



TRONDHEIM KOMMUNE

Leendert Wienhofen, Data Scientist

# Selvlærende, selvoppdaterende chatbot, basert på åpen kildekode



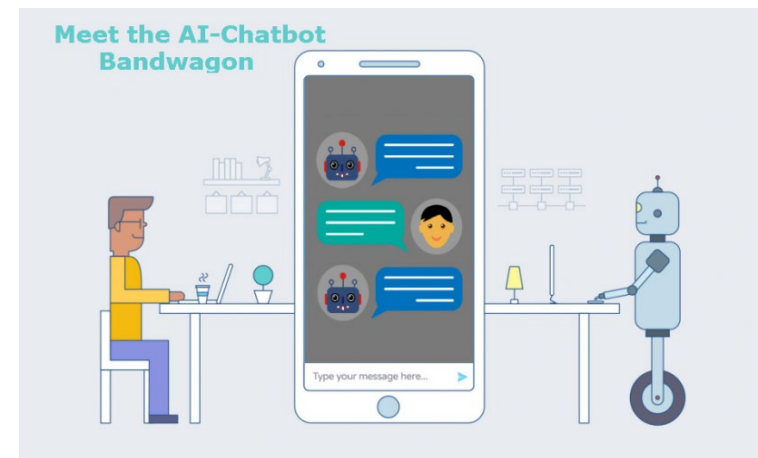
# Agenda

- Chatbot/svarbot
  - Leverandører
  - Utprøving av Dialogflow
- Samarbeid med NTNU
  - PoC: Selvlærende, selvoppdaterende chatbot
- Veien videre
- Muligheter for din virksomhet
  - Kildekode og installasjonsmanual



# Chatbot/svarbot

- Hovedoppgaven til en chatbot/svarbot å tolke teksten som kommer inn og knytte dette til et relevant og oppdatert svar som så blir presentert for brukeren.
- Nøkkelordbasert søk kan være en utfordring
  - Chatbots byr på flere muligheter ettersom vanlig språk kan bli anvendt.



# Leverandører

Flere chatbot-  
leverandører har  
presentert  
sine løsninger

## Fellestrekk

- Fine brukergrensesnitt
- Tidkrevende prosess med (mye) menneskelig innsats
- Gir ofte mangelfulle/feil/utdaterte/ingen svar
  - Kan evt peke videre til en URL



# Intern testing Dialogflow

- Det er forholdsvis lett å lage en enkel chatbot
  - Integrasjoner er predefinert og enkel å sette opp.
  - Svarene blir lagret i Dialogflow og blir fort utdatert.
    - En mulighet for å unngå det er bruk av “webhooks”. Vi har ikke undersøkt dette nærmere.
- Det å sette opp en chatbot med både eksempelspørsmål og reelle svar tar en del tid.
  - Mye klipping og liming, samt strukturering
    - Brukte menneskeressurser veier ikke opp for nytteverdi



# Må det virkelig være slik???

Jeg har bedre ting å gjøre enn å lære opp og vedlikeholde en chatbot!!!



# Samarbeid med NTNU

- Innsikt fra leverandørmøter og egen eksperimentering ledet til definisjon av en studentoppgave ved NTNU.
- Mål: lag en selvlærende chatbot som har svar på alt som finnes på [www.trondheim.kommune.no](http://www.trondheim.kommune.no), og som oppdaterer seg selv
- En gruppe bestående av 6 bachelorstudenter valgte oppgaven og har brukt vårsemesteret 2019 på å sette seg inn i problemstillingen samt ferdigstille en prototype.



# Hovedmål til en kommunal chatbot

Svare korrekt på spørsmål som angår kommunen og kommunens tjenester



Minst mulig bruk av mennesker som ressurs

Riktig bruk programvaredesignprinsipper fra starten  
(gjen)brukbarhet, modularitet, åpent





# Proof of Concept:

## Selvlærende, selvoppdaterende chatbot

- For studentoppgaven ble det definert at chatboten skulle være en slags foredlet søkemotor, og ikke en “menneskelignende samtalepartner”. En svarbot.
- Basere seg på hovedmålene
- Kunnskapsbase [www.trondheim.kommune.no](http://www.trondheim.kommune.no)



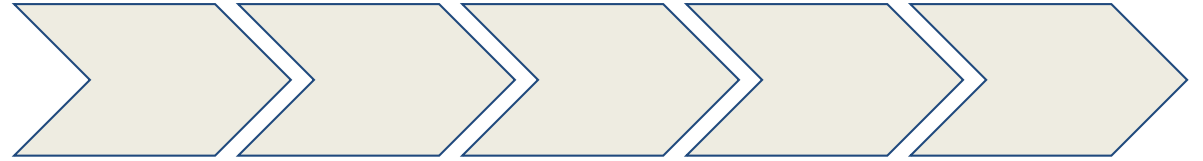
# Språkprosessering

- Naturlig språk er veldig forståelig for mennesker, men er vanskelig å tolke for maskiner.
  - Samme ord/tekst kan ha forskjellige tolkninger i forskjellige kontekst.
- Det er utarbeidet en språkprosesseringsmodul som tolker spørsmålet som brukeren stiller



# Litt mer om NLP

- NLP
  - Tokenisation
  - Part-of-speech tagging
  - Lemmatisering
  - Fjerning av betydningsløse ord
  - Synonymdatabase
- Nøkkelord får forskjellig verdi avhengig av om det er artikkel, adjektiv, verb, osv.



# Søk og gjenfinning: teknologi

- Naturlig språk inn → et eller flere svar tilbake
- scikit-learn
  - TF-IDF for å vurdere hvor viktig et ord er i en samling med dokumente
- Spacy
  - Tokeniser på norsk
- MongoDB search algorithm
- WordNet for synonymer
- Åpen ordbok for stavekorreksjon

---

**Human:** Når åpner husebybadet på fredag?  
**Chatbot:** Jeg har flere mulige svar til deg.

—

Svømmehall - Husebybadet - Ordinære åpningstider:  
Mandag, onsdag, fredag: 15.00–20.00  
Tirsdag, torsdag: 15.00–17.30 og 19.30–22.30  
Lørdag, søndag: 10.00–16.00  
Hentet fra: [ link ]

—

Svømmehall - Husebybadet:  
Husebybadet er stengt onsdag 1.mai  
Hentet fra: [ link ]

---



# MVC-arkitektur

- 3 moduler
  - informasjonssamleren
  - datamodellen
  - språkprosesseringsmodulen
- Alle moduler er generiske
  - Ikke spesifikt laget for Trondheim Kommune sitt nettsted og derfor gjenbrukbar

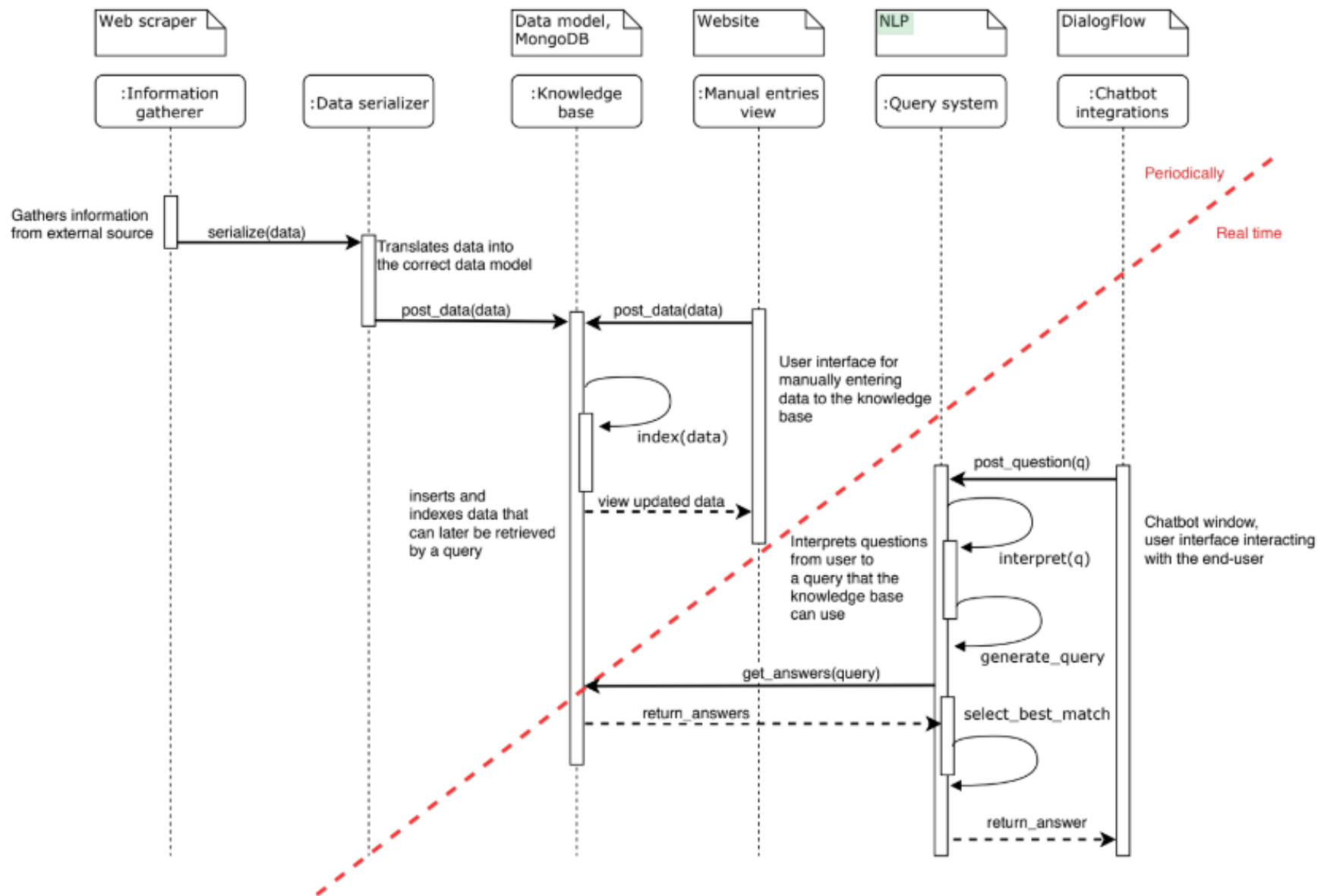


# 6 hovedkomponenter

1. En web scraper som har som oppgave å trekke ut tilgjengelig informasjon fra en informasjonskilde
2. En kunnskapsbase som lagrer og indekserer informasjon slik at det blir lett tilgjengelig for spørring
3. En språkprosesseringsmodul som interpreterer spørsmål fra sluttbrukere og som gjennomfører søk i kunnskapsbasen
4. En adminwebseite som gir muligheter for manuelle endringer til kunnskapsbasen
5. En chatbot integrasjon, slik at sluttbrukere kan kommunisere med chatboten.
6. Et API som tar seg av dataflyten mellom de forskjellige komponenter



# Sekvensdiagram



# Produktets styrker



- Relativt høy svarpresisjon
  - 79% på de 20 mest stilte spørsmål
- Mulighet til å manuelt oppdatere vektene i kunnskapsbasen
- Produktet er språk-agnostisk
- Gjenbrukbar.
  - Minimalt med konfigurering for en ny kunnskapsbase
- Enkel å forbedre over tid grunnet MVC arkitektur
- Åpen





# Produktets svakheter

- Alle avsnitt er ansett som svar
  - Systemet antar at måten kildeinformasjonen er presentert på gir mening som svar
- Samtalekontekst blir ikke håndtert
  - Det vil si at hvert spørsmål blir sett på individuelt og uten sammenheng med tidligere spørsmål eller svar.
- Token-basert språkprosesseringmodell
  - Delvis avhengig av at brukeren bruker noenlunde samme ordlyd som svaralternativet
    - Bruk av wordnet/synonymer gjør dette litt mindre svak



# Veien videre

- Vi har deltidsansatt en av gruppens studenter til å gjøre bot'en produksjonsferdig.
- Testing med Trondheim eiendom pågår
- Sluttbrukergrensesnitt tilgjengelig på Bytorget
  
- Andre enheter vil bli testet snart
- Bot'en bli tilgjengeliggjort for befolkningen



# Muligheter for din virksomhet

- Systemet kan bli installert på få timer og med minimal konfigurering
  - Docker containers
- Kildekode med GPL3.0 lisens og installasjonsmanual
  - <https://github.com/Trondheim-kommune/chatbot>
- Kunnskapsdeling
  - <https://sites.google.com/trondheim.kommune.no/kunnskapsdeling/chatbot>



Leendert Wienhofen

Data Scientist

IT-tjenesten, Trondheim kommune

[leendert.wienhofen@trondheim.kommune.no](mailto:leendert.wienhofen@trondheim.kommune.no)

Tlf: +47 982 45 177

