoré umožňujú nastaviť konfiguračné parametre úlohy ako interval opakovania jej spúšťania (element <every>), spustenie iba raz (element <oneRunOnly>), spôsob generovania metadát pre dáta, buď priamo z odpovede *GetCapabilities* (element <useLayer>), alebo použitím URL adresy na metadáta (<useLayerMD>), ak sa taká nachádza v odpovedi. V rámci elementu <option> je možné definovať kategórie metadát pre dáta (<datasetCategory>) pre potreby kategorizácie v procese ich následného vyhľadávania v rámci vyhľadávacieho klienta GeoNetwork. Hodnota v elemente <every> nebola dôležitá, keďže sme nastavili spustenie úlohy len raz (<oneRunOnly>true</oneRunOnly>). Pre metadáta dát sme nastavili parameter <useLayer>true</useLayer> na získanie informácií obsiahnutých priamo v odpovediach *GetCapabilities*. ID kategórie sme nastavili podľa typu služby, ktorá sa harvestovala.
- Element <privileges> – definuje skupiny používateľov a operácie, ktoré môžu aplikovať na metadáta. Zdefinovali sme pre všetky prevzaté metadáta skupinu s názvom *all* (*id=1*), ktorá definuje všetkých používateľov systému (anonymných aj autorizovaných). Inak povedané všetky zdroje GI nájdené pomocou vyhľadávacieho klienta budú prístupné pre zobrazovanie na mape a ukladanie pre všetkých používateľov.
- Element <categories> – definuje ID kategórie, ku ktorej sa priradia metadáta harvestovanej služby.

Po odoslaní každej požiadavky program GeoNetwork posiela odpoveď, v ktorej sa nachádzajú rovnaké informácie ako v požiadavke ale aj identifikátor *id*, ktorý program vytvoril vo svojej databáze pre harvestovaciu úlohu. Tento identifikátor skript vyberie z odpovede a uloží do tabuľky *services* k príslušajúcej OGC službe do stĺpca *id_gn*. Tento identifikátor je dôležitý z viacerých dôvodov, skript ho používa pri spúšťaní harvestovacích úloh, ale aj pri zisťovaní stavu úlohy (či je ešte spustená, či prebehla úspešne alebo zlyhala).

Pomocou skriptu *harvesting.php* sme popísaným spôsobom vytvorili harvestovacie úlohy pre všetky OGC služby, ktoré sme našli v internetovom vyhľadávači Google. Takto boli úlohy pripravené na spustenie.

Na spustenie harvestovacích úloh vytvorených v predchádzajúcom kroku sa použil ten istý skript *harvesting.php*, ktorý spúšťal úlohy pre tie služby, ktoré mali už priradené hodnoty v stĺpcoch *id_gn* a hodnota v stĺpci *status* bola rovná 3. Na spustenie úloh skript poslal jednoduchú XML POST požiadavku na XML službu (*harvesting.run*) v rámci GeoNetwork API na definovanú URL adresu v nasledujúcej štruktúre:

```

<request>
  <id>123</id>
</request>

```

Po odoslaní požiadavky GeoNetwork posiela odpoveď, ktorá definuje, že harvestovacia úloha beží, alebo, že nastal problém, napr. so zadaným *id* pre harvestovaciu úlohu a na základe toho definuje v odpovedi správu *not-found*. Skript *harvesting.php* je naprogramovaný tak, aby spúšťal jednotlivé úlohy postupne, čiže aby nespustil novú, kým predchádzajúca ešte neskončila. Navrhli sme program takto, aby sme predišli zahlcovanému pamäť počítača, ktorá by mohla nastať pri spustení viacerých úloh naraz. Na zistenie stavu úlohy, skript zasiela ďalšiu XML POST požiadavku s rovnakou štruktúrou, ako na spustenie úlohy len na inú URL adresu XML služby GeoNetwork API (*harvesting.get*). V rámci odpovede sa posielajú informácie o harvestovacej úlohe

medzi ktorými sa nachádzajú informácie, či je úloha spustená (elemente *<running>*) a kedy bola úloha naposledy spustená (elemente *<lastRun>*).

Aktualizácia databázy OGC služieb a harvestovacích úloh

Celkový proces je závislý na aktualizácii vstupných dát vo formáte CSV. Skript *import.php* po opätovnom spustení overuje CSV dáta a porovnáva záznamy v databáze na základe URL adresy. Ak nájde novú URL adresu, uloží ju do nového riadku v databáze a priradí jej hodnotu 0 v stĺpci *status*. Následne skript *crawl.php* overí danú URL adresu a priradí hodnotu v stĺpci *status* 1 (čaká na pridanie do GeoNetwork), alebo 2 (zistenie verzie zlyhalo) a verziu služby do stĺpca *version* (pri *status=1*). Ďalej skript *harvesting.php* vytvorí harvestovacie úlohy v programe GeoNetwork pre záznamy, ktoré majú hodnotu 1 v stĺpci *status* a po úspešnom vytvorení zmení túto hodnotu na 3 (vytvorené v GeoNetwork a priradené *id_gn*) a pre tento záznam spustí úlohu. Takto navrhnutá aktualizácia má nevýhodu v tom, že je závislá na manuálnej činnosti aktualizácie zoznamu nájdených OGC služieb v súboroch CSV.

5 Výsledky a diskusia

5.1 Experiment č.1 - kvantitatívne a kvalitatívne hodnotenie metadát vytvorených pre dáta ZB GIS

V rámci praktického experimentu realizovaného podľa navrhutej metodiky popísanej v kapitole 4.2.1 sme dosiahli nasledovné výsledky:

- bolo vytvorených 46 metadátových záznamov pre dátové súbory a 1 záznam pre zobrazovaciu službu typu WMS v štruktúrach ISO 19115/19119 vo formáte XML extrahovaním informácií z odpovede *GetCapabilities* služby WMS,
- bolo navrhnuté mapovanie metadátových záznamov na šablónu ZB GIS a implementované pomocou XSLT technológie, ktoré sa v budúcnosti môže použiť pre transformáciu akéhokoľvek metadátového záznamu vytvoreného podľa navrhutej metodiky,
- transformácia metadátových záznamov priniesla nasledovné kvalitatívne výsledky:
 - 46 metadátových záznamov pre dátové súbory a 1 pre sieťovú službu bolo správnych (neobsahovali chyby) voči ISO schéme,
 - 4 metadátové záznamy boli správne voči INSPIRE požiadavkám na metadáta (INSPIRE, 2008),
 - 42 metadátových záznamov obsahovalo dve chyby, prvou bola chýbajúca hodnota v povinnom INSPIRE elemente *2.1 Tematická kategória* a druhou chýbajúce kľúčové slovo zo všeobecného viacjazyčného environmentálneho tezauru (GEMET), ktoré opisuje príslušnú tému priestorových dát, ako je definované v prílohe I, II alebo III smernice INSPIRE (INSPIRE, 2007). Dôvodom bolo, že tieto záznamy nepatria pod INSPIRE tému Hydrografia, čiže navrhnuté mapovanie im nepriradilo žiadnu hodnotu pre tieto elementy.

5.2 Diskusia k výsledkom experimentu č.1

Rezort geodézie a kartografie disponuje množstvom dát, ktoré aby mohli byť efektívne zdieľané, musia byť popísané metadátami. Proces tvorby metadát je však časovo, a teda aj finančne, veľmi náročný, z čoho vyplýva, že akékoľvek zlepšenie procesu ich vytvárania má veľký význam a môže ušetriť nemalé prostriedky. Výsledky dosiahnuté v experimente č.1 poukazujú na to, že metodika na automatickú extrakciu a tvorbu metadát z existujúcich zdrojov GI akými sú OGC webové služby, môže zvýšiť efektívnosť v procese tvorby metadát. Ukážka praktickej realizácie sa aplikovala len na malej vzorke dát, kde išlo o 46 tried objektov, pre ktoré boli vytvorené metadátové záznamy v štruktúre ISO a z tých len pre 4 triedy objektov, ktoré patrili do INSPIRE témy Hydrografia bolo vytvorené detailné mapovanie na ZB GIS metadátový profil. Toto mapovanie je možné rozšíriť aj o definíciu ďalších podmienok pre triedy objektov, ktoré patria pod ďalšie INSPIRE témy (Dopravné siete, Katastrálne parcely, Správne jednotky, Geografické názvoslovie, Súradnicové referenčné systémy, Systavy súradnicových sietí a Výškové modely). Tieto témy zasahujú okrem databázy ZB GIS aj do katastra nehnuteľností, geodetických základov, ktoré spolu tvoria informačný systém geodézie, kartografie a katastra zadaného zákonom o geodézii a kartografii (Zákon, 1995).

Aby bolo možné aplikovať realizovaný experiment v praxi s vysokou mierou efektívnosti a presnosti jeho výsledkov, je potrebné splňať niekoľko predpokladov zo strany producenta dát:

- dáta sú poskytované pomocou OGC webových služieb, napr. službou WMS,
- názvy, alebo identifikátory jednotlivých vrstiev sú štandardizované, ak ide napr. o WMS službu, ktorá zobrazuje vrstvy z databázy ZB GIS, ich názvy by mali byť tvorené podľa názvov tried objektov definovaných v katalógu tried objektov (ÚGKK, 2008). Tento krok je dôležitý na identifikáciu dát v ďalšom procese automatizovanej tvorby metadát,
- okrem samotnej služby sú aj dáta popísané v rámci OGC modelu množstvom informácií, ako abstrakt, kľúčové slová, priestorový rozsah (geografické ohraničenie), tematický rozsah (atribúty) a ďalšie (ISO, 2005b). Ak sa tieto informácie uvedú už pri definícii vrstiev služby WMS, potom sa môžu použiť pri transformácii na ISO metadáta, a nemusia sa znova vytvárať,
- informácie, ktoré sa nachádzajú v metadátových položkách, ktoré sú povinné z pohľadu legislatívy (INSPIRE), ale aj vyplývajúce z požiadaviek používateľov dát na národnej, či rezortnej úrovni sú zapracované do tzv. metadátového profilu. Tento profil sa potom použije na ďalšiu transformáciu metadát do podoby, ktorá bude splňať tieto požiadavky.

Ak sú tieto predpoklady dodržané, nemalo by byť komplikované vytvoriť metadáta pre popis geopriestorových zdrojov v akejkoľvek doméne s vysokou mierou automatizácie. Ďalšou výhodou takéhoto prístupu je, že ak je proces aktualizácie dát realizovaný na strane producenta v rámci servera, ktorý publikuje OGC služby, mali by sa tieto zmeny premietnuť aj do samotných metadát v rámci geokatalógu a to opakovaným preberaním (harvestovaním) informácií z týchto služieb.

5.3 Experiment č.2 – OGC služby nájdené v Google a poskytované metadáta ku zdrojom GI

5.3.1 Nájdené URL na GetCapabilities požiadavky pre jednotlivé typy OGC služieb

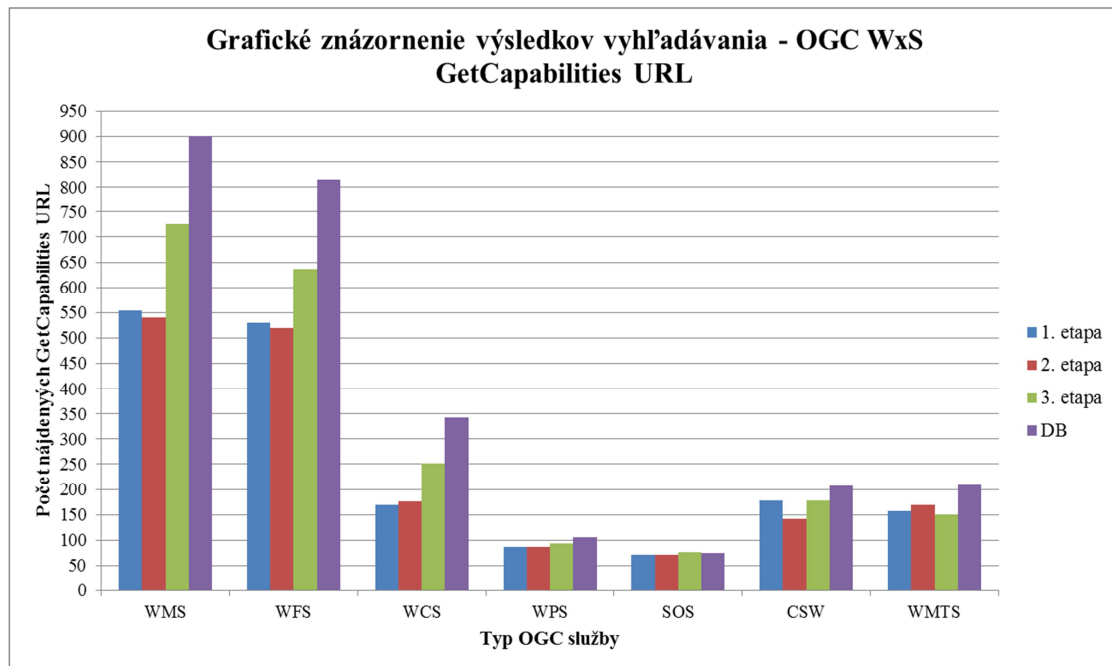
Vyhľadávanie webových adries URL na požiadavky typu *GetCapabilities* (*inurl:request=GetCapabilities*) sa aplikovalo pre sedem typov OGC webových služieb (*inurl:service=WxS*) a poskytlo výsledky sumarizované v tabuľke. 5.1.

Tabuľka 5.1 URL adresy na *GetCapabilities* pre OGC webové služby nájdené v Google

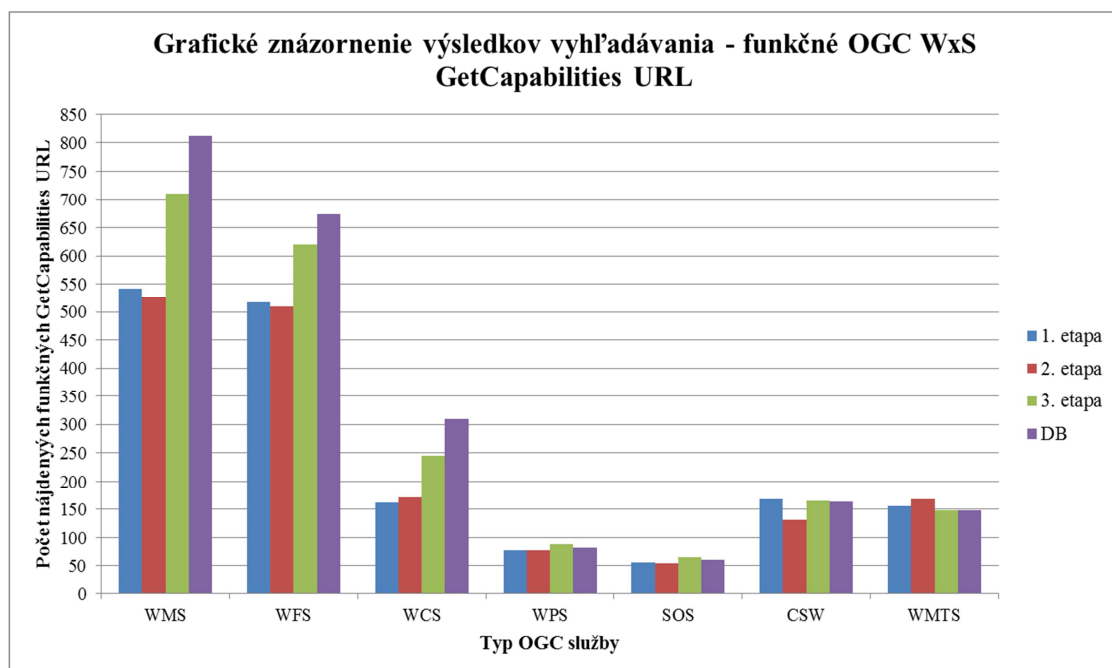
Typ OGC služby	1. etapa vyhľadávania (10/06/2012)	2. etapa vyhľadávania (26/06/2012)	3. etapa vyhľadávania (13/08/2012)	Uložené v databáze (30/08/2012)
WMS	554/541	540/527	725/711	901/812
WFS	530/517	520/509	636/620	815/674
WCS	169/162	176/171	251/245	342/310
WPS	86/77	86/78	94/88	106/82
SOS	71/56	71/55	76/65	74/60
CSW	178/169	142/132	177/166	207/163
WMTS	157/156	169/168	151/149	209/149
Σ	1745/1678	1704/1640	2110/2044	2654/2250

V tabuľke 5.1 sa nachádzajú počty nájdených URL na *GetCapabilities* pre jednotlivé typy OGC služieb (1. stĺpec) v troch etapách vyhľadávania (stĺpce 2 až 4). V poslednom stĺpci je uvedený počet služieb uložených v databáze po spustení scriptov *index.php* a *crawl.php* postupne pre všetky výsledky z jednotlivých etáp. Pre každý typ OGC služby sú v tabuľke uvedené dve hodnoty, kde prvá (hodnota pred lomkou) definuje celkové množstvo nájdených URL adries a druhá (hodnota za lomkou) tie, ktoré prešli testovaním nájdenej URL adresy. Testovanie URL adresy sa v rámci jednotlivých etáp vyhľadávania realizovalo len z pohľadu pravidiel správneho zápisu ľubovoľnej URL adresy pri procese importovania výsledkov do databázy (*index.php*). To, či je testovaná URL adresa aktívna, čiže server poskytuje určitú odpoveď, sa overovalo v rámci výsledkov uložených v databáze pri zisťovaní verzií jednotlivých OGC služieb (*crawl.php*). Táto skutočnosť je viditeľná aj na výsledkoch, kde priemerné množstvo správne zapísaných URL adries v rámci výsledkov vyhľadávania po troch etapách bolo 93,89% na jednu etapu. Na druhej strane skutočne funkčných URL adries, ktoré poskytovali

dokumenty *GetCapabilities* ako odpovede bolo 84,78% (2 250 odpovedí). Grafické znázornenie výsledkov sa nachádza v grafoch 5.1 a 5.2.



Graf 5.1 Grafické znázornenie výsledkov – počet nájdených OGC WxS *GetCapabilities* URL



Graf 5.2 Grafické znázornenie výsledkov – počet nájdených funkčných OGC WxS *GetCapabilities* URL

Na základe grafického znázornenia môžeme rozdeliť výsledky podľa množstva nájdených *GetCapabilities* URL adries do štyroch skupín:

- služby typu WMS a WFS majú najväčšie zastúpenie vo výsledkoch vyhľadávania. Až 66% funkčných URL adries na *GetCapabilities* patrilo službám WMS (36%), alebo WFS (30%),
- služby typu WCS tvoria druhú skupinu s viac ako 13,8% funkčných *GetCapabilities* URL adries,

- služby typu CSW a WMTS tvoria ďalšiu skupinu v množstve nájdených funkčných URL adries. Z celkového počtu nájdených a uložených URL adries patrilo 14% službám CSW (7,4%), alebo WMTS (6,6%),
- služby typu WPS a SOS tvoria najmenej zastúpenú skupinu vo výsledkoch. 6,1% nájdených funkčných URL patrilo WPS (3,5%), alebo SOS (2,6%).

Z pohľadu prírastku nájdených a uložených URL adries OGC služieb po realizovaní troch etáp vyhľadávania môžeme výsledky rozdeliť do nasledovných skupín:

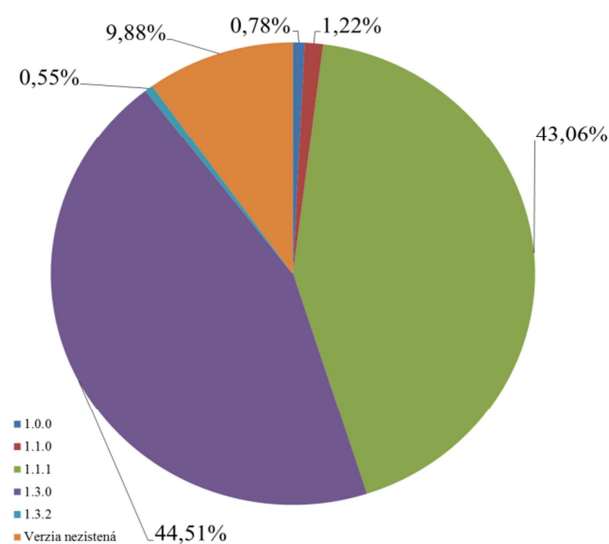
- prírastok nájdených URL pre služby typu WMS, WFS a WCS tvoril v priemere 60% nájdených nových služieb medzi priemerom služieb nájdených v rámci jednotlivých etáp a celkovým množstvom uloženým v databáze. Najväčší prírastok predstavovali služby typu WCS (77,51%), WMS (51,94%) a WFS (48,54%),
- prírastok pre ostatné typy OGC služieb nebol taký výrazný a tvoril v priemere 31%. Najväčší prírastok mali služby typu CSW (32,98%), WMTS (32,56%), WPS (30,86%) a SOS (26,14%).

Pre funkčné URL adresy sa tieto prírastky znížili, ale dve hlavné skupiny ostali zachované s nasledovnými výsledkami:

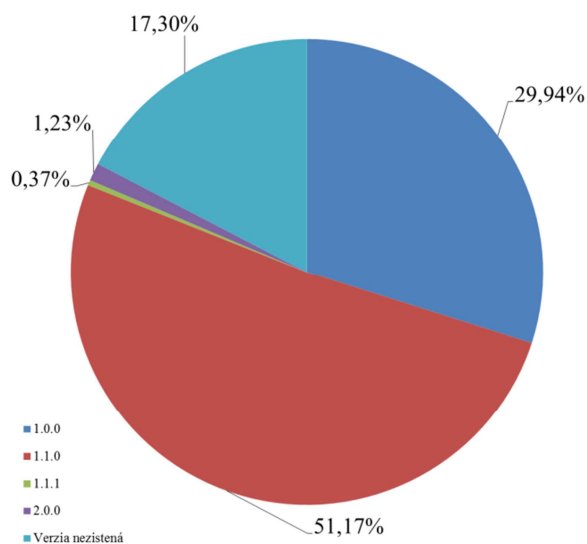
- prírastok nájdených a zároveň aj funkčných URL pre služby typu WMS, WFS a WCS tvoril v priemere 40%, medzi priemerom služieb nájdených v rámci jednotlivých etáp a celkovým množstvom uloženým v databáze. Najväčší prírastok funkčných služieb predstavovali služby typu WCS (60,90%), WMS (36,93%) a WFS (22,84%),
- prírastok funkčných URL pre ostatné typy služieb bol už na minimálnych až záporných hodnotách. Najviac to bolo pre služby typu CSW (4,71%), služby WMTS (2,74%), WPS (1,23%) a pre služby WMTS to bol úbytok (-5,50%).

5.3.2 Výsledky vyhľadávania podľa verzií jednotlivých typov OGC služieb

Po aplikácii skriptu *crawl.php* na *GetCapabilities* URL adresy uložené v databáze *geoserver* sme okrem overenia funkčnosti danej URL adresy, získali aj informácie o verziách jednotlivých OGC služieb. Program vyhľadal túto informáciu v rámci atribútu *version* v koreňovom elemente XML dokumentu, ktorý by mala služba vrátiť späť v odpovedi. Percentuálne zastúpenie nájdených verzií pre WMS a WFS služby je graficky znázornené v grafoch 5.3 – 5.4.



Graf 5.3 Zistené verzie WMS služieb



Graf 5.4 Zistené verzie WFS služieb

Grafické znázornenia ostatných typov OGC služieb (WCS, SOS, WPS, WMTS a CSW) sú uvedené v dizertačnej práci. Najväčší počet verzií bol nájdený pre služby typu CSW (6 verzií), WCS a WMS (5 verzií), WFS (4 verzie), WPS a SOS (3 verzie) a WMTS (2 verzie). Zastúpenie verzií pre jednotlivé typy OGC služieb bolo:

- pre služby WMS bolo nájdených 401 *GetCapabilities* URL verzie 1.3.0 (44,51%), 388 verzie 1.1.1 (43,06%), z celkového počtu 901 uložených *GetCapabilities* URL adries v databáze. Zostatok tvorilo 11 služieb verzie 1.1.0 (1,22%), 7 verzie 1.0.0 (0,78%) a 5 verzie 1.3.2 (0,55%). Pre takmer desatinu WMS služieb (89) sa verzia nezistila,
- pre WFS služby bolo nájdených 417 *GetCapabilities* URL verzie 1.1.0 (51,17%), 244 verzie 1.0.0 (29,94%), 10 verzie 2.0.0 (1,23%) a 3 verzie 1.1.1 (0,27%). Pre 141 WFS služieb (17%) nebola zistená verzia,
- pre WCS služby bolo nájdených 250 *GetCapabilities* URL verzie 1.0.0 (73,10%), 41 verzie 1.1.1 (11,99%), 7 verzie 1.1.2 (2,05%) a po 6 pre verzie 2.0.0 a 1.1.0 (1,75%). Pre 32 WCS služieb (9,36%) nebola zistená verzia,
- pre SOS služby bolo nájdených 55 *GetCapabilities* URL verzie 1.0.0 (74,32%) a po 2 verzií 2.0.0 a 0.0.31 (2,70%). Pre 15 SOS služieb (20,27%) nebola zistená verzia ,
- pre WPS služby bolo nájdených 71 *GetCapabilities* URL verzie 1.0.0 (66,98%), 10 verzie 0.4.0 (9,43%) a 1 verzie 1.1.0 (0,94%). Pre 24 WPS služieb (22,64%) nebola zistená verzia,
- pre WMTS služby bolo nájdených 147 verzie 1.0.0 (70,33%), 2 verzie 1.1.1 (0,96%) a pre 60 WMTS služieb (28,71%) sa verzia nezistila,
- pre CSW služby bolo nájdených 138 verzie 2.0.2 (66,67%), po 7 pre verzie 2.0.0 a 2.0.1 (3,38%), 6 verzií 1.0.0 (2,90%) a po jednej verzii 1.0.1 a 0.9.0 (0,48%). Pre 47 CSW služieb (22,71%) nebola zistená verzia.

5.3.3 Výsledky po ukončení procesu *harvesting.php*

Po spustení skriptu *harvesting.php* sa postupne automaticky vytvárali a spúšťali harvestovacie úlohy v geokatalógu GeoNetwork pre OGC služby uložené v tabuľke *services* v rámci databázy *geoserver*. Profily sa vytvárali len pre tie služby, ktoré mali úspešne zistenú verziu, čiže tie, ktoré mali v stĺpci *status* hodnotu 1 (Graf 5.2). Výsledky procesu sa ukladali do databázy *metadata* v rámci inštancie GeoNetwork použitej v experimentoch.

Po aplikácii skriptu *harvesting.php* sa automaticky aktualizovali hodnoty v tabuľke *services* pre jednotlivé OGC služby v stĺpci *status*. Službám, pre ktoré GeoNetwork nevedel vytvoriť harvestovaciu úlohu, čiže odpovedal chybovým hlásením, sa v tabuľke *services* priradila hodnota 4. Službám, ktorým vytvoril a spustil úlohu sa priradila hodnota 5 a pre ostatné hodnota 6. Výsledky tohto procesu sú uvedené v tabuľke 5.2.

Tabuľka 5.2 Výsledky v databáze *services* po ukončení procesu *harvesting.php*

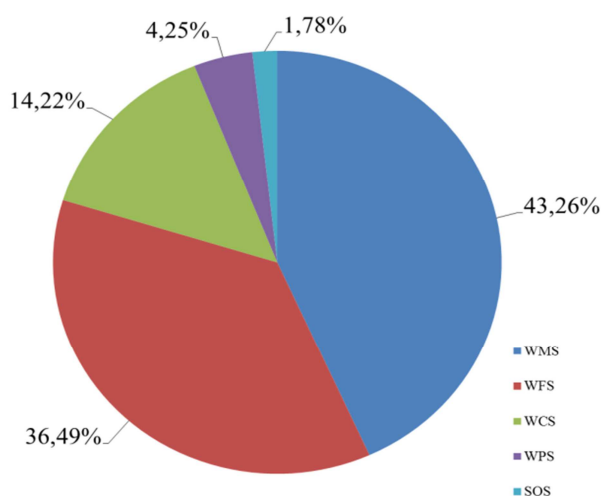
Typ OGC služby	Typ úlohy	status=4	status=5	status=6
WMS		7	801	4
WFS		7	666	1
WCS	ogcwx	2	308	0
WPS		0	82	0
SOS		2	58	0
CSW	csw	3	153	7
WMTS	nepodporované	149	0	0

5.3.4 Kvantitatívne a obsahové vyhodnotenie vytvorených metadát

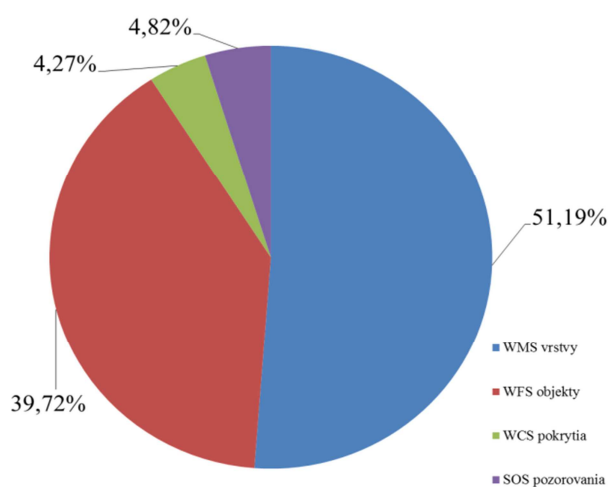
Po spustení skriptu *harvesting.php* sme naplnili databázu geokatalógu GeoNetwork metadátami. Išlo o metadáta k OGC službám, ktoré sme našli v rámci indexu Google a ktoré boli po overení skriptom *crawl.php* funkčné a bolo možné zistiť ich verziu. Okrem týchto metadát sa z jednotlivých služieb vyextrahovali aj metadáta ku zdrojom GI, ktoré tieto služby poskytujú. Jednalo sa o metadáta ku vrstvám (WMS), objektom (WFS), pokrytiam (WCS) a pozorovaniám (SOS). Výsledky sú sumarizované v tabuľke 5.3 a v grafoch 5.5 - 5.6.

Tabuľka 5.3 Metadáta vytvorené pre OGC služby a zdroje GI pre harvestovací profil ogcwxs

OGC služba	WMS	WFS	WCS	WPS	SOS
Vytvorené metadáta pre služby	703	593	231	69	29
Vytvorené metadáta pre WMS vrstvy	36835	-	-	-	-
Vytvorené metadáta pre WFS objekty	-	28579	-	-	-
Vytvorené metadáta pre WCS pokrytia	-	-	3074	-	-
Vytvorené metadáta pre SOS pozorovania	-	-	-	-	3469
Σ	37538	29172	3305	69	3498
			73582		



Graf 5.5 Metadáta vytvorené pre OGC služby



Graf 5.6 Metadáta vytvorené pre zdroje GI poskytované službami OGC

Pre OGC služby typu CSW (katalógové služby) bol použitý iný harvestovací profil (*csw*) špecifikovaný na priamy zber metadát, ktoré služby tohto typu poskytujú. Nešlo o transformáciu informácií obsiahnutých v odpovedi *GetCapabilities*, ale priamo o prevzatie celých metadátových záznamov, ktoré katalógové služby poskytujú. Z celkového množstva 68935 prebratých metadátových záznamov bolo 65364, ktoré popisujú dátové súbory (*dataset*), 449 série dátových súborov (*series*), 1886 pre služby (*services*) 1236 pre iné typy zdrojov GI (ISO, 2003).

Celkový počet metadátových záznamov vytvorených v databáze geokatalógu GeoNetwork bol ku dňu **22.9.2012** **142515** metadát popisujúcich zdroje GI.

Pomocou vyhľadávacieho klienta a katalógovej služby CSW implementovanej programom GeoNetwork sme vykonali rôzne analýzy obsahu získaných metadát. Ďalšia časť práce poskytuje výsledky týchto analýz.

Metadáta popisujúce zdroje pre INSPIRE témy priestorových dát

Smernica INSPIRE (INSPIRE, 2007) definuje 34 tém priestorových dát. Tieto sme vyhľadávali ako kľúčové slová v rámci prevzatých metadátových záznamov pre kategórie, ku ktorým sa priradili výsledky procesu realizovaného skriptom *harvesting.php*. Výsledky vyhľadávania pre témy z príloh I a II sú uvedené v tabuľke 5.4.

Tabuľka 5.4 Metadátové záznamy vyhľadávané podľa INSPIRE tém ako kľúčových slov

INSPIRE témy	WMS	WFS	WCS	WPS	SOS	CSW
Príloha 1 smernice						
1 - Coordinate reference systems (8)	0	0	0	0	0	dataset (5) series (3)
2 - Geographical grid systems (10)	0	0	0	0	0	dataset (7) series (3)
3 - Geographical names (691)	service (2)	0	0	0	0	dataset (652)

	layers (15)					series (11) service (11)
4 - Administrative units (68)	service (1) layers (6)	features (4)	0	0	0	dataset (29) series (7) service (23)
5 - Addresses (26)	service (1)	0	0	0	0	dataset (13) series (5) service (7) iné (2)
6 - Cadastral parcels (46)	0	0	0	0	0	dataset (28) series (1) service (3) iné (14)
7 - Transport networks (96)	service (1) layers (9)	features (3)	0	0	0	dataset (68) series (1) service (9) iné (5)
8 - Hydrography (356)	service (2) layers (1)	features (1)	0	0	0	dataset (328) series (6) service (18)
9 - Protected sites (117)	service (1)	0	0	0	0	dataset (63) series (3) service (13) iné (37)
Príloha 2 smernice						
1 - Elevation (86)	service (1) layers (2)	0	0	service (1)	0	dataset (71) series (2) service (6) iné (3)
2 - Land cover (153)	service (5) layers (0)	0	0	0	0	dataset (22) series (6) service (6) iné (114)
3 - Orthoimagery (12604)	layers (1)					dataset (12591) series (7) service (3) iné (2)
4 - Geology (1072)	service (11) layers (49)	service (7) features (6)				dataset (1020) series (9) service (62) iné (8)

Výsledky vyhľadávania pre témy z prílohy III sú uvedené v dizertačnej práci. Celkový počet metadátových záznamov, ktoré obsahovali názov INSPIRE témy priestorových dát bol 16601, čo v čase vyhľadávania (22.9.2012) tvorilo 11,64% z celkového počtu vytvorených metadát. Väčšina metadát patrila do kategórie CSW (16364), z ktorých bolo 15604 pre typ zdroja *dataset*, 106 typ *series*, 263 typ *service* a 391 pre iné typy (*other*).

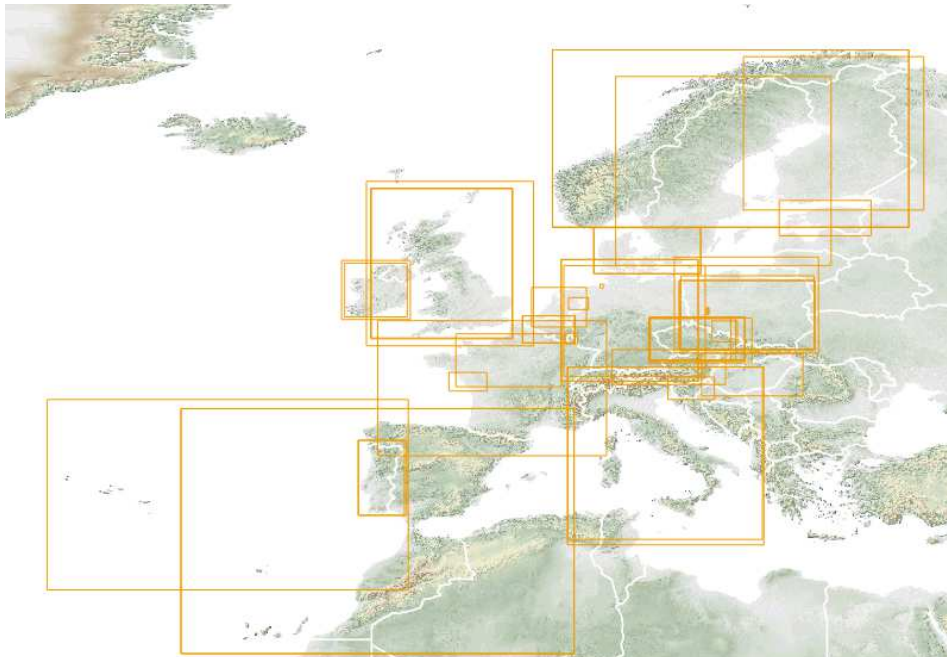
Metadáta popisujúce INSPIRE sieťové služby

Smernica INSPIRE (INSPIRE, 2007) definuje päť typov sieťových služieb. Tieto typy sme vyhľadávali pomocou rozhrania katalógovej služby CSW programu GeoNetwork s použitím filtra *ServiceType*, ktorý je definovaný v špecifikácii katalógových služieb (Nebert et al., 2007). Výsledky sú zosumarizované v tabuľke 5.5.

Tabuľka 5.5 Metadátové záznamy pre INSPIRE sieťové služby

Typ INSPIRE sieťovej služby	ServiceType	Počet MD
Vyhľadávacia	discovery	10
Zobrazovacia	view	1527
Ukladacia	download	27
Transformačná	transformation	1
Služba umožňujúca spustenie služieb priestorových dát	invoke	6
Ostatné služby	other	23
Σ		1594

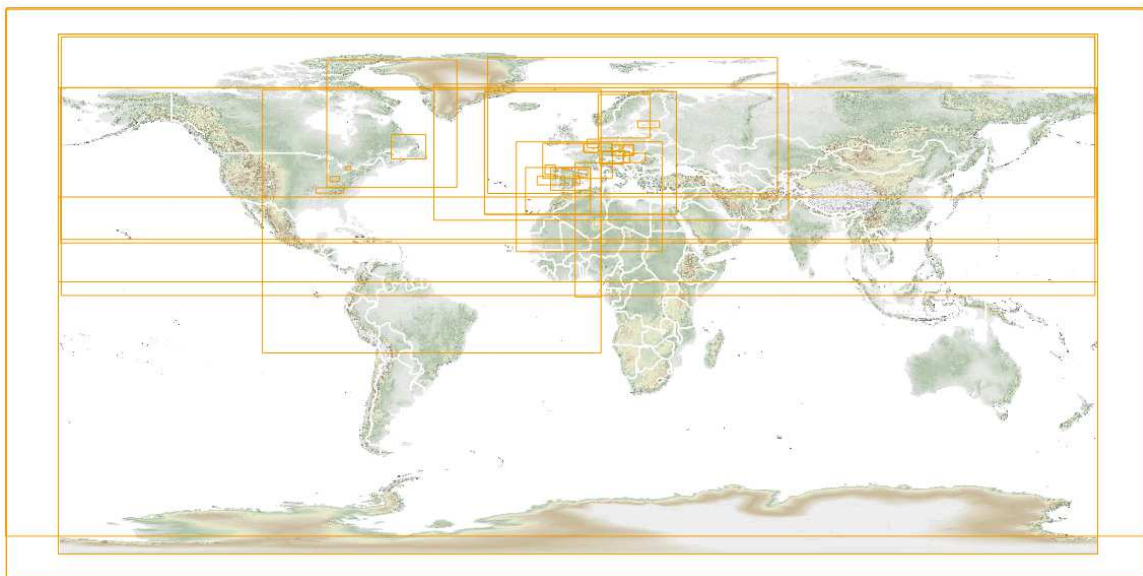
Všetky nájdené metadátové záznamy patrili do kategórie typu *Catalogue Service for the Web*, čiže boli prebraté z katalógových služieb. Z celkového množstva metadát prebratých z CSW, ktoré popisujú zdroje GI typu *service* (1886 záznamov) tvorili v čase analýzy INSPIRE sieťové služby väčšinu výsledkov, až 84,5% (1594 záznamov). Grafické znázornenie ohraničujúcich rámov nájdených metadát popisujúcich INSPIRE sieťové služby je zobrazené na obr.5.1.



Obr. 5.1 Geografické ohraničenia zdrojov GI typu service v metadátach prevzatých zo služieb CSW

Geografické ohraničenie zdrojov GI

Pre grafické znázornenie geografického ohraničenia sme použili funkciu geokatalógu GeoNetwork, ktorá zobrazuje ohraničujúce rámy (*bounding box*) výsledkov vyhľadávania. V princípe ide o prenesenie informácií o rozsahu zdroja v geografickom priestore uvedenom v metadátach vo forme súradníc do mapy. Príklad grafického znázornenie ohraničujúcich rámov metadát uložených v databáze geokatalógu je zobrazený na obr. 5.2 pre prvých 50 metadátových záznamov nájdených pre kategórie WMS služby a vrstvy.

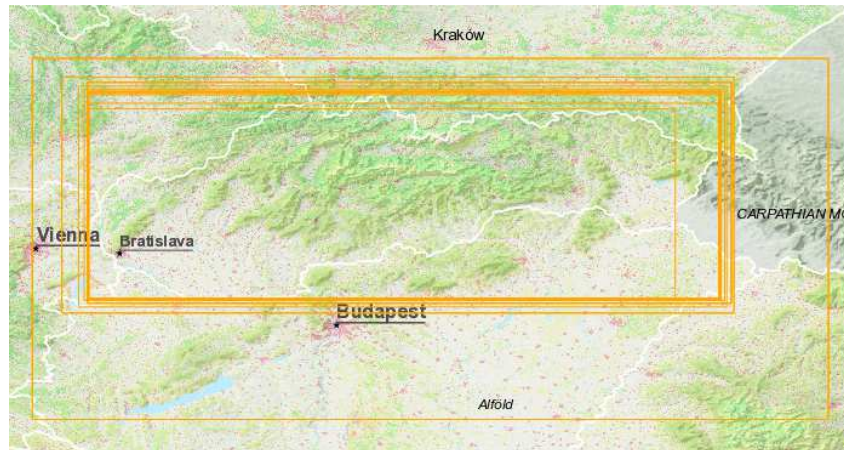


Obr. 5.2 Geografické ohraničenie WMS služieb a vrstiev definované v ich metadátach

Metadáta zdrojov GI pre oblasť Slovenskej republiky

Na zistenie rozsahu zdrojov GI poskytovaných pre oblasť SR sme aplikovali priestorové vyhľadanie s definovaním obdĺžnika ohraničujúceho územie SR a kľúčového slova *slovakia*. Získali sme 117 metadátových

záznamov, z ktorých 4 popisovali WMS služby, 87 WMS vrstvy, 1 WFS objekt, 1 WCS službu, 23 dátové súbory získané z CSW služieb. Nájdené zdroje GI sú poskytované organizáciami zo SR ako Slovenská agentúra životného prostredia (16 metadátoých záznamov), Štátna ochrana prírody (71), Národné lesnícke centrum (1) a Štátny geologický ústav Dionýza Štúra (3), ktoré patria medzi veľkých správcov a poskytovateľov zdrojov GI. Ostatné sú poskytované Európskymi organizáciami ako European Environmental Agency (14), Eurographics (1), Centre for Ecology & Hydrology (1) a svetovými ako World Soil Information Data Centre (2) a Food and Agriculture Organization of the United Nations (1). Grafické znázornenie ohraničujúcich rámov nájdených metadát je zobrazené na obr.5.3.



Obr. 5.3 Geografické ohraničenia metadát pre oblasť SR

5.4 Hodnotenie výsledkov experimentu a vyplývajúca diskusia

Vyhľadávanie OGC služieb v Google

Celkový počet nájdených URL adries na GetCapabilities požiadavky pre OGC služby predstavoval 2654, z ktorých pre 2250 bolo po aplikácii skriptu *crawl.php* možné zistiť ich verziu (tab. 5.1). Tieto výsledky potvrdzujú to, čo aj dostupná literatúra, že geopriestorové webové služby je už v súčasnosti možné nájsť pomocou internetových vyhľadávačov. Porovnaním našich výsledkov s výsledkami výskumu popísané v Pellicer et al. (2010), zistíme, že sme našli takmer o 1 000 OGC služieb menej. Jednou príčinou je, že my sme použili len jeden internetový vyhľadávač a autori použili tri. Ďalšou príčinou môže byť práve metóda, ktorú sme použili pri vyhľadávaní v Google ako aj skutočnosť, že my sme k databázam Google pristupovali cez webové rozhranie a nie s využitím API. Použili sme filter, ktorý vyhľadával textové reťazce *request=getcapabilities* a *service=wxs* v rámci URL adresy zdrojov uložených v databázach Google, čím sme hlbšie špecifikovali požiadavku a tým aj zúžili rozsah možných výsledkov. Dôvodom bolo, že v rámci experimentu nebolo realizované ďalšie spracovanie výsledkov tak ako v Pellicer et al. (2010). Výhodou takého prístupu je, že výsledky neobsahujú veľa šumov (pravdepodobnosť nájdenia URL na GetCapabilities pre OGC služby je vysoká), na strane druhej však tento spôsob vyhľadávania môže prehliadnúť zdroje, ktoré mohli obsahovať URL adresy OGC služieb. V rámci experimentu sme považovali za funkčné OGC služby tie, pre ktoré bolo možné zistiť ich verziu z dôvodu použitia v rámci ďalších krokov experimentu. Nemusí to však odzrkadľovať skutočnosť z pohľadu funkčnosti služieb, pretože zisťovanie verzie bolo realizované vyhľadávaním jej hodnoty v rámci atribútu *version*, umiestneného v koreňovom elemente odpovede. To znamená, že GetCapabilities URL adresy tých služieb, ktoré neobsahovali tento atribút v koreňovom elemente boli označené ako také, pre ktoré zistenie verzie zlyhalo a ďalej sa už s nimi nepracovalo.

Typy OGC služieb a ich a verzie

Najväčšie zastúpenie vo výsledkoch mali služby typu WMS a WFS (66%). Toto môže byť zapríčinené viacerými faktormi ako:

- Funkcionalita služby – WMS služby poskytujú základnú funkcionality pre zobrazovanie geopriestorových dát vo forme georeferencovaných rastrových obrázkov, ktoré generuje mapový server v požadovanom súradnicovom systéme a priestorovom ohraničení. Služba umožňuje prinajmenšom zobrazit', navigovať, priblížiť/vzdialiť, sledovať alebo prekryvať prehliadateľné súbory geopriestorových dát a zobrazit' informácie o legende, čo je z pohľadu používateľa GI požadovaná funkcionality hneď po nájdení relevantných zdrojov. WFS služba vo svojej základnej verzii nadväzuje

na funkcionalitu WMS tým, že umožňuje zobrazené GI, alebo ich časti ukladať ako vektorové dáta pre potreby ďalšieho spracovania.

- Dostupné implementácie a podpora v GIS softvéroch – v poslednom období bolo vyvinutých množstvo softvérov, ktoré poskytujú implementácie služieb WMF a WFS, ktoré sa zhodujú s implementačnými špecifikáciami OGC. Zo skupiny open source sú to najmä GeoServer (OGC certifikované implementácie pre WMS 1.1.1 a WFS 1.0), MapServer a deegree (WMS 1.1.1, WMS 1.3.0 a WFS 1.1.0) a zo skupiny proprietárnych softvérov ArcGIS Server (WMS 1.1.1, WMS 1.3.0 a WFS 1.0.0), GeoMedia SDI Pro (WMS 1.1.1, WMS 1.3.0 a WFS 1.0.0), Bentley Geo Web Publisher a ďalšie. Na druhej strane existuje množstvo softvérov, ktoré podporujú používanie WMS a WFS služieb ako zdrojov dát GIS. Ide o desktop GIS aplikácie ako QuantumGIS, uDig, MapGuide OpenSource, GRASS GIS, či gvSIG zo skupiny open source, alebo GIS produkty od komerčných firiem ako ESRI, Bentley, GeoMedia, Autodesk a ďalšie.
- Súvis s legislatívou v oblasti GI – technické usmernenia pre zobrazovacie a ukladacie služby INSPIRE popisujú príklad praktickej implementácie s využitím ISO štandardov, ISO 19128 (WMS 1.3.0) pre zobrazovacie a ISO 19142 (WFS 2.0) pre ukladacie služby. Ide o sieťové služby pracujúce s dátami, ktoré sa v rámci INSPIRE implementujú ako prvé po vyhľadávacích službách.

Z pohľadu verzií pre WMS služby tvorili väčšinu výsledkov WMS 1.1.1 a WMS 1.3.0 čo predstavovalo spolu viac ako 87%. Z toho vyplýva, že nosnými implementáciami sú tieto dve verzie, čo potvrdzuje aj podpora v dostupných softvéroch. Pre služby typu WFS tvorili viac ako polovicu výsledkov WFS 1.1.0 a menej ako jednu tretinu WFS 1.0.0. Tu môžeme tiež vidieť závislosť na podpore týchto štandardov v existujúcich softvéroch. Najnovšia verzia WFS 2.0, ktorá je aj medzinárodným štandardom ISO 19142 (ISO, 2010) je ešte málo podporovaná v súčasnosti, čo potvrdzujú aj výsledky (1,23%).

Najmenšie zastúpenie vo výsledkoch mali služby WFS a SOS (6,1%). Špecifikácia pre SOS služby patrí medzi najmladšie z rady OGC špecifikácií pre geopriestorové webové služby a v súčasnej dobe ešte nie je dostupné také množstvo softvérov, ako napr. pre služby WMS, WFS alebo WCS. Služba WPS slúži na realizáciu rôznych typov spracovania geopriestorových dát (*geoprocessing*) pre rôzne vstupné a výstupné dátové formáty. V praxi ešte nie je používaná v takom rozsahu ako WMS alebo WFS služby, čo potvrdzujú výsledky. Napriek tomu sa predpokladá vysoký nárast implementácií služieb tohto typu, ktorý súvisí aj so smernicou INSPIRE, ktorá definuje typy sieťových služieb, pre ktoré bude použitie štandardu WPS vyhovujúce.

Metadáta vytvorené harvestovaním OGC služieb

Po ukončení procesu realizovaného skriptom *harvesting.php* bolo v databáze geokatalógu GeoNetwork vytvorených viac ako 142 tisíc metadátových záznamov ku dňu 22.9.2012 pre kategórie zadané pre zdroje GI. Táto hodnota sa bude meniť, pretože nie všetky úlohy sa zrealizovali úspešne a tak sa budú opakovať. Tie úlohy, ktoré sa zrealizovali úspešne sa opakovať už nebudú, pretože boli nakonfigurované bez opakovaní. Môžeme predpokladať vysoký nárast počtu metadátových záznamov najmä pre kategóriu typu CSW, pretože bolo úspešných len 41 z celkového počtu 160 vytvorených harvestovacích úloh, z ktorých 138 by malo byť zrealizovaných, pretože boli verzie 2.0.2, ktoré GeoNetwork podporuje. Problém nastal pri komunikácii geokatalógu GeoNetwork s konkrétnou službou, ktorého výsledkom boli chybové hlásenia typu *bad-xml-response, operation-aborted*, alebo len *error*.

Metadáta pre INSPIRE zdroje GI

Pozitívny vplyv smernice INSPIRE na krajiny v Európe v oblasti poskytovania GI je viditeľný aj z výsledkov experimentu. Veľké množstvo metadátových záznamov najmä pre kategóriu CSW obsahovalo názvy INSPIRE tém priestorových dát v kľúčových slovách. Ak porovnáme množstvá metadát pre dátové súbory a série s množstvom metadát v aplikácii INSPIRE Geoportal, tak zistíme, že sa podstatne líšia. INSPIRE Geoportal obsahuje radovo väčšie množstvá metadát. Toto môže byť zapríčinené skutočnosťou, že v čase vyhľadávania bolo úspešne zrealizovaných harvestovacích úloh len pre 41 služieb typu CSW. Môžeme predpokladať, že po úspešnej realizácii úloh pre ďalšie CSW služby sa naše výsledky priblížia tým v INSPIRE Geoportal. Zaujímavým ukazovateľom z pohľadu INSPIRE sieťových služieb je, že výsledky potvrdzujú súčasný stav ich implementácie, ktorý môžeme nazvať obdobím vyhľadávacích a zobrazovacích služieb. Túto skutočnosť potvrdzuje najmä hodnota 1 527 získaných metadátových záznamov pre služby, ktoré mali hodnotu *view* v časti, ktorá definuje typ služby (*ServiceType*) a ktorá bola zadaná na identifikáciu INSPIRE zobrazovacích služieb. INSPIRE Geoportal poskytuje podobné číslo vo výsledkoch, 1541 služieb typu *View* je možné v súčasnosti vyhľadať.

6 Záver

GI môžeme v súčasnosti vyhľadávať vo viacerých zdrojoch prostredníctvom internetu. Najviac používaným zdrojom pre vyhľadávanie geopriestorových dát a služieb sú infraštruktúry pre priestorové informácie. Avšak bežný používateľ internetu uprednostňuje používanie všeobecných internetových vyhľadávačov ako napr. Google. Predtým ako je akýkoľvek zdroj GI možné vyhľadať na internete, tak by mal spĺňať viacero podmienok. Z pohľadu IPI musí mať tento zdroj vytvorené popisné informácie, tzn. metadáta, ktorých tvorba je v súčasnosti ešte vo veľkej miere manuálna. Poskytovateľ dát a služieb potrebuje vytvoriť metadátové záznamy, ktoré ich popisujú a publikovať v geokatalógu, aby ich bolo možné vyhľadať. Pokiaľ chceme zabezpečiť, aby bolo možné nájsť zdroje GI v internetových vyhľadávačoch, tak je potrebné uviesť URL adresy (napr. URL adresa na GetCapabilities požiadavku alebo koncový bod OGC služby) v rámci webových aplikácií, ktoré používame na interakciu používateľov s dátami prostredníctvom internetu (napr. geoportál). V práci sme navrhli metodiku, ktorou je možné proces tvorby a aktualizácie metadát automatizovať s využitím existujúcich popisných informácií uvedených v rámci OGC metadátového modelu pre popis služieb a špecifických informácií uvedených v rezortných metadátových profiloch. Na praktickú realizáciu tejto metodiky sme využili vstavanú funkcionálnu geokatalógu GeoNetwork, ktorý umožňuje pretransformovať metadáta kódované v modeloch OGC do ISO modelov. Na dosiahnutie interoperability, ktorá je v IPI definovaná medzinárodnými štandardmi a osobitnými špecifickými pravidlami závislými od úrovne IPI, bolo potrebné vytvoriť doplnkovú transformáciu, aby sme tieto požiadavky splnili. Konkrétny výstup tejto transformácie môže byť použitý na zdroje z databázy ZB GIS, ktorej výrez sme použili v rámci nášho experimentu. V druhej časti práce sme navrhli metodiku vyhľadávania OGC služieb v Google a extrahovania popisných informácií do geokatalógu GeoNetwork. Na základe navrhnutej metodiky sme realizovali praktický experiment, ktorý tvrdenie, že tzv. mainstream web už v súčasnosti poskytuje zdroje GI vo forme OGC služieb potvrdil. Vyhľadávaním v internetovom vyhľadávači Google sme našli 2654 URL adries na požiadavky typu GetCapabilities pre súčasné OGC služby. Po extrahovaní popisných informácií k týmto službám a dátam, s ktorými pracujú sme získali viac ako 142 tis. metadátových záznamov ku dňu 22.9.2012. Táto hodnota sa s postupom času bude zvyšovať, pretože tie harvestovacie úlohy, ktoré zlyhali sa budú opakovať až pokiaľ nebudú úspešné. Nájdene metadáta ku GI sú poskytnuté pre vyhľadávanie a ďalšie analýzy pomocou vyhľadávacieho klienta geokatalógu GeoNetwork na URL adrese <http://tokenbros.com:8082/geonetwork>, alebo pomocou rozhrania katalógovej služby CSW na adrese <http://tokenbros.com:8082/geonetwork/srv/eng/csw>.

7 Prínos pre vedu a prax v oblasti geodézie a geoinformatiky

Rezort geodézie a kartografie disponuje množstvom dát, ktoré aby mohli byť efektívne zdieľané, musia byť popísané metadátami. Proces tvorby metadát je však časovo, a teda aj finančne, veľmi náročný, z čoho vyplýva, že akékoľvek zlepšenie procesu ich vytvárania má veľký význam a môže ušetriť nemalé prostriedky. Proces navrhnutý a realizovaný v rámci experimentu č. 1 slúži na automatickú tvorbu metadát s využitím transformácie existujúcich popisných informácií kódovaných v modeloch OGC služieb na ISO modely požadované smernicou INSPIRE a metadátových šablón vytvorených pre špecifický popis zdrojov v rámci rezortu. Navrhnuté a overené postupy môžu významne zefektívniť proces tvorby a aktualizácie metadát, čo je prínosom pre vedu a prax v oblasti geodézie a kartografie.

Výsledky dosiahnuté v rámci experimentu č. 2 poukazujú na veľké množstvá GI, ktoré už je možné v súčasnosti vyhľadať v rámci tzv. mainstream webu s využitím internetového vyhľadávača Google. Táto skutočnosť naznačuje, že bariéra medzi IPI a mainstream webom sa znižuje, čo môže súvisieť aj s tým, že smernica INSPIRE je už tretí rok vo svojej implementačnej fáze. Zapojené organizácie z toho dôvodu poskytujú viac zdrojov GI vo forme OGC služieb v rámci rôznych aplikácií geoportálov, alebo webových informačných stránok, na ktorých adresy týchto služieb uvádzajú. Ich komplexná analýza je prínosom pre vedu v oblasti geoinformatiky.

Vzhľadom na stanovené ciele a dosiahnuté výsledky v praktických experimentoch dizertačnej práce realizovaných podľa navrhnutých metodík, môžeme zosumarizovať prínosy pre vedu a prax do nasledovných bodov:

- metodika automatizovanej tvorby metadát pre zdroje GI pomocou transformácií medzi rôznymi metadátovými modelmi s využitím existujúcich zdrojov GI vo forme OGC služieb a metadátových šablón navrhnutých pre popis v rámci určitej domény,
- metodika vyhľadávania OGC služieb v internetovom vyhľadávači Google, ich ďalšieho spracovania a následnej extrakcie metadát ku zdrojom GI, ktoré služby poskytujú,

- vytvorenie nástroja MDValidator na realizáciu hromadnej validácie metadátových záznamov voči INSPIRE požiadavkám pre metadáta,
- návrh mapovania metadátového záznamu, vytvoreného transformáciou z OGC modelu na ISO model, na rezortný metadátový profil ZBGIS a jeho implementácia s využitím XSLT technológie,
- naplnenie databázy geokatalógu GeoNetwork metadátami popisujúcimi zdroje GI, automatizovaným vytvorením a spustením harvestovacích úloh pre OGC služby nájdené v internetovom vyhľadávači Google, pomocou vytvorených PHP skriptov *crawl.php* a *harvesting.php*,
- poskytovanie získaných metadát cez grafické používateľke rozhranie vyhľadávacieho klienta a katalógovú službu CSW pre potreby vyhľadávania a ďalších analýz na získaných zdrojoch GI.

8 Summary

The World Wide Web (WWW) has geospatial extent today. Web pages make use of geospatial information as a positional reference of their content to the Earth. Web portals like Google Maps use geospatial data of various themes (Transportation Network, Forests, Hydrography, etc.) produced by several authoritative providers in the world (Department of Transportation in the US, European Space Agency, Environmental Information Centre in Indi, etc.). The ability of the users and providers related to management of the geospatial information via the Internet forms the shape of the current geospatial web. Geospatial web may be defined as a collection of geospatial web services, data and metadata, which supports discovery, evaluation and use of the geospatial data and services within a particular domain (Biodiversity, Floods, Landscape Planning, Transportation, etc.). Thesis deals with a problem of the geospatial information resources discovery via the Internet with emphasis on Spatial Data Infrastructures and mainstream web represented by search engine Google. To find the user specific information should be the most important task within any information system.

We have proposed within the first part of the thesis a methodology to an automation of the metadata creation and actualization processes based on a descriptive information provided by OGC services capabilities models and specific information included in the domain specific metadata profiles. We have performed practical experiment in order to apply proposed methodology using a functionality of geocatalogue GeoNetwork called *harvesting*, which implements metadata crosswalks between OGC capabilities and ISO metadata models. To ensure interoperability level defined by the legislation (INSPIRE), related standardization (ISO) and domain specific user requirements, an additional transformation was designed and developed in order to fulfil all these requirements. Specific output of this work can be used to perform a metadata transformation for the resources of the basic database of GIS (ZB GIS) of Slovak republic.

The second part of the thesis proposed a methodology to a discovery of OGC services in Google search engine web page and retrieval of descriptive information into the GeoNetwork. Practical experiment based on proposed methodology has been designed and performed. The results of the experiment showed that the mainstream web provides today a wide range of geospatial information in a form of OGC services and related resources (WMS Layers, WFS Features, WCS Coverages, SOS Observations, etc.). 2654 GetCapabilities URL for current types of OGC services have been discovered on Google after three searches. The descriptive information (metadata) has been harvested from the discovered OGC services for both services and resources they provide. More than 142 thousand metadata records have been discovered by the date of 22nd September, 2012. The number of metadata records may vary according to the results of the harvesting tasks that have failed at their first run, and therefore are being repeated till they succeed. All geospatial metadata are available for discovery and further analysis of the geospatial information they describe and provide through the discovery client interface of GeoNetwork instance running on <http://tokenbros.com:8081/geonetwork>, or discovery service endpoint on address <http://tokenbros.com:8081/geonetwork/srv/eng/csw>.

9 Zoznam použitej literatúry

- ALAMEH, N. S. (2001).** Scalable and extensible infrastructure for distributing interoperable geographic information services on the Internet. Cambridge, MA: Massachusetts Institute of Technology.
- AL-MASRI, E., MAHMOUD, Q. (2008).** Investigating Web Services on the World Wide Web. In: Proceedings of the 17th international conference on World Wide Web. WWW'08. ACM, New York, NY, USA, pp. 795–804. Dostupné na WWW: <http://dx.doi.org/10.1145/1367497.1367605>.
- BATCHELLER, J., (2008).** Automating geospatial metadata generation – An integrated data management and documentation approach. *Computers & Geosciences*, 34: 287 - 398.
- BERGMAN, M. K. (2001).** The Deep Web: Surfacing Hidden Value. *Journal of Electronic Publishing* 7 (1). Dostupné na WWW: <http://dx.doi.org/10.3998/3336451.0007.104>
- FLORCZYK, A. J. (2012).** Search improvement within the geospatial web in the context of spatial data infrastructures. PhD dissertation, Computer Science and Systems Engineering Department, Universidad de Zaragoza, 2012.
- GREEN, D. (2000).** The evolution of Web searching. *Online Information Review* 24 (2), 124–137. Dostupné na WWW: <http://dx.doi.org/10.1108/14684520010330283>.
- INSPIRE (2007).** SMERNICA EURÓPSKEHO PARLAMENTU A RADY 2007/2/ES zo 14. marca 2007, ktorou sa zriaďuje Infraštruktúra pre priestorové informácie v Európskom spoločenstve (Inspire).2007.
- INSPIRE (2008).** NARIADENIE KOMISIE (ES) č. 1205/2008 ktorým sa vykonáva smernica Európskeho parlamentu a Rady 2007/2/ES, pokiaľ ide o metaúdaje. 2008.
- INSPIRE (2010).** INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119. 2010
- INTHIRAN, A., ALHASHMI, S. M., AHMED, P. K. (2010).** A Reflection of Search Engine Strategies. *Communications of the IBIMA*. Dostupné na WWW: <http://dx.doi.org/10.5171/2010.126850>.
- ISO (2003).** ISO 19115:2003 Geographic information - Metadata. ISO, Switzerland, 2003
- ISO (2005b).** ISO 19128:2005 Geographic information -- Web map server interface. ISO, Switzerland, 2005.
- ISO (2007a).** ISO/TS 19139:2007 Geographic information -- Metadata -- XML schema implementation. ISO, Switzerland, 2007.
- ISO (2009).** ISO 15836:2009 - Information and documentation – The Dublin Core metadata element set. ISO, Switzerland, 2009.
- ISO (2010).** ISO 19142:2010 - Geographic information -- Web Feature Service. ISO Switzerland 2010.
- KLIMENT, T., OGGIONI, A. (2011).** Bringing Eco Biological Metadata to the INSPIRE metainformation world. In: Inter-Regional GG2011-X-BORDER-SDI/GDI Symposium: 23.-24.5.2011, Děčín, Česká republika. -: European Commission, 2011. - ISBN 1801-6480. - S. 52.
- KLIMENT, T., TUCHYŇA, M., KLIMENT, M. (2012).** Methodology of conformance testing of spatial data infrastructure. In. *Slovak journal of Civil Engineering*. ISSN 1210-3896. - Vol. XX, 2012, No. 1, 10 – 20, s. 10-20.
- LARSON, R.R. (1996).** Geographic information retrieval and spatial browsing. In L. C. Smith and M. Gluck, editors, *Geographic information systems and libraries: patrons, maps, and spatial information*, pages 81–124, 1996.
- LI, W., YANG, CHA., YANG, CHO. (2010).** An active crawler for discovering geospatial Web services and their distribution pattern – A case study of OGC Web Map Service. In *International Journal of Geographical Information Science*, 24:8, 1127-1147.
- MICHENER, W.K., BRUNT, J.W., HELLY, J.J., KIRCHNER, T.B., STAFFORD, S.G. (1997).** Non geospatial Metadata for the Ecological Sciences. *Ecological Applications* 7(1):330-342.
- NEBERT D. (2009).** The SDI Cookbook. Technical report, Global Spatial Data Infrastructure, 2009. URL: http://memberservices.gsdi.org/files/?artifact_id=655

- NEBERT D., WHITESIDE A., VRETANOS, P. (2007).** Open GIS Catalogue Services Specification. OpenGIS Publicly Available Standard OGC-07-006r1, Open GIS Consortium Inc., February 2007. Version 2.0.2.
- NOGUERAS-ISO, J., ZARAZAGA-SORIA, F., AND MURO-MEDRANO, P.R. (2005).** Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and Information Retrieval. Springer.
- PELLICER, L., F., J. (2011).** Semantic Linkage of the Invisible Geospatial Web. PhD thesis, Universidad de Zaragoza, 2011.
- POWELL, A., NILSSON, M., NAEVE, A., JOHNSTON, P., BAKER, T. (2007).** DCMI Abstract Model. Dostupné: URL <http://dublincore.org/documents/2007/04/02/abstract-model/>
- PRICE, G., SHERMAN, C. (2001).** The Invisible Web: Uncovering Information Sources Search Engines Can't See. Information Today, Inc. Dostupné na WWW: <http://cdigital.uv.mx/bitstream/123456789/10046/2/Invisible%20web.pdf>.
- SHVAIKO, P., IVANYUKOVICH, A., VACCARI, L., MALTESE, V., FARAZI, F. (2010).** A semantic geo-catalogue implementation for a regional SDI. INSPIRE Conference 2010, Krakow, Poland.
- ÚGKK. (2008).** KATALÓG TRIED OBJEKTOV ZB GIS, verzia 2008. ÚGKK SR, TOPÚ Banská Bystrica. 2008. Dostupné na WWW: http://www.skgeodesy.sk/index.php?www=sp_file&id_item=396
- ZÁKON (1995).** Z Á K O N NÁRODNEJ RADY SLOVENSKEJ REPUBLIKY z 12. septembra 1995 o geodézii a kartografii. Dostupné na WWW: <http://katasternehnutelnosti.sk/106-zakon-2151995-o-geodezii-a-kartografii>
- ZHAO, P., CHEN, A., LIU, Y., DI, L., YANG, W., LI, P. (2004).** Grid metadata catalog service-based OGC Web registry service. 12th ACM International Workshop on Geographic Information Systems (ACM-GIS 2004), Washington DC, USA.
- ZHAO, P., YU G., DI, L. (2007).** Geospatial Web Services.

10 Zoznam publikovaných prác doktoranda

- KLIMENT, T. - ĎURAČIOVÁ, R.:** Vyhľadávanie geodát v rámci infraštruktúry priestorových informácií. In: Geodetický a kartografický obzor. - ISSN 0016-7096. - Roč. 54 (96), č. 11 (2008), s. 229-235.
- KLIMENT, T.:** Katalógové služby v doméne metadát pre geodáta. In: Juniorstav 2008: 10.Odborná konferencia doktorského studia.Brno,23.1.2008. - Brno: Vysoké učení technické v Brně, 2008. - ISBN 978-80-86433-45-5. - S. 369.
- KLIMENT, T.:** Possibilities of geometadata distribution between open source and commercial GI applications. In: GI 2008: Symposium Inter-regional.8th Saxonian GIS-Forum.Proceedings.Dresden,15.-16.5.2008. - Dresden: IGN, 2008. - S. 121.
- KLIMENT, T.:** Súčasný stav metadátovej vybavenosti rezortu Geodézie a kartografie v Slovenskej republike. In: XII.mezinárodní vědecká konference.Brno,20.-22.4.2009: sekce 11:Geodézie a kartografie. - Brno: CERM, 2009. - ISBN 978-80-7204-629-4. - S. 53 – 56.
- KLIMENT, T.:** Použitie OPEN SOURCE nástrojov pre implementáciu metainformačnej zložky národnej infraštruktúry pre priestorové informácie. In: ENVIRO I FÓRUM 2009: Odborné fórum o environmentálnej informatike. Zborník prezentácií 5.ročník.Zvolen,SR,9.-11.6.2009. - Bratislava: Slovenská agentúra životného prostredia, 2009. - ISBN 978-80-88850-87-8. - S. 37 – 42.
- KLIMENT, T.:** Možnosti vyhľadávania metainformácií priestorových dát prostredníctvom WEBU na Slovensku. In: Juniorstav 2009: 11. Odborná konferencia doktorského studia.Brno,4.2.2009. - Brno: Vysoké učení technické v Brně, 2009. - ISBN 978-80-214-3810-1. - S. 334.
- KLIMENT, T.:** Metainformation infrastructure for geospatial information. In: Slovak Journal of Civil Engineering. - ISSN 1210-3896. - Vol. 18, No. 2 (2010), s. 19-25.
- GÁLOVÁ, L. - KLIMENT, T.:** Harmonizácia údajov - základný prvok interoperability v GIS krízového manažmentu. In: XVI. medzinárodné slovensko-poľsko-české geodetické dni: Zborník anotácií. Tatranská Lomnica,SR,13.-15.5.2010. - Bratislava: Slovenská spoločnosť geodetov a kartografov, 2010. - ISBN 978-80-969692-5-8. - nestr.
- KLIMENT, T. - KLIMENT, M.:** Implementácia STN EN ISO 19115 v projektoch PÚ. In: Veda mladých 2007: Zborník z V. vedeckej konferencie s medzinárodnou účasťou.Nitra,SR,19.11.2007. - Nitra: SPU v Nitre, 2010. - ISBN 978-80-552-0370-6. - S. 74-80.
- KLIMENT, T.:** Návrh a pilotná realizácia testovacieho procesu pre vyhľadávaciu službu podľa požiadaviek INSPIRE. In: ENVIRO I FÓRUM 2010: Odborné fórum o environmentálnej informatike.6.ročník podujatia.Zvolen,SR,8.-9.6.2010. - , 2010. - S. 38-42.
- GÁLOVÁ, L. - KLIMENT, T.:** Harmonizácia údajov - základný prvok interoperability v GIS krízového manažmentu. In: XVI. medzinárodné slovensko-poľsko-české geodetické dni: Zborník anotácií. Tatranská Lomnica,SR,13.-15.5.2010. - Bratislava: Slovenská spoločnosť geodetov a kartografov, 2010. - ISBN 978-80-969692-5-8. - S. 35-36.
- KLIMENT, T.:** Návrh a pilotná realizácia testovacieho procesu pre vyhľadávaciu službu podľa požiadaviek INSPIRE. In: ENVIRO I FÓRUM 2010: Odborné fórum o environmentálnej informatike.6.ročník podujatia.Zvolen,SR,8.-9.6.2010. - , 2010. - S. 23.
- GÁLOVÁ, L. - KLIMENT, T.:** Heterogénne dáta v povodňovom krízovom manažmente. In: Geodetická a kartografická podpora protipovodňovej ochrany: Pedagogické listy,zošit 16/2010. - Bratislava: STU v Bratislave SvF, 2010. - ISBN 978-80-227-3424-0. - S. 40-50.
- KLIMENT, T. - GÁLOVÁ, L.:** Vyhľadávanie, hodnotenie a používanie povodňových dát v infraštruktúre priestorových informácií. In: Geodetická a kartografická podpora protipovodňovej ochrany: Pedagogické listy,zošit 16/2010. - Bratislava: STU v Bratislave SvF, 2010. - ISBN 978-80-227-3424-0. - S. 64-73.
- FENCÍK, R. - KLIMENT, T. - TUCHYŇA, M.:** Current state of building the spatial data infrastructure in the Slovak republic. 2010. - s.30-33 s. (<http://GDI-SN.blogspot.com>).
- KLIMENT, T.:** Discovery service testing according to INSPIRE implementing rules (poster). 2010. (<http://bizbiz.ccscz.cz/storage/95/files/GI2010-Symposium-Kliment.pdf>).
- VACCARI, L., ANNONI, A., ANGELIS, C., KLIMENT, T., MORELLI, F.** GEOSS testing procedure Phase1: Testing report. Technical report of the results from Phase 1 of the testing procedure related to the

- GEOSS Component and Seervice Registry and GEOSS Clearinghouses, 2010, Joint Research Centre in Ispra, Italy.
- VACCARI, L. - KLIMENT, T. - ANNONI, A. - ANGELIS, C. - PRADO, H. - CRAGLIA, M.** GEOSS testing procedure Phase2: Testing report. Technical report of the results from Phase 2 of the testing procedure related to the GEOSS Geo Web Portals, 2010, Joint Research Centre in Ispra, Italy.
- GÁLOVÁ, L. - KLIMENT, T.:** Návrh a realizácia používateľského scenára pre mapy povodňového ohrozenia v SDI. In: Juniorstav 2011: 13. Odborná konferencia doktorského studia.Brno,VUT,4.2.2011. - Brno: Vysoké učení technické v Brně - Fakulta stavební, 2011. - ISBN 978-80-214-4232-0. - nestr.
- KLIMENT, T. - CIBULKA, D.:** Testovanie vyhľadávacích a zobrazovacích služieb podľa INSPIRE požiadaviek: Zborník. In: GIS.Ostrava 2011: Sborník symposia.Ostrava,ČR,23.-26.1.2011. - Ostrava: VŠB-Technická univerzita Ostrava, 2011. - ISBN 978-80-248-2366-9. - nestr.
- KLIMENT, T. - ĎURAČIOVÁ, R. - GÁLOVÁ, L. - CIBULKA, D. - TUCHYŇA, M.:** Theoretical aspects and open issues in SDI establishment within geodetic and cadastral domain at the Slovak University of Technology. In: GIS.Ostrava 2011: Sborník symposia.Ostrava,ČR,23.-26.1.2011. - Ostrava: VŠB-Technická univerzita Ostrava, 2011. - ISBN 978-80-248-2366-9. - nestr.
- KLIMENT, M. - FRAŠTIA, M. - KLIMENT, T.:** Zber údajov pre účelové mapovanie nemetrickou digitálnou kamerou. In: Kartografia a geoinformatika vo svetle dneška: 19. kartografická konferencia.Bratislava, SR, 8.-9.9.2011. - Bratislava: Kartografická spoločnosť SR, 2011. - ISBN 978-80-89060-19-1. - S. 92-101.
- KLIMENT, T. - OGGIONI, A.:** Bringing Eco Biological Metadata to the INSPIRE metainformation world. In: Inter-Regional GG2011-X-BORDER-SDI/GDI Symposium: 23.-24.5.2011, Děčín, Česká republika. -: European Commission, 2011. - ISBN 1801-6480. - S. 52.
- KLIMENT, T. - TUCHYŇA, M. - KLIMENT, M.:** Testing of SDI components a fundamental interoperability element within INSPIRE and national SDI's (abstract). In: Inter-Regional GG2011-X-BORDER-SDI/GDI Symposium: 23.-24.5.2011, Děčín, Česká republika. -: European Commission, 2011. - ISBN 1801-6480. - S. 51
- KLIMENT, T. - TUCHYŇA, M. - OGGIONI, A. - CARRARA, P. - PUGNETTI, A.:** EnvEurope&INSPIRE use case for collaboration within the EU projects and legislation driven initiatives (abstrakt). In: ENVIRO I FÓRUM 2011: Odborné fórum o environmentálnej informatike.7.ročník podujatia.Zvolen,SR,7.-8.6.2011. – Banská Bystrica: Slovenská agentúra životného prostredia, 2011. - ISBN 978-80-89503-17-9. - S. 15-16
- KLIMENT, T. - CIBULKA, D. - TUCHYŇA, M. - KOŠKA, M. - MOZOLÍK, P. - TÓBIK, J.:** Príklad spolupráce akademického a verejného sektora pri testovaní komponentov infraštruktúr priestorových informácií (abstrakt). In: ENVIRO I FÓRUM 2011: Odborné fórum o environmentálnej informatike.7.ročník podujatia.Zvolen,SR,7.-8.6.2011. – Banská Bystrica: Slovenská agentúra životného prostredia, 2011. - ISBN 978-80-89503-17-9. - S. 34-35
- DANEK, L. - GÁLOVÁ, L. - KLIMENT, T. - FENCÍK, R.:** Model webového povodňového portálu pre mestá a obce. In: Acta hydrologica Slovaca. - ISSN 1335-6291. - Roč.13, č. 1 (2012), s. 186-195.
- KLIMENT, M. - FRAŠTIA, M. - KLIMENT, T.:** Priestorové mapovanie "low cost" fotogrametriou pre projektovanie v krajine. In: Acta horticulturae et regiecturae. - ISSN 1335-2563. - Roč.15, mimoriadne číslo (2012), s. 23-26
- KLIMENT, T. – TUCHYŇA, M. – KLIMENT, M.:** Methodology for conformance testing of spatial data infrastructure components including an example of its implementation in Slovakia. In: Slovak Journal of Civil Engineering. - ISSN 1210-3896. - Vol. 20, No. 1 (2012), s. 10-20.
- TUCHYŇA, M. - KLIMENT, M.:** INSPIRE z implementácie do operačnej prevádzky (abstrakt). In: ENVIRO I FÓRUM 2012: Odborné fórum o environmentálnej informatike.8.ročník podujatia.Zvolen,SR,12.-13.6.2012. – Banská Bystrica: Slovenská agentúra životného prostredia, 2012. - ISBN 978-80-89503-21-6. - S. 32.
- KLIMENT, M. - TUCHYŇA, M. - CIBULKA, D. – KLIMENT, M.:** Testovanie ako predpoklad dosiahnutia interoperability prostredníctvom štandardizácie v oblasti geopriestorových informácií (abstrakt). In: ENVIRO I FÓRUM 2012: Odborné fórum o environmentálnej informatike.8.ročník podujatia.Zvolen,SR,12.-13.6.2012. – Banská Bystrica: Slovenská agentúra životného prostredia, 2012. - ISBN 978-80-89503-21-6. - S. 39.