**Tanulmány a   
„Természetes nyelvek Inferenciája”  
című témakiíráshoz.**

# Bevezetés

A téma lényege a természetes nyelven leírt mondatok reprezentációja, illetve az ezen mondatok közti kapcsolat meghatározása. A feladat folyamán a számítógépnek három féle lehetőség közül kell eldöntenie, hogy milyen kapcsolat van a neki megadott mondatpárok között. Az egyik mondat lehet az másik következménye (*Entailment*), ellentmondhatnak egymásnak (*Contradiction*), vagy nincs semmilyen jelentésbeli kapcsolat köztük (*Neutra*l).

# Kísérleti körülmények

A mondatok szavait egy szóbeágyazásos módszerrel előállított adatbázisból származó szóvektorokkal ábrázolom. Kísérletemhez a MultiNLI előállított 40 dimenziós sűrű vektorait, és 100 dimenziós ritka (sparse) vektorait használtam. A mondatokat az NYU MultiNLI adatbázisából használom tanítópéldaként, és az általuk szolgáltatott tesztelő adatbázison tesztelem a logisztikus regresszióval előállított modellt. A modell kiértékelése végén egy 3\*3-as konfúziós mátrixot kapunk, amelynek oszlopaiban a modell által előrejelzett típusok vannak, soraiban meg a mondatpárok valódi típusai. A mátrix főátlójában találhatóak a modell által helyesen megjósolt étékek száma.

# Eredmények