**Pravděpodobnostní odhad založený na ontologii pro hodnocení sémantické podobnosti systémů klasifikace využití půdy/půdního krytu**

**Ontology-Based Probabilistic Estimation for Assessing Semantic Similarity of Land Use/Land Cover Classification Systems**

Zhou, X.; Xie, X.; Xue, Y.; Xue, B. Ontology-Based Probabilistic Estimation for Assessing Semantic Similarity of Land Use/Land Cover Classification Systems. *Land* **2021**, *10*, 920.

**Klíčová slova:** ontologie; sémantická podobnost; systémy klasifikace využití půdy/pozemkového krytu

**Dostupné z:** https://www.mdpi.com/2073-445X/10/9/920

Využití půdy a půdního krytu (Land Use and Land Cover - LULC) je důležité pro zobrazení historie a současného stavu pozorování půdy, hospodaření, analýzy půdy, monitorování stanovišť a plánování krajiny. Klasifikační mapy LULC jsou k dispozici v mnoha prostorových a časových měřítkách vytvořených mnoha typy klasifikačních standardů. V současné době byly různými zeměmi, institucemi a organizacemi vytvořeny desítky klasifikačních systémů LULC.

Tyto systémy klasifikace půdního krytu a jejich legendy se však liší, což ztěžuje porozumění studiím a sdílení výsledků, a proto byly vyvinuty přístupy, jak se vypořádat se sémantickou heterogenitou. Obecně tyto přístupy postrádají doménové ontologie, což může být významnou překážkou pro implementaci těchto přístupů z hlediska hodnocení sémantické podobnosti. V tomto článku výzkumníci navrhují ontologický přístup k hodnocení doménové podobnosti klasifikačních systémů a standardů LULC. Vyvíjejí doménové ontologie, které explicitně definují popisy a kódy různých klasifikačních systémů a standardů LULC jako sémantické informace, a formálně organizují tyto sémantické informace jako logická pravidla uvažování. Sémantickou podobnost termínů ve dvou různých systémech klasifikace půdního krytu pak lze vypočítat pomocí algoritmů (např. Bayesova algoritmu).

Souhrnně lze říci, že klasifikační systémy LULC čelí dvěma důležitým výzvám:

První z nich je polysémie: pozemek může být různými klasifikačními systémy LULC definován jako různé typy LULC.

Druhým je mnohoznačnost: stejný pojem LULC může být různými klasifikačními systémy LULC definován odlišně.

Polysémie a mnohoznačnost jsou součástí sémantické heterogenity, jejímž cílem je řešit problém záměny výrazů při zpracování přirozeného jazyka.

Pro vypořádání se se sémantickou heterogenitou navrhla řada prací přístupy k sémantické harmonizaci s cílem integrovat informace a rysy z různých zdrojů do koherentního celku.

V této práci výzkumníci vyvinuli doménové ontologie, které explicitně definují popisy a kódy různých klasifikačních systémů a standardů LULC jako sémantické informace a logická pravidla uvažování. Na základě sémantiky a pravidel použili Bayesův algoritmus k vytvoření podmíněného pravděpodobnostního modelu pro výpočet sémantické podobnosti kategorií LULC v různých klasifikačních systémech a standardech LULC. Zkušenosti ukazují, že sémantickou podobnost lze efektivně měřit začleněním pravděpodobnostního modelu založeného na obsahu ontologie.

Existuje několik možných rozšíření tohoto výzkumu, která se zaměřují na integraci obsahu různých klasifikačních systémů a standardů LULC. Pro explicitní reprezentaci skrytých sémantických informací je třeba ještě prozkoumat fúzi různých doménových ontologií pro klasifikační systémy a standardy LULC.

**Zpracovali:** Vincent Onckelet (Plan4all z.s.) a Hana Kubíčková (Plan4all z.s.)