

Západočeská univerzita v Plzni



Fakulta aplikovaných věd
Katedra matematiky

Semestrální práce z Počítačové kartografie

Vyhledávání geoprostorových informací pomocí katalogových služeb

Plzeň, 2011

Bc. Lenka Vondrašová

Obsah

Úvod	1
Geoprostorové informace	1
Principy a způsoby práce katalogových vyhledávačů	1
Problémy a možná řešení ve vyhledávání geoprostorových informací.....	2
Závěr	4
Literatura a internetové odkazy	5

Úvod

V semestrální práci se budu zabývat vyhledáváním geoprostorových informací pomocí katalogových služeb. Níže uvedu stručnou historii internetových vyhledávačů a rozdíl mezi fulltextovým a katalogovým vyhledávačem. Kapitola „Geoprosotorové informace“ je rozdělena do dvou částí. V první části „Principy a způsoby práce katalogových vyhledávačů“ detailněji popíšu práci katalogových vyhledávačů a v druhé části „Problémy a možná řešení ve vyhledávání geoprostorových vyhledávačů“ se zmíním o možných řešeních na úrovni standardizace metadat a úrovni poskytování informací z katalogu (vyhledávání). V závěru shrnu všechny informace a zformuluji svůj názor na řešenou problematiku.

V druhé polovině 90. let 20. století se začaly vytvářet vyhledávače, které odpovídají jejich současné podobě. Tyto vyhledávače jsou však pro geoprostorové informace nedostačující. Většinou podporují tzv. fulltextové vyhledávání, které je jen jednou složkou geografického vyhledávání. Proto bylo třeba do vyhledávačů přidat možnost prostorového a časového vyhledávání. Další vývoj směřoval ke snaze o standardizaci vyhledávaných dat. Tu lze rozdělit do dvou úrovní: standardizace na úrovni poskytování informací z katalogu, standardizace na úrovni metadat. Pro katalogové služby jsou nezbytná metadata, která dodávají dotazovatelné prvky pro katalogové služby. Proto se často setkáváme s kombinací metadatového editoru a katalogu (např. u softwaru MICKA, GeoNetwotk). Podrobnější historii naleznete v literatuře [1].

V současné době lze informace na internetu hledat pomocí katalogových nebo fulltextových vyhledávačů. V praxi jsou katalogy a vyhledávače často částečně propojeny, ale jejich práce je v principu odlišná. Fulltextové vyhledávače hledají požadované informace na základě klíčového slova, které formuluje uživatel. Vyhledávač tedy prochází celý text webových stránek a pak vypíše seznam odkazů na stránky, které obsahují hledanou informaci. Katalogové služby umožňují uživateli strukturované vyhledávání podle určitých parametrů, jako jsou např. typy zdrojů, prostorový rozsah. Katalogové služby lépe vymezují množinu získaných údajů, neboť v katalogových vyhledávačích jsou shromažďovány odkazy na jiné webové stránky a portály spolu se stručnými popisy. Tematicky rozříděné odkazy do kategorií (např. cestování, sport) vedou na další nižší úrovně. Katalogové vyhledávače lze přirovnat ke kartotékám v knihovně nebo k telefonnímu seznamu. Tedy: „Katalogové služby umožňují publikovat a vyhledávat metadata dat příbuzných informačních modelů.“ Citace [1]

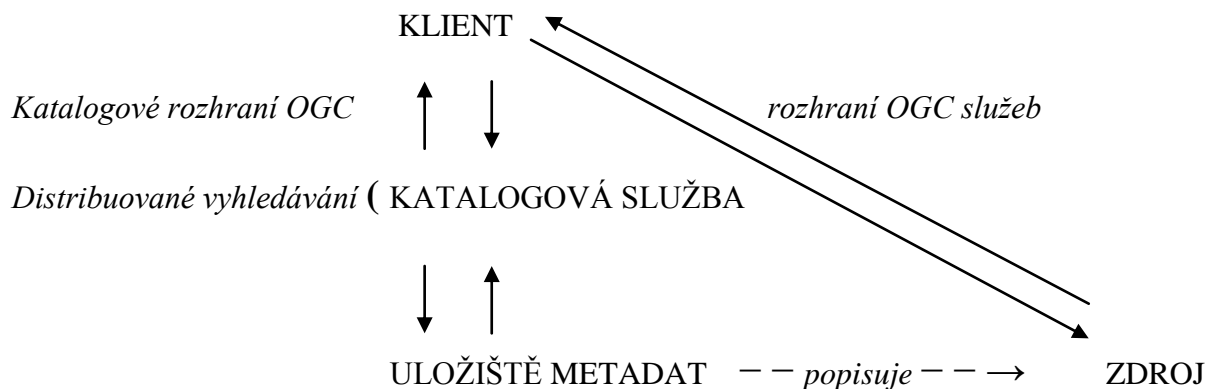
Geoprostorové informace

Principy způsoby práce katalogových vyhledávačů

Katalogové služby zprostředkovávají vztah mezi klienty a lokálním skladem metadat. Dochází k nepřímému vztahu mezi informačními zdroji a klienty. Také je možné zprostředkovat vztah mezi klienty a vzdálenými uložišti metadat - viz. Obr. 1.

Zjednodušenou architekturu na Obr. 1 je nutno uvažovat v mnohonásobném propojení klientů a serverů. Pomocí katalogového rozhraní OGC komunikuje klient s konkrétní katalogovou službou. Katalogová služba se smí ptát jen na tři zdroje: metadatové uložiště této katalogové služby, služby o zdroji (přes uložiště metadat) nebo na jinou katalogovou službu. Tedy každá katalogová služba může být samostatnou katalogovou službou poskytující požadované

odpovědi, ale také klientem dotazu na jinou katalogovou službu.



Obr. 1 – Architektura katalogových služeb podle OGC [4]

Jádro pro budování systémů tvoří podle [1] formální schéma pro geografická data, které vytvořila ISO Technical Committee 211. Schéma zahrnuje následující dokumenty: normu ISO 19115 (pro popis geografických dat), ISO Draft Technical Specification 19139 (pro definování formálního kódování a strukturu ISO metadat pro výměnu) a normu ISO 19119 (pro popis metadatových služeb). Podle [4] je standardizovaná distribuce výsledkem standardizovaných systémů a služeb, které bez problémů podporují interoperabilitu.

Problémy a možná řešení ve vyhledávání geoprostorových informací

Ve druhé kapitole [1] „Úvod metadatového popisu“ je uveden význam metadatových standardů, metadatové schéma pro popis služeb, základní komponenty prostorových infrastruktur a funkce metadat v nich. Následují ontologie pro podporu prostorových datových infrastruktur a shrnutí. V desáté kapitole [1] „Katalogové služby“ je probrána historie vyhledávačů, standardizace katalogových služeb. Dále jsou rozebrány katalogové služby podle OGC a stručné shrnutí. V těchto dvou částech je uveden důvod mísení geografických databází s odlišnými souřadnicovými systémy, měřítka, obsahem aj. - Dochází k tomu, protože geografické informace jsou vytvářeny různými skupinami kartografů, geografů, fotogrammetrů, geodetů, atd. Jako řešení uvádí, že pro dostupnost uživatelů ke geografickým databázím je třeba, aby geografická skupina měla konceptuální a metodické základy a plnila legislativní omezení, standardy databází včetně formátů pro jejich výměnu. Nicméně je nutné implementovat i metody, hardwarové a softwarové nástroje a techniky pro tvorbu a přenos metadatových sad k uživatelům. Dále se části zabývají kvalitním popisem dat, bez kterého dochází ke ztrátě jejich hodnoty. Nekompatibilitu geografických datových sad provází různá struktura. Pokusy o globální o interoperabilitu probíhají (např. International Organization for Standardisation - ISO, Dublin Core Metadata Initiative - DCMI). V roce 2003 byl dovršen proces tvorby metadatového standardu pro jednotný popis všech zdrojů vyskytujících se na internetu tím, že standard DCMI se stal ISO normou. Vyhledávání a získávání metadat z datových zdrojů uvnitř i mimo oblast knihovnictví tlačí organizace k uvažování o implementaci popisu podle Dublin Core. Komunikačními rozhraními pro publikaci popisů geografických dat (metadat) ve standardizované podobě jsou katalogy geografických dat. Nejvýznamnější specifikací pro tvorbu katalogů je OGC (Open Geospatial Consortium) verze 2.0.2. (Zatím není schválena jako řádný standard.) Také je zde vyzdvížena důležitost přístupu

ke geografickým datům (k vizualizaci a zhodnocení prostorových dat) a následně na žádost uživatele dat umožnit jejich distribuci i v komprimované formě. Jako druhý úhel pohledu práce nabízí to, že katalog geografických dat je příkladem služby, která spravuje a publikuje popisy geografických dat. Jako důsledek rozdílných formátů a systémů práce uvádí to, že se objevil problém rozdílnosti geografických databází. Hlavním cílem je dosáhnout interoperability. Základy pro syntaktickou interoperabilitu a kategorizaci geografických dat a informací poskytují standardy konsorcia OGC a Mezinárodní organizace pro standardy (ISO). Avšak využití jedné informace ve více kontextech vyvolává sématickou heterogenitu. Základem pro vyhledávání a sémantický překlad v prostorových datových infrastrukturách jsou ontologie. Základní způsoby využití ontologií patří: podpora porozumění mezi lidmi (experty a znalostními inženýry apod.), podpora vzájemné komunikace mezi počítačovými systémy, usnadnění návrhu znalostně-orientovaných aplikací. Metadata zde představují nenahraditelný základ spojující jednotlivé komponenty infrastruktur. Dále je zde zmíněn názor, že metadata mohou sloužit jako vstupy do katalogových služeb. Ontologie jsou podle [1] používány pro sdílení a znovuvyužití znalostí včetně jejich správného významu.

Ve čtvrté kapitole [2] „Rozdělení internetových vyhledávačů“ je provedeno nejprve obecné rozdělení a popis katalogových, fulltextových vyhledávačů a metavyhledávačů. Následně kapitola analyzuje konkrétní vyhledávače: Google, Jyxo a Yippy (metavyhledávač). Je zde vyzdvížena rychlost a efektivnost vyhledávání Googlu a pochválena aplikace Google Scholar. Jyxo je doporučeno užívat pro rozsáhlé množství dat. Jeho předností je české skloňování a časování. Yippy vyhledává a automaticky třídí nalezené odkazy do kategorií podle obsahu. V této kapitole chybí jakékoli porovnání obecně vyhledávačů a konkrétních produktů. Obsahem a strukturou informačních zdrojů se práce nezabývá. Není zde ani nastíněno možné řešení ve vyhledávání informací.

V práci [3] je řešena otázka významových rozdílů při vyhledávání geoprostorových informací. Rozdílnost ve významu (sémantická heterogenita) během vyhledávání může být podle [3] překonána ontologickými popisy informačních zdrojů. To znamená, že sémantika obsahu informačních zdrojů je strojově interpretovatelná a uživatelé budou moci klást dotazy stručně a výstižně. Text upozorňuje na řadu problémů, které je třeba řešit dříve, než ontologické přístupy dosáhnou širokého přijetí. Tedy je nejprve třeba podle [3] integrace ontologií do stávající infrastruktury prostorových dat, standardizace domény slovníků (ontologický popis), uživatelská podpora pro vytváření sémantického popisu geografických informací, skrytí ontologií a jejich užívání ze strany uživatele a podpora poloautomatického složení komplexního zpracování služeb. Text se nezabývá samotným vyhledáváním informací na internetu.

Ve [4] je stručně zmíněna funkce metadat. Následně se hovoří o standardizaci metadat podle již výše zmíněných norem. Poslední část seznamuje čtenáře s katalogovými službami a jejich standardizací. Na závěr příspěvku [4] je uvedena myšlenka, že interoperabilní systémy a služby jsou základem geoinformační struktury, na jejímž konci jsou dostupné geoinformace pro jakéhokoli uživatele a to v požadované kvalitě a kvantitě. Text nenabízí žádné nové řešení ve vyhledávání ani ve struktuře informací.

V příspěvku [5] se projednávají požadavky interoperability v distribuovaných geografických datech a službách, popisuje se stav techniky v oblasti výzkumu sémantické interoperability v infrastruktuře prostorových dat. Diskutuje se zde i o rámci pro sdílení různorodých geografických informací shromážděných z různých zdrojů. Příspěvek [5] se názorově prolíná s příspěvkem [6]. V příspěvku [6] je heterogenita v oblasti GIS definována jako rozdíly v

datové podobě (syntaktická heterogenita), rozdíly ve schématech (strukturální heterogenita) a rozdíly v zamýšleném významu výrazu ve specifických kontextech (sémantická heterogenita). Ke klíčovému rozšíření této práce patří diskuse o požadavcích a otázkách interoperability v distribuovaných geografických datech a popis sémantické integrace do geoslužeb. Dále tyto dvě práce uvádí, že společný rámec infrastruktury prostorových dat je založen na obecné softwarové platformě, která podporuje celou řadu geografických datových souborových typů, tak i komplexní nástroje pro správu dat, editaci, analýzu a vizualizace. Geografická data a geografické služby jsou zatíženy problémy s interoperabilitou. Jedna z klíčových služeb poskytovaná infrastrukturou prostorových dat je možnost získat geografické datové sady poskytované heterogenními zdroji. Otázky heterogenity jsou společné pro distribuovaný systém, ale geografické datové sady mají specifické vlastnosti, lišící se od ostatních typů dat. Vzhledem k tomu, že logická struktura infrastruktury prostorových dat je založena na souboru heterogenních datových zdrojů, heterogenní geografické informace musí být integrovány. Každý producent geodat přijímá vnitřní pravidla pro řízení jeho zeměpisných datových souborů. Různorodost na úrovni dat vzniká z více důvodů: různé výrobní postupy (různé geografické projekce, data získaná v různých měřítkách, data vyrobená za použití různých topografických zdrojů), různá reprezentace a rozlišení dat (např. silnice by měla být reprezentována pomocí polygonů nebo linií), různá schémata a ontologie. Heterogenita dat zahrnuje strukturální heterogenitu (schematickou různorodost) a sémantickou heterogenitu (datovou heterogenitu).

Závěr

Všechny zmíněné odborné texty se shodují v tom, že všechny problémy by mohly vyřešit pevně dané hranice pro tvorbu geografických dat a společný rámec pro infrastrukturu prostorových dat. Literatura [1] a [3] se shodují v aplikaci implementace ontologií pro popisy informačních zdrojů. Avšak příspěvek [3] vyzdvihuje myšlenku, že než ontologie dosáhnou širokého přijetí, je třeba integrovat ontologie do stávající infrastruktury prostorových dat, standardizovat doménu slovníků. V článkách [5] a [6] je vyzdvížena myšlenka, že nutnost sdílení informací dává podnět pro nastavení globální, mezinárodní, národní a místní infrastruktury pro sběr a šíření geografických dat. Heterogenita dat zahrnuje jak schematickou různorodost, tak datovou heterogenitu, které se musí řešit propojeně.

Myslím si, že obecně u všech geografických informací dochází ke ztrátě jejich hodnoty v závislosti na struktuře jejich metadat. Různá struktura a podrobnost stěžuje práci uživatelů v tom smyslu, že stejný popis dat (např. měřítko) pro různé zdroje najde v jiné části metadat. Také rozdílný formát geografických databází a systémů naráží na problémy ohledně zobrazení a následného užívání geografických informací. Uživatelé dat omezuje vybavenost jeho počítače (hardware, software) a schopnost se naučit více programů, ve kterých geografické informace může prohlížet, upravovat, mazat a vytvářet. Myslím si, že největší problémy nastávají při propojování geografických informací na mezinárodní úrovni. Zde se ke všem zmíněným problémům v textu přidávají i kartografické vlastnosti geografických informací jako je měřítko, podrobnost mapy, souřadnicový systém, výškový systém, způsob vyhotovení mapy, stáří mapy, atd. Tyto kartografické vlastnosti se vyskytují samozřejmě i v jednom státě, ale uživatel je většinou lépe obeznámen. Standardy, ontologie jsou cestou k dosažení interoperability jak syntaktické tak technické, avšak si myslím jen pro nově vznikající geografické informace. Pro stávající či staré mapy jsou jen pro lepší orientaci, porozumění a zařazení do katalogových služeb, které je mohou dále šířit.

Co se týká problémů internetových vyhledávačů je řešením standardizace na úrovni poskytování informací z katalogu. Tedy poskytují standardy konsorcia OGC a Mezinárodní organizace pro standardy (ISO). Podpora vzájemné komunikace mezi počítačovými systémy usnadní vyhledávání a vzájemné propojení katalogových vyhledávačů.

Literatura a internetové odkazy

[1] ŘEZNÍK, T. *Metadatový tok v krizovém řízení*. Brno, 2008. Disertační práce na Přírodovědecké fakultě Masarykovy univerzity. Vedoucí práce Doc. RNDr. Milan Konečný, CSc. Dostupné z: <http://is.muni.cz/th/52529/prif_d/>

[2] ŘEHA, J. *Analýza internetových vyhledávačů metodami formální konceptuální analýzy*. Zlín, 2010. Diplomová práce na Fakultě aplikované informatiky Univerzity Tomáše Bati ve Zlíně. Vedoucí práce RNDr. Jiří Klimeš, CSc. Dostupné z: <<http://dspace.knihovna.utb.cz/handle/10563/11763>>

[3] LUTZ, Michael. KLIEN, Eva. *Overcoming Differences of Meaning during the Discovery and Retrieval of Geospatial Information in Spatial Data Infrastructures*. Germany, Münster, 2008. Dostupné z: <http://ifgi.uni-muenster.de/~klien/publications/ncgia_04_position_lutz_klien.pdf>

[4] KLIMENT, Tomáš. Katalógové služby v doméne metadát pre geodáta. *Juniorstav 2008*. Brno: Fakulta stavební Vysokého učení technického v Brně, 2008. Dostupné z: <http://www.fce.vutbr.cz/veda/juniorstav2008_sekce/pdf/6_1/Kliment_Tomas_CL.pdf>

[5] Vaccari, L. Shvaiko, P. Marchese, M. *An emergent semantics approach to semantic integration of geo-services and geo-metadata in Spatial Data Infrastructures*. Itálie, 2008. University of Trento. Dostupné z: <<http://www.gsdi.org/gsdiconf/gsdi10/papers/TS36.4paper.pdf>>

[6] Vaccari, L. Shvaiko, P. Marchese, M. *A geo-service semantic integration Spatial Data Infrastructures*. Itálie, 2008. University of Trento. Dostupné z: <http://disi.unitn.it/~p2p/RelatedWork/Matching/IJSDIR_113.pdf>