

Content Generation in German and Dutch from Structured Data



HELLO!

I am Dipl. ling. Andreas Madsack

AX Semantics

NLG as business

- There is a market for commercial Data to Text
- A few companies are already active (Arria, Automated Insights, AX Semantics, ...)
- Some inhouse solutions in bigger companies

NLG Pipeline

In difference to Reiter/Dale:2000 our customers need control over document structure. No restrictions in resulting texts.

The surface realiser and parts of the microplanning are realised in code.

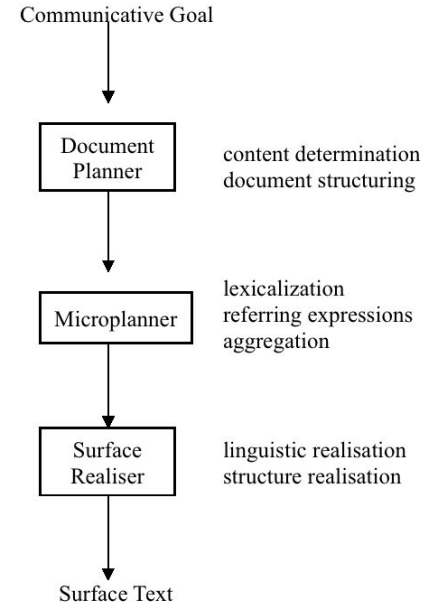


Figure 2: A typical NLG system architecture. From Reiter and Dale p.60.

Data to Text



German and Dutch Generation

German

Die Grundstücksfläche beträgt
[DATA_basearea] m².



Die Grundstücksfläche beträgt 10 m².

Dutch

Het erf omvat [DATA_basearea] m².



Het erf omvat 10 m².

[DATA_basearea] is a kind of variable with the value set by the user data.
Sentence is activated when [DATA_basearea] is set.

Grammatical Agreement

[CountOfPorts;id=count] microUSB [G:noun=port;grammar-from=count]
[G:verb=be;grammar-from=count] available.



CountOfPorts=1 1 microUSB port is available.

CountOfPorts=2 2 microUSB ports are available.

Rule based versus lexicon based

- verb=be is lexicon based
- noun=port is rule based

- rules for adjectives, determiner, cardinals, ordinals, nouns, verbs
(can be) different for every language

- i.e. for Dutch:
 - rules for adjectives - overwrite via lexicon
 - noun pluralisation using 'en' - overwrite via lexicon

German and Dutch Generation

German

[VOC_objecttype_random,det=definite] **kann** [Text:**bis zu** ;On,false=TRIGGER_occupancy_max]
[VOC_occupancy_min][VOC_occupancy_max,id=maxbel;preceding: **bis**
;On,true=TRIGGER_occupancy_max] [G:noun=**Gast**,grammar-from=maxbel] **aufnehmen**.

Die Unterkunft kann 2 bis 4 Gäste aufnehmen.

Dutch

[VOC_objecttype_random,det=definite] **neemt** [Text:**maximaal** ;On,false=TRIGGER_occupancy_max]
[VOC_occupancy_min][VOC_occupancy_max,id=maxbel;preceding: **tot en met**
;On,true=TRIGGER_occupancy_max] [G:noun=**gast**,grammar-from=maxbel] **op**.

De accommodatie neemt 2 tot en met 4 gasten op.

Aiding the User editing the Content

[VOC Maisonette] [GROUP_Objekttyp_Attribute.Best(1)] [GROUP_Benutzte_Attribute.All(),conj=und,prep=mit] - hier beleeft u een mooie [GROUP_Spez_Urlaub_A_V1.Best(1);Alt:UrMODIFIERlaub].

Laat u door dit landhuis [GROUP_Benutzte_Attribute.All(),conj=en,prep=met] en zijn charme op uw [GROUP_Spez_Urlaub_A_V1.Best(1);Alt:vaMODIF

[VOC_Bauernhaus_Adj_NL,adj=yes,det=demonstrative;Alt:Bauernhaus] [GROUP_Objekttyp_Attribute.Best(1)] [GROUP_Benutzte_Attribute.All([GROUP_Spez_Urlaub_A_V1.Best(1);Alt:vaMODIFIERkantie].

Zin in luxe? Laat u betoveren door de flair van deze mooie villa [GROUP_Benutzte_Attribute.All(),conj=en,prep=met] en geniet van een exclusieve [GROUP_Spez_Urlaub_A_V1.Best(1);Alt:vaMODIFIERkantie].

Zin in een ongecompliceerde [GROUP_Benutzte_Attribute.All(),conj=en,prep=met]? Geniet van de onafhankelijkheid van een vakantiewoning [GROUP_Benutzte_Attribute.All(),conj=en,prep=met] een vakantiepark - een ideale combinatie voor iedereen die zijn [GROUP_Spez_Urlaub_A_V1.Best(1);Alt:vaMODIFIERkantie] voorzien van alle gemakken en

In dit vrijstaande vakantiehuis [GROUP_Benutzte_Attribute.All(),conj=en,prep=met] beschikt u over alle gemakken van [GROUP_Spez_Urlaub_A_V1.Best(1);Alt:vaMODIFIERkantie] heeft.

[GROUP_Spez_Urlaub_A_V1.Best(1);Alt:vaMODIFIERkantie] op maat: MODIFIERmet een [GROUP_Spez_Urlaub_A_V1.Best(1);Alt:vaMODIFIERkantie] zoals u wilt!

Tijd voor een [GROUP_Spez_Urlaub_A_V1.Random(1);Alt:vaMODIFIERkantie]! [VOC_Objekttyp_Attribute.Best(1)] [VOC_Ort] is ideaal om aan de alledaagse drukte te ontsnappen.

[GROUP_Zimmer.All(),conj=en] - verheug u op een [VOC_Wohnflaeche.value()] m² grote [VOC_Maximalbelegung.value(),id=maxbel;preceding: tot en met ;On,TRIGGER_Bele

U bewoont hier [VOC_Anzahl_Schlafzimmer.value(),id=anzahlschlafzimmer] [G:noun=schlafzimmer] [VOC_Anzahl_Badezimmer.value(),id=anzahlbadezimmer;On,TRIGGER_Anzahl_Badezimmer=num=anzahlbadezimmer;On,TRIGGER_Anzahl_Badezimmer=true;preceding:]

The Rail Yard Ale (2009) is one of 10 beers brewed by the Wynkoop Brewing Company is one of 10 breweries located in Denver, Colorado. This beer has 5% alcohol and weighs 12oz. The Rail Yard Ale (2009) has a bitterness of 65 units. Colorado has 10 other breweries besides the Wynkoop Brewing Company. The style of this beer is Belgian dark ale. It would be utterly shameful to not have enjoyed this fine beverage on a warm summer evening.

Scale


- 250.000 different texts
- 4 languages (DE, NL, FR, EN)
- Example:
<https://www.vacatello.com/nl/578715/>

Korte beschrijving

Soort accommodatie	Vakantiehuis
Overig	Niet-roken huis
Woonkamer	50 m ²
Minimale bezetting	3 personen
Maximale bezetting	3 personen
Huisdier toegestaan	Huisdieren niet toegestaan
Aantal slaapkamers	1
Aantal badkamers	1
Buitenfaciliteiten	Terras, Tuin

Keuken

Keuken	
Keukenvoorzieningen	Vaatwasser, Oven

Accommodatie beschrijving 

Beoordeling

Kaart en omgeving

Prijzen

Beschikbaarheid

De hoogtepunten van het vakantiehuis

Een mooie vakantie in Amsterdam zal met dit mooie vakantiehuis met WiFi zeker lukken. Aan maximaal 3 personen wordt hier plaats geboden: 1 slaapkamer en 1 badkamer zijn verdeeld over 50 m² woonoppervlak. In het gehele huis is roken niet toegestaan.

- ✓ Een vakantiewoning, die van klanten over alle facetten goede beoordelingen krijgt.
- ✓ Op aangename afstand van het meer: slechts 200 m (hemelsbreed).
- ✓ U bent snel verbonden met de wereld dankzij WiFi.
- ✓ De accommodatie beschikt over een terras.

”

Conclusion

- Our abstraction layer allows any language for the surface realisation part.
- Customers can have 100.000+ datasets to generate a different text each.
- The generated texts can be in another language than the data.



THANKS!

Any questions?

Or try our platform: <https://nlg.ax/>