

Vincze Veronika

tudományos munkatárs
Szegedi Tudományegyetem
Természettudományi és Informatikai Kar
MTA-SZTE Mesterséges Intelligencia Kutatócsoport

A Mesterséges Intelligencia Kutatócsoport projektjeinek nyelvészeti vonatkozásait felügyeli és koordinálja. Érdeklődési körébe tartozik a korpusz- és ontológiaépítés, a jelentéségyértelműsítés és a többszavas kifejezések számítógépes kezelése. Ezenkívül foglalkozik számítógépes morfológiával és szintaxissal, valamint információkinyeréssel, kiemelten a tagadott és/vagy bizonytalan szövegrészek gépi felismerésével. Az évente megrendezett Magyar Számítógépes Nyelvészeti Konferencia szervezőbizottságának tagja. A konferencia előadói mind a felsőoktatási és tudományos intézményekben, mind a vállalatoknál zajló legfrissebb beszéd- és nyelvtechnológiai kutatások eredményeiről beszámolnak. A Szegedi Tudományegyetemen a nyelvtechnológiával és a számítógépes nyelvfeldolgozással kapcsolatos számos kurzust tartott.

Bevezetés a korpuszok és nyelvészeti adatbázisok világába

Az előadás célja, hogy a hallgatóságot megismertesse néhány korpusszal és egyéb nyelvészeti adatbázissal, továbbá a korpusznyelvészet alapjaival. A legfontosabb alapfogalmak után ismertetjük a különféle korpusztípusokat, létrehozási módjukat, továbbá néhány példán keresztül megmutatjuk, milyen nyelvészeti jellegű információkat (annotációkat) tudunk a szövegekben kódolni. Arra is hozunk példát, hogy a nyers szövegállományból miként tudunk automatikusan annotált adatbázist előállítani. A korpuszok gyakorlati felhasználására is külön figyelmet fordítunk: bemutatjuk, hogy a korpuszokból származó adatokat hogyan lehetséges kigyűjteni, majd azokat nyelvészeti vagy más bölcsészettudományi kutatásra felhasználni.

kulcsszavak: korpusz, nyelvészeti adatbázis, annotálás, korpusznyelvészet