



PIO – ABB – Slimme Regelgeving
Workshop 4: Overzicht State-of-the-art
30/04/2021

Agenda

1. Context & Objectieven

2. State-of-the-art

3. Referentiearchitectuur

4. Conclusie

Context en objectieven

Context

Het Agentschap Binnenlands Bestuur (ABB) is verantwoordelijk voor de lokale regelgeving en besluiten. De lokale regelgeving en besluiten bevatten een schat aan informatie, echter deze is momenteel heel moeilijk te exploiteren.

Objectieven project

Inzichten te verkrijgen uit ongestructureerde lokale besluiten en regelgeving alsook het structureren van de ongestructureerde data op basis van nieuwe technologie zoals AI & ML. Hierbij worden volgende topics in rekening gebracht:

- Welke data is voorhanden?
- Wat zijn de wensen van de gebruikers?
- Wat zijn de technische mogelijkheden?

Objectieven workshop

Een overzicht geven van de state-of-the-art technologieën die beschikbaar zijn om de wensen van de gebruikers te realiseren.

Opstellen van een referentiearchitectuur waarop de wensen van de gebruikers en beschikbare technologieën gemapt kunnen worden.

Agenda

1. Context & Objectieven

2. State-of-the-art

3. Referentiearchitectuur

4. Conclusie

Wat is Artificiële intelligentie (AI) ?

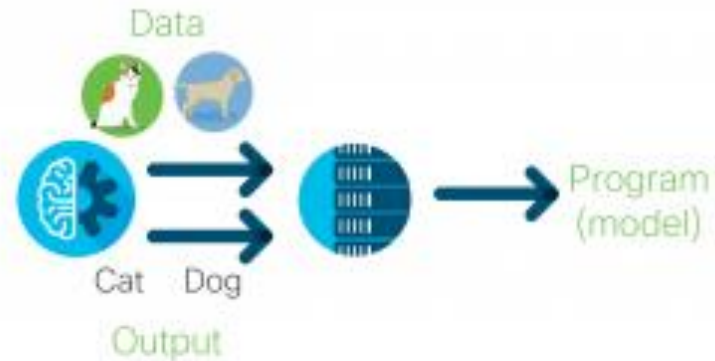
AI/ML Writes Code Based on Examples

Traditional Programming



Programmer describes in code:
"Cats have tail, pointy ears, tail and meow"
"Dogs have tail, pointy snout, and woof"

Machine Learning



Machine Learning generates code:
Based on lots of examples with correct answers

Overzicht state-of-the-art

Automatische verwerking en analyse van “natuurlijke” tekst

Natural language processing (NLP)

- Informatie extractie
- Verbanden leggen tussen teksten
- Natuurlijke taal begrijpen en vragen beantwoorden

Automatische verwerking en analyse van numerieke waarden en media

Machine learning voor patroonherkenning

Audio- en video recognitie

Overzicht state-of-the-art

Automatische verwerking en analyse van “natuurlijke” tekst

Natural language processing (NLP)

- Informatie extractie
- Verbanden leggen tussen teksten
- Natuurlijke taal begrijpen en vragen beantwoorden

Automatische verwerking en analyse van numerieke waarden en media

Machine learning voor patroonherkenning

Audio- en video recognitie

Natural language processing (NLP)

Technologie

Natural language processing verwerkt en analyseert “natuurlijke” tekst (i.e. volle, ongestructureerde tekst)

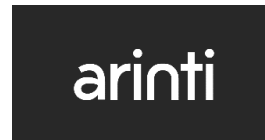
- Laatste jaren sterke evolutie in onderzoeksfase
- Initieel focus op Engels, recent ook initiatieven voor andere talen (multilingual NLP), waaronder Nederlands

Take-aways

- Bruikbaar voor eenduidige taken zoals entiteiten herkennen
- Zware onderzoeksfase voor complexe analyse taken
- Voorgetrainde tools/modellen bruikbaar voor generische cases of als startpunt voor training van specifieke cases
- Te combineren met andere ML-technieken om geëxtraheerde informatie verder te gebruiken

Marktspelers

- Machine learning spelers met ervaring in NLP:



- Internationale softwarespelers (Google, Amazon, Facebook, ...) bieden open-source tools, algoritmes en libraries aan

Overzicht state-of-the-art

Automatische verwerking en analyse van “natuurlijke” tekst

Natural language processing (NLP)

- **Informatie extractie**
- Verbanden leggen tussen teksten
- Natuurlijke taal begrijpen en vragen beantwoorden

Automatische verwerking en analyse van numerieke waarden en media

Machine learning voor patroonherkenning

Audio- en video recognitie

NLP informatie extractie:

Efficiënt voorafbepaalde informatie extraheren uit natuurlijke tekst

1. Named Entity Recognition (NER)

Voorgedefinieerde entiteiten aanduiden/extraheren uit een tekst

- E.g. [Jim]Person bought 300 shares of [Acme Corp.]Organization in [2006]Time.

2. Named Entity Linking (NEL)

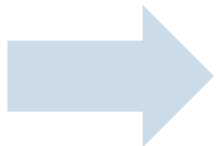
Entiteiten linken met hun corresponderende entiteiten in een bepaalde knowledge base

- E.g. de mandaten databank of Wikipedia
- ~ Vind de juiste “Jan Janssens”

3. Relation Extraction (RE)

Automatisch verbanden leggen tussen entiteiten in één tekst

- E.g. Jim is de vader van Jan Janssens
- Op basis van de informatie in de tekst of externe bronnen



Technisch haalbaar voor veelvoorkomende of standaard entiteiten
Extra research/training nodig voor zelfgekozen entiteiten

NLP informatie extractie:

Efficiënt niet voorafbepaalde informatie extraheren uit natuurlijke tekst

1. Kernwoord extractie

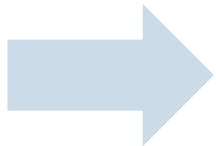
Automatische extractie van kernwoorden die een tekst het best omschrijven

- Werkt typisch met een kandidaat lijst, ofwel vooraf bepaald (semi-supervised) ofwel automatisch gegenereerd door het model (unsupervised)
- Meerdere kernwoorden en eventueel een korte kernzin mogelijk

2. Samenvatten

Automatisch samenvatten van een tekst in een compacte vorm

- Executive: belangrijkste zinnen per alinea hergebruiken
- Abstract: nieuwe zinnen en inzichten verwerken in samenvatting
- Complexiteit hangt sterk af van type samenvatting
- Moeilijk om te generaliseren (e.g. product book vs legal document)



Technisch haalbaar voor specifieke toepassingen en indien er geen hoge eisen zijn op de informatie output, e.g. kernwoord extractie of executive summary.

Zwaar research proces nodig voor complexe, generische toepassingen, e.g. abstracts

Voorbeeld cases NLP informatie extractie

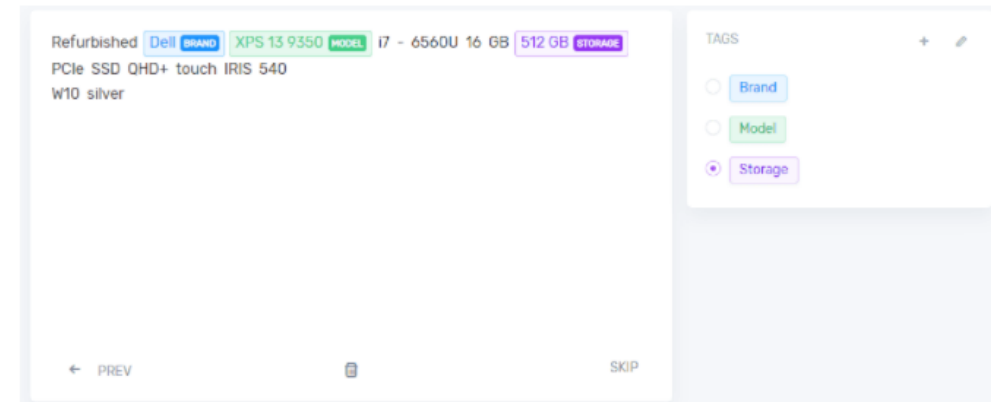
Automatisch detecteren van entiteiten in notariële oprichtingsakte

- NLP Town heeft een model ontwikkeld voor de Federatie van Notarissen (FEDNOT) dat automatisch de meest relevante informatie haalt uit een notariële oprichtingsakte
 - E.g. adres, bedrijf en namen van de oprichters
- [NLP Town](#)



Online Named Entity Recognition tool

- MonkeyLearn is een online platform dat voor verschillende NLP taken gebruikt kan worden, waaronder NER.
 - Voorgetrainde off-the-shelf modellen beschikbaar, e.g. naam extractor voor Engelse teksten
 - Mogelijkheid om zelf modellen te trainen (no code) voor specifieke data, entiteiten en talen, e.g. Nederlands.
 - Labeling: manueel entiteiten taggen op trainings-data
- [MonkeyLearn - Text Analysis](#)



Overzicht state-of-the-art

Automatische verwerking en analyse van “natuurlijke” tekst

Natural language processing (NLP)

- Informatie extractie
- **Verbanden leggen tussen teksten**
- Natuurlijke taal begrijpen en vragen beantwoorden

Automatische verwerking en analyse van numerieke waarden en media

Machine learning voor patroonherkenning

Audio- en video recognitie

Automatisch verbanden leggen tussen verschillende teksten via NLP

1. Topic Classificatie

Inhoudelijk classificeren van teksten in voorgedefinieerde klassen

- Voorgetrainde modellen beschikbaar, e.g. spam versus geen spam
- Voor zelfgekozen, specifieke klassen is training nodig op gelabelde data

2. Clustering

Groeperen/clusteren van inhoudelijk gerelateerde teksten

- Voorgetrainde modellen beschikbaar
- Voor betere en meer specifieke resultaten hertraining nodig

3. Topic modelling

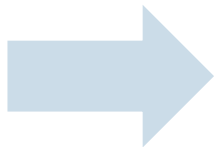
Toekennen van niet-voorgedefinieerde topics aan teksten

- Afhankelijk van complexiteit verschillende technologieën beschikbaar
- Meerdere topics met een confidence level mogelijk

4. Gelijkenischeck

De inhoudelijke gelijkenis tussen teksten uitdrukken d.m.v. een numerieke metriek

- Sterk gerelateerd aan clustering

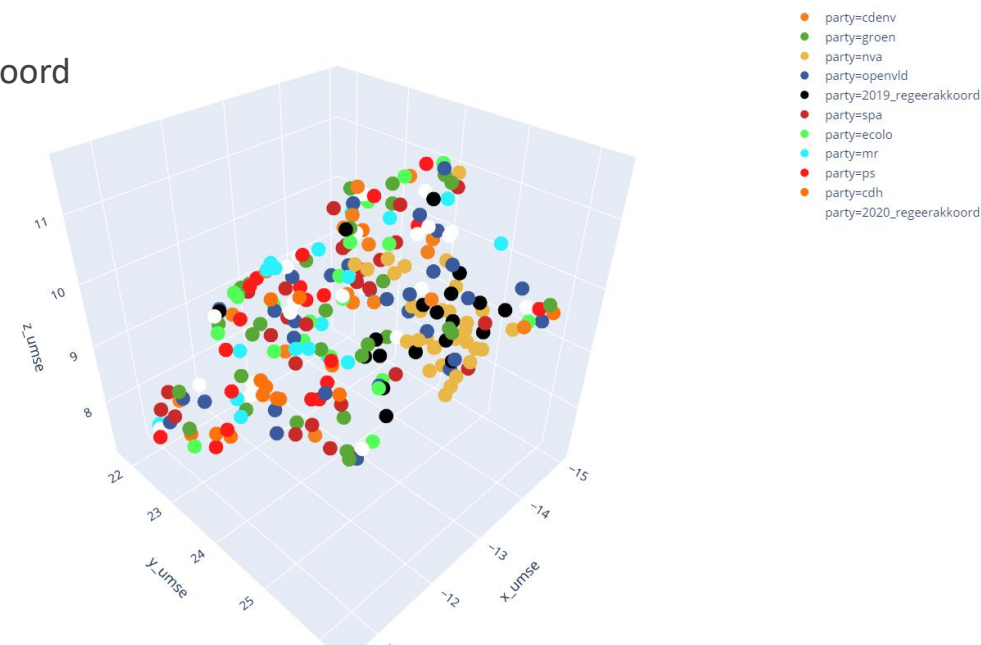


Technisch haalbaar indien:

- Voorgetraind model voldoende is
 - Dataset voldoende ruim en kwalitatief is voor een specifiek domein
- Zwaar research proces nodig, voor complexe, generische toepassingen

Voorbeeld case verbanden leggen tussen teksten via NLP: Correlatie tussen standpunt regeringspartijen en regeerakkoorden

- Gebruik van NLP om het standpunt van Belgische partijen over voorgedefinieerde onderwerpen (e.g. energie) te vergelijken met de visie voor deze onderwerpen in de Vlaamse/Federale regeerakkoorden
 - Doel: analyseren of er een correlatie is tussen de visie in het regeerakkoord en het al dan niet aanwezig zijn van een partij in de Federale of Vlaamse regering.
- Methodiek:
 - **Scrapen** partij websites
 - **Vertalen** Nederlands naar Engels
 - **Samenvatten** van de verschillende teksten per partij/regeerakkoord en onderwerp
 - **Gelijkenischeck en clustering** op de verschillende samengevatte teksten
 - **Analyseren** van de resultaten
- Resultaten:
 - Voor bepaalde onderwerpen: correlatie tussen regeringspartijen en regeerakkoord
 - Voor een beperkter aantal onderwerpen: correlatie minder duidelijk
 - O.a. limitatie door informatieverlies bij samenvatten en vertalen
 - Zelfde setup ook toegepast rechtstreeks op Nederlands, echter met minder kwalitatieve resultaten
- **Onderzoeksvoorbeeld, maar geeft een beeld van de mogelijkheden die NLP biedt**



Overzicht state-of-the-art

Automatische verwerking en analyse van “natuurlijke” tekst

Natural language processing (NLP)

- Informatie extractie
- Verbanden leggen tussen teksten
- **Natuurlijke taal begrijpen en vragen beantwoorden**

Automatische verwerking en analyse van numerieke waarden en media

Machine learning voor patroonherkenning

Audio- en video recognitie

Taal begrijpen: Natural language understanding (NLU) & question answering (QA)

Rule-based

- Vooral gebaseerd op “If-then” logica
- Vrij matuur, maar mogelijks antwoorden die naast de kwestie zijn
→ Typisch (enkel) antwoorden op de “FAQs”

AI-based

- AI/ML/NLP voor: “language understanding”, respons-logica en tekst creatie
- Data-driven, typisch black-box en zwaar researchproces voor specifieke cases

Hybrid

- Start met rule-base voor eenvoudige cases
- Gebruik AI/ML voor specifieke, complexe cases, alsook voor on-the-fly learning

- Variëteit aan internationale platformen voor Q&A (e.g. Chatbot) development
 - Mogelijkheden om te trainen op eigen, specifieke data speelt belangrijke rol om eventuele keuze te maken
- Marktspelers met ervaring in NLU/Q&A/Chatbots:



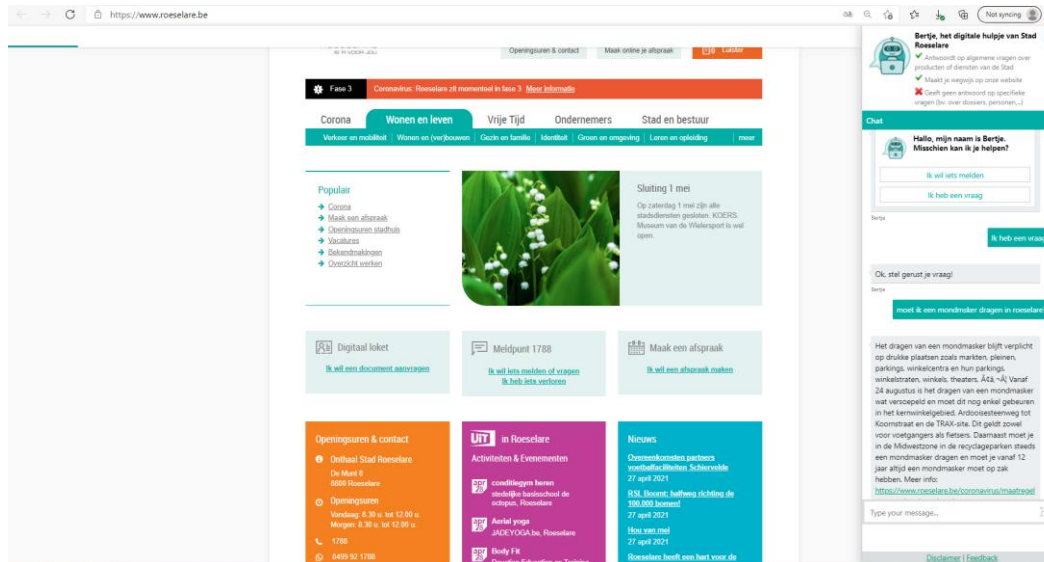
Voorbeeld case NLU en QA: Bertje, the virtual Clerck of the city Roeselare

- Digitale dienstverlening op maat voor zowel de burger als stadsmedewerker, d.m.v. een behulpzame Chatbot:

 **Bertje, het digitale hulpje van Stad Roeselare**

- ✓ Antwoordt op algemene vragen over producten of diensten van de Stad
- ✓ Maakt je wegwijs op onze website
- ✗ Geeft geen antwoord op specifieke vragen (bv. over dossiers, personen,...)

- Trainingsproces: per vraag dienden test users aan te geven hoe goed “Bertje” geantwoord had op hun vraag
- [Bertje, the virtual clerk of the city of Roeselare — Arinti](#)



The screenshot shows the website <https://www.roeselare.be> with a chatbot interface on the right. The chatbot, Bertje, is responding to a question: "moet ik een mondmasker dragen in roeselare?". The response text is: "Het dragen van een mondmasker blijft verplicht op drukke plaatsen zoals markten, pleinen, parkings, winkelcentra en hun parkings, winkelstraten, winkels, theaters, ... Vanaf 24 augustus is het dragen van een mondmasker wat versoepeld en moet dit nog enkel gebeuren in het kernwinkelgebied, Ardooisesteenvweg tot Koorstraat en de TRAX-site. Dit geldt zowel voor voetgangers als fietsers. Daarnaast moet je in de Midwestzone in de recyclageparken steeds een mondmasker dragen en moet je vanaf 12 jaar altijd een mondmasker moet op zak hebben. Meer info: <https://www.roeselare.be/coronavirus/maatregel-en-roeselare/mondmaskers>".



Chat

Bertje

moet ik een mondmsker dragen in roeselare?

ik

Het dragen van een mondmasker blijft verplicht op drukke plaatsen zoals markten, pleinen, parkings, winkelcentra en hun parkings, winkelstraten, winkels, theaters, ... Vanaf 24 augustus is het dragen van een mondmasker wat versoepeld en moet dit nog enkel gebeuren in het kernwinkelgebied, Ardooisesteenvweg tot Koorstraat en de TRAX-site. Dit geldt zowel voor voetgangers als fietsers. Daarnaast moet je in de Midwestzone in de recyclageparken steeds een mondmasker dragen en moet je vanaf 12 jaar altijd een mondmasker moet op zak hebben. Meer info:

<https://www.roeselare.be/coronavirus/maatregel-en-roeselare/mondmaskers>

Moet ik morgen mijn afval buiten plaatsen?

ik

De ophaalkalender kun je online raadplegen via volgende link: <http://www.mirom.be/afval-ophaling/uw-afvalkalender-raadplegen>

Bertje at 5:55:34 PM

Type your message...



Overzicht state-of-the-art

Automatische verwerking en analyse van “natuurlijke” tekst

Natural language processing (NLP)

- Informatie extractie
- Verbanden leggen tussen teksten
- Natuurlijke taal begrijpen en vragen beantwoorden

Automatische verwerking en analyse van numerieke waarden en media

Machine learning voor patroonherkenning

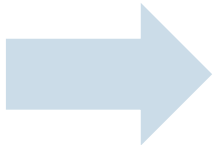
Audio- en video recognitie

Machine learning voor patroonherkenning

1. ML voor patroonherkenning

Patronen detecteren of informatie extraheren uit tabulaire (geen tekst) datasets

- E.g. outlier detectie, clustering, patroonherkenning (en predictie) in time series
- Getrainde modellen beschikbaar voor algemene problemen
- Voor specifieke, complexe taken hertraining nodig met kwalitatieve data



Voor algemene, gekende problemen bestaan getrainde modellen
Researchproces met kwalitatieve data nodig voor specifieke, complexe zaken

Voorbeeld case ML voor patroonherkenning: Fraude detectie in een omgeving waar de overheid subsidies toekent

- Tool die overheidsfunctionarissen helpt bij het detecteren van fraude:
 - Outlier-detectie op data uitvoeren, gebruik makend van standaard regressie modellen
 - Aangevuld met outlier-detectie in tijd
 - Finaal optimaliseren van de resultaten door graaf constructie en analytics op de geconnecteerde data-entiteiten
- [Fraud detection with graph analytics | Data Consultancy — Vectr.Consulting](#)



Audio- en video recognitie

1. Video recognitie/Computer Vision

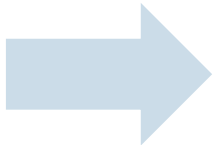
Informatie- en inzicht extractie uit afbeeldingen en video's

- E.g. OCR (optical character recognition) = image-to-tekst, of object detectie/tracking
- Voor simpele taken (bijvoorbeeld OCR) bestaan robuuste oplossingen
- Voor complexe taken en minder kwalitatieve data is researchproces nodig

2. Speech Recognition

Zet spraak om in tekst

- Alomtegenwoordig voor simpele/korte zinnen en commando's
- Zeer uitdagend voor complexe spraak met veel ruis en een lage fouttolerantie
- Vaak in combinatie met NLU om de error rate te verkleinen

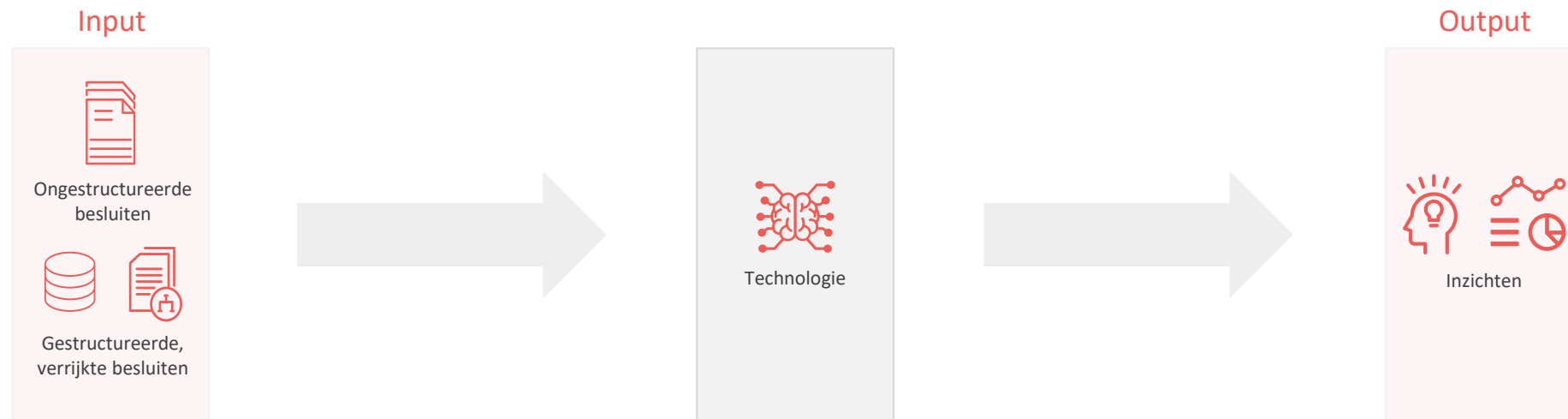
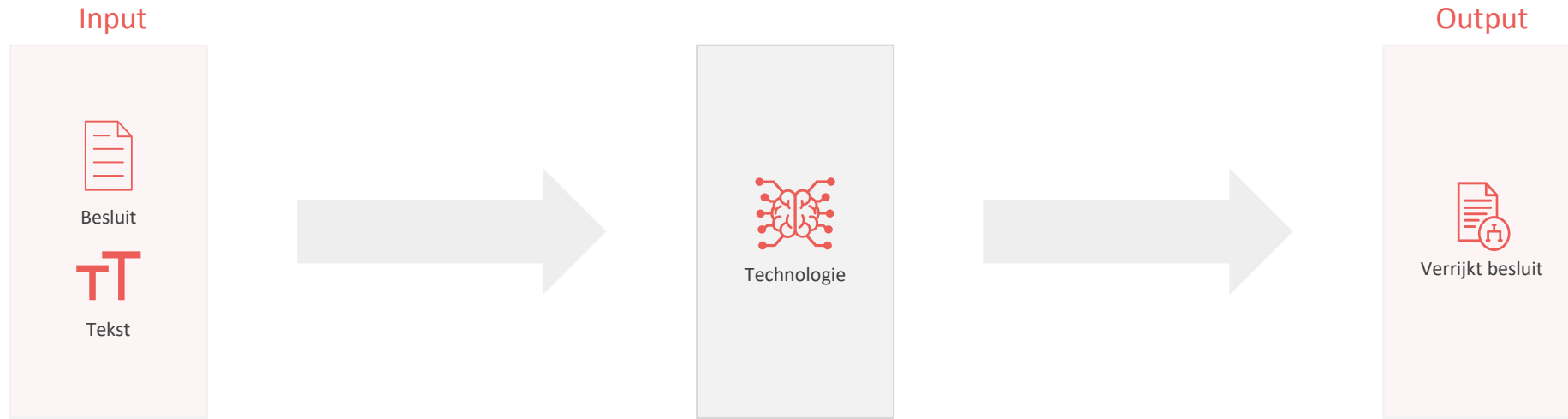


Huidige status van audio- en video recognitie is mogelijks niet voldoende matuur voor de geïdentificeerde databronnen.

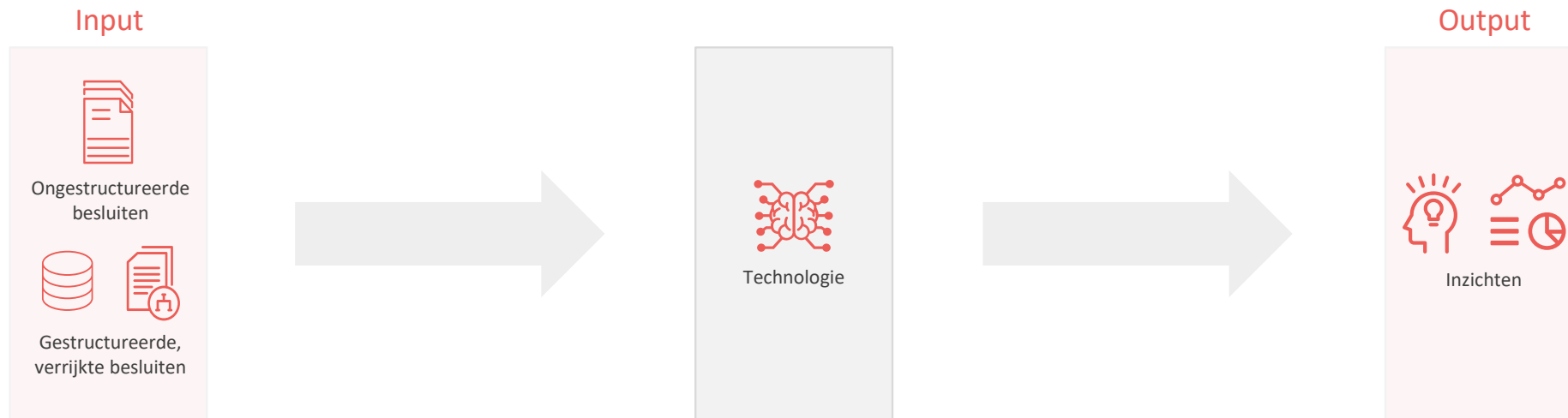
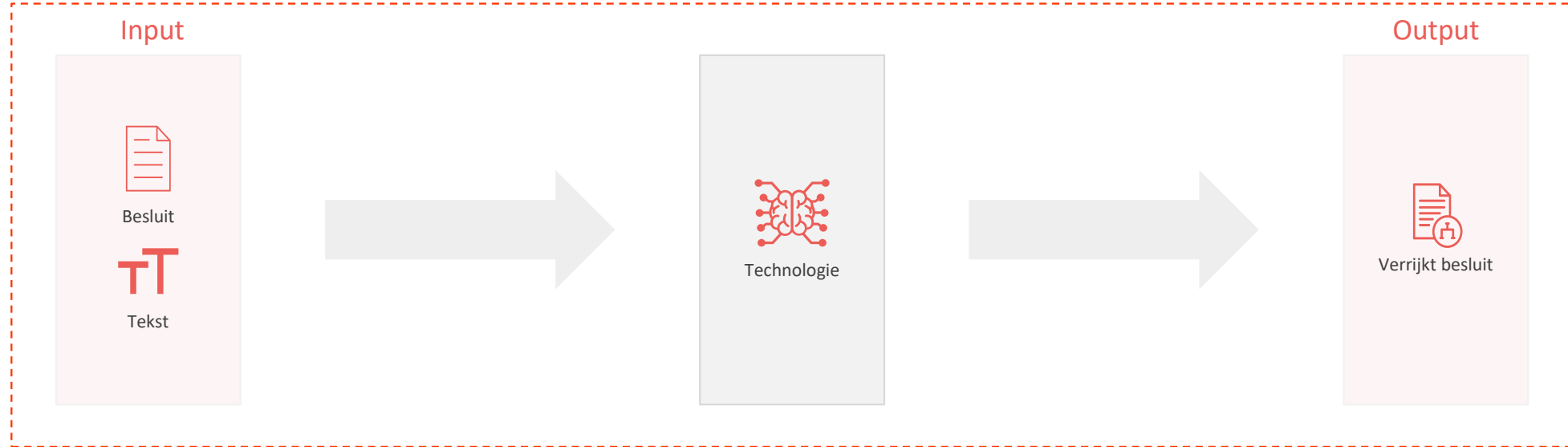
Agenda

1. Context & Objectieven
2. State-of-the-art
3. Referentiearchitectuur
4. Conclusie

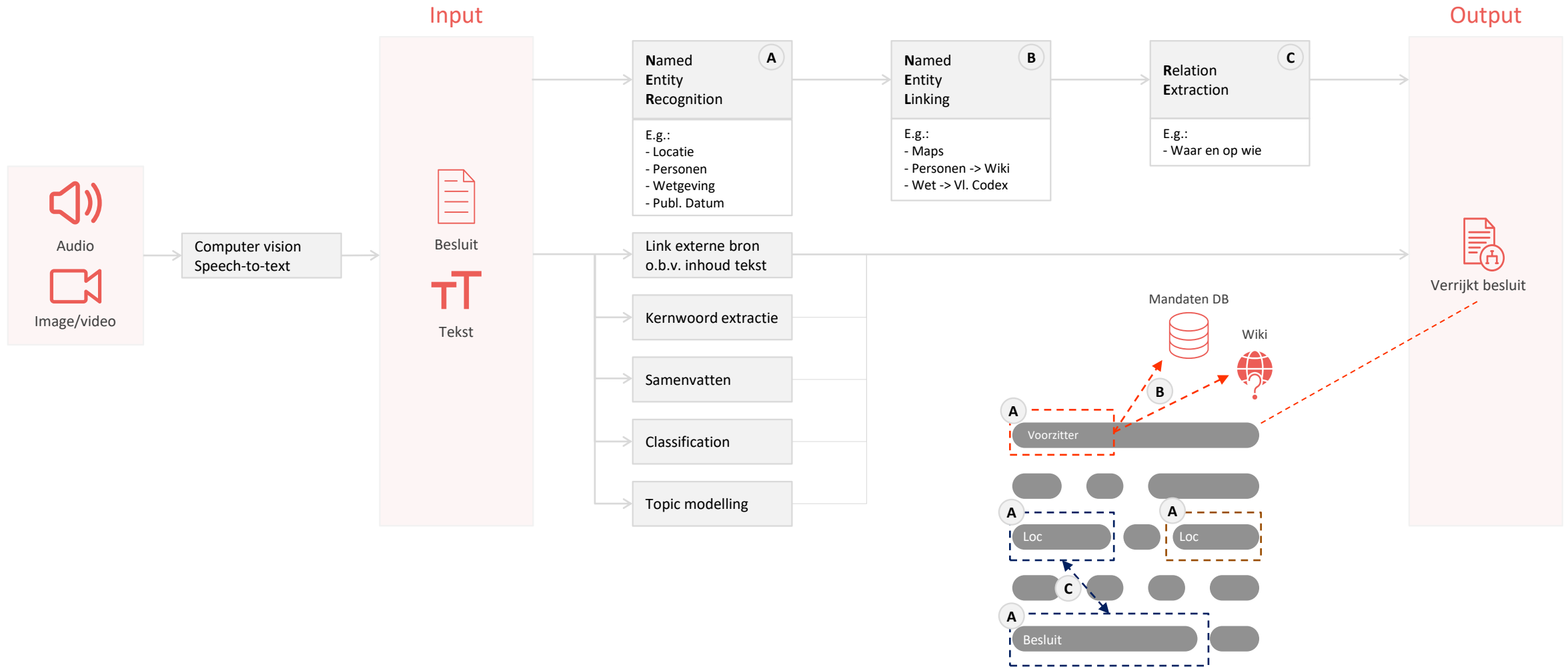
Overzicht referentiearchitectuur



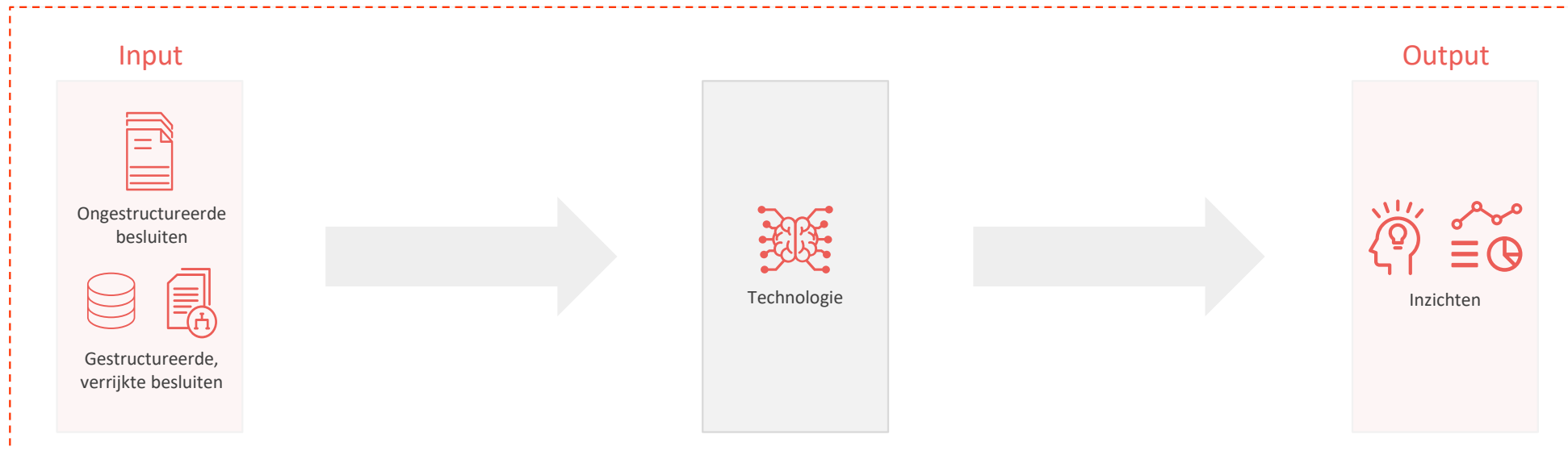
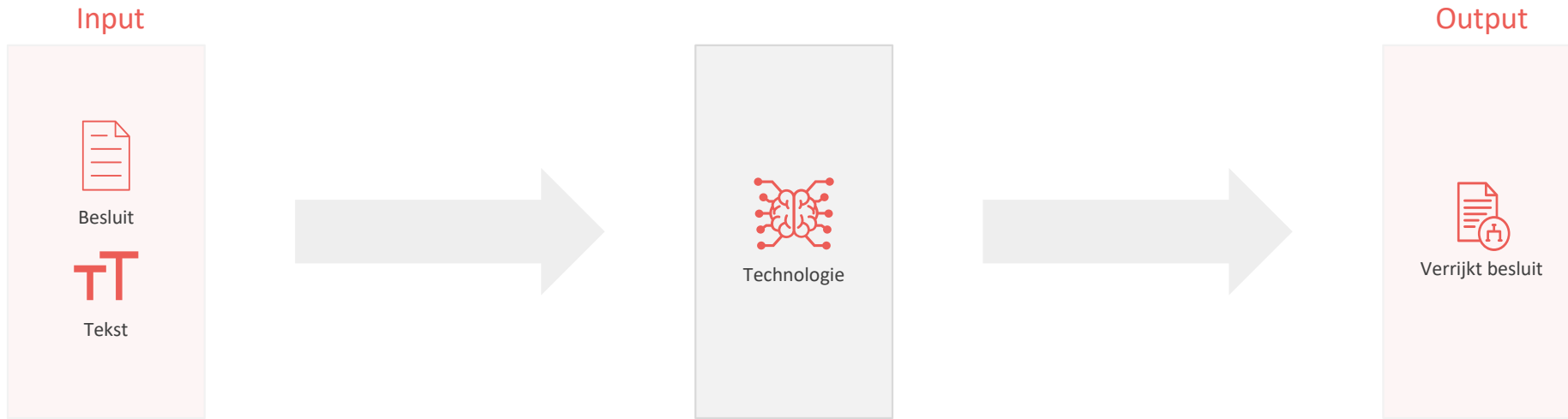
Overzicht referentiearchitectuur



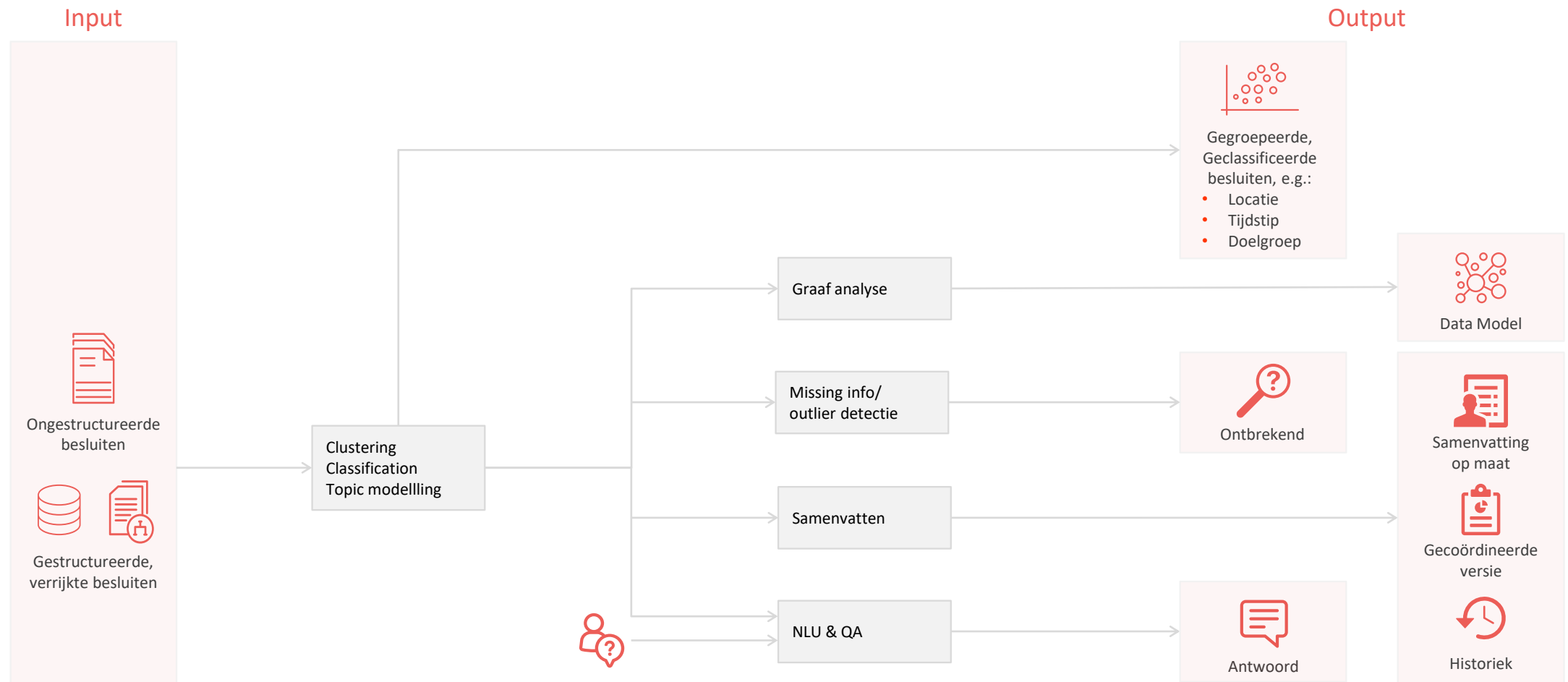
Referentiearchitectuur op besluitsniveau



Overzicht referentiearchitectuur



Referentiearchitectuur op overkoepelend niveau



Agenda

1. Context & Objectieven
2. State-of-the-art
3. Referentiearchitectuur
4. Conclusie

Conclusie

- Om de gedefinieerde use-cases te implementeren is er in hoofdzaak nood aan NLP
 - Voor de marktconsultaties raden we daarom aan om bedrijven uit te nodigen met ervaring in NLP
 - Door een verschil in maturiteitsniveau tussen het Nederlands en Engels, raden we aan om te focussen op Nederlandstalige bedrijven
- De besproken “niet-NLP” technologieën kunnen later ingekoppeld worden op de referentiearchitectuur
 - Bedrijven met ervaring in NLP hebben vaak ook kennis over de vermelde “niet-NLP” technologieën, bijgevolg kunnen we de use-cases ook met hen valideren tijdens marktconsultatie

addestino[®]
innovation delivered.