

***ENVIRONMENTÁLNE GEOPRIESTOROVÉ INFORMÁCIE ZÍSKANÉ V
PROSTREDÍ „MAINSTREAM“ WEBU
ENVIRONMENTAL GEOSPATIAL INFORMATION RETRIEVED FROM
THE MAINSTREAM WEB***

KLIMENT, Tomáš¹, KLIMENT, Marcel²,

¹Slovenská technická univerzita, Stavebná fakulta, Katedra geodetických základov, email:
tomas.kliment@stuba.sk

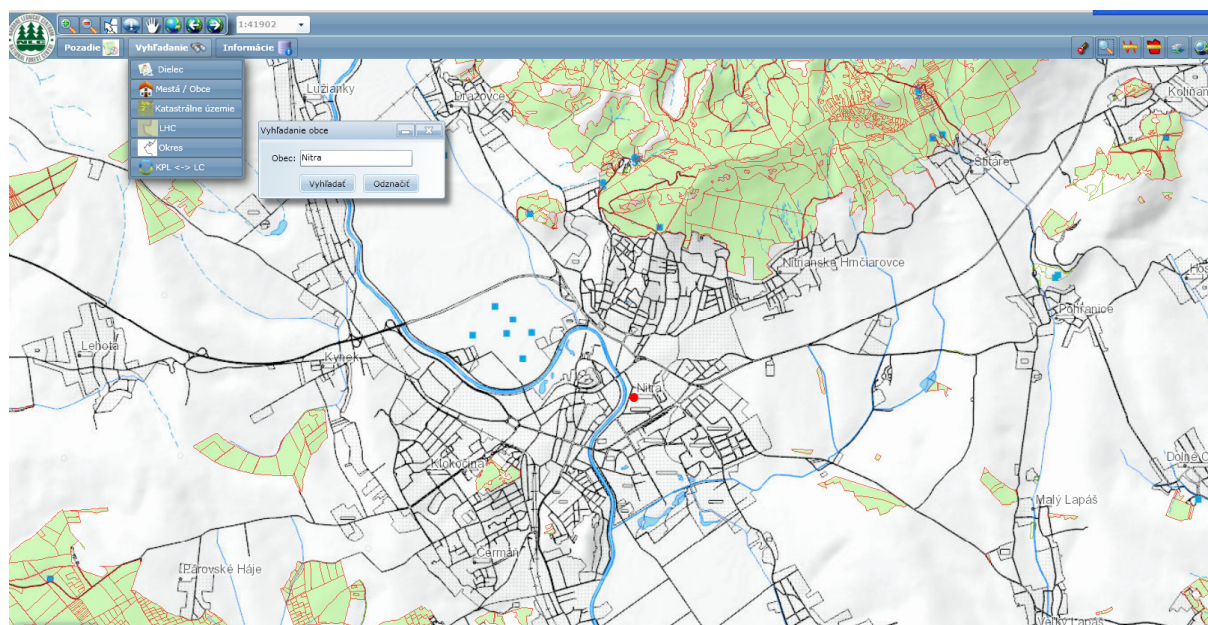
²Slovenská poľnohospodárska univerzita, Fakulta záhradníctva a krajinného inžinierstva,
Katedra krajinného plánovania a pozemkových úprav, email: marcel.kliment@uniag.sk

ABSTRAKT

Geopriestorový web môžeme definovať ako kolekciu webových služieb, geopriestorových dát a metadát, ktoré napomáhajú pri používaní geopriestorových dát ich popisom v rozsahu aplikácie do praxe v rámci konkrétnej domény (krajinné plánovanie, ochrana životného prostredia, obnoviteľné zdroje energie a iné). Vyhľadávanie špecifickej informácie je dôležitá funkcia, ktorú musí akýkoľvek informačný systém podporovať (Florczyk, 2012). Pre geopriestorové informácie (GI) vo webovom prostredí sa v súčasnej dobe na vyhľadávanie ich 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 vo väčšine prípadov preddefinovanú štruktúru. Takto dochádza často k duplicitě informácií, ktoré sú už čiastočne obsiahnuté v pôvodných zdrojoch GI (geopriestorových dátach a službách). Alternatívnym riešením ku geokatalógom môžu byť za určitých podmienok internetové vyhľadávače. Nevyžadujú žiadnu prácu navyše a ako ukazujú výskumy (Kliment, 2012), pri vhodne zvolenej metodike vo významnej miere, poskytujú možnosti na vyhľadanie zdrojov GI na globálnej úrovni a z rôznych domén. Článok popisuje metodiku, ktorá bola použitá na vyhľadávanie zdrojov GI v prostredí internetového vyhľadávača Google a ďalšie kroky aplikované na spracovanie výsledkov vyhľadávania a ich prezentácie v prostredí internetu pomocou geokatalógu GeoNetwork. Na ukážku praktického využitia výsledkov výskumu v praxi sme realizovali tematické vyhľadávanie GI pre doménu krajinného inžinierstva. Oblasť tzv. „mainstream“ webu prezentovaná používateľovi pomocou internetových vyhľadávačov poskytuje veľké bohatstvo informácií, medzi ktoré patria GI. Problémom ešte stále je spôsob, akým tieto informácie vyhľadávať a poskytovať tak, aby boli jednoducho použiteľné pre kohokoľvek, kto potrebuje pracovať s GI.

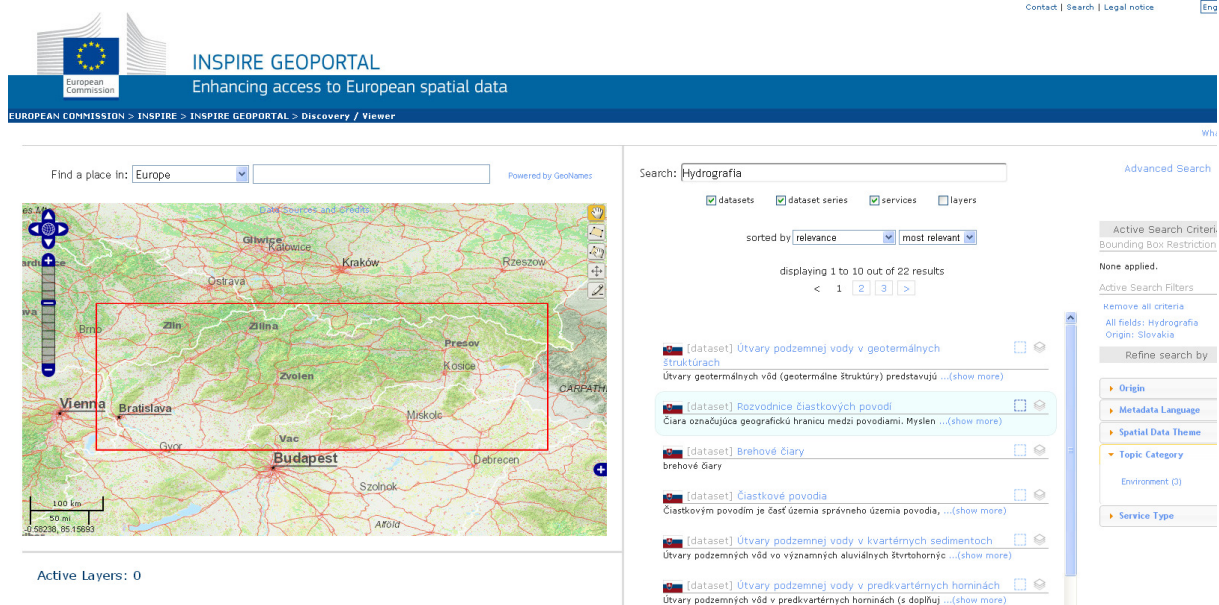
ÚVOD

Geopriestorový web poskytuje v súčasnosti rôzne metódy ako vyhľadávať a pristupovať ku geopriestorovým dátam. Zadefinované sú tri hlavné metódy a to on-line GIS, (IPI) a systémy priestorového prezerania (Pellicer, 2011). On-line 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.



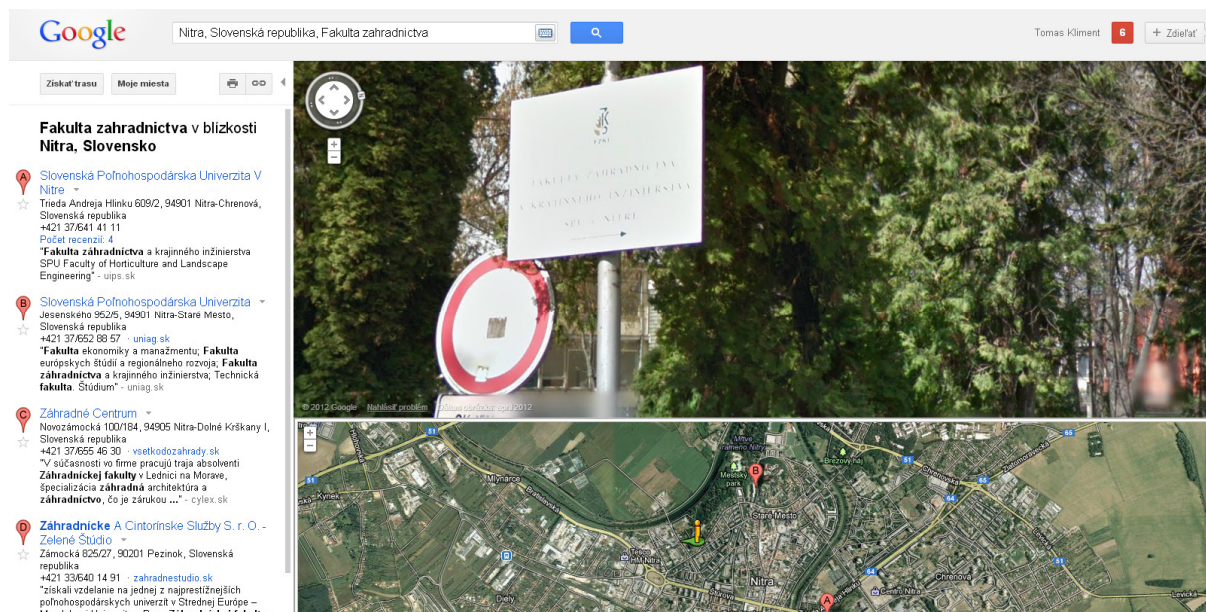
Obr. 1 Vyhľadávanie GI v rámci on-line GIS LGIS (<http://lvu.nlcsk.org/LGIS/>)

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).



Obr. 2 Vyhľadávanie GI v rámci INSPIRE Geoportálu – vstupnej brány do Európskej IPI
(<http://inspire-geoportal.ec.europa.eu>)

Napriek vyššie popísaným skutočnostiam, keď bežný používateľ prehliada web a hľadá informácie, ktoré majú mať priestorovú lokalizáciu, ako napr. adresa fakulty záhradníctva SOU v Nitre, tak nepoužije ani on-line 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 funkcionality 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.



Obr. 3 Vyhľadávanie GI v rámci aplikácie na priestorové prezeranie Google Maps (<https://maps.google.com/>)

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 (Schleidt et al., 2010).

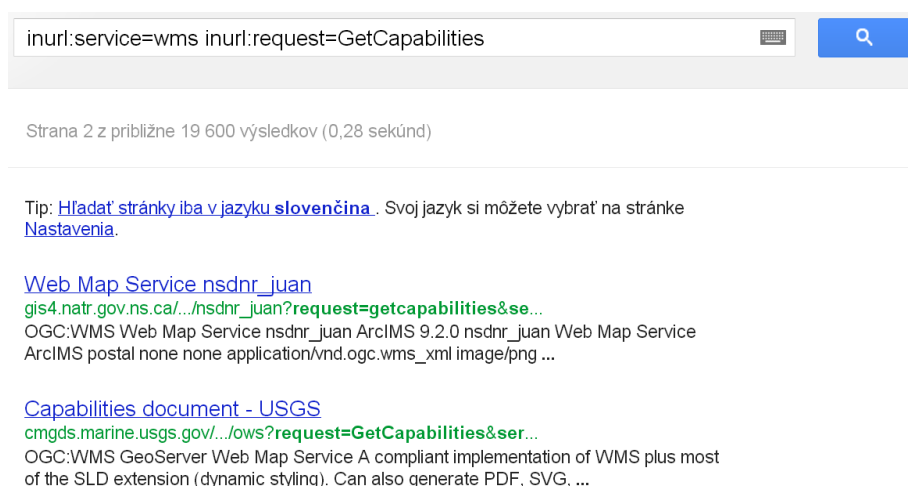
MATERIÁL A METODIKA

Na vyhľadávanie zdrojov GI sme použili webového klienta internetového vyhľadávača Google. Vyhľadávali sme zdroje GI poskytované geopriestorovými webovými službami, ktoré sú definované otvorenými štandardmi vyvíjanými organizáciou OGC (Open Geospatial Consortium). Pre nasledovné typy OGC služieb sme definovali vyhľadávacie kritéria:

- webové mapové služby (WMS) na zobrazovanie dát vo forme rastrových máp,
- webové objektové služby (WFS) na ukladanie dát,
- služby dátových pokrytí (WCS) na ukladanie pokrytí dát,
- služba spracovania dát (WPS) na spracovanie dát a spúšťanie iných služieb,
- služby pozorovaní (SOS) na poskytovanie dát priamo z meraní,
- katalógové služby (CSW) na vyhľadávanie metadát o dátach.

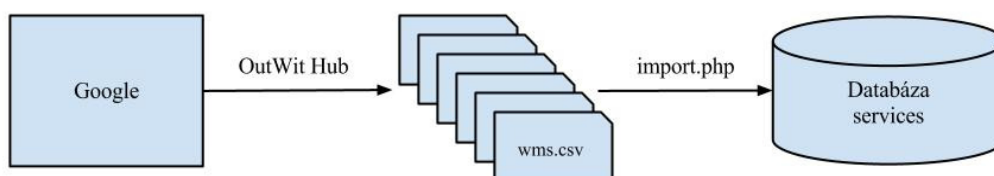
Navrhnutá metodika vyhľadávania, spracovania a poskytovania GI pozostávala z nasledovných krokov:

1. Vyhľadavanie URL adries na OGC webové služby v internetovom vyhľadávači Google – vyhľadávali sme pomocou pokročilého operátora inurl, ktorý umožňuje filtrovať výsledky tak, že vyhľadávaný výraz nachádza v URL adrese výsledku. Vyhľadávané výrazy pozostávali z definície typu služby service=WxS a definície typu požiadavky na službu. My sme použili požiadavku GetCapabilities, ktorá je povinná pre všetky typy OGC služieb a na ktorú služba odpovedá poskytnutím popisných informácií (metadát) o službe a dátach s ktorými operuje. Na obrázku 4 je príklad definície požiadavky na vyhľadanie OGC služieb typu WMS.



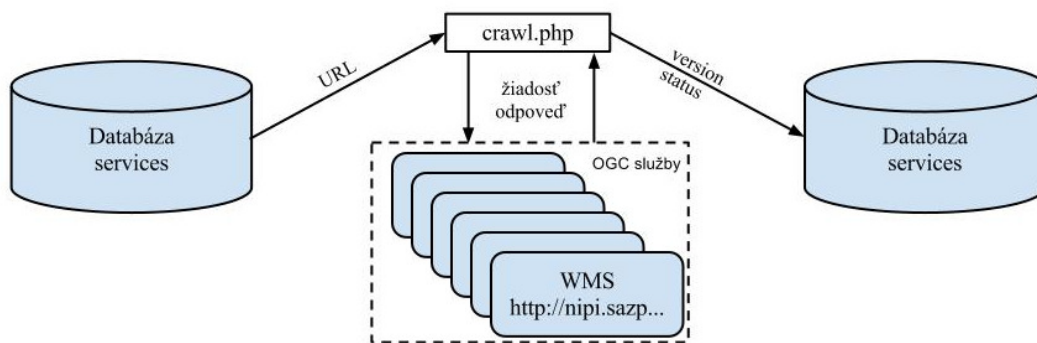
Obr. 4 Vyhľadavanie GetCapabilities URL adries na WMS služby v Google
(<https://www.google.com>)

2. Uloženie výsledkov vyhľadávania do databázy – na extrahovanie výsledkov z Google sme použili voľne dostupnú verziu programu OutWit Hub, pomocou ktorej sme vytvorili súbory typu CSV pre jednotlivé typy služieb, ktoré obsahovali informácie o URL adrese služby, popise a názve. Výsledky sme potom uložili do databázy MySQL pomocou skriptu import.php, ktorý sme naprogramovali v jazyku PHP.



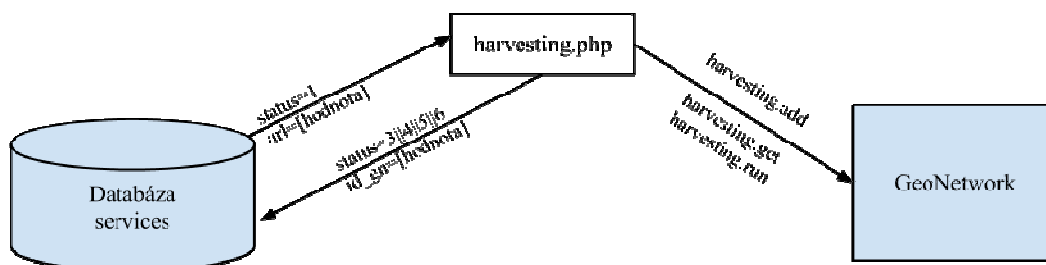
Obr. 5 Schematické znázornenie metodiky - krok č.2

- Overenie funkčnosti nájdených služieb a získanie ich verzií – na overenie funkčnosti a zistenie verzie nájdených služieb sme naprogramovali skript `crawl.php` v jazyku PHP, ktorý postupne zasielal žiadosti a extrahoval informácie o ich verzii z odpovede. Ak sa verzia zistila, tak ju zapísal v databáze do stĺpca `version`, a hodnotu 1 do stĺpca `status` (služba je funkčná a pripravená na ďalšie spracovanie). U služieb, pri ktorých došlo k zlyhaniu, t.j. verziu nebolo možné zistiť, skript zapísal hodnotu 2 v stĺpci `status` a stĺpec `version` zostal prázdny.



Obr. 6 Schematické znázornenie metodiky - krok č.3

- Vytvorenie a spustenie tzv. harvestovacích úloh pre funkčné služby v geokatalógu GeoNetwork – na extrahovanie popisných informácií k nájdeným službám sme použili funkcionality geokatalógu GeoNetwork, ktorý umožňuje vytvoriť tzv. harvestovacie úlohy pre jednotlivé typy OGC služieb. Po spustení sa extrahujú metadáta k službe a dátam s ktorými pracuje. Na automatické vytvorenie harvestovacích úloh pre nájdené služby sme naprogramovali skript `harvesting.php`, ktorý postupne vyberá z databázy služby, ktoré majú hodnotu 1 v stĺpci `status` a spolu s hodnotou ich verzie prebratí zo stĺpca `version` vytvára a spúšťa tieto úlohy v programe GeoNetwork.



Obr. 7 Schematické znázornenie metodiky - krok č.4

VÝSLEDKY A DISKUSIA

Realizovali sme 3 etapy vyhľadávania OGC služieb podľa navrhnutej metodiky. Výsledky sú zosumarizované v tabuľke 1.

Tab. 1 Tabuľka 1 Nájdené OGC služby

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 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). 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. Tieto výsledky potvrdzujú to, čo aj dostupná literatúra, že geopriestorové webové

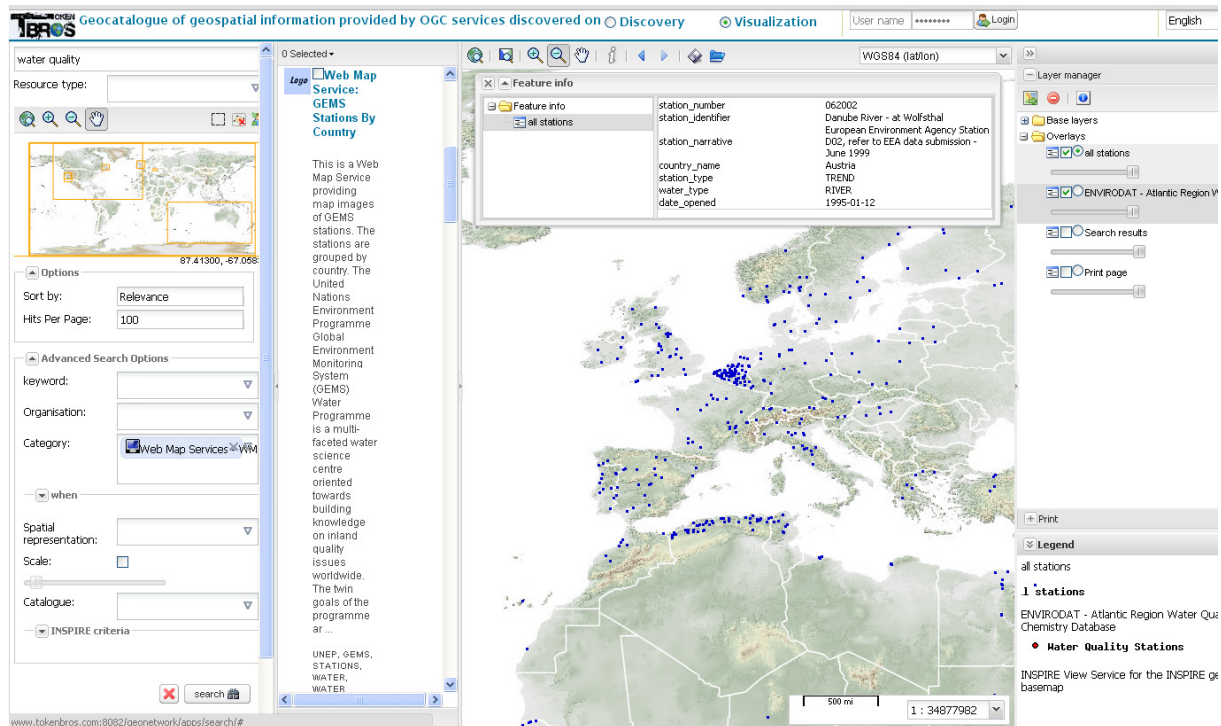
služby je už v súčasnosti možné nájsť pomocou internetových vyhľadávačov. Porovnaním našich výsledkov s výsledkami výskumu popísané v Pellicer (2011), zistíme, že sme našli takmer o 1 000 OGC služieb menej. Jednou príčinou je, že my sme použili len jeden internetový vyhľadávač a autori použili tri. Ďalšou 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.

ZÁVER

Metadáta (viac ako 225 tisíc záznamov) vytvorené k zdrojom GI nájdených v internetovom vyhľadávači Google vo forme OGC služieb sú prístupné cez webové rozhranie geokatalógu GeoNetwork na URL adrese: www.tokenbros.com:8082/geonetwork. Aplikácia môže byť použitá na vyhľadávanie zdrojov GI viacerými spôsobmi:

- jednoduché (Full text) vyhľadávanie v metadátach,
- rozšírené vyhľadávanie s možnosťou kategorizácie výsledkov podľa typu OGC služby a zdrojov GI vo formáte aký poskytujú,
- zobrazovanie dát z WMS služieb v stavanom mapovom prehliadači,
- prehliadanie popisných informácií poskytovaných metadátami,
- integrácia služieb do externých GIS klientov (ArcGIS, Quantum GIS, atď.)

Na jednoduché overenie funkcionality sme realizovali vyhľadavanie kľúčových slov „water quality“ a získali sme 2315 výsledkov. Pre potreby spresnenia výsledkov sme použili kategorizáciu a zvolili kategóriu Web Map Service a získali 12 výsledkov.



Obr. 8 Vyhľadanie a zobrazenie GI na mape v prostredí geokatalógu GeoNetwork

Obrázok 8 zobrazuje grafické rozdanie geokatalógu so zobrazením nájdených environmentálnych dát poskytovaných WMS službou s názvom GEMS Stations By Country. Služba poskytuje mapové zobrazenie globálnych environmentálnych monitorovacích staníc, v našom prípade staníc, ktoré monitorujú kvalitu vody. Takýmto spôsobom je možné vyhľadať množstvo environmentálnych GI, ktoré je možné použiť na prezentáciu prostredníctvom máp, ale aj na ďalšie analýzy a spracovanie.

POUŽITÁ LITERATÚRA

FLORCZYK, A. J. (2012). Search improvement within the geospatial web in the context of spatial data infrastructures. PhD dissertation, Zaragoza (Spain): Computer Science and Systems Engineering Department, Universidad de Zaragoza.

KLIMENT, T. (2012). Vyhľadávanie zdrojov geopriestorových informácií prostredníctvom internetu. Dizertačná práca. Bratislava: Stavebná fakulta STU v Bratislave.

NEBERT D. (2009). The SDI Cookbook. Technical report, Global Spatial Data Infrastructure, 2009. URL: http://memberservices.gsdi.org/files/?artifact_id=655

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.

PELLICER, L., F., J. (2011). Semantic Linkage of the Invisible Geospatial Web. PhD thesis, Universidad de Zaragoza, 2011.

PELLICER, L., F., J., FLORCZYK, J., A., BÉJAR, R., MURO-MEDRANO, P. R., ZARAZAGA-SORIA, F. (2011b). Discovering geographic web services in search engines. Online Information Review, Vol. 35 Iss: 6, pp.909 – 927.

SCHLEIDT, K., KOŠKA, M., MOZOLÍK, P., PRATSCHNER, B. (2010). INSPIRE Biodiversity Mashup. INSPIRE conference 2010. Krakow, Poľsko.

Internetové zdroje

<http://tokenbros.com:8082/geonetwork>

<http://lvu.nlcsk.org/LGIS/>

<http://inspire-geoportal.ec.europa.eu>

<https://maps.google.com/>

<http://www.opengeospatial.org/standards>

<http://www.opengeospatial.org/standards/wms>

<http://www.opengeospatial.org/standards/wfs>

<http://www.opengeospatial.org/standards/wcs>

<http://www.opengeospatial.org/standards/wps>

<http://www.opengeospatial.org/standards/sos>

<http://www.outwit.com/>

<http://www.geonetwork-opensource.org/>