

# Na internetu je dostupný korpus mediálních textů o Národním parku Šumava

Na webové adrese [sumava-corporus.narra.eu](http://sumava-corporus.narra.eu) je od pondělí 14. 9. 2015 veřejně dostupný rozsáhlý korpus téměř 7000 textů týkajících se Šumavy a zejména jejího národního parku. Snadno prohledávatelný soubor převážně mediálních textů z let 1883–2010 poskytne svým uživatelům – novinářům, studentům a ostatním zájemcům – dostupný a přehledný nástroj pro orientaci ve vývoji myšlení a psaní o Šumavě.

Na základě jednoduchého třídění pomocí kategorií „autor“, „respondent“, „zdroj“, „žánr“, „datum“, „titulek“ a za pomoci fulltextového vyhledávání může uživatel nejen rychle najít texty vyjadřující se k dílčímu tématu, ale také získat informace o jejich širším kontextu.

Centrálním korpusem je MediaBank, který sestává z textů vzniklých v letech 1991 až 2010. Převážná většina sebraných textů pochází z českých tištěných médií nebo jejich internetových verzí. Korpus je doplněn přepisem televizních a rozhlasových zpráv a reportáží.

MediaBank obsahuje necelých 6800 textů o rozsahu 2,9 milionu slov.

Druhým (referenčním) korpusem je PrefBank, který je souborem předmluv, doslovů a částečně také kapitol týkajících se obecné charakteristiky pohoří vypsanych z průvodců, fotografických publikací, memoárů a dalších pramenů týkajících se Šumavy. Sestává z téměř 80 textů vzniklých mezi roky 1883 a 2013.

Třetím (referenčním) korpusem je KlosterBank, který je souborem předmluv a doslovů k beletristickým dílům Karla Klostermanna. Celkem 19 textů odráží vývoj vnímání Šumavy od

počátku 20. století.

Všechny tři korpusy vznikly jako materiálová báze pro výzkum, který se věnuje dynamice mediálního diskurzu o Národním parku Šumava. Korpusy lze využívat jako podklad pro kvantitativní analýzu, jejímž účelem je odhalovat frekventovaná sémanticky relevantní lexikálně-gramatická schémata. S pomocí MediaBank je tak například dobře možné odkrývat komunikační stereotypy či argumentační strategie, které jsou pro debatu o Národním parku Šumava klíčové.

Rozhraní korpusu však také umožňuje například velmi rychle a pohodlně zjistit, v jakém období a v jakých médiích ten který aktér (autor) publikoval, s kým konkrétní deník rád dělá rozhovory nebo také to, kdy se o Šumavě či nějakém konkrétním tématu nejvíc psalo. Prostředí MediaWiki navíc nabízí komplexní vyhledávání na standardním „googlovském“ principu.

Sumava-corpus by měl sloužit nejen jako snadno dostupný materiál pro vysokoškolské studenty zabývající se (například ve svých diplomových pracích) analýzou diskurzu, mediální komunikací ap., ale také novinářům, kteří se za pomoci korpusu mohou rychle zorientovat v dílčích rysech dlouholeté debaty o Národním parku Šumava. Korpusy by však měly sloužit také těm, které jednoduše zajímá, jak se o Šumavě v minulosti psalo a mluvilo.

Kontakt:

Michal L. Hořejší, telefon +420 775 220 216, e-mail mhorejsi@ujc.cas.cz