

THINKER - Entity Linking System for Turkish Language

THINKER is a novel entity linking system for Turkish language that automatically maps entity mentions in a text content to the corresponding real world entities defined in Vikipedi or the Turkish dictionary published by Turkish Language Association (TLA).

Source codes of the project can be accessible at this [address](#). We also publish taxonomy represented for entity discovery, pre-trained entity vectors and experiment data sets:

- A. [Turkish Entity Linking Dataset](#); The dataset contains approximately 5 news articles for 112 different meanings from online news papers covering a variety of Turkish ambiguous phases such as pas, petrol, etc..
- B. [Vikipedi Entity Vectors](#): The archive contains two separate json files.
 - **Metadata Vectors**: Entity vectors trained on category, type and infobox information of Vikipedi articles. The model is distributed in 300-dimensional vectors.
 - **Link Vectors**: Entity vectors trained on link information of Vikipedi articles. The model is distributed in 150-dimensional vectors.

The json files can be read using the [Gson](#) Java library that can be used to convert json file into equivalent Java map object. The map object maintains key and value pairs. Vikipedi article ids are the keys and values are their entity vectors. An example java code is given below:

```
try {
    Reader reader = new FileReader("embeddings.json");
    Gson gson = new GsonBuilder().create();
    embeddings = gson.fromJson(reader, HashMap.class);

    nn = new EntityEmbedding();
    nn.load(modelFile);
} catch (FileNotFoundException e) {
    LOGGER.error(e);
}
```

- C. The taxonomy used in the entity discovery process and WordNet mappings are provided in the table below:

| Parent Type | Type | WordNet 3.1 Mapping |
|---------------|------------------|------------------------|
| animal | kuş | bird#1 |
| | bitki | plant#2 |
| | örümcek | spider#1 |
| | memeli | mammal#1 |
| | balık | fish#1 |
| | böcek | insect#1 |
| | sürüngen | reptile#1 |
| | köpek ırkı | dog#1 |
| | yılan | snake#1 |
| | primat | primate#1 |
| | bakteri | bacteria#1 |
| | biyoloji | biology#1 |
| concept | etnik grup | ethnic group#1 |
| | dil | language#1 |
| | kraliyet | kingdom#1 |
| | hastalık | disease#1 |
| | bayrak | flag#1 |
| | mitoloji | mythology#1 |
| | marş | anthem#1 |
| | dil ailesi | language#1 |
| | anatomi | anatomy#2 |
| | ödül | award#3 |
| | islam | islam#1 |
| | harf | letter#2 |
| | programlama dili | programming language#1 |
| | renk | color#1 |
| | kimya | chemistry#1 |
| | sayı | number#5 |
| | fizik | physics#2 |
| | hazır gıda | food#2 |
| | matematik | mathematics#1 |
| creative work | film | movie#1 |
| | albüm | album#1 |
| | tekli | single#2 |
| | kitap | book#1 |
| | televizyon | television#1 |
| | şarkı | song#1 |
| | yazılım | software#1 |
| | video oyunu | video game#1 |
| | tv bölüm | episode#1 |

| | | |
|-------|-------------------------------|--------------------|
| | opera | opera#1 |
| | oyun | game#1 |
| | doctor who bölüm | episode#1 |
| | single | song#1 |
| | işletim sistemi | operating system#1 |
| | sure | prayer#4 |
| | televizyon sezonu | season#1 |
| | müzik türü | music genre#1 |
| | öykü | story#1 |
| | dizi | episode#1 |
| | prison break bölümü | episode#1 |
| | tiyatro oyunu | play#1 |
| | sanatçı diskografisi | discography#1 |
| | roman | novel#1 |
| | simpsonlar bölümü | episode#1 |
| | sanat eseri | artwork#1 |
| | edebiyat | literature#1 |
| | müzik | music#1 |
| | lost | serial#1 |
| | lost karakteri | |
| | seri | serial#1 |
| event | savaş | war#1 |
| | futbol maçı | match#2 |
| | futbol kulübü sezonu | season#1 |
| | uluslararası futbol turnuvası | tournament#1 |
| | futbol ligi sezonu | season#1 |
| | futbol turnuvası | tournament#1 |
| | antlaşma | agreement#1 |
| | sivillere karşı saldırı | attack#1 |
| | seçim | election#1 |
| | konser turnesi | tour#1 |
| | eurovision | contest#1 |
| | deprem | earthquake#1 |
| | nba draft | |
| | olimpiyatlar | Olympics#1 |
| | bayram | holiday#1 |
| | parlamento | parliament#1 |
| | spor derbisi | |
| | olimpiyatlarda etkinlik | Olympics#1 |
| | spor ligi | league#1 |
| | festival | festival#1 |
| | kral tv vmö | award#2 |

| | | |
|----------|-----------------------|-----------------------|
| location | yerleşim | placement#3 |
| | türkiye köy | village#1 |
| | portekiz bucak | district#1 |
| | almanya yerleşim yeri | placement#3 |
| | türkiye ilçe | county#1 |
| | eski ülke | country#1 |
| | portekiz belediye | municipality#1 |
| | ülke | county#1 |
| | nehir | river#1 |
| | göl | lake#1 |
| | eski idari bölüm | |
| | italya il | city#1 |
| | ada | island#1 |
| | türkiye belde | town#1 |
| | iran köy | village#1 |
| | italya belde | town#1 |
| | koruma alanı | |
| | dağ | mountain#1 |
| | türkiye il | city#1 |
| | abd eyalet | state#1 |
| | il | city#1 |
| | şehir | city#1 |
| | türkiye mahalle | vicinity#1 |
| | cadde | avenue#2 |
| | italya komün | commune#1 |
| | uydu | satellite#1 |
| | türkiye | Turkey#2 |
| | ilçe | district#1 |
| | bulgaristan il | city#1 |
| | çin eyalet | state#1 |
| | ukrayna idari birimi | administrative unit#1 |
| | kanton | canton#2 |
| object | otomobil | automobile#1 |
| | ateşli silah | firearm#1 |
| | uçak | airplane#1 |
| | gemi | ship#1 |
| | arma | rigging#1 |
| | cep telefonu | cell phone#1 |
| | amfibi | amphibian#3 |

| | | |
|--------------|--------------------------------|--------------------------------|
| | para | money#1 |
| | uzay aracı | spacecraft#1 |
| organization | futbol kulübü | football team#1 |
| | şirket | company#1 |
| | tv kanalı | channel#7 |
| | üniversite | university#1 |
| | kuruluş | establishment#1 |
| | siyasi parti | political party#1 |
| | gazete | newspaper#1 |
| | dergi | magazine#1 |
| | millî futbol takımı | |
| | website | website#1 |
| | basketbol kulübü | basketball team#1 |
| | futbol ligi | football league#1 |
| | radio istasyonu | radio station#1 |
| | milli futbol cemiyeti | association#1 |
| | rusya federasyonu idari birimi | administrative unit#1 |
| | silahlı kuvvet | armed forces#1 |
| | havayolları | airline#2 |
| | hükümet kurumu | institution#1 |
| | basketbol ligi | basketball league#1 |
| | voleybol kulübü | team#1 |
| | azerbaycan rayon | |
| | siyasi makam | agency#1 |
| | nba takımı | team#1 |
| | plak şirketi | record company#1 |
| | bm | |
| | telekomünikasyon şirketi | company#1 |
| | basketbol | basketball#1 |
| | sivil toplum kuruluşu | nongovernmental organization#1 |
| person | futbolcu | football player#1 |
| | müzik sanatçısı | musician#2 |
| | kişi | person#1 |
| | oyuncu | actor#1 |
| | makam sahibi | authority#2 |
| | yazar | writer#1 |
| | bilim adamı | scientist#1 |
| | basketbolcu | basketball player#1 |
| | asker | soldier#1 |
| | hükümdar | monarch#1 |
| | sanatçı | artist#1 |
| | porno yıldızı | |

| | | |
|--------------------|----------------------------|-----------------------|
| | sporcu | athlete#1 |
| | filozof | philosopher#1 |
| | kurgusal karakter | |
| | tenis sporcu | |
| | çizgi roman karakteri | fictional character#1 |
| | güreşçi | wrestler#1 |
| | manken | model#8 |
| | sunucu | presenter#1 |
| | buz patencisi | skater#1 |
| | voleybolcu | volleyball player#1 |
| | karakter | fictional character#1 |
| | hakem | referee#1 |
| | yıldız savaşları karakteri | fictional character#1 |
| | mimar | architect#1 |
| | bisikletçi | cyclist#1 |
| | doctor who karakteri | fictional character#1 |
| | müzik grubu | band#2 |
| | harry potter karakteri | fictional character#1 |
| | avrupa yakası karakteri | fictional character#1 |
| | hanedan | dynasty#1 |
| | | |
| | insan | human#1 |
| physical object | galaksi | galaxy#3 |
| | takımyıldız | constellation#2 |
| | güneş dışı gezegen | planet#1 |
| structure | stadyum | stadium#1 |
| | havalimanı | airport#1 |
| | dini yapı | place of worship#1 |
| | askeri birim | military unit#1 |
| | baraj | dam#1 |
| | yapı | structure#1 |
| | okul | school#1 |
| | müze | museum#1 |
| | yüksek yapı | structure#1 |
| | yol | road#1 |
| | fakülte | faculty#2 |
| | köprü | bridge#1 |
| | askeri yapı | structure#1 |