Gesprek Marijn Eijkelhof en Thomas

Gesprek: 15-03

# Zonder Thomas

Status quo:

* Hoe labelen jullie nu de data?
* Waarvoor gebruiken jullie die labels?
* Gebruiken jullie ook de labels die bij de centrale overheid worden gebruikt?

Wensen:

* Wat is voor jullie fijn als geautomatiseerd wordt?
* Wat zijn de vereisten van zo’n systeem?
  + Snelheid/computerkracht
  + Foutmarges, waarop optimaliseren (recall precision)

Trainingsdata:

* Hebben jullie documenten met meta-data beschikbaar
  + Zo nee, is er een mogelijkheid om te helpen met de labeling van data in gemeentes?
* Is er tussen gemeente overeenstemming in de structuur van dit soort documenten?

# Met Thomas

Ik weet niet zo veel over je onderzoek, dus ik ben benieuwd:

1. Wat ben je aan het onderzoeken?
2. Welke output moet je systeem krijgen?
3. Wat gebruik je als input van je systeem?
4. Welke voorbereidingsstappen gebruik je?
5. Welke algoritmes/modellen gebruik je?
6. Heb je al resultaten?

Zijn er voor jou papers op dit gebied die ik echt gelezen moet hebben?

Veel van de CNN’s zijn gemaakt voor zinnen, hoe zou jij dat naar grotere documenten overbrengen.

Probeer je ook multi-topic te voorspellen, gebruik je threshold? Hoe leer je die? Of splits je documenten op?

Heb je advies voor mijn onderzoek?