

TEZAURY A GAZETEERY

Historie

- Původně vyhledávání (prostorových) dat v (elektronickém) prostředí úkol pro zaškolené profesionály
- Posléze enormní nárůst objemu dostupných informací
 - nově vytvářená digitální data
 - digitalizace papírových (analogových zdrojů)
- Množství dat komplikovalo vyhledávání relevantních informací (zejména pro koncové uživatele)
- Celá řada přístupů o strukturovaný přístup k datům
 - nejčastější řízené slovníky (též tezaury; angl. thesaurus)

Laboratoř geoinformatiky a kartografie



Principy tezurů na Seznam.cz



převzato z: <http://mapa.riziky.seznam.cz/detail.html>

Laboratoř geoinformatiky a kartografie



Principy

- Řízený (kontrolovaný) slovník = seznam dohodnutých termínů převzatých z běžného jazyka
 - podchycena (kontrolována) gramatická forma
 - u cizojazyčných verzí sjednoceno hláskování
 - v hierarchické struktuře
 - + umožňuje indexování, ukádání, vyhledávání
 - + doplňují vyhledávací algoritmus
 - multitezaurové systémy

Laboratoř geoinformatiky a kartografie



Hierarchické (a relační) vazby

- **Ekvivalence** (vztah mezi synonymy)
 - všechna synonyma mohou být rovnocenná
 - obvykle jeden z termínů je zvolen jako preferovaný, další jsou pak alternativy k danému termínu
 - koncový uživatel může použít jakékoli synonymum
 - vzácně vztahy mezi pseudosynonymy; slovy, které mají podobný, ne však totožný význam
- **Hierarchie** (na rozdíl od ekvivalence není symetrická)
 - vytvořena vymezením vztahů mezi podřazenými pojmy (narrower terms; podtřídami) a nadřazenými pojmy (broader terms; nadtřídami)

Laboratoř geoinformatiky a kartografie



Hierarchické (a relační) vazby

- nejčastěji jeden typ této vazby, výjimečně se definuje více
- **Asociace** (symetrický vztah)
 - pro sémantický vztah, který nelze označit pomocí ekvivalence ani hierarchie
 - např. antonyma
- **Mezijazyková ekvivalence**
 - pouze u multilingválních (vícejazyčných) tezurů
 - obvykle pouze u preferovaných termínů
 - slovník = slova, thesaurus = klíčové fráze

Laboratoř geoinformatiky a kartografie



Tvorba tezaurů

- Převzetí (adaptace) již existujícího
 - obvykle nerealizovatelná
- Tvorba tezaurů je řešena v ISO 2788, ISO 5964 (multilingvální) ANSI/NISO Z39.19
- Obecně používané dva přístupy:
 1. postup shora dolů (top-down method)
 - » skupina expertů stanoví rozsah a přibližný okruh termínů – definice vztahů – upřesnění okruhu termínů – typy vazeb – předběžná verze - testování
 2. postup zdola nahoru (bottom-up method)
 - » skupina expertů vybere sadu indexovaných dokumentů – výběr termínů – upřesní se okruh

Laboratoř geoinformatiky a kartografie



Integrace tezaurů

- Vytvíjí se mnoho tezaurů (snaha pokrýt vědní disciplíny, aplikace)
- Propojení možné formou:
 - supertezauru
 - volné integrace
- Varianty:
 - **federace**: kombinace při zachování autonomie, různé DB, platformy zpracování, nekonzistence
 - **sjednocení**: – o stupeň těsnější vztah, odstraněny nekonzistence, někdy vzniká tzv. metatezaurus
 - » např. tezaurus Meta sdružující několik lékařských tezaurů

Laboratoř geoinformatiky a kartografie



Integrace tezaurů

- **spojení**: integrace mikrotezaurů (tj. tezaurů se specializovanými slovníky; nazývané jako subtezaury) s více obecnými makrotezaury. Mikro- a makrotezaury se musí vzájemně přizpůsobit
 - » např. *European Educational Thesaurus*
- Tezaury v metadatech

Laboratoř geoinformatiky a kartografie



Prostorové tezaury neboli gazeteery

- Prostorové tezaury
 - sada organizovaných geografických názvů včetně vztahů mezi těmito názvy
 - » např. administrativní jednotky
- Od 19. století ve Spojeném Království
- Koncem 19. století i v Rakousku – Uhersku
 - navázáno na sčítání lidu, doplněno statistickými daty
 - základ pro rekonstrukci např. osídlení v daném územním celku v historickém vývoji
 - administrativní (soudní a politické), církevní členění
- Další vývoj v 90. letech 20. století – vývoj Webu

Laboratoř geoinformatiky a kartografie



Prostorové tezaury neboli gazeteery

- Geoparsing
 - klíčovému slovu nebo frázi nestrukturalizovaného textu se přidělují geografické identifikátory
 - + souřadnice
 - + kódové informace
 - » např. *GEOnet Names Server*
- Geokódování
 - analyzuje jednoznačnou strukturovanou lokaci (jako např. PSČ) a explicitně formátované numerické souřadnice
 - geoparsing jde za hranice geokódování

Laboratoř geoinformatiky a kartografie



Prostorové tezaury neboli gazeteery

- Obsahují přinejmenším:
 - unikátní identifikátor geografického prvku (objektu)
 - geografický název
 - typ geografického prvku
 - umístění

Laboratoř geoinformatiky a kartografie



PŘÍKLADY TEZAUŘŮ

Getty Thesaurus of Geographical Names (TGN)

- Geografická klíčová slova pro danou oblast zájmu a rozdílné úrovně měřítka (obydlené místo, mikroregion, kraj, stát, kontinent)
- Od roku 1987, elektronicky pak od roku 2000
- <http://www.getty.edu/research/tools/vocabularies/tgn/>
- Obsahuje přes 1 000 000 názvů osídlených míst, administrativních celků, infrastruktury, hydrografie, orografie a vegetace na celé Zemi
- Většinou angličtina a místní jazyk
- Hierarchie

Laboratoř geoinformatiky a kartografie

Screenshot of the Getty Thesaurus of Geographical Names (TGN) website. The search results for 'Brno' are displayed, showing various administrative levels and geographical features associated with the name. The interface includes search filters, a list of results with their IDs and labels, and a detailed record view for 'Brno (Settled place)'.

GeoNet Name Server (GNS)

- Spojená databáze U.S. National Geospatial-Intelligence Agency a U.S. Board on Geographic Names
- Celosvětový tezaurus, přes 4 miliony termínů
 - osídlená místa, administrativní jednotky, infrastruktura, hydrografie, orografie, vegetace, **podmořské názvy**
- Oproti TGN široké spektrum nástrojů pro vyhledávání a aktuálnější údaje
 - např. pro ČR i synonyma kraj Vysočina – Jihlavský
- <http://earth-info.nga.mil/gns/html/>

Laboratoř geoinformatiky a kartografie

Screenshot of the GeoNet Name Server (GNS) website. The search results for 'Křinec nad Ještěvkem' are displayed, showing a table of names with their administrative levels, coordinates, and features. A map of the region is also visible on the right side of the screen.

The World Gazetteer

- Profilovaný pouze na administrativní členění a demografii
 - země, provincie, kraje, ...
 - města včetně výsledků sčítání lidu
- Od roku 2006 nebyl aktualizován
- Projekt uzavřen na konci července 2013
- <http://world-gazetteer.com>

Laboratoř geoinformatiky a kartografie

The World Gazetteer

The World Gazetteer provides a comprehensive set of population data and related statistics

Population data arranged by country

- Current population figures and area size for all countries and important territories
- Population figures and area size of administrative divisions
- Population figures for the largest cities, towns and places as well as for metropolitan areas
- Historical population data (census or estimates, mostly from last or last two censuses)
- Overview maps that show the position of the country or the administrative divisions

Related data

- Summary statistics for countries and territories as well as for the largest cities and agglomerations
- Current national flags
- Summary maps for various subject
- A pronunciation guide for some languages
- There is a downloadable file for offline research
- A population clock that show the current population of the world
- Comprehensive meta informations about statistics and the project

FAQ

Place index

Contacts

Search

Feedback

Links

Last Update:

2013-03-24: Unfortunately the site will be closed end of July

2013-03-22: on people demand 2013 data and a few updates that had been in the pipeline

Laboratoř geoinformatiky a kartografie



Geonames

- <http://www.geonames.org>
- 8 milionů geografických názvů
- Zdarma
- Časté využití pro sociální sítě



...Conferences Home | Planet Codes | Download | WebServices | About

Name	Country	Feature Class	Latitude	Longitude
1. Brno ID	Czech Republic, South Moravia	second order administrative division	N 49° 12' 42"	E 16° 36' 20"
2. Brno-Tulany Airport ID	Czech Republic	airport	N 49° 0' 4"	E 16° 41' 30"
3. Olava Brno-venkov ID	Czech Republic, South Moravia	second order administrative division	N 49° 10' 0"	E 16° 30' 0"
4. Hrádko ID	Czech Republic, South Moravia	second order administrative division	N 49° 12' 38"	E 16° 36' 22"
5. Brno	Czech Republic, Plošňak	mountain	N 49° 49' 23"	E 16° 39' 53"

Laboratoř geoinformatiky a kartografie



GEMET

- GEneral Multilingual Environmental Thesaurus (GEMET)
- Vyvinuto European Environmental Agency (EEA)
- Počátek od roku 1995 v *Umwelt Thesaurus Umweltbundesamt (UBA-A)* ve Vídni, postupně spojení dalších tezaurů
- Celkem takřka 7000 environmentálních termínů
- 32 jazyků
- INSPIRE témata prostorových dat

Laboratoř geoinformatiky a kartografie



GEMET

...Conferences Home | Planet Codes | Download | WebServices | About

EIONET GEMET Thesaurus

Local navigation: Home, About, Search, etc.

Themes list: administration, agriculture, animal husbandry, building, etc.

Find a person: Search box

Account services: Log in, Register, etc.

Footer: Document last modified 2013-03-20 13:27:18 EIONET GEMET Content in this portal is modified daily by a community of providers.

Laboratoř geoinformatiky a kartografie

