1. urratsa: MedMentions corpusean MED diren entitateak detektatu. HPn deskribatuta dagoen bezala. Hobe BERT bat doituta Medmentions-eko train-arekin (laborategia egin duzuela uste dut)

<https://github.com/chanzuckerberg/MedMentions>

2. urratsa: Finago egin detekzioa eta TypeMentionsStats.xlsx taulako klase pare bat soilik detektatu: aurkikuntzak eta gaixotasunak (edo antzekoenak taulan) T184? (sign or sympton) 31 (Disease or Syndrome?)

<https://github.com/chanzuckerberg/MedMentions/blob/master/full/stats/TypeMentionStats.xlsx>

3. urratsa. Behin gaixotasun bat detektatu duzuenean, saiatu lotura egiten Wikidatarekin. Nola? Edizio distantzia batekin? Levenhstein? (egin zitekeen CUI kodeaz baina oso zaila da horiek detektatzea 34.000 direla uste baitut corpus honetan)

Horrela automatikoki testuan gaixotasuna detektatu eta Wikidatara lotuko zenuke.

Ebaluazioa: klase semantiko gutxi horietako entitateena test atalean + Wikidatan zein estaldura duten horiek.