VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA

Hornicko-geologická fakulta

institut geoinformatiky

HODNOTENIE VÝVOJA A ŠTRUKTÚRY SPRÁV S VYUŽITÍM GEOPARSINGU

diplomová práca

Autor: **Peter Nemec**

Vedúci diplomovej práce: **doc. Dr. Ing. Jiří Horák**

Ostrava 2010

Prehlásenie

- *Celú diplomovú prácu vrátane príloh som vypracoval samostatne a uviedol som všetky použité podklady a literatúru.*
- *Bol som zoznámený s tým, že na moju diplomovú prácu sa plne vzťahuje zákon č.121/2000 Zb. – autorský zákon, hlavne § 35 – využitie diela v rámci občianskych a náboženských obradov, v rámci školských predstavení a využitie diela školského a § 60 – školské dielo.*
- *Beriem na vedomie, že Vysoká škola báňská – Technická univerzita Ostrava (ďalej len VŠB-TUO) má právo bez zárobku, k svojej vnútornej potrebe, diplomovú prácu užívať (§ 35 odst. 3).*
- *Súhlasím s tým, že jeden výtlačok diplomovej práce bude uložený v Ústrednej knižnici VŠB-TUO k prezenčnému nahliadnutiu a jeden výtlačok bude uložený u vedúceho diplomovej práce. Súhlasím s tým, že údaje o diplomovej práci, obsiahnuté v abstrakte, budú zverejnené v informačnom systéme VŠB-TUO.*
- *Taktiež súhlasím s tým, že kompletný text diplomovej práce bude publikovaný v materiáloch zaisťujúcich propagáciu VŠB-TUO, vrátane príloh časopisov, zborníkov z konferencií, seminárov apod. Publikovanie textu práce bude zaistené v obmedzenom rozlíšení, ktoré bude vhodné iba pre čítanie a neumožní teda prípadnú transformáciu textu a ďalších súčastí práce do podoby potrebnej pre jej ďalšie elektronické spracovanie.*
- *Bolo dojednané, že s VŠB-TUO, v prípade záujmu z jej strany, uzavriem licenčnú zmluvu s oprávnením užívať dielo v rozsahu § 12 odst. 4 autorského zákona.*
- *Bolo dojednané, že užívať svoje dielo – diplomovú prácu alebo poskytnúť licenciu k jej využitiu môžem len so súhlasom VŠB-TUO, ktorá je oprávnená v takom prípade odo mňa požadovať primeraný príspevok na úhradu nákladov, ktoré boli VŠB-TUO na vytvorení diela vynaložené (až do ich skutočnej výšky).*

V Ostrave dňa 20.4.2010 ***Peter Nemec***

Poďakovanie

Chcel by som poďakovať všetkým, ktorí mi akýmkoľvek spôsobom pomohli pri spracovaní tejto bakalárskej práce. Moje poďakovanie patrí predovšetkým vedúcemu projektu, doc. Dr. Ing. Horákovi, za vedenie, cenné pripomienky a odborné rady.

ANNOTATION

This thesis deals with evaluation of media news considering their structure and time progress using geoparsing. The objective of this work is to find and apply some of the methods used to appreciate a spatial, a temporal and a thematic aspect of media news. The main intention is to discover, if media news are influenced by time or space and to estimate a predominance of specific themes. The introduction of the thesis is devoted to find proper forms to save and archive news. Afterwards, various media news sources are appreciated regarding their reliability, count of emitted news and their quality. A key work piece of the thesis is an appreciation of thematic, temporal and spatial aspect of news acquired from various sources. A high accent is dedicated for a choice and implementation of computerized methods. This thesis brings new insight for evaluating news coverage in the means of its temporal, spatial and thematic balance.

Keywords: RSS, XML, Geocoding, Geoparsing, media news, thesaurus

ANOTÁCIA

Táto práca sa zaoberá hodnotením mediálnych správ z hľadiska ich štruktúry a vývoja v čase s využitím geoparsingu. Cieľom práce je nájsť a pokúsiť sa aplikovať metódy pre posúdenie priestorového, časového a tematického hľadiska textov obsiahnutých v mediálnych správach. Hlavnou myšlienkou je zistiť, do akej miery sú mediálne správy z rôznych zdrojov ovplyvnené časom, priestorom a či v nich výrazne prevládajú určité témy. Práca v úvode monitoruje formy prijímania a ukladania mediálnych správ. Následne hodnotí zdroje správ z hľadiska spoľahlivosti, množstva vysielaných príspevkov či ich kvality. Kľúčovou súčasťou práce je vyhodnotenie tematického, časového a priestorového aspektu správ získaných z rôznych zdrojov. Predpokladom pre toto vyhodnotenie je vyhľadanie a aplikácia vhodných metód a postupov. Dôraz je kladený na výber a implementáciu automatizovaných postupov a použitie metód, ktoré sú už použité a dokumentované. Práca prináša nové pohľady na hodnotenie spravodajstva a jeho zafarbenie či vyváženosť z hľadiska uvedených aspektov.

Kľúčové slová: RSS, XML, Geokódovanie, Geoparsing, mediálna správa, tezaurus

Obsah

1 Úvod

Už niekoľko rokov sa v spoločnosti hovorí o informačnej explózii. Neustálym rozvojom internetu a predovšetkým nástupom technológie World Wide Web sa počet producentov a príjemcov informácií každoročne rapídne zvyšuje. Podľa Prof. RNDr. Milana Mareša, DrSc[1]. sa nejedná o prvú informačnú explóziu v dejinách ľudstva. V jednom zo svojich článkov uviedol, že po vzniku písma približne 4 000 rokov pred n. l. a vynájdení kníhtlače v polovici 15. storočia sme v súčasnosti nástupom internetu svedkami až tretej informačnej explózie. Dnešnú, internetovú, informačnú explóziu charakterizuje Prof. Mareš slovami: *„Postavení účastníka je mnohem aktivnější a nabídka informací, které má k dispozici, je o několik řádů větší.“* Sme teda aktívnymi prispievateľmi obsahu internetu či už v oblasti blogov alebo sociálnych sietí, menej často ako tvorcovia rôznych publikácií a naviac máme voľný prístup k obrovskému množstvu informácií.

V súvislosti s neustálym rastom informácií narastá potreba triedenia týchto informácií a hlavne získavania zmysluplného obsahu z nich. Informácie v elektronickej podobe sú v rozličných formách na rôznych miestach a s rôznym typom zabezpečenia. Veľké množstvo z nich je však voľne prístupné.

Je veľkou výzvou vyhľadávať a získavať zmysluplné informácie a znalosti z dát ukrytých v prostredí internetu. Ak je firma schopná získavať a následne využiť informácie z internetu, môže z toho vyťažiť konkurenčnú výhodu. Vyhľadávanie informácií sa stále spolieha na vyhľadávanie podľa kľúčových slov alebo na katalógové služby. Tieto formy však stále majú určité nedostatky týkajúce sa predovšetkým vyhľadania zbytočne veľkého množstva irelevantných informácií alebo naopak nedostatočným pokrytím nášho zvoleného dotazu.

Táto práca sa bude zaoberať získavaním nových znalostí nad dátami voľne dostupnými z internetu. Cieľom nie je pojať veľké kvantá informácií z rôznych oblastí, ale zamerať sa na užšiu oblasť. Touto oblasťou sú mediálne správy v českom jazyku šírené v prostredí internetu. Producentmi týchto správ sú internetové verzie televíznych staníc (napr. ČT24), internetové verzie tlačených periodík (napr. Deník) prípadne médiá fungujúce len na internete (napr. Politikon.cz).

Pri vnímaní mediálnych správ, odhliadnuc od formy ich šírenia, si často ani neuvedomujeme, koľko informácií v sebe daná správa ukrýva. Zaujíma nás predovšetkým, o čom daná udalosť pojednáva, koho sa týka, aké témy zaberá, aké má dopady, aké má zafarbenie apod. Mimo týchto aspektov ale môžeme sledovať i ďalšie, na prvý pohľad možno menej atraktívne informácie ako štruktúra správy, jej pozícia v priestore a čase.

V nasledujúcom texte sa pokúsime o hodnotenie správ pochádzajúcich z rôznych médií. Zameriame sa na porovnanie jednotlivých médií, porovnanie obsahu a štruktúry správ, ktoré šíria v prostredí

[1] Prof. RNDr. Milan Mareš, DrSc je riaditeľom Ústavu teórie informácia a automatizácie Akadémie vied ČR

internetu. Bude nás zaujímať, aké typy správ sa vyskytujú v rôznych médiách, ako sú správy distribuované v priestore i čase, o akých témach pojednávajú. Budeme sa snažiť získať z distribuovaných správ čo najviac nových informácií, ktoré nám poslúžia nielen pre zhodnotenie spravodajstva v jednotlivých médiách, ale môžu poslúžiť aj ďalším účelom.

Pre hodnotenie priestorovej distribúcie sa ako vhodná voľba javí použitie geoparsingu (viď.6). Problematiku tematického zaradenia správ alebo vo všeobecnosti textu zachycuje kapitola 9. Časovej distribúcií udalostí sa venuje kapitola 9. Zložením ohodnotení týchto dielčích aspektov si môžeme vytvoriť celkový obraz o správach šírených danými médiami.

2 Ciele práce

Hlavným cieľom práce je ohodnotiť vývoj a štruktúru správ vybraných médií s využitím geoparsingu. Hlavnú myšlienku môžeme rozdeliť do dielčích cieľov:

1. Oceniť priestorovú zložku spravodajstva vybraných spravodajských kanálov. Účelom je predovšetkým nájsť a otestovať rôzne stratégie geoparsingu, určiť, aká stratégia je výhodná pre určitý účel ako sú na jednej strane aplikácie pre informačný servis obcí alebo na druhej strane aplikácie hodnotiace regionálne rozdiely médií či správ.
2. Nájsť a pokúsiť sa aplikovať možnosti analýzy štruktúry spravodajstva. Zvoliť vhodnú metódy pre analyzovanie tematického zamerania spravodajstva, aké témy prevažujú a ktorým je naopak v médiách venovaná menšia pozornosť. Aké sú možnosti rozdelenia správ podľa určitých tém a sledovanie početného zastúpenia správ prislúchajúcich daným témam.
3. Pokúsiť sa demonštrovať hustotu správ v čase. Pokúsiť sa odpovedať na otázky typu: *Existujú nejaké správy alebo typy správ, ktoré sa vyskytujú skôr sezónne? Hovorí sa o rasizme v poslednom období častejšie ako v minulosti?* Ďalej sa pokúsiť objasniť, či existujú v čase nejaké hluché miesta, kde sa správ vyskytuje menej, či je viac alebo menej správ v určitý typický deň (napríklad cez víkend) apod. Posúdiť, kedy sa jedná skôr o zhlukový charakter distribúcie správ v čase alebo je distribúcia rovnomerná.

3 Navrhovaný postup riešenia

Celá koncepcia práce je založená na využívaní postupov, ktoré je možné automatizovať, aby rola človeka bola minimalizovaná (skôr v oblasti zberu a prípravy dát, vlastná analýza a interpretácia je uskutočnená na základe expertných znalostí a skúseností). Dôraz je kladený na implementáciu štandardizovaných riešení, najlepšie voľne dostupných. Ako sme už spomenuli, naším zámerom je vydolovať z dát obsiahnutých v mediálnych správach rôzne informácie, ktoré môžu poslúžiť ďalším účelom. Postup riešenia pozostáva z niekoľkých krokov:

1. **Výber vhodných médií.** Nevyhnutným predpokladom pre hodnotenie spravodajstva viacerých zdrojov je samotný výber určitej vzorky médií, ktoré budú zahrnuté do analýzy. Výber médií musí podliehať stanoveným kritériám ako bezplatnosť poskytovania obsahu, poskytovanie vhodného distribučného kanálu či spoľahlivosť publikovania správ.
2. **Posúdenie priestorovej zložky správ.** Jednou zo zložiek celkového posúdenia spravodajstva bude ohodnotenie priestorovej zložky. Takmer každej mediálnej správe môžeme či už celkom jednoznačne alebo približne stanoviť jej lokalizáciu. Inými slovami odpovedať na otázku: *„kde sa udalosť odohrala“*. Práve v tejto časti uplatníme metódu zvanú *geoparsing*. Metóda stavia na identifikácii geografických entít v prirodzenom texte, kde je zmienka o lokalizácii často vyjadrená nejednoznačne (na rozdiel od geokódovania, ktoré predpokladá štruktúrované určenie geografickej entity – napríklad adresa). Vychádzať budeme z použitia tejto metódy na texty správ v [12] . Zaujímať nás bude spoľahlivosť použitia tejto metódy. Spoľahlivosť budeme kontrolovať proti náhodne vybranej vzorke 100 správ, ktorá bude skontrolovaná ľudskou silou. Ďalej sa pozrieme na použitie geoparsingu v iných aplikáciách a navrhneme rôzne stratégie geoparsingu.
3. **Monitoring správ a spracovanie historických správ.** Po výbere vhodných zdrojov môžeme pristúpiť k ich priebežnému sledovaniu a zberu správ z nich. Ako najvhodnejšia forma pre zber správ sa javí sťahovanie obsahu RSS kanálov jednotlivých zdrojov. RSS kanál je vo svojej podstate jednoduchý XML súbor, ktorého obsah je štruktúrovaný. Obsah súboru sa v pravidelných intervaloch mení – nové príspevky pribúdajú, staré sa mažú. Pre spracovanie historických správ nemáme bohužiaľ veľké možnosti. Jediným spôsobom, ako sa dostať k správam z minulosti, je prístup k www stránkam archívov jednotlivých serverov. Nevýhodou je, že sa

môžeme dostať len k správam, ktoré sú v HTML formáte. Výhodnejší by určite bol XML formát alebo forma jednoduchých textových dokumentov. Extrakcia jednotlivých správ z HTML stránok je pomerne problematická, hlavným úskalím je jednoznačne oddeliť časť, ktorá obsahuje samotnú správu od zvyšku HTML stránky obsahujúcej jednak iný text, obrázky, animácie či videá.

4. **Porovnanie informačných zdrojov.** Ako už bolo spomenuté, za vhodný typ distribúcie a prijímania mediálnych správ na internete bol zvolený RSS kanál. Za informačný zdroj budeme v tomto zmysle považovať vždy jeden RSS kanál z daného média (napr. RSS kanál dopravy z média ČT24, RSS kanál kultúra z média Denik.cz apod.). Pri porovnávaní jednotlivých RSS kanálov môžeme brať do úvahy niekoľko faktorov, charakterizujúcich samotný kanál. Navyše, každému z týchto faktorov môžeme prisúdiť rôznu váhu (rôzny podiel na celkovom hodnotení). Aj z týchto dôvodov budeme hodnotiť kvalitu informačných zdrojov postupmi multikriteriálneho vyhodnotenia a využitím *Saatyho metódy* (pri stanovení váh kritérií) [5] .
5. **Analýza tém správ, hodnotenie vývoja v čase.** Pri analýze tém správ budeme znovu vychádzať z riešení, ktoré už existujú. Nebudeme tvoriť akýsi nový zoznam tém a správu po správe analyzovať, ktorej témy sa týka. Využijeme postupy známe zo spracovania prirodzeného jazyka, konkrétne z oblasti spracovania textu (text processing).

4 Výber informačných zdrojov

Správy z internetu môžeme čerpať automatizovane niekoľkými spôsobmi:

- Extrakciou z HTML stránok – jedná sa o výber správ z HTML stránok. Zdá sa byť veľmi obtiažne riešenie. HTML stránka okrem samotného textu, ktorý nás zaujíma, obsahuje ešte množstvo ďalších elementov, ktoré nie je úplne jednoduché jednoznačne oddeliť od samotného textu (napr. menu, obrázky, oddeľovacie prvky, reklama apod.)
- Z RSS kanálov – RSS kanál predstavuje XML súbor. Je štruktúrovaný, nie je problém z neho extrahovať jednotlivé časti.
- SMS spravodajstvom – posielanie určitých správ do mobilného telefónu
- Nechať si posielať správy na e-mail
- Začlenením správ do vlastného webu – jedná sa o začlenenie časti zdrojového kódu web-stránok publikátora článkov do vlastnej web stránky
- V podobe samostatnej aplikácie napríklad pre populárny telefón iPhone. Takúto aplikáciu poskytuje napríklad ČT24.
- Prostredníctvom webovej služby. Takéto riešenie nebolo v čase tvorby materiálu nájdené na českých spravodajských serveroch. Mohlo by byť však ideálnou voľbou, keďže klient webovej služby by si mohol priamo nadefinovať, aký typ správ chce v danom čase odoberať.

Pre analyzovanie väčšieho množstva správ zdá byť najjednoduchšia a najosožnejšia voľba prijímania správ z RSS kanálov. RSS kanál je ľahko čitateľný a presne štruktúrovaný podľa špecifikácie. Pre čítanie informácií z RSS kanálu sa najčastejšie používajú RSS čítačky. V našom prípade si však pripravíme jednoduchý skript v jazyku PHP, ktorý bude čítať správy z RSS a jednotlivé časti zapisovať do databázovej tabuľky.

Výber informačných zdrojov môžeme rozdeliť do 2 etáp. Najprv je treba určiť, z ktorých médií budeme správy čerpať a následne vybrať vhodné RSS kanály jednotlivých serverov. Jedným z prvých kritérií je teda samotná existencia a poskytovanie správ v podobe RSS kanálu. Väčšina všeobecne známych spravodajských serverov túto službu poskytuje. Pri výbere vhodných informačných zdrojov zohráva rolu ďalej bezplatnosť poskytovania obsahu. Na internete existuje niekoľko platených spravodajských elektronických magazínov (napr. Politikon.cz). Vybraný informačný zdroj by okrem toho mal byť všeobecne známy, aby sme mali záruku, že poskytuje dostatočné množstvo správ a tematickú pestrosť poskytovaných kanálov. Prihliadnuc na spomenuté kritériá bol vybraný zoznam týchto RSS kanálov:

4-1 Zoznam vybraných RSS kanálov

Kanál (skratka použitá v ďalšom texte)	Tematické zameranie	Dostupný na adrese
Server ČeskéNoviny.cz		
cn_cestovani	Cestovanie	http://www.ceskenoviny.cz/sluzby/rss/cestovani.php
cn_domov	Domáce spravodajstvo	http://www.ceskenoviny.cz/sluzby/rss/domov.php
Server ČT24.cz		
ct24_cestovani	Cestovanie	http://www.ct24.cz/rss/cestovani/
ct24_domaci	Domáce spravodajstvo	http://www.ct24.cz/rss/domaci/
ct24_doprava	Doprava	http://www.ct24.cz/rss/doprava/
ct24_kultura	Kultúra	http://www.ct24.cz/rss/kultura/
ct24_regionalni	Regionálne spravodajstvo	http://www.ct24.cz/rss/regionalni/
c24_sport	Šport	http://www.ct24.cz/rss/sport/
Server Denik.cz		
denik_kultura	Kultúra	http://www.denik.cz/rss/kultura.html
denik_domov	Domáce spravodajstvo	http://www.denik.cz/rss/z_domova.html
Server iDnes.cz		
mf_zpravodaj	Správy z domova i zo sveta	http://servis.idnes.cz/rss.asp?c=zpravodaj
Server Novinky.cz		
novinky_cestovani	Cestovanie	http://www.novinky.cz/rss/cestovani/
novinky_domaci	Domáce spravodajstvo	http://www.novinky.cz/rss/domaci/
novinky_krimi	Kriminálne činy	http://www.novinky.cz/rss/krimi/
novinky_kultura	Kultúra	http://www.novinky.cz/rss/kultura/
novinky_sport	Šport	http://www.sport.cz/rss2/

Je zrejmé, že zdroje môžeme rozdeliť do 5 kategórií podľa spravodajských serverov. V každom z nich je niekoľko RSS kanálov (minimálne jeden) obsahujúcich niekoľko všeobecne známych tém ako kultúra, doprava, šport, regionálne spravodajstvo apod.

Výber kanálov pravdepodobne nebude absolútne objektívny. Z veľkého množstva RSS kanálov len zo sféry spravodajstva je možné vybrať niekoľko variantov. Cieľom nie je vybrať čo najväčší počet informačných zdrojov, ale aplikovať metódy priestorového, tematického či časového posúdenia na určitej vzorke. Výber tejto vzorky je ovplyvnený subjektívnym pohľadom autora výberu. Na druhej strane, uvedené metódy môžu byť aplikované na ľubovoľné informačné zdroje. Hodnotenie jednotlivých vybraných informačných kanálov je uvedené v kapitole 8.

5 Použitý software a technológie, vstupné dáta

5.1 Vstupné dáta

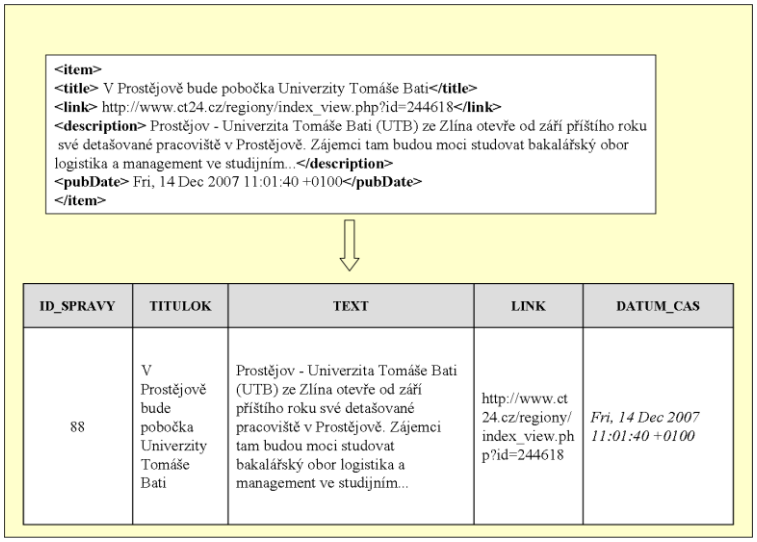
Vstupné dáta tvoria:

1. Územný identifikačný register základných sídelných jednotiek (UIR-ZSJ). Slúži pre získanie informácií o územnom členení a ďalej informácií o jednotlivých obciach – počet obyvateľov, rozloha, geografické koordináty
2. RSS kanály spravodajských serverov. Sú vo forme XML dokumentov.
3. Archívy správ. Sú spravidla v HTML formáte.
4. Vrstvy administratívneho usporiadania ČR (hranice obcí, okresov a krajov) vo formáte ESRI Shapefile. Dáta poskytol Inštitút geoinformatiky a pochádzajú z RSO ČSÚ[2].

Pri zbieraní a archivovaní správ z jednotlivých kanálov musíme mať určené, akým spôsobom budeme správy zbierať a archivovať. K tomuto účelu bola na serveri *gislinb,* slúžiaceho pre študentské účely, vytvorená MySQL databáza *nem272*, nazvaná podľa identifikátora autora v školskom systéme. Databázu tvorí niekoľko tabuliek, ktoré môžeme rozdeliť do 4 kategórií:

- Tabuľky obsahujúce správy z RSS kanálov. Pre každý RSS kanál je vytvorená jedna tabuľka, ktorej záznam je naplnený obsahom a atribútmi jednej správy (Obr. 5-1). Pre kontrolu nových príspevkov a ich následné pridanie do tabuľky sa stará PHP skript,pomenovaný podľa pravidla: *parser_nazovkanalu.php* (napríklad parser_ct24_regionalni.php). Dôležité je periodické spúšťanie tohto skriptu, ktoré zabezpečuje *Cron*. Cron je program, ktorý zabezpečuje automatické spúšťanie iných programov či skriptov v definovaných intervaloch, v našom prípade každú hodinu. Tabuľky sú spravidla pomenované s predponou *news*. Bližšie v [12] kapitola 6.2.

[2] RSO ČSÚ predstavuje register sčítacích obvodov a budov Českého štatistického úradu



Obr. 5-1 Prevod správy z RSS kanálu do záznamu v tabuľke

- Tabuľky obsahujúce správy z archívov. Spravidla majú rovnakú štruktúru ako tabuľky obsahujúce správy z RSS kanálov. Rozdiel je v tom, že majú plný text správy, nielen jeho časť, ktorá je zachytená v RSS kanáli.
- Tabuľky týkajúce sa tezauru GEMET.
- Ďalšie tabuľky. Slúžia pre podporu analýz a zachytenie výsledkov analýz nad vstupnými dátami.

Pre tvorbu a administráciu tabuliek sa využíva prostredie *phpMyAdmin*. Pre analýzy pracujúce nad dátami a tabuľkami a pre zber údajov sa využívajú skripty vytvorené v jazyku PHP. Ďalej sa pri spracovaní tabuliek využíva MS Access 2007 a pre písanie PHP skriptov voľne dostupný PSPad.

6 Spoľahlivosť lokalizácie správ a porovnanie stratégií geoparsingu

6.1 Geoparsing

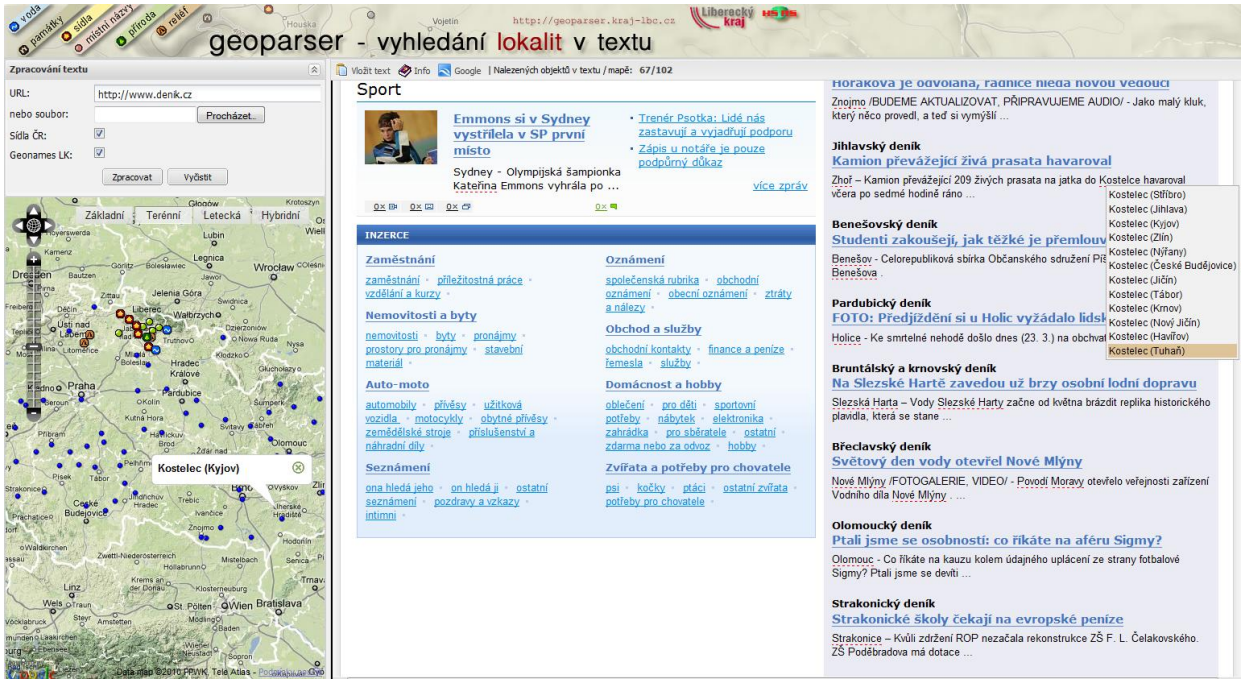
Geoparsing je proces podobný geokódovaniu. Rozdiel je v tom, že geokódovanie určí polohu objektu na základe presne štruktúrovanej adresy (16 Nádražní ulice), pričom geoparsing sa snaží určiť polohu na základe neštruktúrovaného určenia v texte. Informácie o polohe nie sú v texte štruktúrované, ale sú zapísané prirodzeným jazykom, napr. 40 km južne od Prahy; Petrovice v okrese Blansko; Ostrov na Karlovarsku; Bělkovice-Lašťany neďaleko Olomouca apod. Metódy geoparsingu idú za hranice geokódovania. Geoparsing je možné použiť aj pre jednoznačnú identifikáciu obce, ktorej názov je viacnásobný. Napríklad Petrovice, ktoré sú súčasťou názvu niekoľkých obcí, môžeme podľa dodatočných informácií v texte priradiť jednoznačnú lokalizáciu [1] . Kúzlo geoparsingu je ukryté v 2 krokoch. 1) extrakcia entít a 2) odstránenie nejednoznačnosť, tiež označované ako geotagging. Extrakcia entít spočíva v označení tých reťazcov, ktoré sú geografickými názvami. Odstránenie nejednoznačnosti priraďuje názov miesta správnej lokalite. Výsledky geoparsingu sa často vkladajú do pôvodného dokumentu, prípadne sú naformátované do určitého výstupu vhodného pre geopriestorovú aplikáciu [2] .

6.2 Stratégie geoparsingu

Z obecného popisu geoparsingu môžeme odvodiť niekoľko konkrétnych implementácií. Pre lokalizáciu správ je možné využiť niekoľko stratégií, líšiace sa podľa účelu, prípadne podľa koncového užívateľa. Z hľadiska účelu aplikácie pre geoparsing obcí môžeme stratégie rozdeliť do 2 hlavných prúdov:

1. **Optimistický prístup.** Účelom je nájsť čo najviac výskytov danej obce v texte správ, a to i za cenu toho, že nebude vysoká spoľahlivosť. Užívateľ má za cieľ nájsť všetky výskyty jeho zadanej hľadanej obce v texte a nevadí mu, ak sú identifikované i iné obce s podobným názvom. Podstatné preňho je, že chce mať vo výstupe **všetky** výskyty hľadanej obce. Optimistický prístup je zrejmý aj v aplikácii geoparser Libereckého kraja (Obr. 6-1). Pre reťazec Kostelec (v pravej časti obrázku) je

identifikovaných 12 objektov v rôznych častiach územia ČR.



Obr. 6-1 12 objektov názvu Kostelec identifikovaných v databáze Geonames. Ako zdroj textu použitá adresa: http://www.denik.cz

Je však nutné dodať, že optimistické riešenie môže spočívať aj v tom, že sa neidentifikujú len názvy Kostelec, ale aj napríklad Kostelec nad Vltavou, Kostelec u Křížků, Kostelec na Hané…

2. **Pesimistický prístup.** Účelom je do čo najvyššej možnej miery potlačiť chyby. Oproti optimistickému prístupu je cieľom v texte identifikovať len jeden, správny názov obce. V predchádzajúcom príklade by sa teda nenašlo niekoľko výsledkov pre Kostelec, ale len jeden správny, Kostelec na Jihlavsku.

Stratégie, ktoré sa dajú použiť pre geoparsing, môžu využívať optimistický prístup, pesimistický prístup, prípadne ležať niekde medzi. Stratégie sa môžu líšiť jednak označením geografickej entity v texte a jednak spôsobom vylúčenia nejednoznačnosti. Označením geografickej entity sa myslí označenie tých slov (reťazcov), ktoré môžu reprezentovať geografický objekt. Geografické entity sa najčastejšie v zahraničných zdrojoch označujú využitím techník spracovania prirodzeného jazyka. Odstránenie nejednoznačnosti sa líši od aplikácie, prakticky neexistuje akýsi štandardný spôsob pre túto úlohu. Skúsme si nadefinovať niekoľko stratégií, ktoré je možné využiť:

1. **Využitie pravdepodobnosti výskytu:** Podľa predchádzajúcich využitých výsledkov geoparsingu je možné priradiť určitým obciam pravdepodobnosti spomenutia v textoch správ. Napríklad obce, ktorých názov je nejednoznačný by automaticky mali zníženú pravdepodobnosť. Podobne by na tom boli obce, ktoré sú tzv. konfliktné – napríklad názvy, ktoré sú sami alebo obsahujú slová bežného jazyka ako Bílá, Bystrá, Časy, Klec. Ďalej by bolo nutné znížiť pravdepodobnosť u obcí, ktorých názov je konfliktný s ďalšími geografickými či turistickými objektmi ako Dyje (názov obce i rieky), Vítkov (názov obce i mestskej časti v Prahe). Nakoniec by sa pravdepodobnosť mala znížiť obciam, ktorých jednoslovný názov je časťou názvu iných obcí. Napríklad Opatovice je samostatná obec ale i časť názvu Opatovice nad Labem, podobne Osek a Osek nad Bečvou.
2. **Využitie rôznych priorít:** Pri geoparsingu bol pôvodne využitý prístup, že pri odstraňovaní nejednoznačnosti názvu sa postupuje podľa troch kritérií: *podľa oblasti* – v texte sa hľadá napríklad názov okresu a podľa neho sa určí, o ktorú konkrétnu obec sa jedná. Ďalej *podľa blízkosti k iným obciam* – v texte sa hľadajú iné lokality a nejednoznačnosť sa odstráni tým, že sa vyberie obec najbližšia k ďalším spomenutým v texte. A nakoniec *podľa počtu obyvateľov* – ak sa nevyužije žiadne z predchádzajúcich kritérií, vyberie sa obec s najvyšším počtom obyvateľov. Ako rôzne stratégie je možné spomenuté kritériá rozlične meniť – to znamená, môžeme pri nejednoznačnosti vybrať najvhodnejšieho kandidáta priamo podľa počtu obyvateľov. Tento prístup využíva i aplikácia geoparser Libereckého kraja, kde sa výsledky zoradia podľa dôležitosti a za dôležitosť sa považuje počet obyvateľov. Tento prístup bol vyskúšaný aj na vzorke dát (v kapitole spoľahlivosť lokalizácie), oproti pôvodnému prístupu nepriniesol žiadne zmeny. Je však nutné dodať, že kritérium *podľa oblasti* je pravdepodobne účinnejšie ako kritérium *podľa počtu obyvateľov*.

6.3 Spoľahlivosť lokalizácie

Spočíva v ohodnotení spoľahlivosti lokalizácie správ na základe porovnania s manuálnym hodnotením. Z celkového počtu 7 054 správ bolo náhodne vybraných 100 správ, kde bola vykonaná manuálna lokalizácia. Automatická lokalizácia bola pri jednotlivých správach vyhodnotená ako nesprávna, ak minimálne jedna obec nebola identifikovaná alebo ak minimálne jedna obec bola identifikovaná nesprávne. Správna lokalizácia – všetky obce v správe – v domicile i v texte – boli lokalizované správne. V stĺpci spresnenie sa zaznamená,

či bolo použité spresnenie lokalizácie, ak áno, na základe akého kritéria (podľa oblasti, podľa vzdialenosti od iných spomenutých obcí, podľa počtu obyvateľov). V poznámke sú zmienené aj ďalšie geografické názvy, ktoré zatiaľ v mechanizme nie sú použité, v budúcnosti je však možné ich využiť. Základné štatistické ukazovatele analýzy zhŕňa Tab. 6-1. Výsledky analýzy sú uvedené v Tab. 11-1.

Tab. 6-1 Základné štatistické ukazovatele použitej metódy geoparsingu

Celkový počet správ	7 054
Vzorka analyzovaných správ	100
Správne lokalizovaných	92
Nesprávne lokalizovaných	8
Spresnenie lokalizácie (použitie geoparsingu)	6
- podľa oblasti	4
- podľa počtu obyvateľov	1
- podľa vzdialenosti	1

Chyby lokalizácie vyplývajú jednak z nedostatočného porozumenia textu správ strojovému spracovaniu, ale aj nepresností zo strany producenta správ. Chyby môžeme zoskupiť do týchto kategórií:

- Chyby vyplývajúce z viacslovných názvov. Pre väčšinu viacslovných názvov nie sú definované pády. Pre viacslovné názvy obcí boli definované pády len pre významné obce ako Karlovy Vary, Hradec Králové, České Budějovice a pre viacslovné názvy s počtom obyvateľov nad 10 000. Chyby lokalizácie vyplývajú z toho, že sa mylne vyberie obec, ktorej názov je totožný s jedným slovom z viacslovného názvu. Napríklad názov obce Opatovice je súčasťou názvu Opatovice nad Labem. Riešením môže byť pridanie všetkých pádov ku všetkým viacslovným názvom. Týchto názvov je však 1 140. Ďalším môže byť priradenie nižšej pravdepodobnosti (nižšej šance) ku všetkým jednoslovným názvom, ktoré môžu byť súčasťou názvov viacslovných.
- Chyby vyplývajúce z konfliktov. Jedná sa o konflikty medzi názvami obcí a názvami iných geografických či negeografických názvov. Príkladom je rieka Dyje a obec Dyje na Znojemsku. Ďalej obec Vítkov na Opavsku a mestská časť Vítkov v Prahe. Alebo

ulica Výsluní v Plzni a obec Výsluní na Chomutovsku. Ďalším konfliktom bol konflikt medzi názvom obce Lipno nad Vltavou a názvom skiareálu Lipno.

- Iné chyby. Vyplývajú zo zlého pochopenia kontextu automatizovaným spôsobom. Príkladom je slovné spojenie *Mariánské a Františkovy Lázně*. Mechanizmus geokódovania identifikoval len Františkovy Lázně. Riešením môže byť implementácia rôznych pravidiel do mechanizmu geoparsingu. Pravidlá by museli vychádzať z metód spracovania textu alebo obecnejšie z umelej inteligencie.
- Chyby producenta. Príkladom je napríklad, že autor článku v texte uviedol slovné spojenie *Věstonice (pri Mikulove)*, ktoré de facto neexistujú. Existujú 2 obce – Horné Věstonice a Dolné Věstonice. Podobne je na tom chybný text – v správe sa objavilo Velké Meziříčí namiesto Velké Meziříčí. Riešením uvedených problémov môže byť pripustenie určitej chyby, inak povedané, že zhoda názvu obce v databáze a textového reťazca v texte nemusí byť 100-percentná.

7 Monitoring správ, spracovanie historických správ

Pri spracovaní historických správ vybraných RSS kanálov by bolo ideálne, keby sme mali k dispozícii archív vybraného kanálu (vybranej rubriky), uložený v ucelenej podobe, najlepšie vo forme jednoduchých textových súborov alebo vo forme XML. Poskytnutie takéhoto archívu zo strany vydavateľov je však problematické. Bolo oslovených niekoľko vydavateľov s prosbou o poskytnutie archívu, ale neúspešne.

Jedinou možnosťou, ktorou sa môžeme dostať k historickým správam, je hľadanie na webových stránkach. Našťastie sú archívy na stránkach pomerne prehľadné a dobre usporiadané. Čitateľnosť pre človeka je dobrá. Nás však zaujíma získanie jednoduchých textov správ. A najväčší problém predstavuje práve extrakcia čistého textu správ z formátu HTML.

Skúsme si popísať, v čom tkvejú nevýhody formátu HTML. Jazyk HTML bol navrhnutý pre zobrazovanie dát, v čom je hlavná nevýhoda oproti XML, ktorý bol navrhnutý pre prenos a uchovávanie dát. HTML sa zameriava na to, ako dáta vyzerajú [3] . V HTML neexistuje skutočné členenie alebo hierarchia medzi jednotlivými tagmi. Extrakcia určitých častí alebo logických celkov je z tohto dôvodu náročná. Ďalším významným obmedzením HTML je, že webové stránky napísané v HTML často nie sú validné (v súlade s normou, v súlade s DTD). Prehliadače si však tieto chyby nevšímajú a stránku vykreslia dobre. A nakoniec, webové stránky v HTML často nie sú ani správne naformátované, dochádza napríklad ku kríženiu značiek, ktoré je v samotnej špecifikácii HTML zakázané, napriek tomu ho prehliadače opäť dobre zobrazia. Príklad kríženia značiek:

```
<i>Dnes je<b>pekný deň!</i></b>
```

Všetky spomenuté obmedzenia nám sťažujú získavanie textov z HTML stránok, v ktorých sú napísané archívy. Cieľom je vždy vytiahnuť zo zdrojového kódu stránky len určitú časť (ostatné prvky nás nezaujímajú) a jednoznačné ohraničenie tejto časti býva problémom. Nepríjemným vedľajším efektom pri extrakcii textov z webových stránok býva aj nechcené začlenenie tých častí, ktoré s naším cieľovým textom ani nesúvisia (text mimo text správy, obrázky, videá, animácie, reklamné bannery, skripty a ďalšie prvky).

Tab. 7-1 Zoznam zdrojov, z ktorých je možné čerpať historické správy

Zdroje	
cn_cestovani	denik_kultura
cn_domov	denik_domov

ct24_cestovani	mf_zpravodaj
ct24_domaci	novinky_cestovani
ct24_doprava	novinky_domaci
ct24_kultura	novinky_krimi
ct24_regionalni	novinky_kultura
ct24_sport	novinky_sport

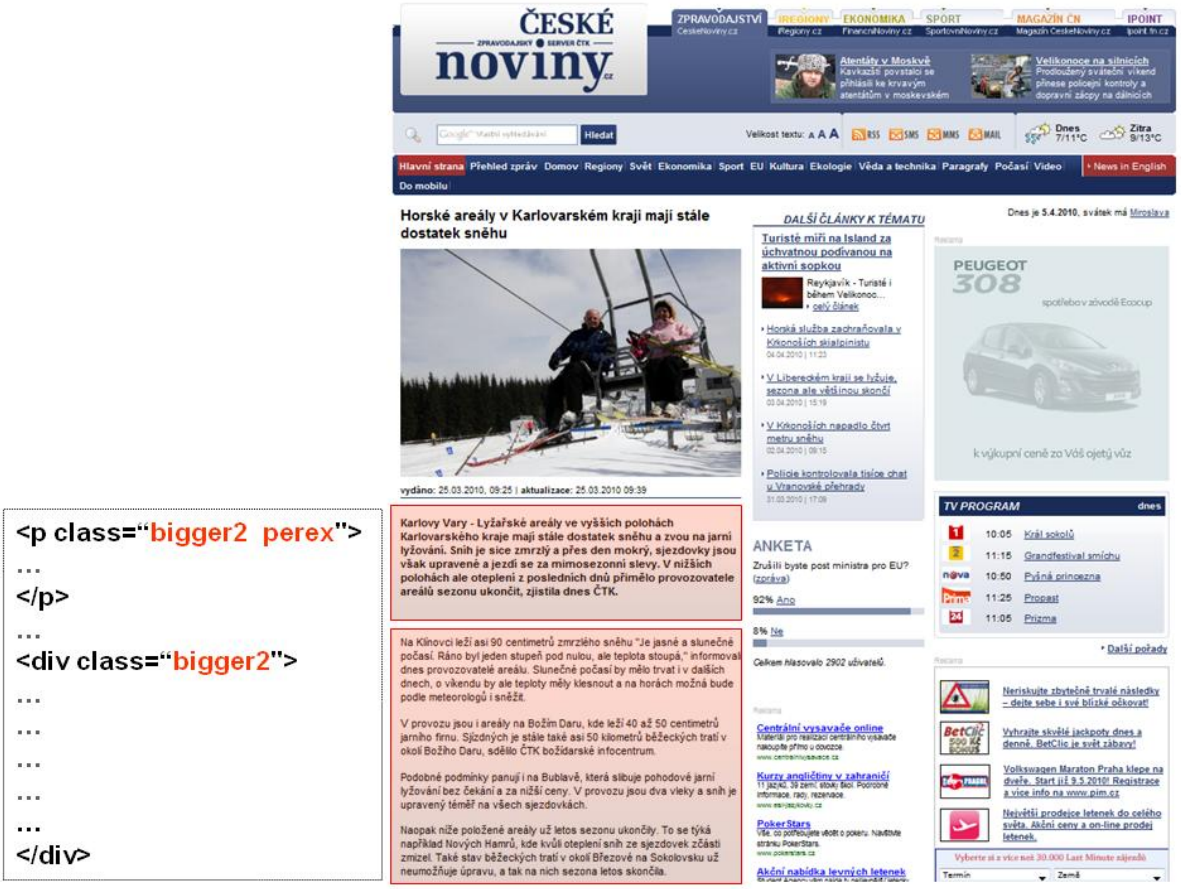
Zoznam zdrojov, z ktorých budeme čerpať zachycuje Tab. 7-1. Výhodou je, že nie každý zdroj má špecifickú štruktúru archívu. Zdroje, ktoré majú rovnakú štruktúru archívu začleníme do skupín. Z uvedených zdrojov zostavíme 6 skupín. V popise každej skupiny zdrojov si budeme všímať tagy vymedzujúce text správy a ďalej dátum v minulosti, do ktorého je možné správy získať. Tento dátum však treba bližšie špecifikovať. Často sa totiž stáva, že archív siaha napríklad až po dátum 12. 7. 1998. Avšak z roku 1998 je tam povedzme 1 správa, 3 správy z roku 1999. Je nutné teda rozlíšiť dátum najstaršej správy a dátum, po ktorý vedie kontinuálne rozmiestnenie správ v čase (teda najstarší dátum, v okolí ktorého sú správy „nahusto“). Z tohto hľadiska budeme rozlišovať dátum najstarší a dátum smerodatný.

Archívy fungujú na všetkých serveroch na podobnom princípe. Ak v príslušnej rubrike užívateľ klikne na odkaz typu „starší zprávy“, dostane sa do archívu. Archív je vždy rozdelený na niekoľko častí, nazvime ich „náhľady“ (Obr. 7-1). Náhľady zobrazujú prehľad niekoľkých (napríklad 10, 20) správ, prípadne prehľad správ za jeden mesiac (server Novinky.cz). Kliknutím na odkazy pri jednotlivých správach v náhľadoch sa dostaneme k ich plnému zneniu. Počet náhľadov a počet správ v nich sa líši u jednotlivých zdrojov.

Obr. 7-1 Ukážka náhľadu z archívu cn_domov. Zdroj: server České Noviny

7.1 Skupina České Noviny

Do skupiny patria zdroje *cn_cestovani, cn_domov*. Ukážka náhľadu archívu *cn_domov* je na Obr. 7-1. Z hľadiska spracovania tém správ je určite zaujímavá časť *témata*, uvedená pod každou správou, ktorá zobrazuje zoznam tém, ktoré sa v správe objavujú. Bohužiaľ, celkový zoznam všetkých tém nie je verejne prístupný, ani sa nepodaril získať. V každom náhľade sa zobrazuje 11 správ. Každá správa v sebe skrýva odkaz na plné znenie, krátky úvod, čas publikovania a témy. Plné znenie správy a jej štruktúru zachycuje Obr. 7-2.



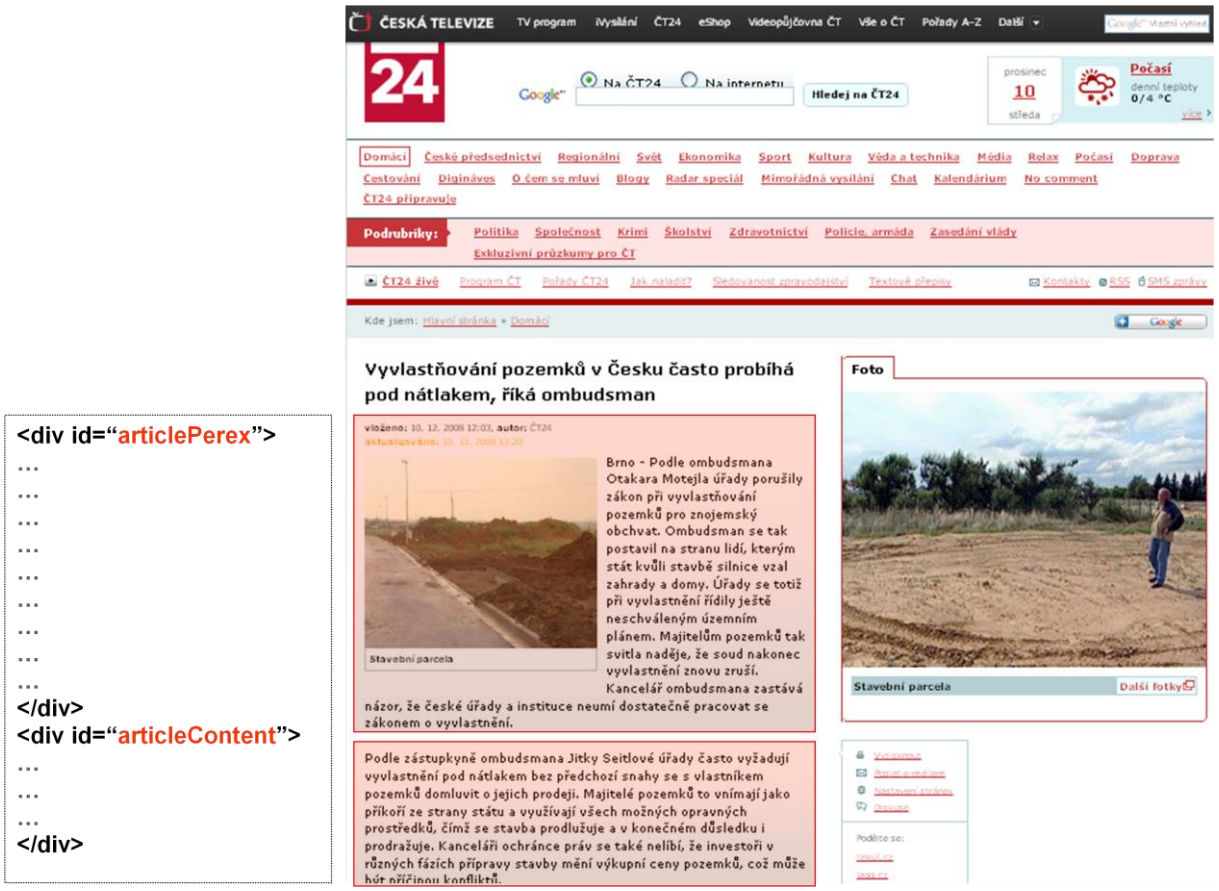
Obr. 7-2 Plné znenie správy z archívu a štruktúra HTML. Zdroj: server České Noviny

Plné znenie správy sa skladá z úvodného odseku (perexu), ohraničeného tagom *<p class="bigger2 perex">*. Rozširujúce znenie správy je obsiahnuté v tagu * *.

Archív *cn_cestovani* siaha do dátumu 19. 5. 2006, posledný dátum je zhodný so smerodatným dátumom. Archív *cn_domov* má posledný dátum 5. 5. 2009, posledný dátum je zhodný so smerodatným.

7.2 Skupina ČT24

Do skupiny patria zdroje *ct24_cestovani, ct24_domaci, ct24_doprava, ct24_kultura, ct24_regionalni, ct24_sport*. Náhľad archívu obsahuje 20 správ, každá správa v sebe obsahuje odkaz na plné znenie, krátky úvod a čas publikovania. Plné znenie správy a jej štruktúru zachycuje Obr. 7-3.



Obr. 7-3 Plné znenie správy z archívu a štruktúra HTML. Zdroj: server ČT24

Server ČT24 používa na svojich stránkach pre perex tag <*div*> s atribútom *id* a hodnotou atribútu *articlePerex*. Rozširujúce znenie správy je obsiahnuté v tagu <*div*> s atribútom *id* a hodnotou atribútu *arcicleContent.* Časové ohraničenie archívu zobrazuje Tab. 7-2.

Tab. 7-2 Hraničné dátumy pre skupinu ČT24

Archív	Posledný dátum	Smerodatný dátum
ct24_cestovani	9. 1. 2007	9. 1. 2007
ct24_domaci	21. 9. 2006	6. 8. 2007
ct24_doprava	4. 10. 2006	2. 7. 2007
ct24_kultura	4. 10. 2006	5. 2. 2007
ct24_regionalni	17. 8. 2006	14. 9. 2006
ct24_sport	12. 7. 2007	14. 12. 2007

7.3 Deník – domáce spravodajstvo

Do tejto kategórie patrí len jeden zdroj *denik_domov*. Archív siaha do dátumu 26. 3. 2007, čo je zároveň i dátum kontinuálny. V každom náhľade je 15 správ, obsahujúcich krátky úvod, odkaz na plné znenie a dátum a čas publikovania. Štruktúra HTML je zachytená na Obr. 7-4.

Obr. 7-4 Plné znenie správy z archívu a štruktúra HTML. Zdroj: server Denik.cz, rubrika domáce spravodajstvo

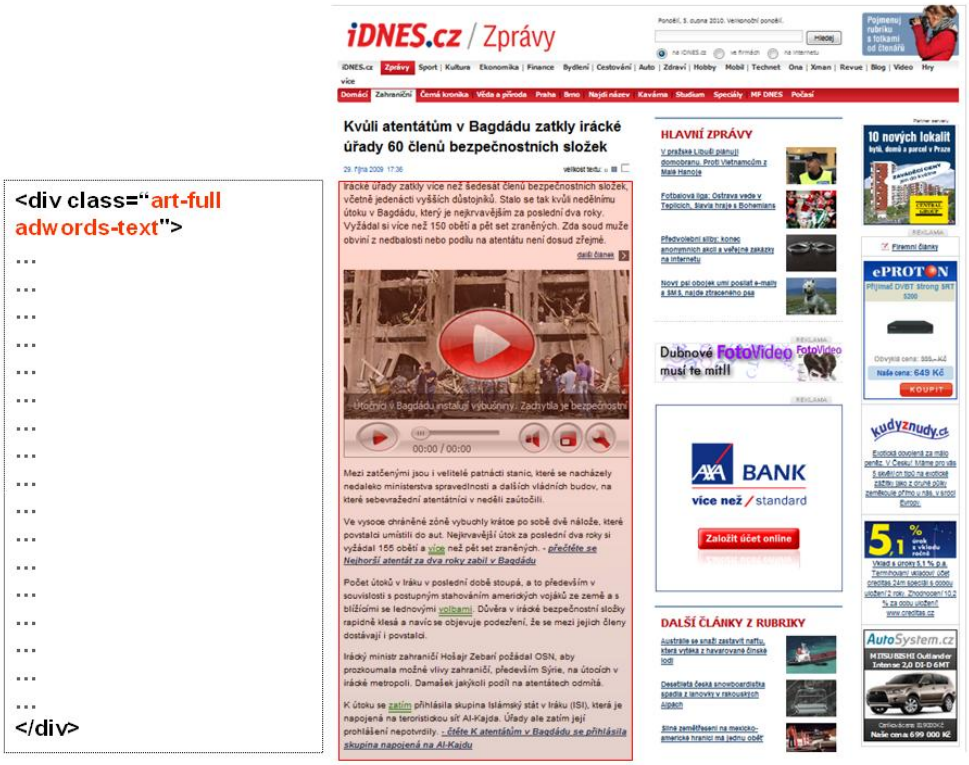
Na stránkach serveru Denik.cz archívu domáceho spravodajstva je znenie správy obsiahnuté v 2 tagoch. Tag *<p>* s atribútom *class* a hodnotou atribútu *info* v sebe obsahuje perex. Rozširujúce znenie správy sa ukrýva v tagu * * s atribútom *class* a hodnotou atribútu *bbtext*.

7.4 Deník – kultúra

Patrí sem jediný zdroj denik_kultura. Pre kultúru má server Denik.cz zriadený samostatný portál. Je možné pristupovať do archívu, ale len po jednotlivých dielčích sekciách kultúry ako: hudba, film – TV, divadlo, knihy, umenie, festivaly, vstupenky. Náhľady obsahujú 10 správ.

7.5 iDNES – spravodaj

Patrí sem kanál *mf_zpravodaj*. Náhľady obsahujú 25 správ, u každej správy je odkaz na plné znenie, krátky úvod a dátum a čas publikovania. Plné znenie správy a štruktúru HTML zachycuje Obr. 7-5.



Obr. 7-5 Plné znenie správy z archívu a štruktúra HTML. Zdroj: server iDNES.cz, rubrika spravodajstvo

Plné znenie správ z tohto zdroja nie sú rozdelené na úvodnú časť a rozšírené znenie. Text celej správy je ohraničený v rámci tagu <*div*> s atribútom *class* a hodnotou atribútu *art-full adwords-text*. Už na obrázku Obr. 7-5 vidíme množstvo iných prvkov okrem textu. Pri extrakcii textu nie je problém preskočiť video. Komplikácie však prinášajú v texte vnorené odkazy typu: *„- čtěte K atentátům v Bagdádu se přihlásila skupina napojená na Al-Kájdu“*.

7.6 Skupina Novinky.cz

Patria sem zdroje uvedené v Tab. 7-3. Od predchádzajúcich skupín sa táto skupina líši v náhľadoch archívu. V náhľade nie je pevne stanovený počet správ, ale vždycky správy za jeden mesiac. Užívateľ sa môže upravením URL adresy ľahko dostať k správam v určitom mesiaci a roku. Náhľad obsahuje pri každej správe odkaz na správu, krátky úvod, dátum a čas publikovania. Posledný dátum je vždy zhodný s kontinuálnym dátumom. Plné znenie správy a štruktúru HTML zachycuje Obr. 7-6.

Obr. 7-6 znenie správy z archívu a štruktúra HTML. Zdroj: server Novinky.cz

Znenie správy je rozdelené na úvodnú časť (perex) a rozširujúce znenie. Perex je ohraničený tagom <*p*> s atribútom *class* a hodnotou atribútu *perex*. Rozširujúce znenie sa skrýva v tagu <*div*> s astribútom *class* a hodnotou atribútu *content*. Dátumy pre jednotlivé zdroje uvádza Tab. 7-3.

Tab. 7-3 Hraničné dátumy pre skupinu Novinky.cz

Archív	Posledný dátum	Kontinuálny dátum
novinky_cestovani	31. 8. 2006	31. 8. 2006
novinky_domaci	1. 1. 2003	1. 1. 2003
novinky_krimi	1. 1. 2003	1. 1. 2003
novinky_kultura	1. 2. 2008	1. 2. 2008

7.7 Novinky.cz – rubrika šport

Server Novinky.cz má vo svojej ponuke RSS kanálov i šport. Správy pre tento RSS kanál sa však čerpajú z iného serveru (Sport.cz). Prístup do archívu na tomto serveri je pomerne problematický a vyriešený dosť nešikovne. Nie je možné sa dostať do archívu ani po jednotlivých sekciách ako napr. *futbal, hokej, motoršport*, ale až na úrovni tém ako *majstrovstvá sveta 2010, Gambrinus liga, Tipsport Cup* apod.

7.8 Zhodnotenie spracovania historických správ

Asi najprehľadnejší archív udržiava na svojich stránkach server Novinky.cz. Užívateľ sa môže dotazovať priamo na mesiac, ktorý ho zaujíma. Okrem toho, plné znenia správ (texty) sa dajú vyextrahovať najjednoduchšie, čiže ich logická štruktúra HTML je prehľadná. Ďalšou nespornou výhodou je možnosť dostať sa k správam siahajúcim pomerne hlboko do histórie – január 2003.

Plné znenia správ na serveri iDNES.cz sú zaťažené množstvom vedľajších prvkov, rôznych odkazov apod. Tieto prvky značne znepríjemňujú extrakciu čistých textov. Podobný problém sa v zmenšenej miere vyskytuje v plných zneniach správ v archívoch serveru Denik.cz, kde v rámci textu správy sa vyskytne napríklad ponuka na diskusiu k článku.

Okrem plného znenia správ sprístupnených v archívoch nás určite zaujíma aj distribúcia správ v čase. Dátumy a časy publikovania správ je vo všetkých prípadoch (až na Novinky.cz – rubrika šport) možné získať z náhľadov archívov. Nie je teda nutné pristupovať k plným zneniam. Pre účely analýzy časového rozloženia udalostí budeme ako hraničný dátum v minulosti brať kontinuálny dátum. Správy za touto hranicou až po posledný dátum by nám výrazne skreslili výsledku. Tieto správy však môžu byť užitočné napríklad v prípade, že chceme získať kompletný mediálny servis pre dané termíny prípadne pre danú obec.

8 Porovnanie náročnosti, využiteľnosti a spoľahlivosti rôznych informačných zdrojov

8.1 Postup

Pre kvalitatívne porovnanie vybraných informačných zdrojov (RSS kanálov) bola zvolená skupina kritérií, podľa ktorých môžeme zhodnotiť ich vhodnosť a využiteľnosť. Kritérium RSS kanálu odráža určitú jeho vlastnosť, ocenenú hodnotou (hodnota udáva mieru splnenia / nesplnenia vlastnosti). Každé z kritérií má svoj podiel na celkovom rozhodovaní. Tomuto podielu budeme hovoriť *váha* a celkové zhodnotenie môžeme považovať za *multikriteriálne ocenenie (MCE).*

Prvou podstatnou úlohou je výber vhodných kritérií. Akými rysmi môžeme charakterizovať RSS kanál a čím sa môže kvalitatívne líšiť jeden kanál od druhého?

8.2 Výber kritérií a ich ohodnotenie

Každé kritérium by malo odzrkadľovať určitú kvalitatívnu vlastnosť RSS kanálu. Mieru splnenia kritéria bude oceňovať hodnota (skóre), vybraná zo škály od 0 do 4. Vyššie skóre znamená lepšie splnenie kritéria. Každý kanál bude mať teda pri každom jeho kritériu hodnotu od 0 do 4 v závislosti na tom, do akej miery toto kritérium spĺňuje či nespĺňuje.

RSS je vo svojej podstate XML dokument vystavený na webe. Každý XML dokument by mal byť v súlade s jeho deklarovanou štruktúrou. Deklarácia štruktúry sa nachádza na začiatku XML súboru, prevažne odkazom na ňu. Keď je v súlade, hovoríme, že je validný. **Validita** kanálu je prvým kritériom pre jeho hodnotenie.

RSS kanál má verejne známu a prístupnú štruktúru (špecifikáciu), a to pre všetky verzie. Nie je teda problém vziať ľubovoľný RSS kanál a posúdiť, či je alebo nie je v súlade s danou štruktúrou. Tento prístup je síce presný, na druhej strane však príliš prácny, najmä ak chceme zahrnúť do analýzy väčší počet kanálov. Preto využijeme už navrhnutý nástroj - validátor, ktorý poskytne podrobné informácie o validite RSS kanálu. Existuje množstvo validátorov, dostupných online na webe. Nie všetky sú celkom spoľahlivé, to znamená, že nemusia odhaliť všetky nesúlady s deklarovanou štruktúrou. Ďalej sa odlišujú množstvom informácií poskytnutých o zdroji. Niektoré z nich vypíšu len informáciu typu validný / nevalidný. Kvalitnejšie validátory ponúknu okrem základnej informácie o validite i zoznam chýb zdroja, počet výskytov danej chyby a tzv. odporučenia – recommendations (Obr. 8-1). Odporučenia

znamenajú určité tipy pre vylepšenie zdroja, napríklad publikovať zhodný titulok pre logo i adresu zdroja apod.

Ako najvhodnejší validátor bol vybraný Feed Validation Service vytvorený medzinárodným konzorciom W3C, zaoberajúcim sa tvorbou webových štandardov. Validátor je vystavený na adrese: http://validator.w3.org/feed/. Podľa výsledkov validácie môžeme kanál ohodnotiť stanovenou škálou od 1 do 4. Validný kanál s malým množstvom odporúčaní bude mať hodnotu 4, s vysokým množstvom odporúčaní 3. Kanál nevalidný s malými chybami skóre 2 resp. 1, so závažnými chybami a množstvom odporúčaní skóre 0. Testom validity neprešli kanály spadajúce pod České Noviny a pod Deník. Chyby však neboli závažné, týkali sa len mediálneho obsahu (napr. obrázky nemali udanú veľkosť v bajtoch). Tieto nedostatky síce nespôsobovali žiadne chyby pri čítaní kanálu, ale samotný fakt, že sú nevalidné, môže vyústiť k ich menšej rozšírenosti.

Obr. 8-1 Feed Validation Service od konzorcia W3C pre kanál ct24_domaci.

Ďalším kritériom pre rozhodovanie je **zmena adresy či štruktúry.** Počas zberu správ zo zdrojov sa vyskytol prípad, kedy sa zmenila adresa kanálu. Problémom je, že samotná špecifikácia RSS neponúka nástroje k tomu, aby v ňom mohol vydavateľ zverejniť informáciu o zmene adresy či štruktúry. Zmenou štruktúry sa myslí napríklad to, že vydavateľ po nejakom čase pridá resp. uberie niektoré elementy. Zmena adresy kanálu sa týkala v testovanej vzorke len jedného kanálu – denik_kultura. Zmena adresy sa udiala 29. Novembra 2009. Naviac sa zmenil i spôsob publikovania, kedy sa pôvodný kanál so všetkými správami z oblasti kultúry rozčlenil na 7 sekcií (hudba, film-tv, divadlo, knihy, umenie, festivaly, vstupenky). Kanál denik_kultura dostane skóre 1, kanály spadajúce pod gesciu ČT24 dostanú skóre 3, pretože zmenili čiastočne štruktúru (pridanie elementov s multimediálnym obsahom) a ostatné zdroje dostanú skóre 4.

Kritérium **počet správ v kanáli**. Pre odberateľov článkov, ale predovšetkým pre producentov, môže byť veľkým problémom preťažovanie serveru. Každá položka RSS kanálu je ohraničená v elemente <item>. Počet týchto elementov však nie je obmedzený. To môže priniesť problémy – ak je položiek príliš veľa (stovky, tisíce), dochádza k problematickému čítaniu (z hľadiska času), okrem toho sú v kanáli príliš staré príspevky. Spracovanie takéhoto kanálu môže trvať príliš dlho a dokonca môže spôsobiť i zahltenie serveru [10] . Na druhej strane príliš malý počet príspevkov (10 a menej) zvyšuje riziko, že pri určitom intervale odoberania nemusia byť odobraté všetky príspevky (ak je napríklad interval odoberania 1 hodina, počet príspevkov 10 a v priebehu polhodiny sa všetky príspevky vymenia za nové, tak hrozia veľké straty dát na strane odoberateľa). V rámci zaradených kanálov sa počet správ pohyboval od 10 do 92. Kanály pod správou ČT24 obsahovali 24 príspevkov, pod Deníkom 20, pod Mladou Frontou 92, pod Novinkami 10 (okrem sekcie šport, kde je 24 príspevkov), pod Českými Novinami 50 príspevkov. Počet správ 92 nespôsobuje pri odoberaní veľké problémy, napriek tomu je poznať dlhšiu dobu spracovania. Počet správ 10 však môže spôsobiť problémy, najmä v domácom spravodajstve, kde sa príspevky môžu meniť pomerne často. Kanály s počtom správ 10 boli ocenené hodnotou 1, s počtom správ 92 hodnotou 3, 4 body dostali kanály s počtom správ 20-50.

Kritérium **priemerný počet správ za deň.** Udáva, koľko priemerne príspevkov je uverejnených v jednom dni. Čím vyšší tento počet, tým lepšie. Je potreba ale brať do úvahy fakt, že niektoré kanály už svojou povahou budú publikovať menej príspevkov – napríklad cestovanie či kultúra. Neplatí to však obecne. Počet príspevkov na deň nám dosť kolíše, takmer každý kanál má inú hodnotu, preto je nutné nájsť metódu pre pridelenie skóre (Tab. 8-1).

Pre pridelenie skóre využijeme minimálnu hodnotu, maximálnu hodnotu a štandardizovaný rozsah. Minimálna hodnota je 1, maximálna 18, štandardizovaný rozsah vypočítame ako rozdiel maximálneho a minimálneho možného skóre, teda 4 – 0 = 4. Výpočet sa riadi podľa vzťahu:

$$x_i = \frac{R_i - R_{min}}{R_{max} - R_{min}} * STROZ$$

R pôvodné skóre

x_i pridelené skóre

STROZ štandardizovaný rozsah

R_{min}, R_{max} minimálna, maximálna hodnota pôvodného skóre [5]

Tab. 8-1 Priemerný počet správ za deň a pridelené skóre

Kanál	Priemerný počet správ za deň	Pridelené skóre
cn_cestovani	7	1
cn_domov	18	4
ct24_cestovani	1	0
ct24_domaci	17	4
ct24_doprava	9	2
ct24_kultura	8	1
ct24_regionalni	12	3
c24_sport	11	3
denik_kultura	14	4
denik_domov	14	4
mf_zpravodaj	15	4
novinky_cestovani	5	0
novinky_domaci	15	4
novinky_krimi	11	4
novinky_kultura	11	4
novinky_sport	10	4

Kritérium **obsah** zahŕňa kvalitatívne hodnotenie náplne elementu <description>. Tento element obsahuje časť príspevku. Väčšinou sa jedná o úvodný odsek správy, v žurnalistike označovaný ako perex[3]. Z hľadiska ocenenia tohto kritéria bude hrať rolu hlavne dĺžka textu, poskytnutie celého znenia správy v textovej podobe a spôsob publikovania úvodného odseku. Najvyššie skóre (4) dosiahli kanály pod správou ČT24, pretože vždy publikujú celý perex, naviac v elemente <content:encoded> je uverejnené i celé znenie správy v textovej podobe[4]. Ostatné kanály dostali 3 body za striktné dodržiavanie uverejňovania prvého odseku, až na kanál od Mladej Fronty, ktorý publikuje len časť perexu zakončenú tromi bodkami (empiricky zistené približne 150 prvých znakov). Tento kanál dostal 1 bod za príliš krátky obsah.

Kritérium **domicil**. Domicil je miesto opisovanej udalosti alebo náhradné označenie miesta. Za domicil môže byť považovaná aj určitá oblasť (Jesenicko, Žďársko...). Prítomnosť domicilu môžeme považovať za výhodu oproti ostatným kanálom. Jednoznačnou výhodou je

[3] Podľa Wikipedie pojem perex označuje krátky text (obvykle 2-5 viet), který rozvádza titulok či podtitulok a kladie si za cieľ upútať čitateľovu pozornosť.

[4] Neplatí to úplne doslovne, boli identifikované prípady, kedy nebolo uverejnené celé znenie, ale len akési rozšírenie perexu.

to pre následné potreby geoparsingu. Kanály, ktoré spravuje ČT24 mali najvyššiu kvalitu domicilu, dostali 4 body, najmenej bodov kanály patriace Novinkám – 0 bodov. 2 body dostal kanál od Mladej Fronty – niekedy sa vyskytoval, ale väčšinou nie. 3 body získali kanály z Deniku a Českých Novin.

Kritérium **potreba špeciálnych nastavení**. Toto kritérium sa priamo dotýka kanálov od ČT24. Pre zber správ z týchto kanálov bolo nutné použiť špeciálny bezpečnostný reťazec, poslaný priamo z redakcie. Pre jazyk PHP je tento príkaz v tvare `ini_set('user_agent', 'Mozilla/4.0')`. Kanály spravované ČT24 boli ocenené hodnotou 0, ostatné kanály hodnotou 4.

Kritérium **georss.** Stále častejšie sa začínajú objavovať príspevky, ktoré majú okrem štandardných informácií aj informáciu o svojej polohe. V Českej republike boli v oblasti spravodajstva zatiaľ takéto informácie o polohe identifikované len v niektorých RSS kanáloch od ČT24. Poloha je umiestnená v tagu <georss:point>. Jedná sa teda zatiaľ o implementáciu bodového určenia polohy. Iné ako bodové určenie polohy udalosti by mohlo byť problematické, možno však v oblasti dopravy, kde by bol napríklad úsek cesty, kde sa vyskytla nejaká nepríjemnosť, označená líniou, teda elementom <georss: line>. Pri oceňovaní tohto kritéria sa berie v úvahu jednak prítomnosť elementu označujúceho polohu ale aj jeho kvalita. Prítomnosť takéhoto elementu bola zistená u všetkých kanálov patriacich ČT24 okrem kultúry. Tieto kanály boli ocenené hodnotou 2, ostatné hodnotou 0. Hodnota 2 odzrkadľuje i kvalitu určenia polohy a prítomnosť v rámci samotných príspevkov. Vyskytli sa totiž správy obsahujúce v domicile napríklad Praha a časť z nich mala informáciu o polohe, druhá časť nie. Podobne na tom boli i niektoré správy zo zahraničia, kde i pomerne známe lokality ako Atény, Soul či Bratislava nemali v sebe svoju lokalizáciu obsiahnutú.

Kritérium **kvalita formátovania**. Zahŕňa v sebe informácie o kvalite formátovania textu vo vnútri elementov, hlavne v rámci elementu <description>. Pri spracovávaní textov z internetu dochádza často k určitým komplikáciám. Nie je problém spracovávať jednoduchý text bez špeciálnych znakov. Ťažkosti nastávajú najmä pri spracovaní úvodzoviek, apostrofov, pomlčiek či znaku „ampersand“. Ďalším problémom bolo napríklad zahrnutie niekoľkých znakov medzera na úvod každej správy v kanáloch od Českých Novin. 4 body za kvalitu formátovania dostali kanály od Deniku a Noviniek, kde neboli zachytené žiadne komplikácie. 3 body získali kanály z Mladé Fronty a ČT24, kde sa vyskytli drobné komplikácie pri

využívaní rôznych znakov pre pomlčky či apostrofy. 2 body získali kanály od Českých Novin, kde sa prejavili problémy s pomlčkami a odsadzovaním prvého odseku medzerami. Chyby v pomlčkách sú pomerne významné, keďže v následnom procese geoparsingu správ je nutné jednoznačne oddeliť domicil od zvyšku správy a deliacim znakom je práve pomlčka.

Kritérium **metadáta**. Metadáta sú zjednodušene dáta o dátach. V našom prípade sú to doplňujúce informácie o RSS kanáli, uvedené v jeho úvode. Špecifikácia RSS k tomuto účelu umožňuje uplatniť niekoľko elementov. Pravdou je, že len malý zlomok z nich sa prakticky využíva. Napríklad pomerne významný element <ttl>. Určuje počet minút, počas ktorých môže byť kanál „kešovaný“ pred ďalšou aktualizáciou zo strany zdroja. Inak povedané je to minimálny interval aktualizácií. S aktualizáciou súvisí i element <lastBuildDate>, udávajúci dátum a čas poslednej aktualizácie zdroja. K ďalším metadátam môžeme zaradiť napríklad logo, informácie o poskytovateľovi, webmastrovi, apod. Najlepšie metadáta boli obsiahnuté v kanáli od Mladé Fronty, preto mu boli pridelené 4 body. V ostatných kanáloch boli metadáta približne na rovnakej úrovni, ocenené boli hodnotou 2.

Ocenenie jednotlivých kritérií pre každý RSS kanál zhrňuje Tab. 8-2.

Tab. 8-2 Priradenie skóre ku každému z kritérií

RSSkanál	Validácia	Zmena adresy a štruktúry	Počet správ v kanáli	Priemerný počet správ za deň	Obsah	Domicil	Potreba špeciálnych nastavení	Georss	Kvalita formátovania	Metadáta
cn_cestovani	1	4	4	1	3	3	4	0	2	2
cn_domov	1	4	4	4	3	3	4	0	2	2
ct24_cestovani	4	3	4	0	4	4	0	2	3	2
ct24_domaci	4	3	4	4	4	4	0	2	3	2
ct24_doprava	4	3	4	2	4	4	0	2	3	2
ct24_kultura	4	3	4	1	4	4	0	0	3	2
ct24_regionalni	4	3	4	3	4	4	0	2	3	2

ct24_sport	3	3	4	3	4	4	0	2	3	2
denik_kultura	2	1	4	4	3	3	4	0	4	2
denik_domov	2	4	4	4	3	3	4	0	4	2
mf_zpravodaj	3	4	3	4	1	2	4	0	3	4
novinky_cestovani	4	4	1	0	3	0	4	0	4	2
novinky_domaci	4	4	1	4	3	0	4	0	4	2
novinky_krimi	4	4	1	4	3	0	4	0	4	2
novinky_kultura	4	4	1	4	3	0	4	0	4	2
novinky_sport	4	4	4	4	3	0	4	0	4	2

8.3 Stanovenie váh pre jednotlivé kritériá – Saatyho metóda

Druhou podstatnou časťou pri multikriteriálnom ocenení je stanovenie váhy pre každé kritérium. K dispozícii je množstvo metód. Obecne platí, že váhy volíme tak, aby ich súčet bol 1. Ak chceme brať do úvahy preferencie (dôležitosti) určitých kritérií, tak týmto kritériám budú priradené vyššie váhy. V našom prípade bola zvolená Saatyho metóda, ktorá porovnáva jednotlivé páry kritérií.

Saatyho metóda patrí medzi jednu z najčastejších metód pre voľbu váh, používa sa napríklad i pri AHP[5]. Porovnáva páry kritérií a výsledky ukladá do tzv. Saatyho matice $S = (s_{ij})$ podľa schémy:

$$(s_{ij}) = \begin{cases} 1 - i\ a\ j\ sú\ rovnocenné \\ 3 - i\ je\ slabo\ preferované\ pred\ j \\ 5 - i\ je\ silno\ preferované\ pred\ j \\ 7 - i\ je\ veľmi\ silno\ preferované\ pred\ j \\ 9 - i\ je\ absolutne\ preferované\ pred\ j \end{cases}$$

Musí platiť, že $s_{ji} = \frac{1}{s_{ij}}$ pre všetky *i*.

Podľa [4] zahrňuje výpočet týchto 5 krokov:

1. Vyplníme Saatyho maticu: Na diagonále budú samé jednotky. Stačí vyplniť spodnú časť pod diagonálou. Matica je symetrická, za zvyšné hodnoty môžeme dosadiť

[5] AHP predstavuje analytický hierarchický proces

reciproké hodnoty príslušných buniek. Saatyho metóda patrí medzi subjektívne metódy stanovania váh. Matica bola vytvorená na základe subjektívneho vnímania tvorcu v pohľade na páry jednotlivých kritérií. Z matice je napríklad zrejmé, že validácia je preferovaná pred zmenou adresy a štruktúry a ďalej je veľmi silno preferovaná pred kvalitou formátovania. Počet správ v kanáli je slabo preferovaný pred metadátami apod.

Tab. 8-3 Saatyho matica. Obsahuje vzťahy medzi všetkými pármi kritérií

S_{ij}	Validácia	Zmena adresy a štruktúry	Počet správ v kanáli	Priemerný počet správ za deň	Obsah	Domicil	Potreba špeciálnych nastavení	Georss	Kvalita formátovania	Metadáta
Validácia	1	3	5	3	1	1	3	7	7	5
Zmena adresy a štruktúry	1/3	1	1/3	1/3	1/5	1/5	3	3	1	3
Počet správ v kanáli	1/5	3	1	1/3	1/3	1/3	3	5	3	3
Priemerný počet správ za deň	1/3	3	3	1	1	1	5	3	5	5
Obsah	1	5	3	1	1	1	5	3	5	5
Domicil	1	5	3	1	1	1	5	3	5	5
Potreba špeciálnych nastavení	1/3	1/3	1/3	1/5	1/5	1/5	1	1/3	1	1
Georss	1/7	1/3	1/5	1/3	1/3	1/3	3	1	3	1
Kvalita formátovania	1/7	1	1/3	1/5	1/5	1/5	1	1/3	1	1
Metadáta	1/5	1/3	1/3	1/5	1/5	1/5	1	1	1	1

2. Pre každé i spočítame hodnotu $s_i = \prod_{j=1}^{10} s_{ij}$
3. Pre každé i spočítame hodnotu $R_i = (s_i)^{1/10}$
4. Ďalej spočítame $\sum_{i=1}^{10} R_i$
5. Posledným krokom je určenie váh podľa vzťahu $v_i = R_i / \sum_{i=1}^{10} R_i$

Tab. 8-4 Výpočet váh pre kritériá

	$s_i = \prod_{j=1}^{10} s_{ij}$	$R_i = (s_i)^{1/10}$	$v_i = R_i / \sum_{i=1}^{10} R_i$
Validácia	33075,000000	2,831066	0,214
Zmena adresy a štruktúry	0,040000	0,724780	0,055
Počet správ v kanáli	3,000000	1,116123	0,084
Priemerný počet správ za deň	1125,000000	2,018902	0,153

Obsah	5625,000000	2,371441	0,179
Domicil	5625,000000	2,371441	0,179
Potreba špeciálnych nastavení	0,000099	0,397613	0,030
Georss	0,003175	0,562560	0,043
Kvalita formátovania	0,000127	0,407732	0,031
Metadáta	0,000178	0,421685	0,032
		$\sum_{i=1}^{10} R_i$ = 13,223342	$\sum_{i=1}^{10} v_i$ =1

8.4 Multikriteriálne ocenenie

Jednou z najpoužívanejších procedúr pri multikriteriálnom rozhodovaní je vážená lineárna kombinácia WLC. Vstup do procedúry tvoria ocenenia kritérií a ich váhy. Kritériá sú kombinované podľa pridelených váh a suma udáva mapu vhodnosti.

$$S = \sum_i w_i . x_i$$

S suitability

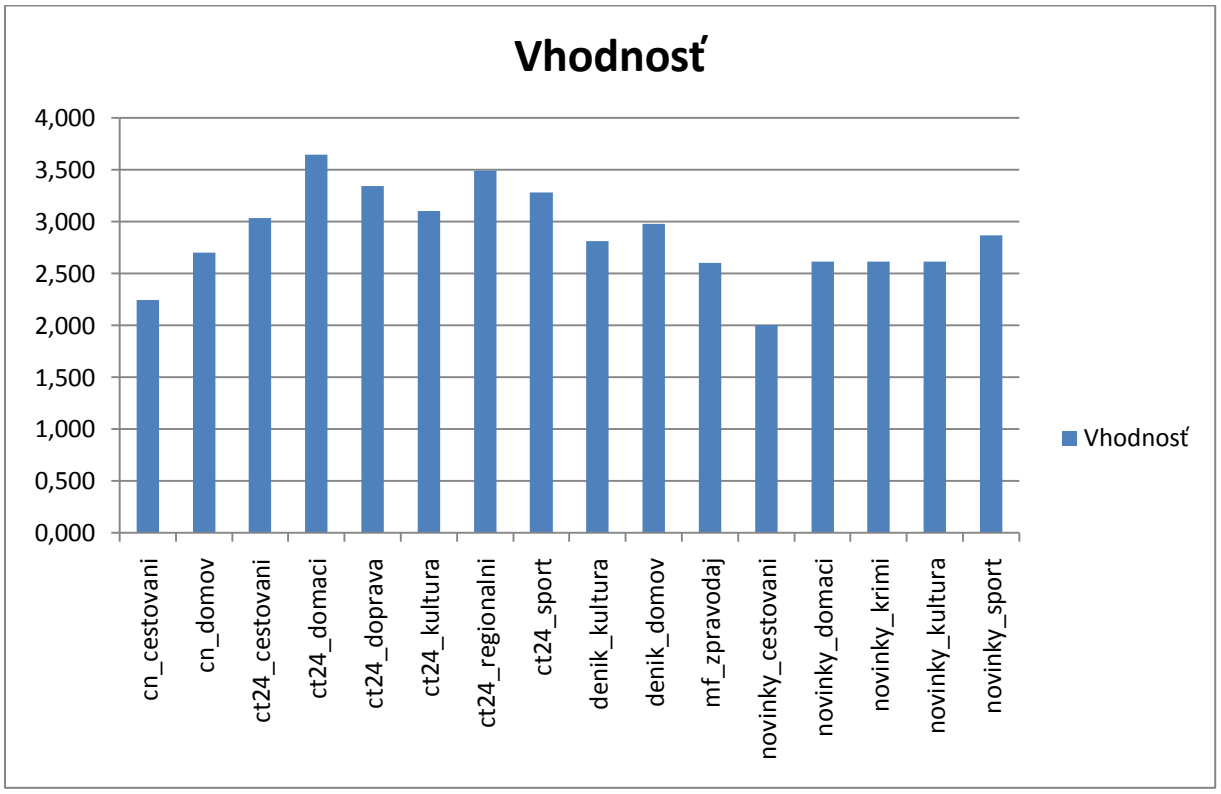
w_i váha kritéria i

x_i kriteriálne skóre kritéria i [5]

Výsledky analýzy dokumentuje Tab. 8-5 s pripojeným grafom. Vhodnosť využitia RSS kanálu pre zber správ a poťažmo vhodnosť kanálu pre lokalizáciu správ by sa mala pohybovať v rozmedzí stanovenej lineárnej škály, čiže medzi 0 a 4, čo je potvrdené. Najnižšiu vhodnosť vykazuje kanál novinky_cestovani (zo serveru Novinky.cz). Naopak najvyššiu vhodnosť má kanál ct24_domaci, nasledovaný ct24_doprava (server ČT24.cz). Rozdiely medzi kanálmi nie sú nijak extrémne. Všetky hodnoty vhodnosti sa pohybujú nad hodnotou 2, žiadny kanál nedosiahol blízko maxima – hodnoty okolo 4. Môžeme teda usúdiť, že kvalita vybratých českých spravodajských RSS kanálov by mohla byť lepšia. Nedostatky sú predovšetkým vo validite kanálov, v poskytovaní metadát a v kvalite formátovania. Plusom kanálov od ČT24 je zaradenie elementu <georss: point>, slúžiaceho pre presnú lokalizáciu správy. Je nutné ale dodať, že v tejto oblasti je ešte potreba vývoja a skvalitnenia.

Tab. 8-5 Vhodnosť vybraných RSS kanálov

RSS kanál	Popis	Vhodnosť (suitability)
cn_cestovani	Cestovanie	2,245
cn_domov	Domáce spravodajstvo	2,703
ct24_cestovani	Cestovanie	3,035
ct24_domaci	Domáce spravodajstvo	3,645
ct24_doprava	Doprava	3,340
ct24_kultura	Kultúra	3,102
ct24_regionalni	Regionálne	3,493
ct24_sport	Šport	3,278
denik_kultura	Kultúra	2,815
denik_domov	Domáce spravodajstvo	2,979
mf_zpravodaj	Spravodajstvo	2,604
novinky_cestovani	Cestovanie	2,005
novinky_domaci	Domáce spravodajstvo	2,616
novinky_krimi	Kriminálne činy	2,616
novinky_kultura	Kultúra	2,616
novinky_sport	Šport	2,869



Ako najvhodnejšie sa javí použitie domáceho spravodajstva z kanálu ČT24 s vhodnosťou 3,645. Najmenšiu vhodnosť má kanál serveru Novinky.cz rubrika cestovanie. Nízke číslo vhodnosti je v tomto prípade spôsobené hlavne malým priemerným počtom správ v kanáli a absenciou domicilu či elementu <*georss*> v zdrojovom kóde RSS kanálu. Z grafu nemôžeme

jednoznačne zoskupovať kanály podľa príslušnosti k danému serveru. I keď sú kanály poskytované z jedného serveru, môžeme v nich vidieť značné rozdiely – napríklad kanál novinky_cestovani a zvyšné kanály zo serveru Novinky.cz

9 Analýza tém správ, ich podobnosti a hodnotenie vývoja v čase

Každá správa, ktorú máme zachytenú v monitoringu správ, pochádza z určitého zdroja. Na základe povahy zdroja (RSS kanálu) vieme o správe povedať, akú tému spracováva. Napríklad ak pochádza zo zdroja *cn_cestovani*, bude takmer s istotou pojednávať o záležitostiach týkajúcich sa cestovania. Avšak nebude jednoduché o udalosti pochádzajúcej z kanálu *ct24_domaci* jednoznačne prehlásiť, akú tému bude zachycovať. Jediné, čo vieme, že bude hovoriť o udalosti z domova. Bližšie zaradenie bez nahliadnutia do obsahu nie sme schopní identifikovať. Naviac, do úvahy prichádza i možnosť, že zdroje *cn_cestovani* a *ct24_domaci* môžu v určitých prípadoch publikovať správu pojednávajúcu o tej istej udalosti, obecnejšie o tej istej téme.

Pri analýze tém správ narážame na 2 hlavné problémy: stanoviť množinu tém a metódu, pomocou ktorej priradíme správu k téme. Jedným z riešení je stanoviť vlastný zoznam tém a správu po správe analyzovať, či patrí alebo nepatrí do danej témy. Tento prístup však naráža na jedno úskalie – zoznam tém nebude určite kompletný a pravdepodobne nebude zachycovať danú oblasť, na ktorú sa chceme zamerať. Okrem toho by bol tento prístup časovo príliš náročný a vyžadoval by tím ľudí, vzdelaných najmä v oblasti lingvistiky.

Našou úlohou je spracovať tematické rozdelenie správ. Správy sú uložené ako text v elektronickej podobe, a teda je možné tento text spracovať automatizovanými postupmi. Informatika nám k tomu ponúka spracovanie textu (text processing) alebo obecne spracovanie prirodzeného jazyka (natural language processing).

9.1 Spracovanie prirodzeného jazyka

Spracovanie prirodzeného jazyka patrí k jedným z najnáročnejších úloh umelej inteligencie. Okrem spracovania samotného textu (text processing) sem patrí [6] :

- Syntéza a rozpoznávanie reči
- Generovanie prirodzeného jazyka (natural language generation)
- Strojový preklad
- Odpovedanie na otázky (question answering)
- Získavanie informácií (information retrieval)
- Extrakcia informácií (information extraction)
- Korektúra textu
- Výťah z textu (automatic summarization)

Cieľom práce nie je obsiahnuť celú tému spracovania prirodzeného jazyka, nakoľko problematika je sama o sebe pomerne rozsiahla. Cieľom je vybrať a otestovať určitú metodiku vhodnú pre analýzu tém správ.

Pri spracovaní budeme využívať databázu správ, uložených vo forme voľného, neštruktúrovaného textu. Budeme využívať hlavne metódy vyhľadávania v týchto textoch. V rámci spracovania môžeme hovoriť o textovej databáze.

9.2 Textové databázy

„*Textovou databázou obecne nazývame databázu, v ktorej sú všetky údaje alebo ich podstatná časť vo forme voľného textu*“ [7] . Hlavným rozdielom medzi textovými a relačnými databázami je v spôsobe vyhľadávania. V relačných databázach poznáme klasické formulácie podmienok selekcie, zamerané na hľadanie zhody celého reťazca alebo podreťazca. V textových databázach je potrebné naviac formulovať dotaz obecnejšie: v našom prípade napríklad *nájdi všetky správy pojednávajúce o znečisťujúcich látkach*. Dotaz sa potom formuluje pomocou množiny slov a ich kombinácií, postihujúcich oblasť znečisťujúcich látok [7] .

Základným prvkom textových databáz je slovo, s ktorým sa nepracuje len ako s reťazcom znakov, ale aj so slovnými kmeňmi a pádmi. Okrem toho sa využíva i výslovnosť slov, predovšetkým v angličtine. V textových databázach sa používajú okrem logických operátorov i *distančné* či *kontextové operátory*, ktoré označujú vzájomnú pozíciu slov (v jednom odseku, nie viac ako n slov za sebou apod.) [7] .

9.3 Vyhľadávanie, indexovanie

Pri vyhľadávaní v rámci textových dokumentov činí obecne problém nájsť vhodný spôsob, ako informácie obsiahnuté v textových databázach indexovať tak, aby bol čo najvernejšie zachytený ich význam. Pre popis obsahu sa používa systém vybraných slov. *Proces vyjadrenia obsahu dokumentu pomocou prvkov selekčného jazyka, obvykle s cieľom umožniť spätné vyhľadávanie sa nazýva heslovanie (indexácia)* [13] . Predpokladom pre indexovanie je analýza obsahu dokumentu. Výsledkom analýzy je minimálna skupina hesiel, ktoré danú tému popisujú. Následne sa z hesiel vyberú jednoduché výrazy, ktoré sa označujú *kľúčové slová (keywords)*. Výber kľúčových slov musí spĺňať niekoľko kritérií ako: výrazy sa musia vyskytovať s určitou frekvenciou, musia byť spojené s určitými slovnými druhmi, musia byť obsiahnuté v kľúčových miestach textu (názov, úvod, názvy kapitol apod.) [13] .

Podľa použitých metód môžeme indexáciu rozdeliť na intelektuálnu indexáciu a automatickú indexáciu:

a) **Intelektuálna indexácia.** Rola človeka je vyššia ako u automatizovanej, zapája svoje mentálne schopnosti. Z hľadiska kvality je lepšia ako automatizovaná. Na druhej strane, podľa [14] má intelektuálna indexácia aj svoje nevýhody. Prvou z nich je, že človek nedokáže indexovať a vyhodnocovať dokumenty rovnakou rýchlosťou ako stroj. Navyše sa tu objavujú subjektívne intervencie každého človeka, ktoré sú u každého jedinca rozdielne [13] .
b) **Automatická indexácia.** Delí sa na *automatickú extrakciu* (slovná indexácia), kde sa indexačné termíny získavajú z plného znenia dokumentov. Druhou je *automatické priraďovanie*. Tiež nazývané pojmová indexácia. Využíva sa porovnanie termínov z určitého slovníku alebo znalostnej bázy s výrazmi z plného textu dokumentov [13] .

Okrem zmienených indexácií sa ešte využíva poloautomatická indexácia, ktorá je založená na kombinácii automatickej a intelektuálnej indexácii.

Veľkú rolu pri vyhľadávaní zohrávajú kľúčové slová. Jedná sa o jedno alebo viacvýrazové podstatné mená, ktoré charakterizujú dokument po obsahovej stránke. Kľúčové slová sú buď prevzaté zo slovníku alebo ich vytvorí odborník na základe riadeného slovníku [13] . Vyhľadávanie podľa kľúčových slov výrazne uľahčuje identifikáciu tých dokumentov, o ktoré má užívateľ záujem. Na správnom zadaní kľúčových slov je založené aj vyhľadávanie na internete.

9.4 Selekčné jazyky

ČSN ISO 5127-6 definuje selekčný jazyk ako: *„formalizovaný jazyk používaný k charakterizovaní obsahu dokumentu alebo údajov pre potreby ich ukladania a vyhľadávania“*. Selekčný jazyk je teda určený pre identifikáciu údajov obsiahnutých v dokumentoch, pre lepšiu formuláciu požiadaviek užívateľa na vyhľadávanie a pre vhodné ukladanie dokumentov [13] .

Podľa [13] sa selekčné jazyky delia na:

1. **Systematické.** Vznikli pre potreby usporadúvania dokumentového fondu na báze prirodzeného jazyka. Patrí sem MDT – medzinárodné desatinné triedenie, DDT – Deweyho desatinné triedenie, LCC – klasifikácia Kongresovej knižnica a ďalšie.

2. **Predmetové.** Vznikli na základe potrieb prístupu ku knižnému fondu z predmetového hľadiska. Sem patria jazyky predmetových hesiel v prostredí lístkových katalógov a jazyky deskriptorového typu (odborové tezaury jedno a mnohojazyčné).

9.5 Tezaurus

Tezaurus sa využíva v indexových metódach vyhľadávania. Z hľadiska indexovania sa jedná o riadené indexovanie. Tezaurus je vždy tematicky zameraný, okrem vlastnej témy však v mnohých prípadoch spracováva čiastkovo aj iné témy, ktoré sú si blízke. Tezaurus pre svoju činnosť potrebuje indexový jazyk, čo je obmedzený počet slov určený pre indexovanie. Z týchto slov sa zostaví slovník, ktorý má okrem zoznamu slov aj hierarchické vzťahy, asociatívne vzťahy a vzťahy rovnosti medzi slovami.

Tezaurus je *„riadený a meniteľný slovník povolených štandardizovaných a spravidla hierarchicky usporiadaných výrazov, ktorý sa zakladá na bežnej terminológii príslušného vedného odboru alebo skupiny odborov (napr. geovied)“* [13] .

Tezaurus môžeme rozdeliť podľa niekoľkých kritérií. Podľa jazyku rozlišujeme jednojazyčný a viacjazyčný, podľa témy špeciálny, polytematický a univerzálny, podľa štruktúry na fasetový[6] a tematický [13] .

Z hľadiska štruktúry tvoria tezaurus 2 stavebné prvky – *deskriptory* a *nedeskriptory*:

9.5.1 Deskriptor

ČSN ISO 5964 uvádza deskriptor ako *„lexikálnu jednotku užívanú záväzne pri indexovaní k vyjadreniu určitého pojmu“*. Deskriptor je teda preferovaný výraz zaradený do tezauru. Môže sa skladať z jedného alebo viacerých slov. Deskriptory je možné zoskupovať do skupín alebo tried vzhľadom na danú tematiku. Názov skupiny môže priamo vyjadrovať i hierarchické zaradenie deskriptoru, napr [13] :

Trieda	hutníctvo železa	12
Tematická skupina	rudy Fe a Mn	1202
Deskriptor	siderit	120271

Výber deskriptoru sa riadi určitými doporučenými pravidlami[13] :

- Používajú sa výhradne podstatné mená

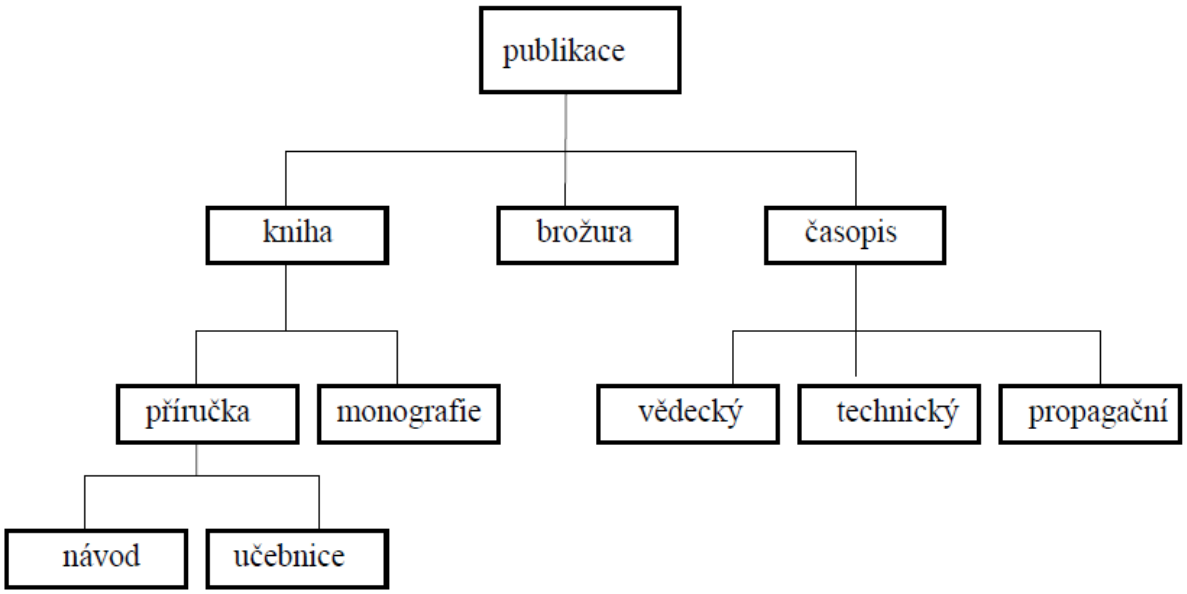
[6] Fasetová klasifikácia umožňuje zaradenie objektov do viacerých tried

- Pri výbere sa volí najpoužívanejší tvar slova
- Volí sa domáci výraz oproti cudzojazyčnému
- Výraz je spravidla v jednotnom čísle
- Homonymá sú doplňované zátvorkou, v ktorej je vysvetlenie jednotlivého termínu, napr: Napätie (elektr.), napätie (chem.), napätie (mechan.)

Ako sme už spomenuli, jedným zo základných rysov tezauru je, že je štruktúrovaný a hierarchicky usporiadaný. Medzi deskriptormi existujú 3 skupiny väzieb (vzťahov):

1. **Vzťah ekvivalencie.** Ekvivalentné sú výrazy, ktoré musia byť priradené k jednému pojmu. Pre tento typ vzťahu sa využívajú skratky UŽI a UM (užité miesto), anglicky USE a UF (use for). USE odkazuje na deskriptor z nedeskriptoru. Naopak UF uvodzuje u deskriptoru nepreferovaný výraz. Príklad z tezauru AGROVOC:
   teplota koreňov.....USE *pôdna teplota*
   pôdna teplota.....UF *teplota koreňov*
2. **Hierarchické vzťahy.** Jedná sa o väzby nadradenosti a podradenosti. Využívajú sa 2 typy vzťahov: BT (broadar term) označuje širší výraz. NT (narrower term) označuje výraz užší. Príklad tezauru GEMET:
   ihličnan.....BT *rastlina nahosemenná*
   rastlina nahosemenná.....NT *ihličnan*
3. **Vzťah afinity.** Ide o vzťah príbuznosti. Je to úzky pojmový vzťah medzi deskriptormi. Označuje sa RT (related term). Príklad z tezauru GEMET:
   fytopatológia.....RT *fyziológia rastlín*

Obr. 9-1 demonštruje konštrukciu jednoduchého tezauru nad oblasťou publikácií (použitie napríklad v knižničnom systéme).



Obr. 9-1 Schéma tezauru pre oblasť publikácií. Prevzaté od [7]

Skupinu *publikace* tvoria 3 štandardné slová. *Kniha* je BT ku *příručka*, takisto *kniha* je BT ku *monografie*. *Brožura* je NT k *publikace*.

9.5.2 Prehľad tezaurov, výber vhodného tezauru

Existuje množstvo tezaurov, voľne stiahnuteľných z internetu. Často sa vyskytujú v rôznych jazykových mutáciách. Typickým znakom každého tezauru je, že je zameraný na určitú oblasť. Podľa [15] sa v oblasti geografie používajú tieto tezaury: *„At the state of the art, different thesauri, vocabularies or taxonomies about Nature Geographic data are already available such as GEMET, EARTH, CORINE, EUNIS, NATURA2000, etc. Each of them represents a partial solution covering specific aspects in the four domains“*. Napríklad spomeňme tezaury pre český jazyk a ďalšie tezaury (bez podpory češtiny):

- **Agrovoc.** Producentom je organizácia FAO (Food and Agriculture Organization of the United Nations), spadajúca pod OSN. Vznikol v 80-tych rokoch minulého storočia. Česká verzia obsahuje v súčasnosti 38 662 lexikálnych jednotiek.
- **Český pedagogický tezaurus.** Vznikol v roku 1994. Vznikol prekladom Európskeho pedagogického tezauru.
- **Český teologický tezaurus.** Je interkonfesionálny tezaurus určený predovšetkým pre využitie v odborových knižniciach. Zameriava sa na oblasť teológie, religionistiky, filozofie a etiky. V súčasnosti obsahuje 4 992 lexikálnych jednotiek [9] .

- **EUROVOC.** Je viacjazyčný tezaurus zameraný na oblasť práva a legislatívy Európskej únie. Je polytematický, takže okrem práva pokrýva aj radu ďalších oblastí ako doprava, priemysel, vzdelávanie apod. Súčasná verzia 4.2 obsahuje 6 645 deskriptorov. EUROVOC spravuje Skupina pre správu Eurovocu (Eurovoc Maintenance Team) pri Úrade pre úradné dokumenty EU. Významným používateľom a do 1. 5. 2004 i správcom českej verzie bola Parlamentná knižnica. Parlamentná knižnica používa EUROVOC vo svojom knižničnom systéme v 4 jazykových verziách 0.
- **GEMET.** Tezaurus zameraný na oblasť životného prostredia. Producentom je Eionet (European Environment Information and Observation Netword). Obsahuje viac ako 5 000 deskriptorov, preložený je do 27 jazykov.
- **Ďalšie tezaury:** Múzejný tezaurus (producentom je Slovenské národné múzeum), Tezaurus dátových typov (od W3C), NASA Thesaurus (zameraný na kozmický výskum), MeSH (Medical Subject Heading, zameraný na obor zdravotníctva, lekárstva a príbuzných oborov, Music Thesaurus (producentom je Spindrift Music Company, zameraný na oblasť hudby) a mnoho ďalších.

Ponuka voľne dostupných tezaurov je pomerne široká. Menej je tých, ktoré majú českú jazykovú verziu. Otázkou je, ktorý tezaurus by bol vhodný použiť pre analyzovanie tém mediálnych správ. Máme niekoľko zdrojov (RSS kanálov), rôzne zameraných. Limitujúcim faktorom je nutnosť českej verzie tezauru. Ďalej je určite vhodnejšie implementovať tezaurus, ktorý je bohatý na lexikálne jednotky. Asi najlepšou voľbou bude použitie EUROVOCu alebo GEMETu. Tezaurus GEMET má však určité prednosti – je často používaný v oblasti geoinformatiky, a to aj vďaka faktu, že je aplikovaný v smernici INSPIRE[7] pri vytváraní metadát. Proti EUROVOCu hovorí aj zložitá cesta jeho získania.[8] V ďalšom texte sa budeme zaoberať popisom tezauru GEMET a jeho použitím v oblasti mediálnych správ.

9.6 Tezaurus GEMET

GEMET (the GEneral Multilingual Environmental Thesaurus) bol zostavený z niekoľkých slovníkov, jeho slová pochádzajú z niekoľkých zdrojov. Bol vyvinutý pre potreby organizácií ETC/CDS[9] a EEA[10]. Zmyslom vývoja bolo získať a využiť to najlepšie z vtedy vyvinutých

[7] INSPIRE je smernica Európskeho parlamentu a Rady 2007/2/ES zo dňa 14. 3. 2007 o zriadení Infraštruktúry pre priestorové informácie v Európskom spoločenstve

[8] Pre získanie EUROVOCu je nutné dojednať licenčnú dohodu, podpísať ju a zaslať poštou na Publications Office. Následne získame opäť poštou CD-ROM s tezaurom.

[9] ETC/CDS značí European Topic Centre on Catalogue of Data Sources

mnohojazyčných tezaurov pre potreby ušetrenia času, energií a fondov. Výber pojmov uskutočnili národní experti z jednotlivých odvetví, ktoré tvoria tezaurus.

GEMET je voľne dostupný tezaurus. Verejnosti je dostupný vo viacerých formátoch. Užívateľ môže prehľadávať tezaurus on-line zadaním kľúčového slova, prípadne listovaním v abecednom či hierarchickom zozname. Navigácia je veľmi jednoduchá. Ďalšou možnosťou je prístup cez webové služby. Posledným spôsobom je stiahnutie celého tezauru vo formáte HTML alebo SKOS.[11]

GEMET používa 2 systémy usporiadania deskriptorov: **skupinový, tematický**.

9.6.1 GEMET – skupinové usporiadanie deskriptorov

Pozostáva z klasifikačnej schémy 4 super-skupín (Tab. 9-1), a 32 skupín (Tab. 9-2). 4 super-skupiny rozdeľujú oblasť životného prostredia na 4 hlavné, najobecnejšie tézy. Ďalšie rozdelenie do 32 skupín odráža systematické rozčlenenie na kategórie, vedné disciplíny. Každá skupina patrí pod jednu super-skupinu, každý pojem patrí práve do jednej skupiny.

Tab. 9-1 Zoznam super-skupín a ich popisu

Id	Type	SubGroupOf	Label
5306	SuperGroup		PŘÍSLUŠENSTVÍ
4044	SuperGroup		LIDSKÉ AKTIVITY A PRODUKTY, DOPADY NA ŽIVOTNÍ PROSTŘEDÍ
5499	SuperGroup		PŘÍRODNÍ PROSTŘEDÍ, ANTROPICKÉ PROSTŘEDÍ
2894	SuperGroup		SOCIÁLNÍ ASPEKTY

Tab. 9-2 Zoznam skupín, ich zaradenie do super-skupín a popis každej skupiny

Id	Type	SubGroupOf	Label
96	Group	2894	SPRÁVA, ŘÍZENÍ, POLITIKA, POLITICKÉ PRAKTIKY, INSTITUCE, PLÁNOVÁNÍ
234	Group	4044	ZEMĚDĚLSTVÍ, LESNICTVÍ, VÝROBA ŽIVOČIŠNÁ, RYBÁŘSTVÍ
618	Group	5499	ATMOSFÉRA [ovzduší, klima]
893	Group	5499	BIOSFÉRA [ORGANISMY, EKOSYSTÉMY]
1062	Group	5499	ANTROPOSFÉRA
1349	Group	4044	CHEMIE, LÁTKY, PROCESY
1922	Group	2894	INFORMACE, VZDĚLÁVÁNÍ, KULTURA, EKOLOGICKÁ OSVĚTA
2504	Group	2894	EKONOMIKA, FINANCE
2711	Group	4044	ENERGIE
3875	Group	2894	ZDRAVÍ, VÝŽIVA
4125	Group	5499	HYDROSFÉRA [sladká voda, mořská voda, vodstvo]
4281	Group	4044	PRŮMYSL, ŘEMESLA, TECHNOLOGIE, ZAŘÍZENÍ
4630	Group	5499	PŮDA [krajina, geografie]
4750	Group	2894	LEGISLATIVA, NORMY, ÚMLUVY
4856	Group	5499	LITOSFÉRA [půda, geologické procesy]
6237	Group	4044	FYZIKÁLNÍ ASPEKTY, HLUK, VIBRACE, ZÁŘENÍ
7007	Group	4044	REKREACE, TURISTIKA

[10] EED značí European Environment Agency

[11] SKOS značí Simple Knowledge Organization System. Je rodina formátov pre reprezentáciu tézaurov, klasifikačných schém, taxonómii apod. Často sa používa pre sémantický web, je vyvíjaný konzorciom W3C. Vychádza z jazyka XML.

7136	Group	2894	VÝZKUM, VĚDY
7243	Group	2894	RIZIKA, BEZPEČNOST
7779	Group	2894	SPOLEČNOST
7956	Group	5499	PROSTOR
8575	Group	4044	OBCHOD, SLUŽBY
8603	Group	4044	DOPRAVA, PŘEPRAVA
9117	Group	4044	ODPADY, ZNEČIŠŤUJÍCÍ LÁTKY, ZNEČIŠTĚNÍ
10111	Group	5499	ŽIVOTNÍ PROSTŘEDÍ [přírodní prostředí, prostředí člověka]
10112	Group	4044	VÝROBKY, MATERIÁLY
10114	Group	4044	ÚČINKY, DOPADY
10117	Group	5306	OBECNÉ TERMÍNY
10118	Group	4044	ZDROJE [VYUŽITÍ ZDROJŮ]
13109	Group	2894	POLITIKA EKOLOGICKÁ
14979	Group	5499	DOBA [chronologie]
14980	Group	5306	FUNKČNÍ TERMÍNY

Pojem však nemusí byť priamym potomkom skupiny, ale pomocou vzťahov BT a NT sa môže dostať až niekoľko úrovní pod skupinami. Takže napríklad pojem *infiltrace vody do podzemí* je niekoľko úrovní pod skupinou *HYDROSFÉRA [sladká voda, mořská voda, vodstvo]* (Obr. 9-2).

Obr. 9-2 Online prostredie pre hierarchické listovanie v pojmoch

Pojem *infiltrace vody do podzemí* je NT k pojmu *infiltrace*. Naopak pojem *infiltrace* je BT k pojmu *infiltrace vody do podzemí*. Oba tieto pojmy patria do skupiny *HYDROSFÉRA [sladká voda, mořská voda, vodstvo]*. Zvláštnu úlohu v GEMETe má väzba RT (related term). Pojmy, ktoré majú medzi sebou tento typ väzby nemusia patriť do tej istej skupiny ani do tej istej témy. Napriek tomu medzi sebou nejakým spôsobom súvisia, napríklad slová paplón – perie, antonymá (riziko – bezpečie), nástroje a procesy (maľovanie – štetka) apod.

9.6.2 GEMET – tematické usporiadanie deskriptorov

Tematické rozdelenie odpovedá usporiadaniu deskriptorov do 40 tém. Boli zostavené podľa požiadaviek organizácie EEA, zohľadňujúc aktuálne informačné potreby. Zoznam tém berie do úvahy informačné zdroje, ktoré boli základom pre tvorbu GEMETu, ako The Dobris Assessment, DPSIR Dataflow Scheme apod. Každý deskriptor môže byť zaradený do viacerých tém. Deskriptory s príliš obecným obsahom sú zaradené do špeciálnej témy nazvanej *všeobecně*.

Tab. 9-3 Zoznam tém

Id	Type	Label
1	Theme	ADMINISTRATIVA
2	Theme	ZEMĚDĚLSTVÍ
3	Theme	OVZDUŠÍ
18	Theme	ŽIVOČIŠNÁ VÝROBA
4	Theme	BIOLOGIE
5	Theme	VÝSTAVBA
6	Theme	CHEMIE
7	Theme	KLIMA
32	Theme	RIZIKA, HAVÁRIE
9	Theme	EKONOMIKA
10	Theme	ENERGIE
11	Theme	ENVIROMENTÁLNÍ POLITIKA
12	Theme	RYBÁŘSTVÍ
13	Theme	POTRAVINY, PITNÁ VODA
14	Theme	LESNICTVÍ
15	Theme	VŠEOBECNĚ
16	Theme	GEOGRAFIE
17	Theme	LIDSKÉ ZDRAVÍ
19	Theme	PRŮMYSL
20	Theme	INFORMACE
21	Theme	LEGISLATIVA
27	Theme	MATERIÁLY
22	Theme	VOJENSKÉ ASPEKTY
23	Theme	PŘÍRODNÍ ÚZEMÍ, KRAJINA, EKOSYSTÉMY
8	Theme	PŘÍRODNÍ DYNAMIKA
24	Theme	HLUK, VIBRACE
25	Theme	FYZIKA
26	Theme	ZNEČIŠĚNÍ
28	Theme	RADIACE
30	Theme	VĚDA A VÝZKUM
31	Theme	ZDROJE
34	Theme	SOCIÁLNÍ ASPEKTY, POPULACE
35	Theme	PŮDA
36	Theme	PROSTOR
29	Theme	TURISTIKA
33	Theme	OBCHOD, SLUBY
37	Theme	DOPRAVA
38	Theme	URBANIZOVANÉ PROSTŘEDÍ
39	Theme	ODPADY
40	Theme	VODA

9.6.3 Implementácia tezauru GEMET

Asi najjednoduchším spôsobom implementácie GEMETu je jeho stiahnutie do vlastného prostredia. Na web-stránkach je k dispozícii skupina dokumentov vo formáte jednoduchých HTML tabuliek a vo formátoch RDF či SKOS. SKOS je vybudovaný na formáte RDF a RDF vychádza z XML. V ďalšom texte budeme hovoriť o XML dokumentoch a nebudeme rozlišovať, či sa jedná o RDF alebo SKOS, pri čítaní týchto dokumentov si vystačíme so základnou znalosťou XML technológie.

Celý tezaurus GEMET možno zhrnúť do 4 vystavených XML dokumentov. Z obsahu alebo z atribútov jednotlivých elementov sme schopní porozumieť celému tezauru. V týchto dokumentoch sú zachytené všetky pojmy, skupiny, super-skupiny, témy, vzťahy medzi pojmami, príslušnosť skupín do superskupín a príslušnosť pojmov do skupín či tém:

1. Dokument **gemet-groups.rdf**. Okrem koreňového elementu obsahuje 3 skupiny elementov, zamerané na popis super-skupín, skupín a tém.
   a. Element <*SuperGroup*>. V atribúte *rdf:about* sa ukrýva ID super-skupiny. V sub-elemente <*rdfs:label*> je názov superskupiny.

```
<SuperGroup
rdf:about="http://www.eionet.europa.eu/gemet/supergroup/4044">
<rdfs:label>
LIDSKÉ AKTIVITY A PRODUKTY, DOPADY NA ŽIVOTNÍ PROSTŘEDÍ
</rdfs:label>
</SuperGroup>
```

   b. Element <*Group*>. V atribúte *rdf:about* sa ukrýva ID skupiny. V sub-elemente <*rdfs:label*> je názov skupiny.

```
<Group rdf:about="http://www.eionet.europa.eu/gemet/group/618">
<subGroupOf
rdf:resource="http://www.eionet.europa.eu/gemet/supergroup/5499"/>
<rdfs:label>ATMOSFÉRA [ovzduší, klima]</rdfs:label>
</Group>
```

   c. Element <*Theme*>. V atribúte *rdfs:label* je názov témy. Atribút *rdf:about* skrýva ID témy.

```
<Theme rdfs:label="administrativa"
rdf:about="http://www.eionet.europa.eu/gemet/theme/1"/>
```

2. Dokument **gemet-definitions.rdf.** Obsahuje identifikátory a popisy pojmov. Okrem koreňového elementu obsahuje: element <*rdf:Description*>, ktorý v atribúte *rdf:about* ukrýva ID pojmu. Elementu <*rdf:Description*> obsahuje sub-element

<*skos:prefLabel*>, ktorý obsahuje popis pojmu.

```
<rdf:Description
rdf:about="http://www.eionet.europa.eu/gemet/concept/7">
<skos:prefLabel>místo průmyslové opuštěné</skos:prefLabel>
</rdf:Description>
```

3. Dokument **gemet-backbone.rdf**. Definuje príslušnosť pojmov k skupinám a k témam. Každý pojem patrí k jednej skupine, každý pojem môže patriť k niekoľkým témam. Súbor okrem koreňového elementu obsahuje element <*rdf:Description*> so sub-elementmi <*gemet:theme*> a <*gemet:group*>. Element <*rdf:Description*> v atribúte *rdf:about* ukrýva ID deskriptoru. Sub-element <*gemet:theme*> v atribúte *rdf:resource* ukrýva ID témy, ktorej členom je daný deskriptor. Sub-element <*gemet:group*> v atribúte *rdf:resource* ukrýva ID skupiny, ktorej členom je daný deskriptor. Sub-element <*gemet:group*> sa pri každom pojme vyskytuje jedenkrát, sub-element <*gemet:theme*> sa môže vyskytovať jeden alebo viackrát.

```
<rdf:Description rdf:about="http://www.eionet.europa.eu/gemet/concept/8192">
<gemet:theme rdf:resource="http://www.eionet.europa.eu/gemet/theme/6"/>
<gemet:group rdf:resource="http://www.eionet.europa.eu/gemet/group/1349"/>
</rdf:Description>
```

4. Dokument **gemet-skoscore.rdf.** Súbor zachycuje vzťahy medzi pojmami. Vzťahy môžu byť: BT (broader term), NT (narrower term), RT (related term). Súbor obsahuje okrem koreňového elementu element <*skos:Concept*>, ktorý v atribúte *rdf:about* ukrýva ID pojmu. Element <*skos:Concept*> musí obsahovať minimálne jeden zo skupiny sub-elementov <*skos:broader*>, <*skos:narrower*> *a* <*skos:related*>. Všetky sub-elementy majú v atribúte *rdf:resource* ukryté ID pojmu, na ktorý vedie ich vzťah. Z uvedených sub-elementov musí byť zastúpený minimálne jeden zo skupiny <*skos:broader*>, <*skos:narrower*> a to jeden alebo viackrát. Sub-element <*skos:related*> nemusí byť zastúpený vôbec alebo je zastúpený jeden alebo viackrát.

```
<skos:Concept rdf:about="http://www.eionet.europa.eu/gemet/concept/11">
<skos:broader rdf:resource="http://www.eionet.europa.eu/gemet/concept/2457"/>
<skos:narrower rdf:resource="http://www.eionet.europa.eu/gemet/concept/470"/>
<skos:narrower
rdf:resource="http://www.eionet.europa.eu/gemet/concept/3640"/>
<skos:related rdf:resource="http://www.eionet.europa.eu/gemet/concept/1462"/>
</skos:Concept>
```

Pri analyzovaní tém správ by práca so samotnými XML súbormi bola dosť náročná. XML bol vyvinutý najmä pre účely prenosu dát. Jednou z nevýhod XML je, že jeho syntax je obšírny

a súbory sa tým stávajú pomerne veľké, čo môže spôsobiť problémy na sieti. Z týchto dôvodov boli XML súbory pretransformované na jednoduché tabuľky, ktoré obsahujú jednak všetky zoznamy pojmov, skupín, super-skupín a tém a jednak vzťahy medzi objektmi. K tomuto účelu slúžia 4 navrhnuté skripty v jazyku PHP, ktoré prevedú spomínané XML súbory do podoby MySQL tabuliek v databáze (Tab. 9-4).

Tab. 9-4 Popis transformácie vstupných XML súborov do tabuliek

Názov skriptu	Vstupný súbor	Názov výstupnej tabuľky (tabuliek)	Atribúty výstupnej tabuľky
gemet_groups.php	gemet-groups.rdf	gemet_supergroups	ID, LABEL
		gemet_groups	ID, LABEL
		gemet_themes	ID, LABEL
gemet_definitions.php	gemet-definitions.rdf	gemet_definitions	ID, LABEL
gemet_backbone.php	gemet-backbone.rdf	gemet_backbone	ID, RELATION, OBJECT
gemet_skoscore.php	gemet-skoscore.rdf	gemet_skoscore	ID, RELATION, OBJECT

Úvodné 3 výstupné tabuľky tvoria zoznamy super-skupín, skupín, tém a pojmov. Vzťahy sú zachytené v tabuľke *gemet_backbone* a *gemet_skoscore*.

Tabuľka *gemet_backbone* znázorňuje príslušnosť pojmu do skupiny alebo témy. *ID* predstavuje identifikátor pojmu, *OBJECT* je identifikátor skupiny alebo témy, *RELATION* má hodnotu *theme* alebo *group*, podľa toho, či sa jedná o príslušnosť pojmu do skupiny alebo témy.

Tabuľka *gemet_skoscore* zachycuje vzťahy medzi pojmami. *ID* je identifikátor prvého pojmu, *RELATION* obsahuje hodnoty *broader, narrower* alebo *related*. *OBJECT* je identifikátor druhého objektu. Ak je teda záznam v tvare *ID:7, RELATION:narrower* a *OBJECT:4666*, v tom prípade je pojem s identifikátorom 7 užší (narrower) ako pojem s identifikátorom 4666.

9.6.4 Použitie GEMETu pre analýzu tém správ

Samotná štruktúra tezauru GEMET nám do určitej miery určuje, akým spôsobom ho môžeme využiť pre oblasť spravodajstva. Zhrňme si, že základným prvkom štruktúry je deskriptor. Každý deskriptor patrí jednak práve do jednej skupiny (počet skupín je 32), každá skupina patrí práve do jednej superskupiny (počet superskupín je 4). Každý deskriptor patrí do jednej alebo viacerých tém (počet tém je 40). Navyše, medzi jednotlivými pojmami sú vzťahy hierarchické (širší pojem BT, užší pojem NT) a vzťahy afinity (označované RT). Vzťahy hierarchie a afinity priamo nesúvisia so zaradením pojmov do skupín a tém.

Štruktúra tezauru GEMET nám determinuje niekoľko možností spracovania tém v oblasti spravodajstva. Prvou úrovňou je vyhľadanie jednotlivých pojmov vo fulltextoch správ. Pre každý pojem určíme, v ktorých správach sa objavil, prípadne počet výskytov pojmu v jednotlivých správach. Ďalšiu úroveň tvorí rozdelenie správ (v ktorých boli identifikované pojmy z GEMETu) podľa skupín pojmov alebo podľa tém. Na tejto úrovni teda definujeme, ktoré skupiny pojmov a ktoré témy sa v správach objavili. Agregáciou priradenia správ ku skupinám sa môžeme jednoducho dostať k priradeniu správ do jednotlivých superskupín. Poslednou úrovňou je spracovanie podobnosti (príbuznosti) správ na základe vzťahov medzi jednotlivými pojmami, ktoré sa v týchto správach vyskytujú. Na tejto úrovni môžeme určiť, ktoré správy pojednávajú o príbuzných pojmoch.

9.6.4.1 Vyhľadanie pojmov v správach

Vyhľadanie pojmov v správach je nevyhnutným predpokladom pre ďalšie úrovne spracovania. V texte každej správy sa môže objaviť jeden, niekoľko, prípadne žiadny pojem z tezauru GEMET. Pri vyhľadávaní pojmov musíme vziať do úvahy niekoľko faktorov. Z povahy definície tezauru vyplýva, že pojmy v ňom obsiahnuté sú spravidla podstatné mená, v prvom páde, v najpoužívanejšom tvare, jednotnom čísle (ďalšie obmedzenia v kap. 9.5.1). Vyhľadávanie však uplatňujeme na textoch, ktoré sú písané prirodzeným jazykom. Preto by bolo ideálne využiť aplikáciu, ktorá po zadaní pojmu vráti tento pojem vo všetkých pádoch, prípadne naopak, z pojmu v určitom páde vráti jeho základný tvar. V čase písania práce sa bohužiaľ nepodarilo získať voľne dostupný nástroj, ktorý by umožňoval spracovanie pádov či vzorov pre skloňovanie a ktorý by bol jednoducho možný implementovať. Pomerne zaujímavý je projekt *AJKA* (morfologický analyzátor češtiny), spracovaný na Masarykovej univerzite v Brne. Tento produkt na základe vloženého slova či dlhšieho textu, dokáže z tohto pojmu alebo textu získať slová v základnom tvare. Žiaľ, program nie je voľne dostupný. Je možné spracovanie textu v prostredí webu, text je však obmedzený veľkosťou 2KB. Navyše, program sa už dlhšiu dobu nevyvíja.

Pre ďalšie spracovanie zjednodušíme vyhľadávanie v textoch na priame vyhľadávanie pojmov z tezauru. V prvej fáze budeme vyhľadávať pojmy z tezauru v fulltextoch správ kanálu ČT24, rubrika regionálneho spravodajstva, z obdobia od 1. 1. 2008 do 15. 4. 2010. K tomuto účelu bol vytvorený PHP skript s názvom *gemet_def_spravy_net.php*. Výsledkom skriptu je tabuľka s určením, ktoré pojmy sa nachádzajú v ktorých správach a v akom počte. Skript pracuje podľa nasledujúceho postupu:

1. Načítaj správu z databázy. Odstráň zo správy interpunkciu, špeciálne znaky ako úvodzovky.
2. Rozdeľ správy do slov.
3. Prechádzaj postupne každý deskriptor. Zisti počet slov v deskriptore (označme *n*). Spoj slová v správe do *n-tíc*. Jednu *n-ticu* tvorí n slov, ktoré sú v poradí za sebou.
4. Porovnaj každý deskriptor s každou vytvorenou *n-ticou*.
5. Ak sa nájde zhoda prípadne viac zhôd deskriptoru s *n-ticou*, zapíš do tabuľky *gemet_def_spravy_regio* identifikátor deskriptoru, identifikátor správy, názov deskriptoru a počet zhôd (Tab. 9-5).

Tab. 9-5 Výsledok vyhľadávania deskriptorov vo fulltextoch správ zo zdroja ČT24 regionálne spravodajstvo. Použitý tezaurus: GEMET. Časové vymedzenie správ: 1. 1. 2008 - 15. 4. 2010

id_definition	definition	id_spravy	pocet
25	Havárie	6730	3
25	Havárie	1630	1
243	AIDS	2318	1
1751	Půda kontaminovaná	7397	1
...	...	...	...

Táto tabuľka bude tvoriť jeden zo vstupov pre ďalšie analýzy, kde ju budeme kombinovať s ďalšími, už vytvorenými tabuľkami.

Z výsledkov vyplynulo, že z celkového počtu pojmov (5206), ktoré sa nachádzajú v tezaure GEMET, bolo v správach identifikovaných 762 jedinečných pojmov. Celkový počet spracovávaných správ je 5 250. Zoznam 20 najčastejšie sa vyskytujúcich pojmov ilustruje tab. 9-6.

tab. 9-6 Zoznam najčastejšie sa vyskytujúcich pojmov

Deskriptor	Počet výskytov	Deskriptor	Počet výskytov
Město	729	Sdružení	302
Mapa	713	Obyvatel	293
Policie	503	Společnost	280
Stát	416	Území	256

Práce	402	Ministerstvo	239
Úřad	337	Náměstí	214
Rozhodnutí	337	Soud	213
Projekt	336	Služby	211
Firma	328	Obec	203
Nemocnice	318	Náklady	180

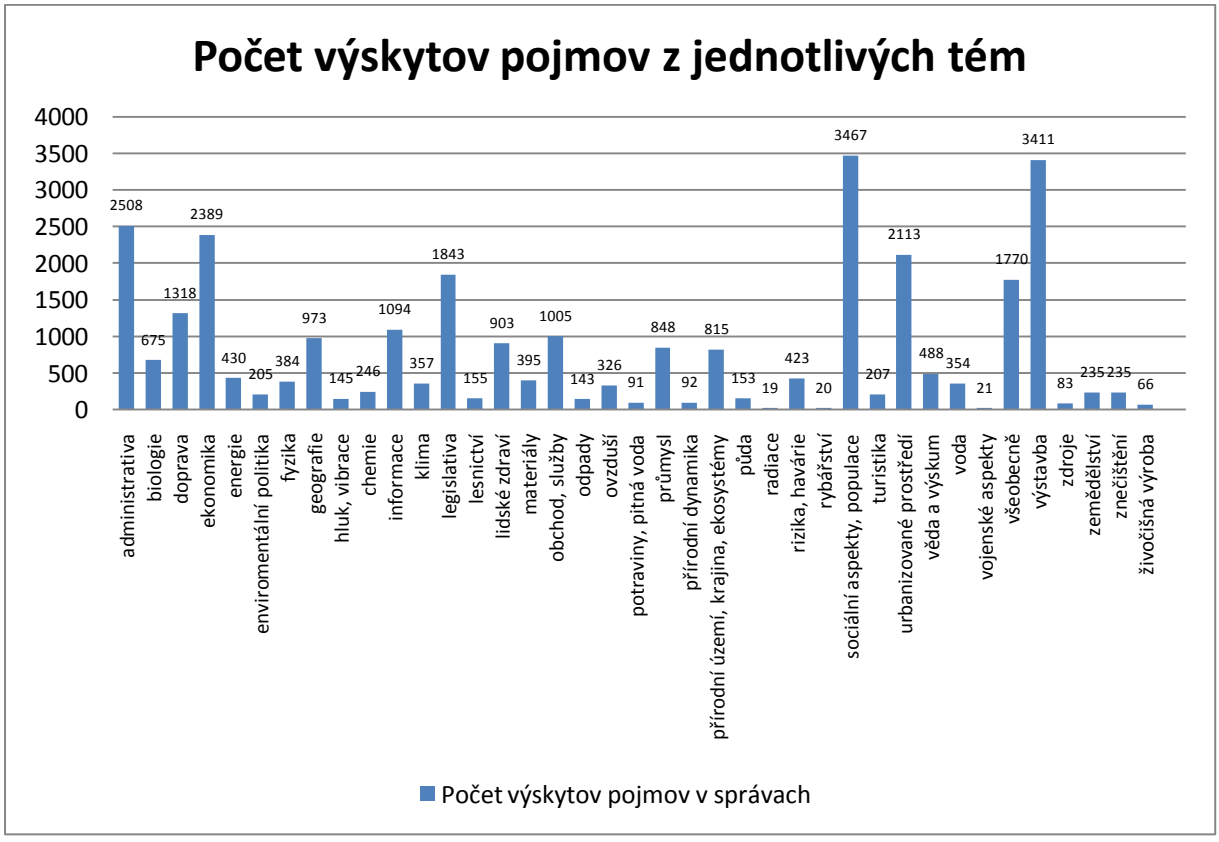
9.6.4.2 Rozdelenie správ podľa skupín pojmov a tém

Vstupom do rozdelenia správ podľa skupín pojmov a tém sú 2 už vytvorené tabuľky. Prvú tabuľku tvoria výsledky vyhľadávania pojmov z predchádzajúcej kapitoly, druhou je už vytvorená tabuľka *gemet_backbone*, obsahujúca zaradenie pojmov do skupín a tém. (vytvorenie a obsah tabuľky *gemet_backbone* popísaný v kap. 519.6.3). Výsledky pre rozdelenie správ podľa skupín zachycuje Obr. 9-3.

Obr. 9-3 Počet pojmov vyskytujúcich sa v správach k danej skupine.

Najvyšší podiel v spravodajstve má skupina antroposféra, ktorej pojmy sa objavili v správach v počte 3 568. Na druhom mieste je skupina *správa, řízení, politika, politické praktiky, instituce, plánování* (2 521) *společnost* (2 051). Kategória *antroposféra* má výrazne vyššie zastúpenie ako ostatné kategórie. Dôvodom je pravdepodobne fakt, že sa jedná o spravodajstvo a spravodajstvo sa obecne zaujíma o ľudské činnosti, čomu nasvedčuje aj veľké zastúpenie kategórií *společnost* a *správa, řízení, politika, politické praktiky, instituce, plánování.* Z kategórií, ktoré majú povahu skôr prírodných, majú mierne vyššie zastúpenie *biosféra [organismy, ekosystémy]* (636) a *výzkum, vědy* (476). Najmenej výskytov sa objavilo zo skupiny *obecné termíny.* Malý počet výskytov tejto skupiny vychádza z faktu, že v samotnom tezaure sa k skupine viažu len 2 termíny.

Rozdelenie správ podľa tém vzniklo obdobným spôsobom ako rozdelenie správ podľa skupín. Výsledky zhrňuje Obr. 9-4.

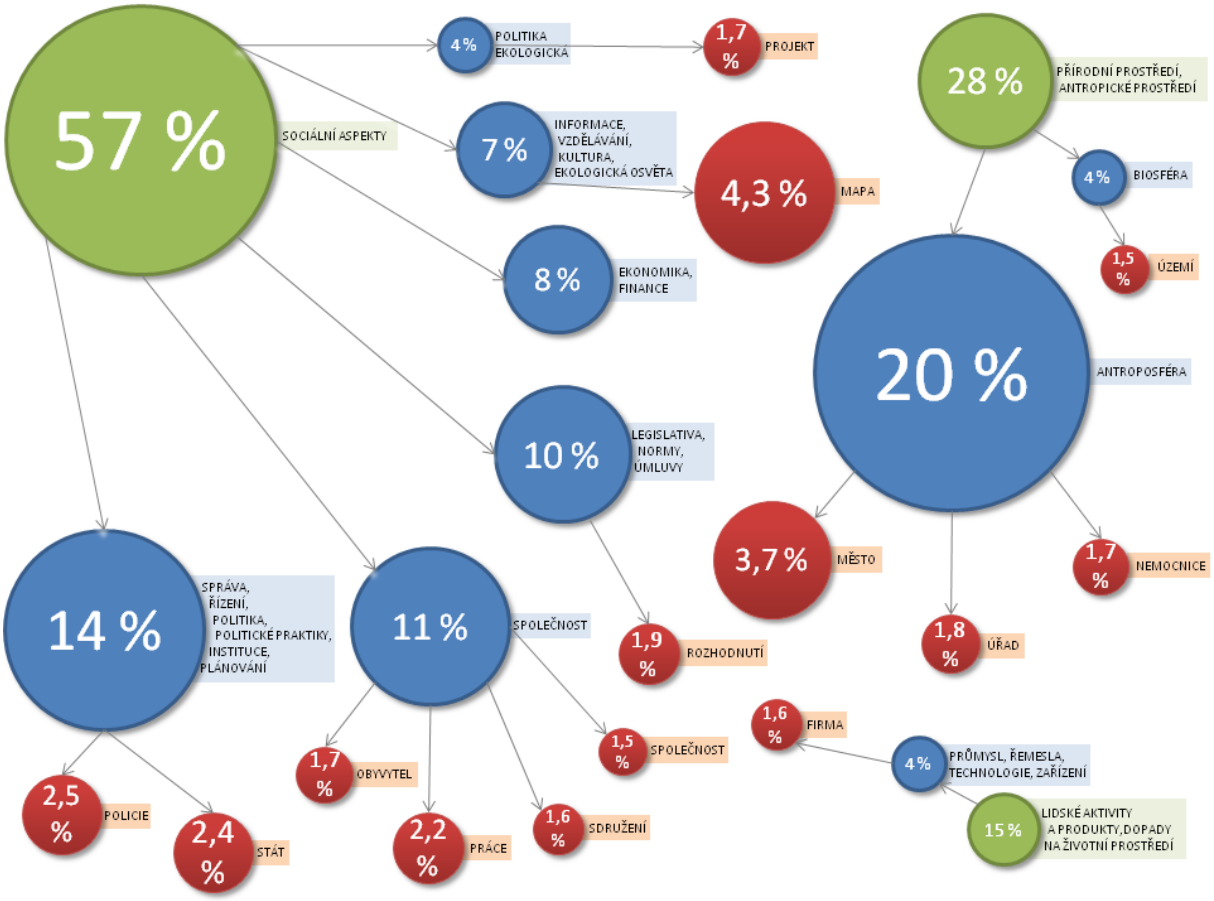


Obr. 9-4 Počet pojmov vyskytujúcich sa v správach k danej téme

Z tematického hľadiska je situácia mierne odlišná. 2 témy sú zastúpené výrazne vyššie ako ostatné. Opäť je potvrdený charakter, že vyšší počet výskytov majú témy týkajúce sa spoločnosti, sociálnych aspektov a dôsledkov ľudskej činnosti. Najvyššie zastúpenie

v správach má téma *sociální aspekty, populace* (3 467). Druhé miesto patrí téme *výstavba* - jej pojmy sa vyskytli v správach 3411 krát. Nad 2 000 výskytov pojmov sa vošli ešte témy administratíva a ekonomika. Témy zachycujúce skôr prírodné aspekty a aspekty životného prostredia majú počet výskytov pod hodnotou 1000, najviac téma *přírodní území, krajina, ekosystémy* (815). Pomerne vysoké zastúpenie má téma *geografie* (973). Naopak najmenšie zastúpenie majú témy *radiace* (19), *rybářství* (20), *vojenské aspekty* (21).

Zaujímavý pohľad na výsledky spracovania nám môže poskytnúť zobrazenie v podobe, ktorá jednak zachycuje hierarchickú štruktúru použitého tezauru, jednak vyjadruje intenzitu výskytov pojmov, vyjadrenú veľkosťou kruhu. Znázornime super-skupiny v podobe zelených kruhov, skupiny modrými kruhmi a jednotlivé pojmy červenými kruhmi. Percentuálne zastúpenie pojmu, skupiny či superskupiny v rámci jej úrovne vyjadríme okrem číselného percentuálneho udania aj veľkosťou kruhu. Dostaneme pomerne rozmanitú štruktúru (Obr. 5-1).



Obr. 9-5 Hierarchicky znázornené zastúpenie pojmov, skupín a superskupín v znení správ. Kanál ČT24 - regionálny, správy od 1. 1. 2008 do 15. 4. 2008

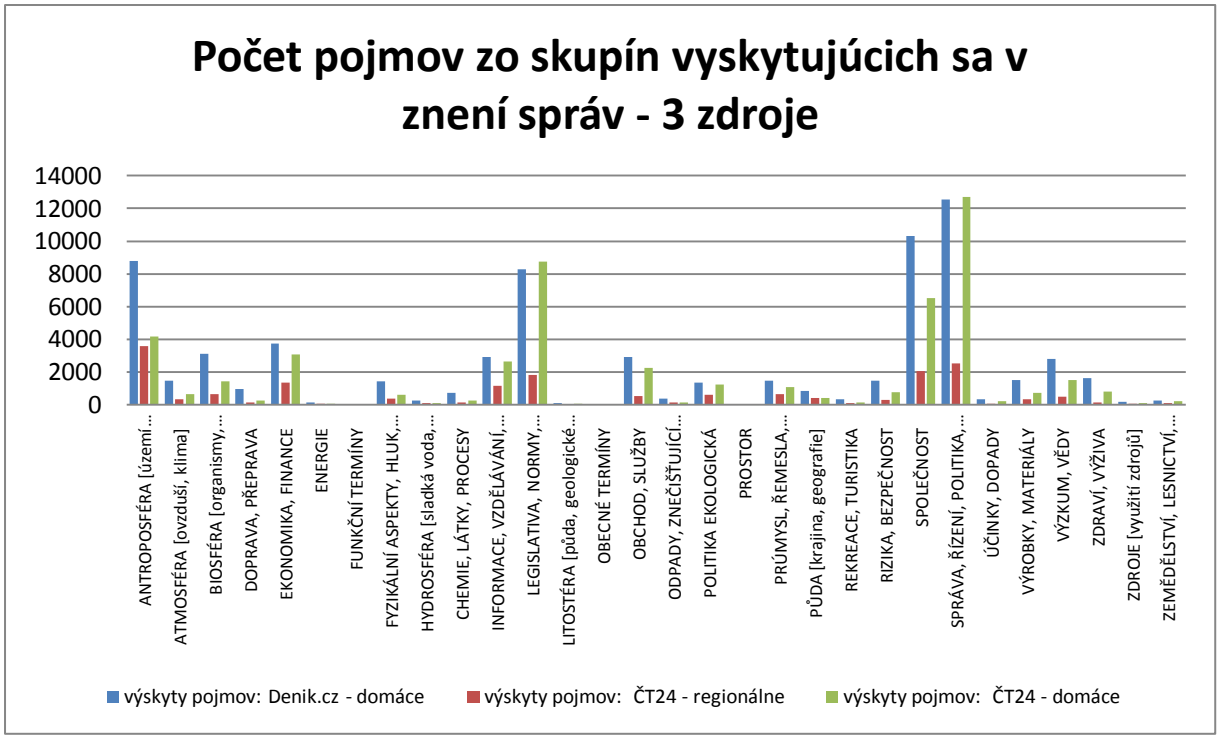
Pri tvorbe schémy sa vždy ku každej super-skupine, skupine či pojmu vyjadrilo jeho percentuálne zastúpenie oproti ostatným pojmom (skupinám či super-skupinám) na jeho úrovni. Pre každú úroveň je stanovená vlastná škála pre veľkosti kruhov, aby bola schéma názornejšia. Čo sa týka samotných pojmov, neboli, opäť kvôli prehľadnosti, zaradené všetky, ale len tie, ktorých zastúpenie je vyššie ako 1,5 %. Zo skupín sa na schému dostali len tie, ktorých podiel je vyšší ako 4 %. Šípky znázorňujú príslušnosť pojmu do skupiny, či skupiny do superskupiny.

Schéma zachycuje 3 super-skupiny. Štvrtá super-skupina (*příslušenství)* mala podiel len 0,5 percenta, preto nebola zaradená. Z pohľadu super-skupín ma najvyšší podiel *Sociální aspekty* (57 %), v ktorej je zaradených najviac zastúpených skupín.

Z hľadiska hierarchie je zaujímavé, že prvky s najvyšším zastúpením nepatria do nadúrovne s najvyšším zastúpením. Napríklad pojem *mapa*, ktorý je najčastejšie spomínaný v správach, patrí do skupiny, ktorej percentuálny podiel vôbec nepatrí k najvyšším (7 %). Ak by sme sa presunuli o úroveň vyššie, podobne je to so skupinou antroposféra (20 %), ktorá nepatrí do najviac zastúpenej super-skupiny.

9.6.4.3 Porovnanie viacerých zdrojov

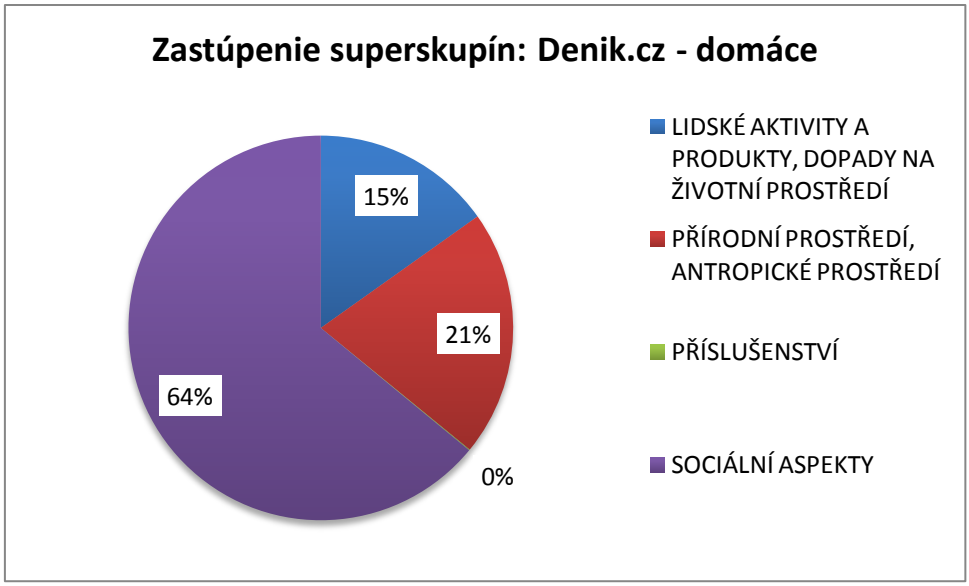
Skúsme z hľadiska skupín pojmov zaradiť do analýzy viac kanálov. Pri výbere kanálov do analýzy sa bral ohľad na výsledky z kap. 8. Brala sa teda v úvahu vhodnosť jednotlivých kanálov pre spracovanie. Výber pozostáva z 3 zdrojov – Denik.cz rubrika domáceho spravodajstva, ČT24 – domáce spravodajstvo a ČT24 – regionálne spravodajstvo. Výber bral ohľad aj na to, aby bolo možné porovnať jednak viac kanálov od rôznych poskytovateľov jednak 2 kanály od jedného poskytovateľa. Najprv sa pozrieme na zastúpenie jednotlivých skupín (Obr. 9-6):



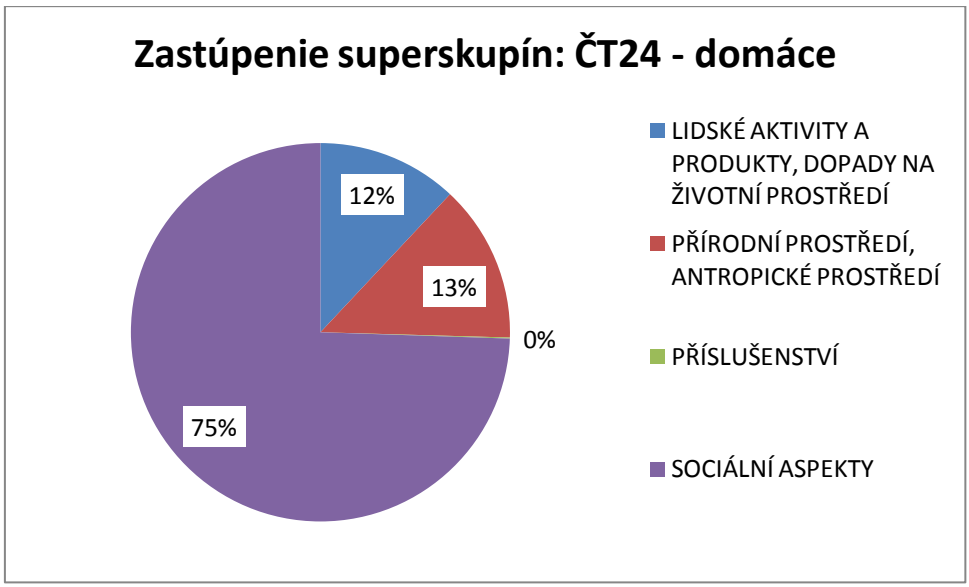
Obr. 9-6 Počet pojmov vyskytujúcich sa v správach k danej téme – súhrn pre 3 rôzne zdroje. Jedná sa o správy z obdobia od 1. 8. 2008 do 15. 4. 2010

Graf zachycuje absolútne počty výskytov pojmov z jednotlivých skupín. Veľmi podobné sú výsledky pre skupinu *správa, řízení, politika, politické praktiky, instituce, plánování* pre zdroje ČT24 – domáce a Denik.cz – domáce (12 702 a 12 539). Opäť sa potvrdil trend, že vyššie podiely majú skupiny týkajúce sa ľudských aktivít a spoločnosti. Jedná sa o skupiny *antroposféra, společnost*, *legislativa*, *normy* a už spomenutá *správa, řízení, politika, politické praktiky, instituce, plánování.* Výrazné rozdiely medzi kanálom ČT24 – regionálne a ostatnými je spôsobený nízkym počtom správ pre tento kanál (asi polovičný počet správ oproti ostatným dvom).

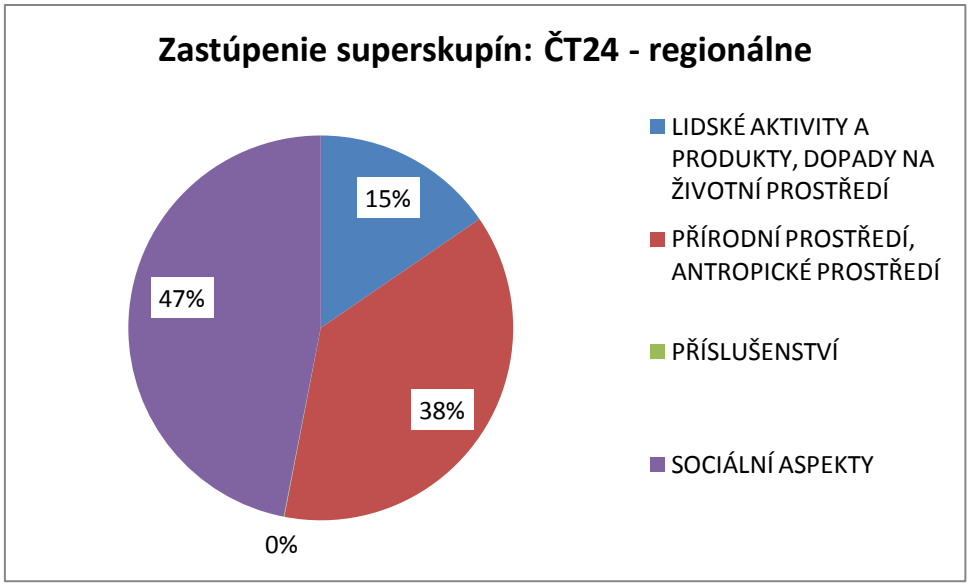
Na úrovni superskupín sa pozrieme na percentuálne zastúpenie jednotlivých super-skupín v spravodajstve vo vybraných kanáloch. Na prvý pohľad je zastúpenie superskupín približne rovnaké vo všetkých troch vybraných kanáloch. Regionálne spravodajstvo má však mierne vyšší podiel super-skupiny *přírodní prostředí, antropické prostředí* (viac ako dvojnásobne oproti kanálu ČT24 – domáce.



9-7 Zastúpenie superskupín tezauru GEMET v spravodajstve kanálu Denik.cz - domáce. Správy pochádzajú z časového rozmedzia od 1. 1. 2008 do 15. 4. 2010



9-8 Zastúpenie superskupín tezauru GEMET v spravodajstve kanálu ČT24 - domáce. Správy pochádzajú z časového rozmedzia od 1. 1. 2008 do 15. 4. 2010



9-9 Zastúpenie superskupín tezauru GEMET v spravodajstve kanálu ČT24 - regionálne. Správy pochádzajú z časového rozmedzia od 1. 1. 2008 do 15. 4. 2010

9.6.5 Ukážka využitia tematického, časového a priestorového štruktúrovania spravodajstva

Skúsme navrhnúť situáciu, v ktorej využijeme poznatky získané behom spracovania spravodajstva. Analýza štruktúry spravodajstva nám môže pomôcť odpovedať na rôzne druhy špecifických otázok, zahrňujúcich v sebe rôzne aspekty. Môže nás napríklad zaujímať, v ktorých obdobiach vystupujú do popredia určité témy a v ktorom čase sú naopak v úzadí. Ďalej nás môže zaujímať, či určitá lokalita je uprednostňovaná oproti iným alebo je jej naopak venovaná neúmerne malá pozornosť. Sme schopní odpovedať na dotazy typu: *Zisti, v ktorom mesiaci sa najviac diskutovalo o poľnohospodárstve v danej lokalite*. Alebo *Má obľúbenosť témy výstavby za posledné obdobie stúpajúci charakter?*. Samozrejmým predpokladom pre odpovedanie na podobné typy dotazov je vhodná reprezentácia dát. V nasledujúcom texte sa pokúsime namodelovať úlohu a vyriešiť ju využitím poznatkov získaných pri spracovávaní mediálnych správ.

9.6.5.1 Téma Ostrava a ovzdušie v mediálnych správach

Bude nás zaujímať, či téma znečistenia ovzdušia v okolí Ostravy je naozaj témou žhavou, či sa v súvislosti so znečistením výrazne častejšie hovorí o Ostrave (napríklad v čase písania práce sa hovorilo o súvise medzi znečistením v Ostrave a zvýšeným výskytom astmy vo veľmi skorom veku detí). Ďalej nás bude zajímať, či sú obdobia, v ktorých sa o znečistení hovorí viac či menej (typický smog, trápiaci Ostravu a okolie najmä v zime).

Pri riešení by sme mohli začať zistením, ktoré témy či skupiny sa môžu eventuálne dotýkať témy znečistenia. Z ponúknutých skupín, ktoré sú obsiahnuté v tezaure GEMET sa

k znečisteniu a atmosfére pravdepodobne najviac viažu témy: *atmosféra; chemie, látky procesy; rizika, bezpečnost; odpady, znečišťující látky, znečištění*.

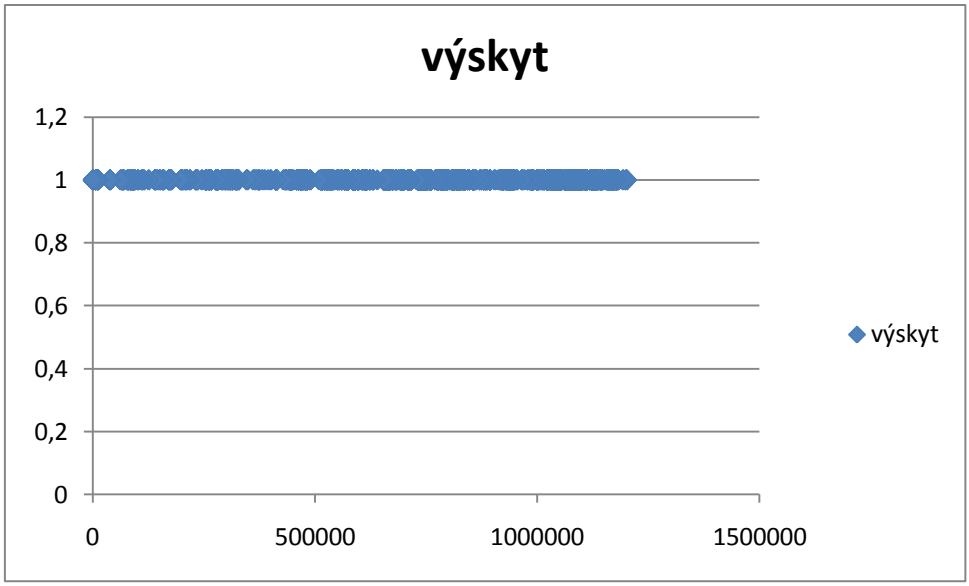
Poďme sa pozrieť na početnosť, resp. zastúpenie jednotlivých tém a obcí v správach Tab. 9-7.

Tab. 9-7 Zastúpenie obcí v správach v súvislosti s témou

SKUPINA	OBEC	POČET SPRÁV
RIZIKA, BEZPEČNOST	Praha	546
ATMOSFÉRA [ovzduší, klima]	Praha	477
ÚČINKY, DOPADY	Praha	171
CHEMIE, LÁTKY, PROCESY	Praha	169
ODPADY, ZNEČIŠŤUJÍCÍ LÁTKY, ZNEČIŠTĚNÍ	Praha	99
RIZIKA, BEZPEČNOST	Brno	52
ATMOSFÉRA [ovzduší, klima]	Brno	52
RIZIKA, BEZPEČNOST	Ostrava	51
ATMOSFÉRA [ovzduší, klima]	České Budějovice	48
ATMOSFÉRA [ovzduší, klima]	Plzeň	33
ATMOSFÉRA [ovzduší, klima]	Ostrava	32
ATMOSFÉRA [ovzduší, klima]	Pec pod Sněžkou	30
ATMOSFÉRA [ovzduší, klima]	Špindlerův Mlýn	29
RIZIKA, BEZPEČNOST	Plzeň	23
RIZIKA, BEZPEČNOST	Vítkov	23
ÚČINKY, DOPADY	Brno	23
ATMOSFÉRA [ovzduší, klima]	Hradec Králové	22
ODPADY, ZNEČIŠŤUJÍCÍ LÁTKY, ZNEČIŠTĚNÍ	Ostrava	21
CHEMIE, LÁTKY, PROCESY	Brno	19
CHEMIE, LÁTKY, PROCESY	Ostrava	19

Z tabuľky je zjavné, že Ostrava nevedie v žiadnej z ponúkaných tém z hľadiska počtu správ, v ktorých sa v súvislosti s danou témou vyskytla. Napríklad Praha sa objavila v 99 správach, v ktorých bola preberaná téma Odpady a znečisťujúce látky. Naproti tomu Ostrava 21 krát. Z tabuľky je však zjavná aj iná záležitosť – napr. v súvislosti s atmosférou sa často nejedná o jej znečistení – v prípade obce Pec pod Sněžkou.

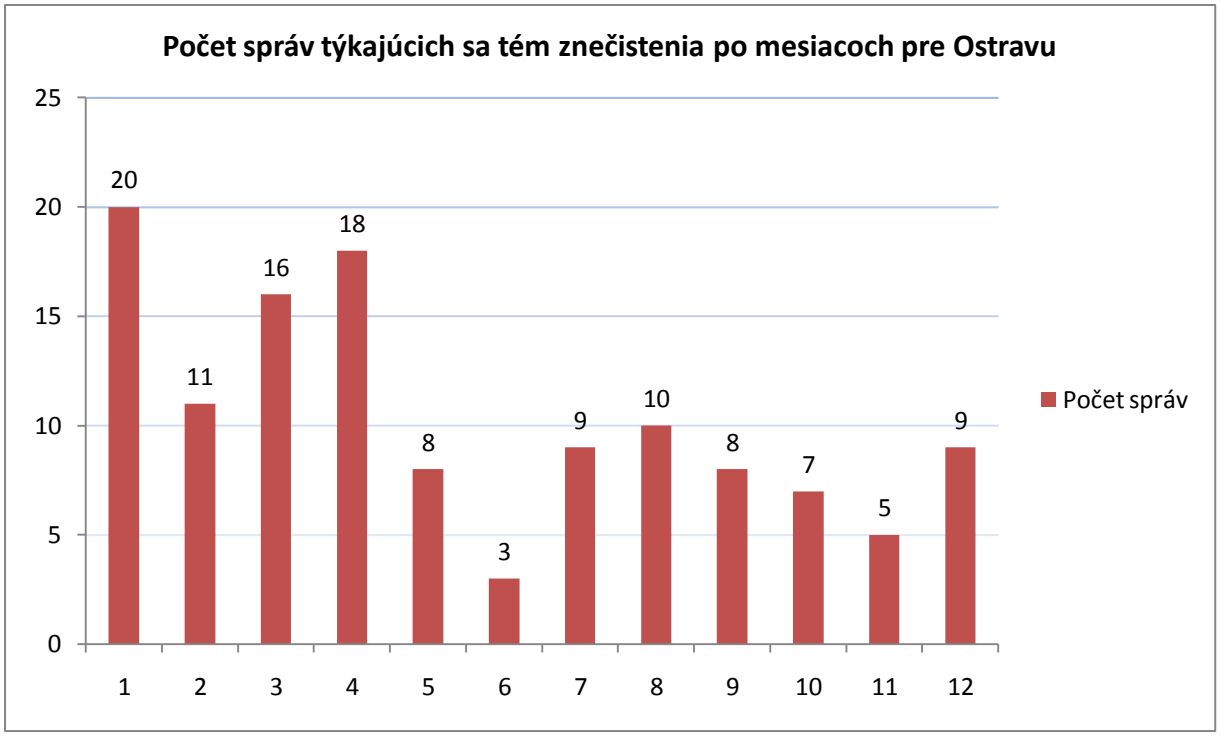
Ďalej sa poďme pozrieť na výskyt obce Ostrava v správach na časovej ose. Ide o výskyt obce v správach súvisiacich s témami znečistenia, spomenutými vyššie .



Obr. 9-10 Ostrava a jej výskyt v čase v správach súvisiacich so znečistením. Čísla na spodnej ose predstavujú počet minút od dátumu 1. 1. 2008 po 15. 4. 2010

Z obrázku je zrejmé, že Ostrava sa vyskytuje s danými témami zcela pravidelne, hlavne od začiatku roku 2009 (signalizovaného hodnotou okolo 500 000). V grafe sa nám asi nepodarí nájsť nejaké charakteristické črty či trendy. Skúsme sa dotázať na dáta v ráde mesiacov či rokov.

Z hľadiska času sa môžeme pozrieť na spravodajstvo o vybranej téme (téme znečistenia) aj podľa mesiacov či dní. Môže nás napríklad zaujímať, v ktorých mesiacoch sa o znečistení v Ostrave hovorilo viac či menej, prípadne či sú medzi mesiacmi rozdiely v rôznych rokoch. Chceme teda vybrané dáta agregovať na úroveň mesiacov Obr. 9-11.



Obr. 9-11 Počet správ, v ktorých boli spomenuté témy týkajúce sa znečistenia a ktoré geograficky prislúchajú k Ostrave. Dáta z kanálu ČT24 domáce z obdobia od 1. 1. 2008 do 15. 4. 2010

Z grafu je zrejmé, že najvyšší podiel v spravodajstve pojednávajúcom o znečistení má Ostrava v zimných mesiacoch – najviac v januári a v marci. Najmenej sa k danej téme vyjadruje vybrané médium v mesiaci jún. Prax dokazuje, že obecne najmenej správ je médiami vyprodukovaných v júni prípadne v máji a decembri (vianočné sviatky).

Pokúsme sa ďalej vysledovať, či intenzita témy znečistenia v Ostrave nejako výraznejšie nesúvisí napríklad s dňom v týždni. Na prvý pohľad možno trochu naivná predstava, ale na druhej strane existujú vo svete výskumy a štatistiky, ktoré berú do úvahy typ dňa. Napríklad na Slovensku zomiera pri autonehodách najviac ľudí v piatok a v pondelok – čiže tesne pred a tesne po víkende, nemeckí vedci sa vážne pohrávali s myšlienkou, že po víkende by malo prísť vyčasenie a krajšie počasie (skúmali intenzitu automobilovej dopravy cez víkend a mimo víkendu v súvislosti so zmenou počasia). My sa pozrieme na počet správ viažúcich sa k téme znečistenia v Ostrave vzhľadom na deň.

Obr. 9-12 Ostrava v správach o znečistení v súvislosti s typom dňa

Z pohľadu typu dňa nemôžeme v žiadnom prípade hovoriť o extrémnych situáciách. Približne polovičný podiel oproti iným dňom má utorok, inak sú hodnoty dosť vyrovnané. Víkend sa takisto neodlišuje od zvyšku týždňa.

Celkovo je teda možné hodnotiť, že téma znečistenia ovzdušia v Ostrave nebola aspoň vo vybranom kanáli potvrdená. Neukázali sa ani výraznejšie nahusťovania v čase. Mohli sme však byť svedkami vyššieho počtu správ o tejto problematike v zimných mesiacoch, čo potvrdzuje zažité predstavy.

9.7 Podobnosť správ

Na podobnosť správ sa taktiež môžeme pozerať z viacerých strán. Jednak sa môže jednať o to, či sú správy si podobné svojím obsahom alebo na druhej strane podobné skôr systematickými charakteristikami ako napríklad dĺžka správy, čas jej publikovania apod. Budeme sa snažiť načrtnúť možný postup pri spracovaní obsahovej podobnosti správ s prihliadnutím na čas. Otázkou je, ako správne pristúpiť ku kvalitatívnemu hodnoteniu podobnosti správ. Jednou z možností by bolo využiť nástroje tezauru GEMET ako sú vzťahy medzi pojmami – RT, BT a NT a na základe nich určitým spôsobom zoskupovať správy. Problémom ale je, že celá sieť týchto vzťahov v tezaure je pomerne zložitá a rozsiahla a často by sme sa pri tom mohli dostať do komplikácií.

Jednou z ďalších možností je pristúpiť k hodnoteniu podobnosti opäť z pohľadu časového, geografického a tematického. Pričom tematickým aspektom myslíme zaradenie správ do jednej zo skupín či tém v tezaure GEMET. Hodnotenie podobnosti budeme vykonávať na základe časovej, geografickej a tematickej blízkosti objektov, v našom prípade blízkosti správ. Ak si vezmeme akékoľvek 2 správy, tak vieme medzi nimi okamžite určiť ich časový rozdiel, teda napríklad počet dní

či hodín, ktoré medzi nimi ubehli. Z hľadiska tematického je takisto pomerne jednoducho možné získať akúsi tematickú blízkosť daných 2 správ. Na základe predchádzajúcich analýz vieme u každej správy určiť, o ktorých témach pojednáva. Ak teda v 2 správach sa vyskytujú rovnaké témy, či už sa všetky zhodujú, alebo sa zhodujú len čiastočne, môžeme o týchto správach hovoriť, že majú určitú tematickú blízkosť. A nakoniec z hľadiska priestoru je taktiež zcela otvorene možné hovoriť o blízkosti správ. Na základe toho, že sa správy odohrali blízko seba, je možné hovoriť aj o ich podobnosti.

Pri pokuse o zahrnutí všetkých spomenutých 3 aspektov bol navrhnutý postup, ako oceniť príbuznosť správ. Majme ľubovoľné 2 správy, o ktorých vieme tieto charakteristiky:

Tab. 9-8 Správy a ich vybrané charakteristiky pre hodnotenie podobnosti

Správa 1		
Dotknuté témy	**Obec**	**Čas**
Biosféra, atmosféra, antroposféra	Praha	13. 5. 2009
Správa 2		
Dotknuté témy	**Obec**	**Čas**
Spoločnost, antroposféra	Olomouc	12. 6. 2009

Z charakteristík ako čas a obec vieme pomerne jednoducho určiť geografickú a časovú vzdialenosť. Geografickú vzdialenosť dvoch obcí napríklad meranú vzdušnou čiarou v metroch a časovú určenú rozdielom časov. Otázne je, ako určiť blízkosť dvoch správ, z ktorých jedna pojednáva o 3 témach, druhá o dvoch a majú jednu spoločnú.

Tento problém bol vyriešený tak, že sa na tematickú podobnosť pristúpi z 2 strán. Najprv zo strany *Správa 1*. Z pohľadu správy 1 môžeme hovoriť o 33,3 percentnej zhode s druhou správou, keďže jena z troch tém sa dotkla aj Správy 2. Z pohľadu Správy 2 zase môžeme hovoriť o 50 percentnej zhode, keďže jedna z dvoch tém sa objavila aj v Správe 1.

Pri spracovaní však nastáva i ďalší problém, a to zjednotenie mier pre všetky charakteristiky, napríklad na percentá (stupeň zhody vyjadrený v percentách). Pri charakteristike obec bola zvolená hranica 50 kilometrov. Táto hranica určuje akési rozmedzie. To znamená, že vzdialenosť medzi obcami, ktorá sa pohybuje medzi 0 až 50 kilometrov, sa prepočíta na percentá v rozmedzí od 0 do 100 percent. 0 percent znamená 50 a viac kilometrová vzdialenosť, 100 percent vzdialenosť 0 km. Podobná hranica bola stanovená aj pre časovú charakteristiku.

Výsledok spracovania môže vyzerať nasledovne:

Tab. 9-9 Možné výsledky hodnotenia podobnosti správ

ID SPRÁVY 1	ID SPRÁVY 2	TEMATICKÁ12	TEMATICKÁ21	PRIESTOROVÁ	ČASOVÁ
1	2	100 %	25 %	75 %	99 %
1	4	100 %	50 %	0 %	98 %

Praktické nasadeni postupu však vôbec nebolo jednoduché. Ukázalo sa, že spracovanie trvá nesmierne dlhú dobu. Uvedomme si, že spracúvame a porovnávame každú správu s každou. Pri vybranom kanáli, ktorý čítal viac ako 10 000 je výpočet zložitý. Navyše si pravdepodobne nemôžeme vybrať len určitý počet náhodne vygenerovaných správ. V tejto oblasti sa ako riešenie javí nasadenie pokročilejších metód ako tých, ktoré sa v rámci práce používali (napr. použitie iného typu databáz či programovacieho jazyka.

10 Záver

Práca zhŕňa a pokúša sa aplikovať rôzne postupy pre zamerané na hodnotenie spravodajstva. Ukázalo sa, že pri analýzach je možné a aj účelné zaradiť rôzne metódy známe z oblasti informatiky a spracovania textu. Vhodným použitím týchto metód a testovaním rôznych parametrov sa môžeme dostať k zaujímavým výsledkom. Kľúčovým impulzom pre ďalší rozvoj práce bolo implementovanie tezauru GEMET. Nielenže pomohol pri určovaní tematického zaradenia jednotlivých správ, ale poskytol podkladové dáta pre ďalšie zaujímavé rozbory najmä v súvislosti s časovou a priestorovou zložkou správ.

Pri určovaní priestorovej zložky sme využili už vyvinutý nástroj pre geoparsing. Odhalili sme v ňom však i chyby, ktoré potrebujú vyriešiť – hlavne konflikty názvov obcí s inými geografickými entitami ako je napríklad rieka, časť obce či pohorie. Na druhú stranu sú však problémy, ktoré by sme zdolávali veľmi obtiažne, napríklad preklepy v správe či použitie zaniknutého či miestneho (nárečového) názvu obce. Ku kvalite určovania lokalizácií v správach a obecne v texte by v oblasti služby RSS najviac pomohlo vyššie začleňovanie značiek <*georss*> alebo obecne využívanie služby GeoRSS. Ponuka GeoRSS kanálov je bohužiaľ stále nedostačujúca, pritom by mohli postihovať ďaleko viac oblastí ako postihujú v súčasnosti. Stále sa potvrdzuje, že je dôležité využívať metadáta, ktoré by v našom prípade výrazne odbremenili od prácneho spracovania vstupných dát, navyše by sa rapídne znížila chybovosť, keby sme informáciu napríklad o lokalite či téme správy mali skrytú na presne označenom mieste a v presnom formáte.

V súvislosti s tematickým zameraním správ by určite bolo vhodné použiť viac tézaurov. Ideálne by samozrejme bolo hlbšia práca s jednotlivými pojmami najmä z pohľadu ich skloňovania a získavania ich významu v súvislosti s inými pojmami v prirodzenom texte. Tu sa už dostávame k spracovaniu prirodzeného jazyka a k oblasti umelej inteligencie.

Praktická realizácia rôznych postupov spomenutých v práci – napríklad vyhľadávanie pojmov z tezauru vo fulltextoch – priniesla nečakané komplikácie a to najmä v súvislosti s dobou výpočtu resp. dobou porovnávania pojmov z tezauru so slovami v databáze správ. V oblasti spracovania daných aspektov spracovania prakticky neexistujú presné postupy. Často bolo nutné navrhnúť možný algoritmus. To však môže viesť k určitej polemike nad výsledkami, prípadne nad nastavením vstupných parametrov. Napríklad pri navrhovaní postupov pre spracovanie podobnosti správ je na nás, ako si nastavíme parametre. Na druhej strane ale máme vyššiu voľnosť a môžeme tým do danej problematiky preniknúť hlbšie.

V rámci postupu riešenia stanovených úloh bolo podmaňujúce stretávať sa s novými poznatkami, ktoré často nečakane prišli v náväznosti na analyzovanie spravodajstva z rôznych uhlov pohľadu. Napríklad sa celkom nečakane ukázalo, že existujú značné rozdiely v RSS kanáloch v rámci jedného spravodajskoho servera.

Zoznam použitej literatúry

[1] CHARVÁT, Karel; KOCÁB, Milan; KONEČNÝ, Milan; KUBÍČEK, Petr. Geografická data v informační společnosti. Praha: VÚGTK, 2007. 280 s., ISBN 970-80-85881-28-8

[2] CALDWELL, Doubles. Geoparsing Maps the Future of Text Documents. 2009, september. Dostupný na WWW: http://www.directionsmag.com/article.php?article_id=3268

[3] W3 SCHOOLS. Introduction to XML. Dostupný na WWW: http://www.w3schools.com/xml/xml_whatis.asp

[4] KALČEVOVÁ, Jana. Vícekriteriální hodnocení variant. 2006. Dostupný na WWW: http://jana.kalcev.cz/vyuka/kestazeni/EKO422-Vahy.pdf

[5] HORÁK, Jiří. Zpracování dat v GIS. 2009. 199 s.

[6] Wikipedie. Zpracování přirozeného jazyka. [cit. 2010-03-31]. Dostupný na WWW: http://cs.wikipedia.org/wiki/Zpracov%C3%A1n%C3%AD_p%C5%99irozen%C3%A9ho_jazyka

[7] ŠARMANOVÁ, Jana. Informační systémy a datové sklady. Ostrava: VŠB – Technická univerzita Ostrava, 2007. 169 s.

[8] TYLOVÁ, Veronika. Struktura tezauru. 2004, máj. [cit. 2010-03-31]. Dostupný na WWW: http://www.phil.muni.cz/kivi/clanky.php?cl=35

[9] Knihovna Evangelické teologické fakulty Univerzity Karlovy v Praze. Český teologický tezaurus. [cit. 2010-03-31]. Dostupný na WWW: http://www.etf.cuni.cz/~library/infoctt.html#2
Parlamentní knihovna Česká republika. Tezaurus EUROVOC. 2006, marec. [cit. 2010-03-31]. Dostupný na WWW: http://www.psp.cz/kps/knih/e_zakinf.htm

[10] KRČÁL, Martin. Využití RSS pro personalizované doručování článků z odborných a vědeckých časopisů. Diplomová práce. Brno: 2008.

[11] NEMEC, Peter; HORÁK, Jiří. The Geographical Balance of Czech TV CT24 News

[12] NEMEC, Peter. Geokódovanie mediálnych správ. Ostrava: VŠB – Technická univerzita Ostrava, 2008. 49 s.

[13] SCHEJBAL, Ctirad; HOMOLA Vladimír; STANĚK František. Geoinformatika. PONT, s. r. o., 2004. 229 s., ISBN 80-967611-8-8

[14] BLAŽEK, Jakub. Srovnání automatické a intelektuální indexace. 2008, apríl [cit. 2010-04-24]. Dostupný na WWW: http://www.inflow.cz/srovnani-automaticke-intelektualni-indexace

[15] HANZLOVÁ, Markéta; et al. NATURE-SDIplus: towards the implementation of the European SDI in nature conservation. Zborník GIS Ostrava 2010. 14 s.

11 Prílohy

11-1 Výsledky analýzy spoľahlivosti

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
19	Hradec Králové - V neděli začíná advent, čas předvánočního rozjímání. Navzdory tomu se dva pořadatelé adventních trhů, které právě dnes začaly v Hradci Králové, dostali do sporu. Muzeum se cítí poškozené a hovoří o tom, že konkurence se na dvacetileté tradici jeho trhů přiživuje. Firma Bohemia Touristik ale na svojí akci nevidí nic špatného.	569810	Hradec Králové	1	Správne	-
94	Plzeň/Zlín - V centru Plzně si lidé dobu před dvaceti lety připomínají úsměvnou formou. Na ulicích totiž postávají veksláci, kteří kolemjdoucím nabízejí bony. Mezi lidmi se proplétají tajní policisté, kteří sbírají podpisy pod prohlášení	554791	Plzeň	1	Správne	-
94	Plzeň/Zlín - V centru Plzně si lidé dobu před dvaceti lety připomínají úsměvnou formou. Na ulicích totiž postávají veksláci, kteří kolemjdoucím nabízejí bony. Mezi lidmi se proplétají tajní policisté, kteří sbírají podpisy pod prohlášení	585068	Zlín	1		
264	Kravaře (Opavsko) - Devětasedmdesátiletá žena uhořela v noci na dnešek ve svém bytě v Kravařích na Opavsku. Dva policisté, kteří po nahlášení požáru na místě zasahovali, se nadýchali zplodin. V bytě ve Vyhlídalově ulici začalo hořet v sobotu kolem jedenácté večer. Příčinu požáru policisté a hasiči ještě vyšetřují.	507580	Kravaře	1	Správne	Podľa oblasti
398	Pardubice - Některé regiony kritizují hromadnou státní zakázku na likvidaci starých ekologických zátěží v Česku. Například Pardubický kraj by si rád o zakázce ve svém regionu rozhodl sám, ministr financí Eduard Janota však o zrušení tendru vůbec neuvažuje. Částku potřebnou k sanaci ekologických zátěží v regionu vyčíslil Pardubický kraj na jeden a půl miliardy korun, podle podmínek zakázky by se však vedení kraje kvůli obchodnímu tajemství ani nemuselo dozvědět, kolik by měl stát za sanaci skutečně zaplatit.	555134	Pardubice	1	Správne	-
430	Praha - Organizátoři mezinárodního muškařského závodu Orvis cup vypustili v Praze do Vltavy zhruba 10 tisíc ryb. Jedná se hlavně o pstruha duhového. Více než hodinu trvalo rybářům z Velkého Meziříčí naplnit kádě. 25 metráků ryb pak vezli do Prahy - žádný pstruh naštěstí během cesty neuhynul. Přes 100 rybářů bude v jejich lovu soutěžit už o tomto víkendu.	554782	Praha	1	Správne, Velké Meziříčí – chyba vydavateľa	-
444	Valašské Meziříčí - Vláda České republiky na začátku ledna dostane Kleopatru na trůně. Tak se totiž jmenuje vzácný gobelín, který bude zdobit jednu z místností úřadu. Více než 400 let starý artefakt nyní mají v rukách pracovnice Moravské gobelínové manufaktury ve Valašském Meziříčí. Při práci používají i speciální počítačový program.	545058	Valašské Meziříčí	1	Správne	-
451	Zlonice (Kladensko) - Ve Zlonicích na Kladensku se dnes pro veřejnost naposledy otevřou brány železničního muzea. Zájemci tak mají od devíti hodin dopoledne poslední příležitost prohlédnout si muzeum v té podobě, do jaké dospělo za 13 let své existence. Naposledy se otevřou také vrata staré výtopny i dveře původní expozice zabezpečovacího zařízení na rohu ulic Nádražní a Tyršova. Na poslední cestu vyjede i vlak polní malodráhy s lokomotivou BN 30. Závěrečná jízda čeká i vláček na modelovém kolejišti. Návštěvníci si mohou prohlédnout i výstavu lokomotiv, které většinou mohli spatřit pouze uvnitř výtopny.	533114	Zlonice	1	Správne	-
491	Ostrava - Severomoravská města Ostrava a Hradec nad Moravicí v těchto dnech hostí mezinárodní konferenci důlních záchranářů z celého světa. Setkání, jehož hlavní náplní je analýza příčin a prevence největších důlních neštěstí za poslední léta, se koná poprvé na evropském kontinentu. Jedním z mimořádných témat letošní konference bude i páteční tragédie, která se stala v důlní šachtě nedaleko polských Katovic.	507270	Hradec nad Moravicí	2	Správne	-
491	Ostrava - Severomoravská města Ostrava a Hradec nad Moravicí v těchto dnech hostí mezinárodní konferenci důlních záchranářů z celého světa. Setkání, jehož hlavní náplní je analýza příčin a prevence největších důlních	554821	Ostrava	1		

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
	neštěstí za poslední léta, se koná poprvé na evropském kontinentu. Jedním z mimořádných témat letošní konference bude i páteční tragédie, která se stala v důlní šachtě nedaleko polských Katovic.					
506	Karlovy Vary – Vedení Karlových Varů chce mít na okraji města lanovku a lyžařskou dráhu. Opoziční zastupitelé ale takový nápad kritizují. Podle nich jde o neúčelně vynaložené prostředky, protože Vary leží nízko a se sněhem mají problém. Investice je to přitom za stovky milionů korun. Lanovka by měla vést na Vítkovu horu z lokality Kouzelné městečko, kde začíná lázeňská čtvrť Karlových Varů. V současné době je zde záchytné parkoviště. Pro plánovanou modernizaci této části města a stavbu sjezdovky s lanovkou bylo třeba schválit změnu územního plánu.	554961	Karlovy Vary	1	Správne	-
545	Uherské Hradiště - Slovácko se představuje v Uherském Hradišti. Začal zde sedmý ročník Slováckých slavností vína a otevřených památek. Nabízí však nejen víno, ale i jihomoravské gastronomické pochoutky a především přehlídku lidových krojů z 56 vesnic celého Slovácka. Za rozmanitostí slováckého folkloru mohou zájemci zavítat i v neděli, město počítá s návštěvou až pětapadesáti tisíc lidí.	592005	Uherské Hradiště	1	Správne	-
610	Pardubice - "Koně v akci" - tak se jmenuje mezinárodní hipologická výstava, kterou najdete přímo na dostihovém závodišti v Pardubicích. Představí se na ní 450 koní čtyř desítek plemen. Výstava chce koně představit jako pomocníka člověka v různých situacích v průběhu jednotlivých období lidského vývoje. Kromě koní se na výstavě předvádějí i zástupci různých řemesel, která s chovem koní souvisejí. Dlouhou tradici mají především kováři a podkováři, kteří zde dokazují, že jejich řemeslo v dohledné době nezanikne.	555134	Pardubice	1	Správne	-
695	Kolová (Karlovarsko) - Lesní požár u Kolové na Karlovarsku, který vypukl dnes dopoledne v těžko přístupném terénu u Stanovické přehrady, se podařilo hasičům zlikvidovat. Se zdoláním ohně, jenž zasáhl plochu velkou zhruba jeden hektar, jim pomohl i vrtulník. Nyní už zbývá dohasit pouze některá drobná ohniska v podzemí. Za předpokládanou příčinu požáru považují hasiči nedbalost lesních dělníků.	555258	Kolová	1	Správne	-
700	Stanovice - Obec Stanovice na Karlovarsku získala dotaci a konečně se dočká vodovodu, o nějž usilovala patnáct let. Paradoxní je, že obec leží necelé dva kilometry pod vodárenskou nádrží. Obyvatelé místní části Dražov si nemohou otočit kohoutkem a napustit si tolik čisté vody, kolik potřebují. Staré studně prosakují a rychle vysychají. K čistotě vody měli výhrady i hygienici.	555550	Stanovice	1	Správne	Podľa oblasti
705	Helfštýn (Přerovsko) - Na třetím nádvoří středověkého hradu Helfštýn začal za doprovodu středověké hudby 28. ročník kovářského fóra. Své umění zde předvede na pět stovek kovářů a teoretiků z dvaceti zemí. Po celý týden budou navíc špičkoví italští umělci pracovat na plastice, kterou pak předají do hradní expozice. Návštěvníci mohou sledovat tvorbu plastiky každý den od 9 do 18 hodin kromě pondělí, kdy je hrad uzavřen.	-	-	-	Správne	-
722	Plzeň - Starosta České Břízy na Plzeňsku Milan Pondělík chce odejít z funkce. Okresní soud mu totiž uložil měsíční podmíněný trest za to, že bagrista najaté firmy překopl dálkový optický kabel, který prochází obcí. Dvě sousední vesnice proto zůstaly tři hodiny bez pevných telefonních linek. Starosta se brání tím, že bagrista tehdy vůbec neměl nařízeno jít lžící pod zem, ale jen rozhrnout přivezenou ornici po pozemku. Podmíněný trest ho zaskočil - nikomu prý neublížil, nikoho neohrozil a šlo jen o náhodu.	564648	Bříza	2	Nesprávne, nie sú definované pády pre Česká Bříza	-
722	Plzeň - Starosta České Břízy na Plzeňsku Milan Pondělík chce odejít z funkce. Okresní soud mu totiž uložil měsíční podmíněný trest za to, že bagrista najaté firmy překopl dálkový optický kabel, který prochází obcí. Dvě sousední vesnice proto zůstaly tři hodiny bez pevných telefonních linek. Starosta se brání tím, že bagrista tehdy vůbec neměl nařízeno jít lžící pod zem, ale jen rozhrnout přivezenou ornici po pozemku. Podmíněný trest ho zaskočil - nikomu prý neublížil, nikoho neohrozil a šlo jen o náhodu.	554791	Plzeň	1		
801	Jičín - Majitelé unikátních veteránů ze samých počátků historie automobilismu si dali sraz v Jičíně. Divákům se na setkání nazvaném příznačně Loukotě a řemeny představily nejstarší vozy vyrobené do roku 1918. Účast na akci přijalo množství nadšenců, kteří přihlížejícím předvedli opravdové lahůdky. Mnoho strojů je v takové kondici, kterou by jim mohli mnozí závidět. Jičínská jízda je první a zároveň největší akcí pro tuto kategorii strojů v České republice.	572659	Jičín	1	Správne	-

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
909	Pasohlávky (Brněnsko) - U nádrže Nové Mlýny v blízkosti Věstonic a Mikulova odkrývají archeologové základy někdejšího polního římského lazaretu. Stavba v Pasohlávkách byla součástí rozsáhlého opevněného komplexu, který si ve druhém století za vlády císaře Marka Aurelia postavila u Jantarové stezky na kopci Hradisko desátá římská legie. Lazaret je největším zařízením svého druhu, které se z tohoto období dochovalo v oblasti na sever od Dunaje.	584649	Mikulov	2	Správne, Věstonice neexistujú samostatne, len Horné a Dolné	-
909	Pasohlávky (Brněnsko) - U nádrže Nové Mlýny v blízkosti Věstonic a Mikulova odkrývají archeologové základy někdejšího polního římského lazaretu. Stavba v Pasohlávkách byla součástí rozsáhlého opevněného komplexu, který si ve druhém století za vlády císaře Marka Aurelia postavila u Jantarové stezky na kopci Hradisko desátá římská legie. Lazaret je největším zařízením svého druhu, které se z tohoto období dochovalo v oblasti na sever od Dunaje.	584762	Pasohlávky	1		-
1010	Ústí nad Labem - Ústecký magistrát se stále potýká se sociálně vyloučenou lokalitou, městskou částí Předlice. Několik desítek rodin zde žije v devastovaných objektech. Přestože se situace zlepšila po generálním úklidu, který magistrát nařídil, stále v jednom zdemolovaném domě v Marxově ulici žije několik rodin. Špatné hygienické podmínky navíc ohrožují sousední mateřskou školu.	554804	Ústí nad Labem	1	Správne, ďalšie názvy: časť Předlice, Marxova ulica	-
1071	Litomyšl - Ocenění z rukou ministryně školství převzali tři studenti z Gymnázia Aloise Jiráska v Litomyšli. Nedávno zvítězili v prestižní mezinárodní soutěži vědeckých a technických projektů pro středoškoláky. Ve Spojených státech předvedli chytrého robota, který dokáže vyhledávat předměty a pak s nimi manipulovat. Výsledek své práce dnes ukázali i svým spolužákům.	578347	Litomyšl	1	Správne	-
1287	Zákupy (Českolipsko) - Obyvatelé Zákup hovoří o svých pozemcích jako o znehodnoceném majetku. Důvodem je názor vodohospodářů, kteří rozsáhlé oblasti v Zákupech prohlásili za záplavové území. V něm se ocitly mnohé parcely a soukromé pozemky, kde se nyní nesmí stavět. Podle obyvatel, kteří zde dlouhodobě žijí, je prohlášení vodohospodářů přehnané. Úředníci však mají jiný názor.	562262	Zákupy	1	Správne	-
1329	Nýrsko - Pětitisícové Nýrsko na Klatovsku zavedlo na začátku letošního roku jako první město v České republice vlastní protikrizová opatření. Tamní radnice odložila plánované zdražování a zmrazila lidem na dva roky nájmy, ceny tepla, vody a odpadu. Místním živnostníkům dokonce srazila z nájmů za městské prostory 20 procent. Nyní se radní i obyvatelé zpětně ohlížejí a účinky opatření hodnotí.	556831	Nýrsko	1	Správne	-
1357	Jaroměř (Náchodsko) - Záchranu topících se vodáků mohli vidět lidé na Úpě nedaleko Jaroměře. Nejednalo se však o ostrou akci, bylo to pouze cvičení, pro hasiče ho uspořádali vodáčtí instruktoři. Uvíznutí ve vodním válci je totiž na českých řekách podle statistik příčinou převážné většiny všech utonutí. Právě na jezech, které na první pohled nevypadají nebezpečně, lidí umírá nejvíc.	574121	Jaroměř	1	Správne	-
1405	Praha - Dělníci dnes kolem poledne dokončili opravu prasklého vodovodního řadu v Praze 3 na křižovatce ulic Malešická a Nad Kapličkou. Dodávka vody se nyní postupně obnovuje. Uvedla to mluvčí Pražských vodovodů a kanalizací (PVK) Marcela Dvořáková. Vodovod praskl v noci na dnešek a bez vody se ocitlo několik tisíc obyvatel Malešic a části Strašnic. PVK zajistila náhradní zásobování voznicemi v ulicích Na Třebešíně, Limuzská, Za Stadionem, Univerzitní a Kounická, občany také zásobuje autocisterna.	554782	Praha	1	Správne, ďalšie názvy: ulica Malešická, Nad Kapličkou, Na Třebešíně, Limuzská, Za Stadionem, Univerzitní, Kounická	-
1415	Praha - Dělníci na Žižkově začali rozebírat jeden ze dvou původních železničních mostů, kterým se říká Malá a Velká Hrabovka. Nahradila je nová moderní estakáda tzv. Nového koridoru mířící do tunelu pod Vítkov. S komplikacemi a zpožděním tu musí počítat především řidiči. Až do nedělního odpoledne bude zavřená křižovatka ulic Trocnovské a Husitské.	554782	Praha	1	Nesprávne – konflikt s Vítkovom na Opavsku, ďalšie názvy: Trocnovská, Husitská ulica	-
1415	Praha - Dělníci na Žižkově začali rozebírat jeden ze dvou původních železničních mostů, kterým se říká Malá a Velká Hrabovka. Nahradila je nová moderní estakáda tzv. Nového koridoru mířící do tunelu pod Vítkov. S komplikacemi a zpožděním tu musí počítat především řidiči. Až do nedělního odpoledne bude zavřená křižovatka ulic Trocnovské a Husitské.	511021	Vítkov	2		
1461	Olomouc - Vojenský výcvikový prostor Libavá na Olomoucku se otevřel veřejnosti. Do jindy nepřístupných míst proudily už od rána stovky turistů. V rámci akce s názvem Bílý kámen mohli absolvovat hned několik různě	503941	Libavá	2	Správne	-

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
	dlouhých tras. Příležitost objevovat krásy dosud uzavřeného prostoru využili jak cyklisté, tak i milovníci pěší turistiky. A nikdo z nich neodešel zklamaný.					
1461	Olomouc - Vojenský výcvikový prostor Libavá na Olomoucku se otevřel veřejnosti. Do jindy nepřístupných míst proudily už od rána stovky turistů. V rámci akce s názvem Bílý kámen mohli absolvovat hned několik různě dlouhých tras. Příležitost objevovat krásy dosud uzavřeného prostoru využili jak cyklisté, tak i milovníci pěší turistiky. A nikdo z nich neodešel zklamaný.	500496	Olomouc	1		
1494	Plzeň - Ani usilovnými protesty nezabránili obyvatelé Výsluní v Plzni stavbě nového vedení vysokého napětí přes svoje pozemky. Plánovaná trasa zůstala a teď už přijela i těžká technika. Za své vzala i žádost přeložit vedení o několik desítek metrů dál do neobydleného lesa. Podle ČEZu nebyla jiná trasa možná. 110 kV místo současných 20 nejsou lidé z Výsluní podle svých slov ochotni tolerovat ani ve veřejném zájmu.	554791	Plzeň	1	Správne	-
1494	Plzeň - Ani usilovnými protesty nezabránili obyvatelé Výsluní v Plzni stavbě nového vedení vysokého napětí přes svoje pozemky. Plánovaná trasa zůstala a teď už přijela i těžká technika. Za své vzala i žádost přeložit vedení o několik desítek metrů dál do neobydleného lesa. Podle ČEZu nebyla jiná trasa možná. 110 kV místo současných 20 nejsou lidé z Výsluní podle svých slov ochotni tolerovat ani ve veřejném zájmu.	563498	Výsluní	2	Nesprávne, konflikt ulica Výsluní a obec Výsluní na Chomutovsku	-
1542	Sokolovsko - Elektřinu začne nová větrná elektrárna, která právě vyrůstá na kopci nad Horním Částkovem na Sokolovsku, dodávat už v květnu. Včera postavili 105 metrů vysoký stožár a dnes osadili 35metrové vrtule. Situace při prosazování větrníků v severozápadních Čechách je ale problémová. Někde se proti nim pořádají referenda, jinde vadí samosprávám.	-	-	-	Správne	-
1649	Karlovy Vary - Necitlivé zásahy v centru Karlových Varů by mohly ohrozit zápis lázeňského trojúhelníku na seznam kulturního dědictví UNESCO. Kandidátem na zápis je tento trojúhelník tří měst od roku 2007. Přístup karlovarského magistrátu k některým stavebním úpravám ale podle ministra Jehličky ohrožuje i Mariánské a Františkovy Lázně.	554529	Františkovy Lázně	2	Nesprávne, chýbajú Mariánské Lázně	-
1649	Karlovy Vary - Necitlivé zásahy v centru Karlových Varů by mohly ohrozit zápis lázeňského trojúhelníku na seznam kulturního dědictví UNESCO. Kandidátem na zápis je tento trojúhelník tří měst od roku 2007. Přístup karlovarského magistrátu k některým stavebním úpravám ale podle ministra Jehličky ohrožuje i Mariánské a Františkovy Lázně.	554961	Karlovy Vary	1		
1704	Břeclav - Novinku, kterou zatím žádné jiné město na jižní Moravě nemá, plánuje na letošek Břeclav. Bude jí městský přívoz. Po Dyji začnou jezdit pravidelné lodní linky se šesti zastávkami v různých částech města. Sloužit mají nejen turistům, ale i všem obyvatelům Břeclavi, kteří nechtějí stát v dopravních zácpách. Na novou sezonu se právě teď připravují i ostatní provozovatelé lodních linek v kraji.	584291	Břeclav	1	Nesprávne, konflikt rieka	-
1704	Břeclav - Novinku, kterou zatím žádné jiné město na jižní Moravě nemá, plánuje na letošek Břeclav. Bude jí městský přívoz. Po Dyji začnou jezdit pravidelné lodní linky se šesti zastávkami v různých částech města. Sloužit mají nejen turistům, ale i všem obyvatelům Břeclavi, kteří nechtějí stát v dopravních zácpách. Na novou sezonu se právě teď připravují i ostatní provozovatelé lodních linek v kraji.	593991	Dyje	2		
1755	Kyselka (Karlovarsko) - Veřejný ochránce práv prověřuje, jestli úřady postupovaly v souladu se zákony v případu zchátralých lázní v Kyselce na Karlovarsku. Památkáři si myslí, že úředníci mohli postupovat razantněji, úřady to však odmítají. Nejasnosti okolo vlastnictví objektu situaci příliš nepomohly a společnost, která budovy po vleklých soudních sporech koupila, říká, že za poslední dva roky firma za přípravy na záchranu lázeňských budov utratila dvacet šest milionů korun. Provedla příslušné průzkumy a vypracovala potřebné zakreslení stávajícího stavu.	555347	Kyselka	1	Správne	-
1775	Jihlava - Pediatři v kraji Vysočina mají málo dávek hexavakcíny, která se používá k základnímu očkování proti šesti dětským nemocem, například záškrtu nebo dávivému kašli. Spolu s epidemilogy proto kritizují ministerstvo zdravotnictví. Oproti objednávkám totiž do kraje dorazilo o 900 dávek méně. Hlavní hygienik odmítá kritiku tím, že si kraje musí v období nouze vypomoci mezi sebou.	586846	Jihlava	1	Správne	-
1796	Karlovy Vary - Rozsáhlé pozemky ve velmi zanedbaném stavu v centru Karlových Varů by se měly postupně proměnit v obchodní a pěší zónu. Ve spolupráci s městem ji vybuduje soukromá firma, která v lokalitě dolního	554961	Karlovy Vary	1	Správne	-

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
	nádraží koupila pozemky od Českých drah. Obnova celé oblasti bude trvat několik let a vyžádá si finanční náklady v řádu miliard korun.					
1942	Karlovy Vary - Při včerejší sněhové vánici musely Ústecký a Karlovarský kraj uzavřít silnici mezi Božím Darem a Vejprty. Nikdo se nemohl dostat do lyžařských areálů. Počasí se pak sice umoudřilo, jenže sjezdovky už zůstaly bez lyžařů.	554961	Karlovy Vary	1	Nesprávne, nie sú definované pády pre Boží Dar	-
1942	Karlovy Vary - Při včerejší sněhové vánici musely Ústecký a Karlovarský kraj uzavřít silnici mezi Božím Darem a Vejprty. Nikdo se nemohl dostat do lyžařských areálů. Počasí se pak sice umoudřilo, jenže sjezdovky už zůstaly bez lyžařů.	563404	Vejprty	2		
2102	Praha - Centrem hlavního města Prahy dnes projely soupravy historických tramvají. Spanilou jízdu připravilo ve spolupráci s pražským dopravním podnikem Národní muzeum v rámci probíhající výstavy Republika. Na výjezd souprav z holešovického Výstaviště zavítaly na 3 stovky zájemců, přes 2 třetiny z nich se poté historickými tramvajemi také projely.	554782	Praha	1	Správne	-
2139	Kadaň - ČSSD se vložila do sporu společnosti ČEZ a nájemníků, kteří v Severních Čechách protestují proti prodeji bytů, v nichž bydlí. Podle sociálně demokratických vyjednavačů lze dojednat, aby lidé o lidé o své byty nepřišli a ČEZ přitom neprodělal.	563102	Kadaň	1	Správne	-
2203	Praha - Záchranné stanici pro handicapovaná zvířata v pražských Jinonicích hrozí podle její provozovatelky po 26 letech zánik. Pražští radní odsouhlasili novou koncepci péče o nemocná a nalezená zvířata, která této stanici omezuje činnost a krátí jí dotace. Město už nechce mít jen jednu záchrannou stanici, ale celou síť poskytovatelů pomoci.	554782	Praha	1	Správne, ďalšie názvy: časť obce Jinonice	-
2208	Plzeň - Do brdských obcí, které leží v oblasti plánovaného amerického radaru, začaly proudit první peníze slíbené vládní pomoci. Městečko Mirošov na Rokycansku díky nim zprovoznilo most a plánuje i rekonstrukci školy. Ne všude jsou ale tak úspěšní. Do 22 obcí Plzeňského a Středočeského kraje má jít 1,25 mld. Kč. Dotčené obce chtěly původně asi třikrát tolik. Komise však jejich plány zredukovala a vybrala priority. 250 milionů má plynout do opravy silnic, stejnou částku dá stát do spolufinancování projektových dokumentací. Zbytek musí obce získat z dotačních programů. Právě v tom však je v mnoha případech kámen úrazu. Zatím první projekt, který se podařilo zrealizovat, je most v Mirošově na Rokycansku.	559997	Mirošov	2	Správne	Podľa oblasti
2208	Plzeň - Do brdských obcí, které leží v oblasti plánovaného amerického radaru, začaly proudit první peníze slíbené vládní pomoci. Městečko Mirošov na Rokycansku díky nim zprovoznilo most a plánuje i rekonstrukci školy. Ne všude jsou ale tak úspěšní. Do 22 obcí Plzeňského a Středočeského kraje má jít 1,25 mld. Kč. Dotčené obce chtěly původně asi třikrát tolik. Komise však jejich plány zredukovala a vybrala priority. 250 milionů má plynout do opravy silnic, stejnou částku dá stát do spolufinancování projektových dokumentací. Zbytek musí obce získat z dotačních programů. Právě v tom však je v mnoha případech kámen úrazu. Zatím první projekt, který se podařilo zrealizovat, je most v Mirošově na Rokycansku.	554791	Plzeň	1		
2540	Liberec - Starostové v Libereckém kraji protestují proti novince, kterou na ně chystá ministerstvo pro místní rozvoj. Chce zrušit 13 stavebních úřadů. Podle starostů to občanům přinese jen další komplikace a nelíbí se to ani těm městům, která by agendu měla převzít. Především se ale nikdo nedozvěděl, proč se tak má stát.	563889	Liberec	1	Správne	-
2684	České Budějovice - O cukrářský rekord se postarali studenti v Českých Budějovicích, vyrobili největší perníkový dort ve střední Evropě. Obří zákusek vážící přes půl tuny ukázali při kulinářské výstavě Gastrofest.	544256	České Budějovice	1	Správne	-
2777	Praha - První ročník soutěže Obec přátelská rodině má své vítěze. Ceny se udělovaly v pěti kategoriích podle velikosti obce. Mezi největšími městy připadlo prvenství Brnu. Z menších měst a obcí první ceny převzali představitelé Jablonce nad Nisou, Dobříše, Darkovic na Opavsku a Dubu u Prachatic. Do soutěže se přihlásilo 218 obcí. Ceny dnes v Poslanecké sněmovně předali ministr práce a sociálních věcí Petr Nečas (ODS) a předseda sněmovní komise pro rodinu Tomáš Kvapil (KDU-ČSL).	582786	Brno	2	Správne	-
2777	Praha - První ročník soutěže Obec přátelská rodině má své vítěze. Ceny se udělovaly v pěti kategoriích podle velikosti obce. Mezi největšími městy připadlo prvenství Brnu. Z menších měst a obcí první ceny převzali představitelé Jablonce nad Nisou, Dobříše, Darkovic na Opavsku a Dubu u	568228	Darkovice	2		

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
	Prachatic. Do soutěže se přihlásilo 218 obcí. Ceny dnes v Poslanecké sněmovně předali ministr práce a sociálních věcí Petr Nečas (ODS) a předseda sněmovní komise pro rodinu Tomáš Kvapil (KDU-ČSL).					
2777	Praha - První ročník soutěže Obec přátelská rodině má své vítěze. Ceny se udělovaly v pěti kategoriích podle velikosti obce. Mezi největšími městy připadlo prvenství Brnu. Z menších měst a obcí první ceny převzali představitelé Jablonce nad Nisou, Dobříše, Darkovic na Opavsku a Dubu u Prachatic. Do soutěže se přihlásilo 218 obcí. Ceny dnes v Poslanecké sněmovně předali ministr práce a sociálních věcí Petr Nečas (ODS) a předseda sněmovní komise pro rodinu Tomáš Kvapil (KDU-ČSL).	540111	Dobříš	2		
2777	Praha - První ročník soutěže Obec přátelská rodině má své vítěze. Ceny se udělovaly v pěti kategoriích podle velikosti obce. Mezi největšími městy připadlo prvenství Brnu. Z menších měst a obcí první ceny převzali představitelé Jablonce nad Nisou, Dobříše, Darkovic na Opavsku a Dubu u Prachatic. Do soutěže se přihlásilo 218 obcí. Ceny dnes v Poslanecké sněmovně předali ministr práce a sociálních věcí Petr Nečas (ODS) a předseda sněmovní komise pro rodinu Tomáš Kvapil (KDU-ČSL).	563510	Jablonec nad Nisou	2		
2777	Praha - První ročník soutěže Obec přátelská rodině má své vítěze. Ceny se udělovaly v pěti kategoriích podle velikosti obce. Mezi největšími městy připadlo prvenství Brnu. Z menších měst a obcí první ceny převzali představitelé Jablonce nad Nisou, Dobříše, Darkovic na Opavsku a Dubu u Prachatic. Do soutěže se přihlásilo 218 obcí. Ceny dnes v Poslanecké sněmovně předali ministr práce a sociálních věcí Petr Nečas (ODS) a předseda sněmovní komise pro rodinu Tomáš Kvapil (KDU-ČSL).	554782	Praha	1		
2777	Praha - První ročník soutěže Obec přátelská rodině má své vítěze. Ceny se udělovaly v pěti kategoriích podle velikosti obce. Mezi největšími městy připadlo prvenství Brnu. Z menších měst a obcí první ceny převzali představitelé Jablonce nad Nisou, Dobříše, Darkovic na Opavsku a Dubu u Prachatic. Do soutěže se přihlásilo 218 obcí. Ceny dnes v Poslanecké sněmovně předali ministr práce a sociálních věcí Petr Nečas (ODS) a předseda sněmovní komise pro rodinu Tomáš Kvapil (KDU-ČSL).	550094	Prachatice	2		
2791	Lipno nad Vltavou/Špindlerův Mlýn/Opavsko - Přípravy na zahájení zimní sezony vrcholí ve Skiareálu Lipno. Jeho modernizace přišla na 200 milionů korun. Tři nové 4sedačkové lanovky, delší a širší sjezdové tratě s bezpečnostními prvky, i dětské vleky - to vše je připravené pro návštěvníky. Tak rozsáhlá modernizace areálu má podle jeho provozovatelů jediný cíl - vybudovat na Lipně nejkvalitnější areál pro rodinné lyžování v České republice, jak sdělil ředitel areálu Luboš Krejza.	566403	Lipno	2		
2791	Lipno nad Vltavou/Špindlerův Mlýn/Opavsko - Přípravy na zahájení zimní sezony vrcholí ve Skiareálu Lipno. Jeho modernizace přišla na 200 milionů korun. Tři nové 4sedačkové lanovky, delší a širší sjezdové tratě s bezpečnostními prvky, i dětské vleky - to vše je připravené pro návštěvníky. Tak rozsáhlá modernizace areálu má podle jeho provozovatelů jediný cíl - vybudovat na Lipně nejkvalitnější areál pro rodinné lyžování v České republice, jak sdělil ředitel areálu Luboš Krejza.	545597	Lipno nad Vltavou	1	Nesprávne Lipno	-
2791	Lipno nad Vltavou/Špindlerův Mlýn/Opavsko - Přípravy na zahájení zimní sezony vrcholí ve Skiareálu Lipno. Jeho modernizace přišla na 200 milionů korun. Tři nové 4sedačkové lanovky, delší a širší sjezdové tratě s bezpečnostními prvky, i dětské vleky - to vše je připravené pro návštěvníky. Tak rozsáhlá modernizace areálu má podle jeho provozovatelů jediný cíl - vybudovat na Lipně nejkvalitnější areál pro rodinné lyžování v České republice, jak sdělil ředitel areálu Luboš Krejza.	579742	Špindlerův Mlýn	1		
2931	Karlovy Vary - Historické památce v centru Karlových Varů - Národnímu domu - možná svítá na lepší časy. Společnost, která chce do jeho rekonstrukce investovat miliardu korun, nyní přišla s řešením, jak rozmotat letité soudní spory kolem tohoto deset let chátrajícího objektu. A mají i podporu města. Jednoznačný zatím ale není postoj bývalého vlastníka Iva Ouřady.	554961	Karlovy Vary	1	Správne	-
3157	Praha - Náročný způsob montáže dvou evakuačních výtahů v pražské Všeobecné fakultní nemocnici zkusila česká firma vůbec poprvé a zřejmě úspěšně. Výtahy smontovat, převézt přes město, pomocí 200tunového jeřábu postupně zvednout a zasadit do výtahových šachet, navíc s milimetrovou přesností - to byl oříšek i pro odborníky.	554782	Praha	1	Správne	-
3242	Ostravsko - Nedostatek pracovníků s technickým vzděláním přinutil firmy v	598917	Karviná	2	Správne	-

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
	Moravskoslezském kraji hledat nové možnosti spolupráce se středními školami. O technické obory totiž stále není velký zájem. Firmy proto nabízejí náborové příspěvky a dotují nové obory. Například v Karviné po 20 letech znovu vznikl obor důlní zámečník.					
3287	Chomutov - Hospodaření chomutovského Podkrušnohorského zooparku provázely závažné nedostatky. Kontrola odhalila chyby v účetnictví, například při investičních akcích a v čerpání pohonných hmot do aut. Kontrola prověřovala období, kdy zoopark vedl odvolaný ředitel Přemysl Rabas. Ten byl poslancem i chomutovským zastupitelem za Stranu zelených. Rabas některá pochybení připouští - závěry zprávy ale považuje za účelové.	562971	Chomutov	1	Správne	-
3306	Pardubice - Obyvatele pardubického sídliště Dubina v noci na dnešek vylekal požár plynové přípojky. Plameny šlehaly až do šestimetrové výšky. Hasiči dostali oheň pod kontrolu za několik desítek minut. Požár vznikl zřejmě při krádeži plynoměru. Nikdo z obyvatel sídliště nebyl zraněn ani nemusel být evakuován. Škoda se odhaduje na stovky tisíc korun. Příčina požáru se vyšetřuje.	555134	Pardubice	1	Správne, ďalšie názvy: sídlisko Dubina	-
3374	Tábor - Nárůst černých skládek, často i s nebezpečným odpadem, řeší odbor životního prostředí táborské radnice už řadu let. Sběrné dvory jsou umístěny na opačných stranách města, lidé k nim tedy mají daleko, proto mnohdy to, co se nevejde k popelnicím, skončí v lese nebo na odpočívadlech. Tábor chce nyní vybudovat dva nové sběrné dvory, kam by lidé mohli nepotřebný materiál odkládat.	552046	Tábor	1	Správne	-
3392	Kerschbaum/Rakousko - Čeští mechanici se opět vyznamenali v tom dobrém slova smyslu. Na koleje koněspřežky v Rakousku znovu vyjela replika mechanického vozu českého vynálezce Josef Božka. Šlapací stroj, předchůdce lokomotiv, který stál na počátku železniční dopravy, se na dráze objevil po 183 letech a opět u toho byli čeští konstruktéři, kteří repliku sami vyrobili.	-	-	-	Správne	-
3399	Praha - V parku u trojského zámku v Praze začal festival divadla a původních řemesel Archa 2008, který připomíná například 330 let od položení základního kamene zámku. V průběhu týdne byl v zahradě zámku sestaven model archy, která festival symbolizuje. quot;Archa tady v Tróji symbolizuje mnohé, nejčerstvěji tedy povodně v roce 2002,quot; uvedl autor scénáře slavností Václav Špale.	554782	Praha	1	Správne, ďalšie názvy: časť obce Trója	-
3611	Veltrusy - Barokní zámek Veltrusy se už šest let vzpamatovává z povodní, tehdy voda zaplavila nejen rozlehlý park, ale i vnitřní prostory zámku. Památka je přes úsilí všech stále ve špatném stavu, i tak je však v mnoha ohledech unikátem v českém i evropském měřítku.	535273	Veltrusy	1	Správne	-
3775	Kopřivnice - Od zítřka začne platit v Kopřivnici vyhláška zakazující pití alkoholu na některých veřejných místech. Město má také novou tzv. prohibiční mapu. Na ní jsou zachycena místa, kde zákaz platí. Jeho dodržování budou pravidelně kontrolovat policisté.	599565	Kopřivnice	1	Správne	-
3776	Praha - V Klánovicích kdysi bylo golfové hřiště. Současné pokusy o jeho obnovení však narážejí na odpor a spor se táhne již několik let. Jeho poslední fáze se však tentokrát netýká ani tak hřiště samotného, ale spíš staré golfové klubovny. Klánovickým se nelíbí její rekonstrukce. Podle nich nebyla povolená.	554782	Praha	1	Správne, ďalšie názvy: Klánovice	-
3872	Ústí nad Labem - Ústecká radnice se nechala inspirovat sousedními městy a hodlá zrušit poplatky za odvoz odpadu. Zadarmo to však nebude. Plánuje totiž zvýšit daň z nemovitostí. Podle svých propočtů na tom město ani majitelé nemovitostí neprodělají. Opozice je tvrdě proti a již dopředu říká, že tuto quot;rošáduquot; nepodpoří.	554804	Ústí nad Labem	1	Správne	-
3997	České Budějovice/Písek - V jižních Čechách řádí vandalové a lupiči, kteří se zaměřují na sochy. V Českých Budějovicích dvě sochy ukradli a jednu zničili. V Písku zase vandalové poškodili obří pohádkové postavy na náplavce, vymodelované z písku. Oba případy vyšetřuje policie, viníky ale zatím nedopadla.	544256	České Budějovice	1	Správne	-
3997	České Budějovice/Písek - V jižních Čechách řádí vandalové a lupiči, kteří se zaměřují na sochy. V Českých Budějovicích dvě sochy ukradli a jednu zničili. V Písku zase vandalové poškodili obří pohádkové postavy na náplavce, vymodelované z písku. Oba případy vyšetřuje policie, viníky ale zatím nedopadla.	549240	Písek	1		
4017	Jindřichův Hradec - Jindřichohradecká úzkokolejka se potýká se zloději železa. Dnes pracovníci dráhy zjistili už třetí krádež kolejových spojek od konce	545881	Jindřichův Hradec	1	Správne	-

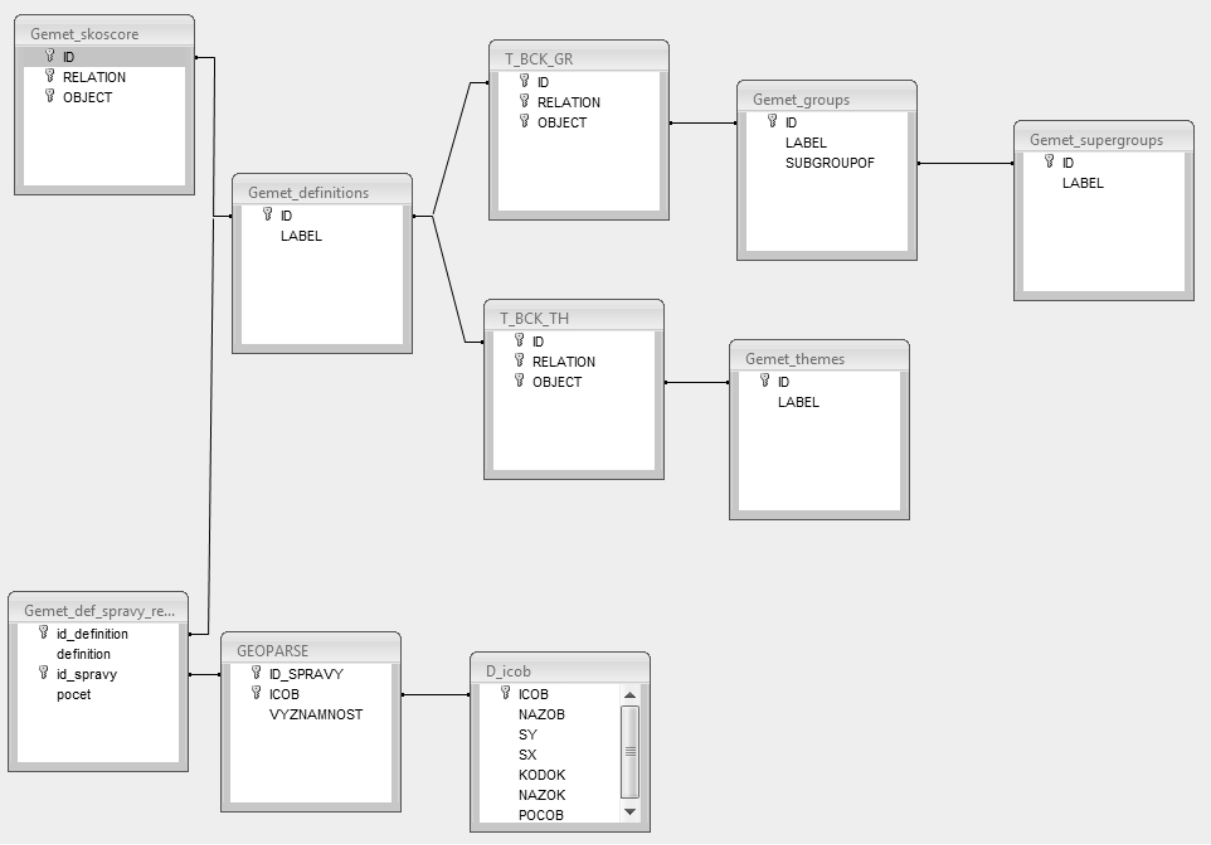
ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
	května. První přitom odhalili těsně před odjezdem vlaku s dětmi a jejich rodiči. Policisté případ vyšetřují jako obecné ohrožení. Vedení dráhy se nyní rozhodlo vypsat na dopadení zlodějů odměnu 20 000 korun.					
4073	Plzeň - Plzeňská teplárenská, největší výrobce tepla i elektřiny v regionu, začíná budovat quot;zelenýquot; blok na spalování biomasy za 850 milionů. V roce 2010 chce vyrábět čtvrtinu energie z obnovitelných zdrojů, loni to bylo 14 procent. Firma chce zdroj spustit 31. března 2010.	554791	Plzeň	1	Správne	-
4122	Praha - Rada Prahy 2 se rozhodla zakázat pití alkoholu na veřejnosti na Karlově a Tylově náměstí. Podle návrhu bude popíjení zapovězeno také v blízkých ulicích. Návrh radnice odešle Magistrátu hlavního města Prahy, který už chystá obecnou vyhlášku o omezení pití alkoholu na veřejných prostranstvích.	554782	Praha	1	Správne, ďalšie názvy Karlovo, Tylovo náměstí	-
4193	Cheb - Nelegální zboží za téměř 190 milionů korun zabavili celníci během víkendové kontrolní akce na tržnici Dragon ve Svatém Kříži u Chebu. Šlo o statisíce neznačených cigaret, padělky textilu i pirátské audio a videonosiče. Tržnice v obci Svatý Kříž zůstává nadále pod stálým dohledem celní správy. Od loňského října je v pořadí 13. tržnicí v západočeském příhraničí, kde celní orgány uplatnily toto dozorové opatření.	554481	Cheb	1	Správne, ďalšie názvy: časť obce Svatý Kříž	-
4197	Horní Jiřetín - Vláda zatím neodepíše zásoby uhlí, které leží pod Horním Jiřetínem. Nedovolí ale, aby město bylo kvůli uhlí zbouráno. To je jedna ze zásadních informací, kterou přineslo dnešní výjezdní zasedání vlády v Teplicích.	567175	Horní Jiřetín	1	Správne	-
4197	Horní Jiřetín - Vláda zatím neodepíše zásoby uhlí, které leží pod Horním Jiřetínem. Nedovolí ale, aby město bylo kvůli uhlí zbouráno. To je jedna ze zásadních informací, kterou přineslo dnešní výjezdní zasedání vlády v Teplicích.	567442	Teplice	2		
4226	Nový Jičín - Provazochodec Manuel Berousek známý jako Berondini spadl při sobotní produkci v Novém Jičíně z lana a po převozu do nemocnice zemřel. Padající artista zranil při svém pádu rovněž sedmiletou dívku. Ta je nyní na jednotce intenzivní péče porubské fakultní nemocnice, ale její stav je stabilizovaný.	599191	Nový Jičín	1	Správne	-
4441	Kašperské Hory - Kalamita, kterou způsobila větrná bouře Emma, turistickou sezonu v jižních Čechách nijak zásadně neomezí. Na rozdíl od loňského roku nechtějí zavírat rozsáhlé oblasti a většina stezek je už průchozích. Vítr v Národním parku Šumava zničil statisíce stromů.	556432	Kašperské Hory	1	Správne	-
4494	Praha - Městský soud v Praze jako jeden z prvních otevřel nové informační centrum. Mělo by pomoci hlavně účastníkům řízení, ale i veřejnosti. Podle ministra spravedlnosti Jiřího Pospíšila je to jeden ze zásadních kroků ke zkvalitnění práce českých soudů. Ulehčí se i úředníkům, kteří doposud nahlížení do spisů zprostředkovávali. Ministr Pospíšil už svolal schůzku s předsedy krajských soudů. Chce, aby se podobná centra otevřela po celé České republice.	554782	Praha	1	Správne	-
4757	Babylon - Český zájezdový autobus, který vezl středoškoláky do Alp, havaroval u obce Babylon na Domažlicku. Sjel do příkopu, kde se převrátil na bok. Z osmačtyřiceti cestujících bylo 12 studentů lehce zraněných, většina z nich byla po ošetření propuštěna. Příčinou nehody byla podle policie nepozornost řidiče, který zřejmě za jízdy manipuloval s videem, najel na krajnici a ta se utrhla. Dechová zkouška vyloučila, že by před jízdou pil. Hmotná škoda se odhaduje na 100 000 korun.	553433	Babylon	1	Správne	-
4758	Ústí nad Labem - Exekutoři v Ústí nad Labem dnes od rána vymáhali dlužné částky od neplatičů jízdného v MHD. Vybrali si šest dlužníků, kteří se zavázali splácet, ale svůj slib nedodrželi, nebo poslali jen jednu splátku. U dlužníků nabylo rozhodnutí o exekuci právní moci a nebyl u nich zjištěn žádný příjem nebo účet, ze kterého by se dala částka strhnout.	554804	Ústí nad Labem	1	Správne	-
4840	Opava - Městské sady v Opavě čeká obnova. Vedení města plánuje vykácet 460 vzrostlých stromů, které by mohly ohrozit návštěvníky. Místo nich v sadech vysadí nové. Některým obyvatelům Opavy se ale záměr radnice nelíbí. Rekonstrukce Městských sadů je podle radnice důležitý krok, podle obyvatel zbytečnost.	505927	Opava	1	Správne	-
4854	Liberec - Liberecký azylový dům Speramus se potýká s finančními problémy. Od státu totiž, stejně jako většina provozovatelů sociálních služeb, zatím nedostal slíbené peníze na provoz. Ministerstvo práce a sociálních věcí předpokládalo, že dotace pošle těmto zařízením do 15. února.	563889	Liberec	1	Správne	-

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
4996	Březová - Starostové měst a obcí, kteří mají na svém území komunální skládky, nesouhlasí s novelou zákona o odpadech a chtějí se bránit. Chystané změny by podle nich snížily příjmy z uloženého odpadu a obce by přišly o milióny korun. Starosty mrzí, že s nimi o novele zákona, kterou chystá ministerstvo životního prostředí a Asociace krajů, nikdo nejednal. Podle ministerstva životního prostředí ale o tolik peněz nepřijdou. Novela počítá s tím, že poplatky za odpad se budou postupně zvyšovat. Ze současných asi 200 až na 1500 korun.	560294	Březová	1	Správne	Podľa počtu obyvateľov
5050	Světlá nad Sázavou - Věznice v zemi jsou přeplněné a ministerstvo spravedlnosti nemá peníze na rozšíření a opravu stávajících věznic. Chystá však otevření nové věznice pro muže v Rapoticích na Třebíčsku. Kapacitní problémy ale mezitím hlásí největší ženská věznice u nás ve Světlé nad Sázavou. Priorit je však mnoho a při rozdělování peněz je třeba postupovat opatrně.	591581	Rapotice	2	Správne	-
5050	Světlá nad Sázavou - Věznice v zemi jsou přeplněné a ministerstvo spravedlnosti nemá peníze na rozšíření a opravu stávajících věznic. Chystá však otevření nové věznice pro muže v Rapoticích na Třebíčsku. Kapacitní problémy ale mezitím hlásí největší ženská věznice u nás ve Světlé nad Sázavou. Priorit je však mnoho a při rozdělování peněz je třeba postupovat opatrně.	569569	Světlá nad Sázavou	1		
5167	quot;Jsme rádi, že je to konečně za námi. Stanovisko životního prostředí je nejdůležitějším podkladem pro stavební povolení,quot; řekl deníku Michal Kestl z olomoucké firmy K3 Sport, která areál postaví.	-	-	-	Správne	-
5197	Lékaři v ústecké poliklinice si zatím nemohou stěžovat. Pro třicet korun sahají pacienti do kapsy hned, jak vstoupí do ordinace. Jednomu z doktorů dokonce přidali k penězům i kasičku. Sami pacienti problém nevidí: quot;Když musíte, tak musíte. Pokud člověk potřebuje doktora, tak k němu musí jít. Zdraví je přednější než peníze.quot; Lékařka Jelena Labuťová potvrzuje: quot;Problémy nemáme, lidé platí.quot;	-	-	-	Správne	-
5283	Vrchlabí - Strážníci ve Vrchlabí si stěžovali zastupitelům, že zaměstnavatel porušuje zákoník práce. Uvedli, že dostali pokyn, aby se snažili přistihnout zastupitele opozice při přestupku. Za to měli dostat odměnu. Zastupitel Rudolf Fiala podal trestní oznámení na velitele a starostu pro podezření ze zneužití pravomoci veřejného činitele. Podle starosty je však tvrzení strážníků nepravdivé.	579858	Vrchlabí	1	Správne	-
5371	Moravskoslezský kraj - Občanské sdružení Adra v Moravskoslezském kraji hledá dobrovolníky, kteří by docházeli do sociálních a zdravotnických zařízení. V současné době pomáhá personálu i klientům nemocnic a domovů důchodců ve svém volném čase přibližně 350 lidí, což je o 150 méně, než by bylo potřeba.	-	-	-	-	-
5459	Brno - Brno možná nepřijde o železniční zastávku v centru města, na kterou přijíždí regionální spoje z okolí města. Brněnský zastupitel Jan Veselý (KDU-ČSL) dnes představil nové řešení rekonstrukce železničního uzlu Brno, které počítá s tím, že z nového hlavního nádraží posunutého o 800 metrů na jih by vedla mimoúrovňová rychlodráha napříč brněnským centrem v trase stávající železnice. Návrh postoupí do pracovní skupiny města, která má do konce listopadu předložit možná řešení přesunu nádraží.	582786	Brno	1	Správne	-
5588	Strakonice - Mentálně postižené děti i dospělí mají ve Strakonicích k dispozici speciální zařízení, takzvaný snoezelen. Místnost vybavená nejmodernější technikou, která pomáhá duševně postiženým, je jednou z mála na jihu Čech. Slouží pro více než pět desítek klientů z celého kraje.	550787	Strakonice	1	Správne	-
5591	Karovy Vary - Krajská knihovna v Karlových Varech otevřela na dvou místech tzv. biblioboxy, do kterých mohou čtenáři vracet vypůjčené knížky. Speciální schránky knihovna vybírá dvakrát týdně a podle čárových kódů knihy odepíše čtenářům z výpůjčních karet. Tato služba je zdarma a knihovna ani do budoucnosti neplánuje její zpoplatnění. Tyto velké kovové schránky jsou v Čechách novinkou.	554961	Karlovy Vary	2	Správne	-
5720	Pardubice - V Rybitví na Pardubicku si připomínají 180 let od vynálezu ruchadla bratranců Veverkových. Na tehdejší dobu převratné zařízení usnadňovalo orbu, která byla jednou z fyzicky nejnamáhavějších zemědělských prací. Kromě dobových dokumentů se v obci dochoval i rodný domek vynálezců.	555134	Pardubice	1	Správne	-
5720	Pardubice - V Rybitví na Pardubicku si připomínají 180 let od vynálezu	575593	Rybitví	2		

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
	ruchadla bratranců Veverkových. Na tehdejší dobu převratné zařízení usnadňovalo orbu, která byla jednou z fyzicky nejnamáhavějších zemědělských prací. Kromě dobových dokumentů se v obci dochoval i rodný domek vynálezců.					
5733	Praha - Pražský Týden mobility má mimo jiné upozornit zodpovědné orgány na přetrvávající problémy, které trápí cyklisty, pohybující se po pražských ulicích. Stále chybí dostatečné množství cyklostezek, ale i vstřícnost Dopravního podniku hlavního města Prahy při přepravě kol v metru, autobusech a tramvajích. Cyklisté mají stále pocit, že pro ně dělá město málo.	554782	Praha	1	Správne	-
5743	Opočno - Bez vědomí a souhlasu zastupitelů zrušil Krajský úřad v Hradci Králové porodnici a gynekologii nemocnice v Opočně. Vedení zdravotnického holdingu, který nemocnice v kraji spravuje, přitom zastupitele trvale ujišťuje, že nad řízením zdravotnictví kontrolu neztratí a že o změnách budou rozhodovat oni.	569810	Hradec Králové	2	Správne	Podľa vzdialenosti od Hradca Králové
5743	Opočno - Bez vědomí a souhlasu zastupitelů zrušil Krajský úřad v Hradci Králové porodnici a gynekologii nemocnice v Opočně. Vedení zdravotnického holdingu, který nemocnice v kraji spravuje, přitom zastupitele trvale ujišťuje, že nad řízením zdravotnictví kontrolu neztratí a že o změnách budou rozhodovat oni.	576590	Opočno	1		
5923	Hradec Králové - Moderní přístroje a důmyslně propracovaná technika. Takové jsou metody dnešních zlodějů aut. Stovky krádeží evidují každý rok policisté na východě Čech. Pravděpodobnost, že se majitelé s autem znovu setkají, je přitom velmi malá.	569810	Hradec Králové	1	Správne	-
6007	České Budějovice - Některé lesy na Novohradsku zůstanou kvůli polomům nepřístupné další dva měsíce. Stromy v porostech stále padají a navíc se kolem cest objevují nedostatečně zajištěné hranice dřeva, které se můžou kdykoliv sesunout. Na nebezpečí upozornili sami lesníci, podle kterých se odvoz kmenů zpožďuje a skládky se zvětšují.	544256	České Budějovice	1	Správne	-
6016	Praha - Zatím 49 stanovišť taxi v Praze bude nejpozději do 15. srpna označeno názvem Fair Place, neboli Poctivé místo. Přesně takový nápis bude mít v Praze necelá polovina taxikářských stanovišť, hlavně na lukrativních místech v centru. Půjde pouze o ta, která mají svého správce, jenž zaručí, že tamní taxikáři nepředražují jízdné a neporušují zákony. A magistrát je bude často kontrolovat.	554782	Praha	1	Správne	-
6068	Praha - Pražský magistrát chystá omezení pro některá alternativní vozidla, například pro rikši. Podle radních Prahy 1 vládne v přepravě turistů v centru města chaos. Tyto jízdy často brzdí automobilový provoz a úřadům chybí možnost, jak je regulovat. Změna by měla přijít s novelou tržního řádu.	554782	Praha	1	Správne	-
6093	Pardubice - Ve Chvaleticích na Pardubicku se znovu rozbíhají práce v nepovoleném skladišti chemikálií. Osmiměsíční přestávku v likvidaci nebezpečných látek způsobil nedostatek peněz, nové výběrové řízení i šetření antimonopolního úřadu. Desítky tun nezákonně uskladněných jedů úřady objevily vloni v červnu. Od té doby se podařilo zlikvidovat zhruba polovinu všech odpadů.	575071	Chvaletice	2	Správne	-
6093	Pardubice - Ve Chvaleticích na Pardubicku se znovu rozbíhají práce v nepovoleném skladišti chemikálií. Osmiměsíční přestávku v likvidaci nebezpečných látek způsobil nedostatek peněz, nové výběrové řízení i šetření antimonopolního úřadu. Desítky tun nezákonně uskladněných jedů úřady objevily vloni v červnu. Od té doby se podařilo zlikvidovat zhruba polovinu všech odpadů.	555134	Pardubice	1		
6138	Liberec - Evropskou archeologickou raritu - šest a půl tisíce let starou vesnici - zničí bagry. Unikátní naleziště z mladší doby kamenné objevili odborníci v Příšovicích. Pozemky chtěli od soukromé firmy, která tu hodlá stavět, vykoupit. Nesehnali ale dost peněz.	563889	Liberec	1	Správne	-
6138	Liberec - Evropskou archeologickou raritu - šest a půl tisíce let starou vesnici - zničí bagry. Unikátní naleziště z mladší doby kamenné objevili odborníci v Příšovicích. Pozemky chtěli od soukromé firmy, která tu hodlá stavět, vykoupit. Nesehnali ale dost peněz.	564354	Příšovice	2		
6280	Plzeň - Dejte nám pár let a Bolevecký rybník v Plzni bude minimálně dvakrát tak čistý než dnes. Slibují to odborníci. Se svým experimentem začali už loni. A první výsledky tu prý už jsou.	554791	Plzeň	1	Správne	-
6292	Lednice (Břeclavsko) - Lednicko-valtický areál o víkendu oslaví 10 let od	584631	Lednice	1	Správne	-

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
	zapsání na seznam památek UNESCO. Během této doby prošel postupnou rekonstrukcí za 300 milionů korun. Jen lednický zámek navštíví ročně více než 300 tisíc turistů, z nichž třetinu tvoř cizinci.					
6306	Jaroměř - Přes dvě stě dravců a dalších ptáků ročně ošetří ochránci přírody v Záchranné stanici v Jaroměři. Mnozí mají popáleniny ze stožárů vysokého napětí. Ekologové proto stupňují tlak na energetiky, aby na stožáry nainstalovali ochrany. Přirozeným lákadlem pro ptáky totiž stožáry elektrického vedení zůstanou i nadále. Montér ČEZ Jiří Hrdinka k tomu říká: „Sedají tam dravci, potom hlavně čápi a volavky. Nejvíce se to projevuje tam, kde ptáci hnízdí.“	574121	Jaroměř	1	Správne	-
6323	Zlín - V centrum filmů pro děti a mládež se promění město Zlín, které letos pořádá už 47. ročník nejstaršího festivalu svého druhu na světě. Lákadly jsou premiéra třetího dílu Shreka a návštěva dětské filmové hvězdy ze snímků Šestý smysl a Umělá inteligence.	585068	Zlín	1	Správne	-
6359	Hradec Králové - Dominantu Hradce Králové, renesanční Bílou věž, budou na jaře a v létě oživovat scénické prohlídky. Nabídnou divadelní i hudební vystoupení. Režisérka a scenáristka projektu Emílie Zámečníková zdůrazňuje, že při vytváření představení čerpala především z příběhů a pověstí, které existují a které se vypráví.	569810	Hradec Králové	1	Správne	-
6369	Pačejov - Obce v Pošumaví se spojily v boji proti plánům na výstavbu úložiště jaderného odpadu u Pačejova. Dokončily společné memorandum, kterým osloví vládu, sněmovnu a ministerstva. Při rozhodování chtějí mít právo veta.	556912	Pačejov	1	Správne, ďalšie Pošumaví	-
6425	Šumava - Na Šumavě se znovu vedou spory o splavnění horního toku Vltavy. Kompromis místních podnikatelů, starostů i správců Národního parku se nelíbí nevládním ekologickým organizacím. Ti požadují minimální výšku hladiny 75 centimetrů, majitelé půjčoven lodí a kempu by zase nechtěli žádný limit.	-	-	-	Správne	-
6496	Ostrava - 34 tisíc tun nebezpečných látek stále zamořuje půdu v Kopřivnici na Novojičínsku. Tamní skládka ohrožuje zdraví lidí už od šedesátých let. Městu se od roku 1991, kdy začal platit nový zákon o odpadech, nepodařilo sehnat potřebných 60 milionů na sanaci skládky.	599565	Kopřivnice	2	Správne	-
6496	Ostrava - 34 tisíc tun nebezpečných látek stále zamořuje půdu v Kopřivnici na Novojičínsku. Tamní skládka ohrožuje zdraví lidí už od šedesátých let. Městu se od roku 1991, kdy začal platit nový zákon o odpadech, nepodařilo sehnat potřebných 60 milionů na sanaci skládky.	554821	Ostrava	1		
6537	Pardubice - Občané protestovali před budovou krajského úřadu v Pardubicích proti stavbě spalovny v Opatovicích nad Labem. Svůj protest proti vznikajícímu zařízení podpořili peticí, kterou podle člena petičního výboru Adama Záruby podepsalo 6460 lidí. Spalovna za dvě miliardy korun by měla vzniknout v areálu místní elektrárny. Odpůrci stavby nechápou, proč Pardubický kraj spalovnu stále prosazuje, i když ji lidé v referendu odmítli.	583553	Opatovice	2	Nesprávne, má byť Opatovice nad Labem – chýbajú pády	-
6537	Pardubice - Občané protestovali před budovou krajského úřadu v Pardubicích proti stavbě spalovny v Opatovicích nad Labem. Svůj protest proti vznikajícímu zařízení podpořili peticí, kterou podle člena petičního výboru Adama Záruby podepsalo 6460 lidí. Spalovna za dvě miliardy korun by měla vzniknout v areálu místní elektrárny. Odpůrci stavby nechápou, proč Pardubický kraj spalovnu stále prosazuje, i když ji lidé v referendu odmítli.	555134	Pardubice	1		
6572	Třebíč - K domovu pro seniory, ubytovně, ke sportovní hale nebo do domova mládeže v Třebíči vede cesta pouze velmi tmavou ulicí. Lidé, kteří tu bydlí, se zlobí, že světla nesvítí už hodně dlouho a náhradní halogenové lampy jim vadí. Město se ale s krajem Vysočina, kterému ulice patří, stále nemůže dohodnout na tom, kdo světla opraví.	590266	Třebíč	1	Správne	-
6732	Šumava - Olympijská vítězka Kateřina Neumannové trénovala v zakázaných místech Šumavy. Upozornili na to ekologičtí aktivisté, píše o tom i dnešní MF Dnes. Ředitel Národního parku Šumava Alois Pavlička potvrdil, že lyžařce umožnil dvouhodinový trénink na trase, kam je vstup pro běžkaře zakázán. Podle aktivistů tím ředitel pochybil. Případ šetří Ministerstvo životního prostředí.	-	-	-	Správne	-
6741	Milionové ztráty počítají hoteliéři v Novém Městě na Moravě. Pořadatelé totiž museli kvůli nedostatku sněhu zrušit největší sportovní akci Vysočiny - závody Světového poháru v běhu na lyžích. Té se každoročně účastní běžecká špička z celého světa.	596230	Nové Město na Moravě	2	Správne	-

ID_GEN_SPRAVY	DESCRIPTION	ICOB	NAZOB	VYZNAMNOST	POZNÁMKA	SPRESNENIE
6884	Severní Čechy - Lyžařská sezona na českých horách právě začala. První vlek spustili v Krkonoších - na Černé hoře v Janských Lázních. Sjezdovky v dalších zimních střediscích zatím zůstávají uzavřené.	579351	Janské Lázně	2	Správne	-
6919	České Budějovice - Činností Veřejných služeb, firmy řízené magistrátem Českých Budějovic, se začala zabývat policie. Měly odtud utíkat nejméně stovky tisíc korun. Policejní vyšetřování odstartovala kontrola, kterou ve Veřejných službách nařídil magistrát v minulých dnech. Připravuje totiž likvidaci firmy. Inspekce podle bývalých náměstků i úřadujícího primátora prokázala v městské firmě rozsáhlé finanční úniky.	544256	České Budějovice	1	Správne	-
6967	Plzeň - Dopravní podnik v Plzni už nebude mít své revizory. Vedení rozhodlo, že kontrolovat pasažéry MHD bude soukromá firma. Zaměstnanci a odboráři s tím nesouhlasí. Podnik se sice podle nich zbaví administrativy spojené s vymáháním pokut, připraví ale o práci své dlouholeté zaměstnance.	554791	Plzeň	1	Správne	-
6976	Karlovy Vary - Ostrov na Karlovarsku hostí Dětský filmový festival Oty Hofmana. Malým divákům nabízí 35 filmů. Lákadlem se letos stala dětská média. Desítky mladých novinářů se na festivalu učí pracovat s hlasem, vytvářet zprávy a informovat lidi o tom, co se v Ostrově děje.	554961	Karlovy Vary	1	Správne	Podľa oblasti
6976	Karlovy Vary - Ostrov na Karlovarsku hostí Dětský filmový festival Oty Hofmana. Malým divákům nabízí 35 filmů. Lákadlem se letos stala dětská média. Desítky mladých novinářů se na festivalu učí pracovat s hlasem, vytvářet zprávy a informovat lidi o tom, co se v Ostrově děje.	555428	Ostrov	2		



Obr. 11-1 Schéma vzťahov medzi tabuľkami. Model použitý v prostredí MS Access pre vyhodnotenie predovšetkým tém správ