

# textkernel

Machine Intelligence for People and Jobs

## **RNN sequence labeling for document parsing in Tensorflow**

Carsten van Weelden  
cvanweelden@textkernel.com

# Document parsing

## Curriculum Vitae

ACHTERNAAM: Van de Ven  
VOORNAMEN: Petra Paula Maria  
ADRES: Hoofdstraat 15-A  
POSTCODE / WOONPLAATS: 6717 AA EDE  
GEBORTE DATUM / PLAATS: 29-02-1972 te Wageningen  
GESLACHT: Vrouwelijk  
NATIONALITEIT: Nederlandse  
BURGERLIJKE STAAT: Gehuwd  
CONTACT: +31(0)643054815 / 0318-204594  
VervdPetra@yahoo.nl

## Ervaring

- 01-11-2000 – heden  
Office Manager bij **CTB Flex Software & Services BV** te Ede.  
Functieomschrijving: o.a. beantwoorden van inkomende telefoongesprekken met behulp van telefooncentrale; notuleren bij (interne) vergaderingen; het bestellen kantoorartikelen; bijhouden van agenda's; archiveren; koffie/thee en lunches verzorgen ten behoeve van (interne) besprekingen; inkomende en uitgaande post inschrijven en verwerken; controleren van de inkomende facturen; distributie van de uitgaande offertes.
- 01-08-1997 – 01-11-2000  
Secretaresse bij **Allure Uitzendbureau BV** te Utrecht.  
Functieomschrijving: o.a. telefoon beantwoorden en zelfstandig correspondentie voeren in de vorm van brieven of e-mail; informatie verstrekken aan uitzendkrachten naar aanleiding van zeer diverse en specifieke vragen Wet Flexibiliteit en Zekerheid; het adviseren van de directie bij het in dienst nemen van werknemers met betrekking tot contracten en salarissen; notuleren en verslagen schrijven; archiveren; mede organiseren van promotionele activiteiten voor klanten en uitzendkrachten.
- 14-10-1996 – 01-08-1997  
Boekhoudkundig medewerkster **Autobedrijf P.P. Langebroek** te Wageningen.  
Functieomschrijving: o.a. postboek bijhouden; coderen en invoeren in administratie van inkomende facturen; kas-, bank-, girostukken; registratie klantgegevens; betaalopdrachten doen met telebanking; debiteuren bewaken; advertenties maken voor lokale kranten; vitrine inrichten met autoaccessoires; zelfstandig correspondentie voeren; agenda's bijhouden; archiveren; telefoon beantwoorden middels kleine telefooncentrale; informatie verstrekken aan klanten.

## Opleidingen

- Aug. 1995 – Nov. 1996 **Algemene Ondernemersvaardigheden**, Avondschool De Driesprong in Ede.  
Vakkenpakket: Personeel & Organisatie, Operationeel Management, Financiële Administratie, Marketing, financieel management, ondernemingsplan.  
Diploma behaald op 12 november 1996
- Aug. 1991 – Nov. 1995 **Opleiding Directiesecretaresse**, HBO Hogeschool Schoevers te Arnhem  
Vakkenpakket: Algemene Economie, Bedrijfseconomie, Bedrijfsorganisatie, Commerciële Calculaties, Commerciële vaardigheden, Dienstenmarketing, Engels, Managementvaardigheden, Marketing voor NIMA-A, Marktonderzoek, MS Office, Nederlands, Project & Praktijk, Public Relations, Sales  
Diploma behaald op 15 november 1995
- Juli 1985 – Mei 1991  
**Atheneum A**, Bisschop Bekkers College te Ede.  
Vakken: Wiskunde A, Economie 1, Frans, Duits, Engels, Nederlands, Geschiedenis.  
Diploma behaald op 13 mei 1991

## Overige diploma's

- 1996 – 1997 Cursus Nima-A, Avondopleiding Schoevers te Arnhem.  
Diploma behaald in 1997.
- 1997 – 1998 Cursus Nima-B, Avondopleiding Schoevers te Arnhem.  
Diploma behaald in 1998.

## Taalbeheersing

- Nederlands: Moedertaal
- Engels: Spreken / lezen en schrijven: Uitstekend
- Duits: Spreken / lezen: Goed. Schrijven: Voldoende.
- Frans: Spreken: Matig. Lezen / schrijven: Voldoende.

## Kennis van Computers

Bekend met grote netwerken, Internet, Word, Excel, Exact, Microsoft Outlook, Outlook Express, WordPerfect, diverse telebankingprogramma's, Publisher, PowerPoint, OEC Carsoft (boekhouding/ database).

## Overige

Vaardigheden: flexibel, bereid om te leren, energiek en creatief.  
In het bezit van rijbewijs B,C.

## Hobbies

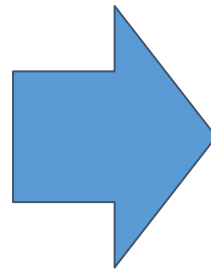
Fitness, waterskiën, lezen en actieve vakantiereizen.

# Document parsing

Curriculum Vitae	
ACHTERNAAM:	Van de Ven
VOORNAMEN:	Petra Paula Maria
ADRES:	Hoofdstraat 15-A
POSTCODE / WOONPLAATS:	6717 AA EDE
GEBORTE DATUM / PLAATS:	29-02-1972 te Wageningen
GESLACHT:	Vrouwelijk
NATIONALITEIT:	Nederlandse
BURGERLIJKE STAAT:	Gehuwd
CONTACT:	+31(0)643054815 / 0318-204594 VervdPetra@yahoo.nl

Ervaring	
• 01-11-2000 – heden	Office Manager bij <b>CTB Flex Software &amp; Services BV</b> te Ede. Functieomschrijving: o.a. beantwoorden van inkomende telefoongesprekken met behulp van telefooncentrale; notuleren bij (interne) vergaderingen; het bestellen kantoorartikelen; bijhouden van agenda's; archiveren; koffie/thee en lunches verzorgen ten behoeve van (interne) besprekingen; inkomende en uitgaande post inschrijven en verwerken; controleren van de inkomende facturen; distributie van de uitgaande offertes.
• 01-08-1997 – 01-11-2000	Secretaresse bij <b>Allure Uitzendbureau BV</b> te Utrecht. Functieomschrijving: o.a. telefoon beantwoorden en zelfstandig correspondentie voeren in de vorm van brieven of e-mail; informatie verstrekken aan uitzendkrachten naar aanleiding van zeer diverse en specifieke vragen Wet Flexibiliteit en Zekerheid; het adviseren van de directie bij het in dienst nemen van werknemers met betrekking tot contracten en salarissen; notuleren en verslagen schrijven; archiveren; mede organiseren van promotionele activiteiten voor klanten en uitzendkrachten.
• 14-10-1996 – 01-08-1997	Boekhoudkundig medewerkster <b>Autobedrijf P.P. Langebroek</b> te Wageningen. Functieomschrijving: o.a. postboek bijhouden; coderen en invoeren in administratie van inkomende facturen; kas-, bank-, girostukken; registratie klantgegevens; betaalopdrachten doen met telebanking; debiteuren bewaken; advertenties maken voor lokale kranten; vitrine inrichten met autoaccessoires; zelfstandig correspondentie voeren; agenda's bijhouden; archiveren; telefoon beantwoorden middels kleine telefooncentrale; informatie verstrekken aan klanten.



```
{
  "personal": {
    "first_name": "Petra",
    "last_name": "Van de Ven",
    ...
  },
  "work_experience": [
    {
      "job_title": "Office Manager",
      "date": "01-11-2000 - heden",
      "company": "CTB Flex Software &
      ...
    },
    ...
  ]
}
```

# Anatomy of a CV

Curriculum Vitae	
ACHTERNAAM:	Van de Ven
VOORNAMEN:	Petra Paula Maria
ADRES:	Hoofdstraat 15-A
POSTCODE / WOONPLAATS:	6717 AA EDE
GEBORTE DATUM / PLAATS:	29-02-1972 te Wageningen
GESLACHT:	Vrouwelijk
NATIONALITEIT:	Nederlandse
BURGERLIJKE STAAT:	Gehuwd
CONTACT:	+31(0)643054815 / 0318-204594 VenvdPetra@yahoo.nl

Ervaring
<ul style="list-style-type: none"><li>01-11-2000 – heden Office Manager bij <b>CTB Flex Software &amp; Services BV</b> te Ede. Functieomschrijving: o.a. beantwoorden van inkomende telefoongesprekken met behulp van telefooncentrale; notuleren bij (interne) vergaderingen; het bestellen kantoorartikelen; bijhouden van agenda's; archiveren; koffie/thee en lunches verzorgen ten behoeve van (interne) besprekingen; inkomende en uitgaande post inschrijven en verwerken; controleren van de inkomende facturen; distributie van de uitgaande offertes.</li><li>01-08-1997 – 01-11-2000 Secretaresse bij <b>Allure Uitzendbureau BV</b> te Utrecht. Functieomschrijving: o.a. telefoon beantwoorden en zelfstandig correspondentie voeren in de vorm van brieven of e-mail; informatie verstrekken aan uitzendkrachten naar aanleiding van zeer diverse en specifieke vragen Wet Flexibiliteit en Zekerheid; het adviseren van de directie bij het in dienst nemen van werknemers met betrekking tot contracten en salarissen; notuleren en verslagen schrijven; archiveren; mede organiseren van promotionele activiteiten voor klanten en uitzendkrachten.</li><li>14-10-1996 – 01-08-1997 Boekhoudkundig medewerkster <b>Autobedrijf P.P. Langebroek</b> te Wageningen. Functieomschrijving: o.a. postboek bijhouden; coderen en invoeren in administratie van inkomende facturen; kas-, bank-, girostukken; registratie klantgegevens; betaalopdrachten doen met telebanking; debiteuren bewaken; advertenties maken voor lokale kranten; vitrine inrichten met autoaccessoires; zelfstandig correspondentie voeren; agenda's bijhouden; archiveren; telefoon beantwoorden middels kleine telefooncentrale; informatie verstrekken aan klanten.</li></ul>

Personal information

Work experience



# Two labeling problems

## Curriculum Vitae

ACHTERNAAM: Van de Ven  
VOORNAMEN: Petra Paula Maria  
ADRES: Hoofdstraat 15-A  
POSTCODE / WOONPLAATS: 6717 AA EDE  
GEBORTE DATUM / PLAATS: 29-02-1972 te Wageningen  
GESLACHT: Vrouwelijk  
NATIONALITEIT: Nederlandse  
BURGERLIJKE STAAT: Gehuwd  
CONTACT: +31(0)643054815 / 0318-204594  
VenvdPetra@yahoo.nl

## Ervaring

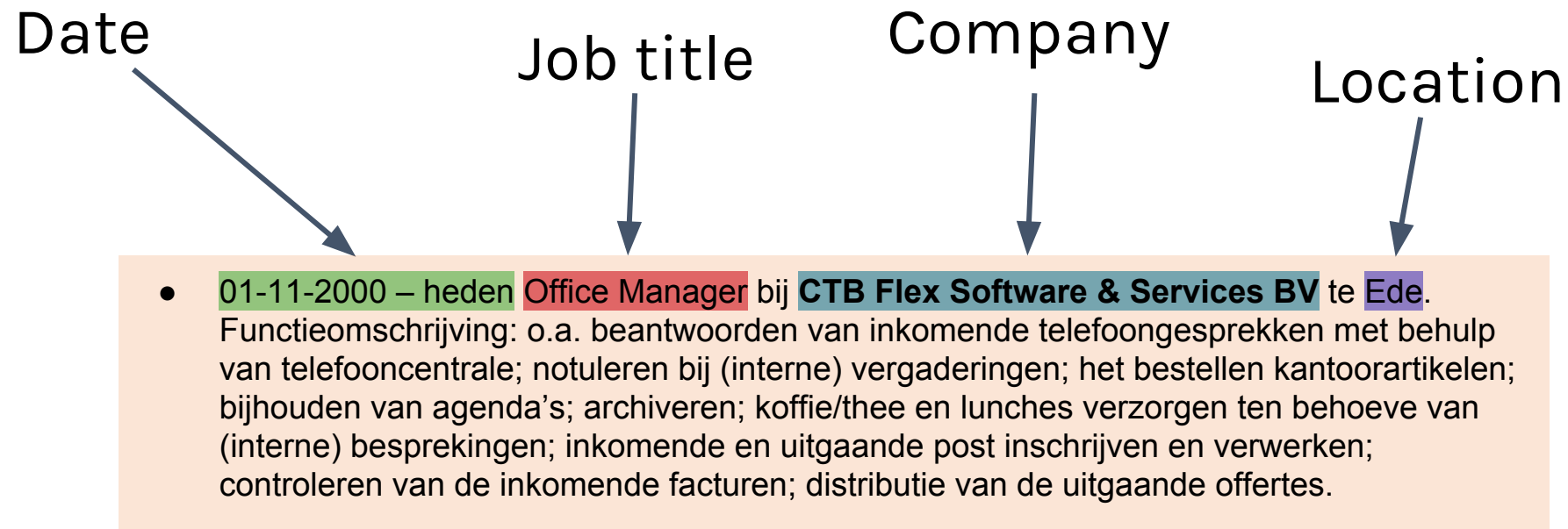
- 01-11-2000 - heden  
Office Manager bij **CTB Flex Software & Services BV** te Ede.  
Functieomschrijving: o.a. beantwoorden van inkomende telefoongesprekken met behulp van telefooncentrale; notuleren bij (interne) vergaderingen; het bestellen kantoorartikelen; bijhouden van agenda's; archiveren; koffie/thee en lunches verzorgen ten behoeve van (interne) besprekingen; inkomende en uitgaande post inschrijven en verwerken; controleren van de inkomende facturen; distributie van de uitgaande offertes.
- 01-08-1997 - 01-11-2000  
Secretaresse bij **Allure Uitzendbureau BV** te Utrecht.  
Functieomschrijving: o.a. telefoon beantwoorden en zelfstandig correspondentie voeren in de vorm van brieven of e-mail; informatie verstrekken aan uitzendkrachten naar aanleiding van zeer diverse en specifieke vragen Wet Flexibiliteit en Zekerheid; het adviseren van de directie bij het in dienst nemen van werknemers met betrekking tot contracten en salarissen; notuleren en verslagen schrijven; archiveren; mede organiseren van promotionele activiteiten voor klanten en uitzendkrachten.
- 14-10-1996 - 01-08-1997  
Boekhoudkundig medewerkster **Autobedrijf P.P. Langebroek** te Wageningen.  
Functieomschrijving: o.a. postboek bijhouden; coderen en invoeren in administratie van inkomende facturen; kas-, bank-, girostukken; registratie klantgegevens; betaalopdrachten doen met telebanking; debiteuren bewaken; advertenties maken voor lokale klanten; vitrine inrichten met autoaccessoires; zelfstandig correspondentie voeren; agenda's bijhouden; archiveren; telefoon beantwoorden middels kleine telefooncentrale; informatie verstrekken aan klanten.

ACHTERNAAM: Van de Ven  
VOORNAMEN: Petra Paula Maria  
ADRES: Hoofdstraat 15-A  
POSTCODE / WOONPLAATS: 6717 AA EDE  
GEBORTE DATUM / PLAATS: 29-02-1972 te Wageningen  
GESLACHT: Vrouwelijk  
NATIONALITEIT: Nederlandse  
BURGERLIJKE STAAT: Gehuwd  
CONTACT: +31(0)643054815 / 0318-204594  
VenvdPetra@yahoo.nl

## Ervaring

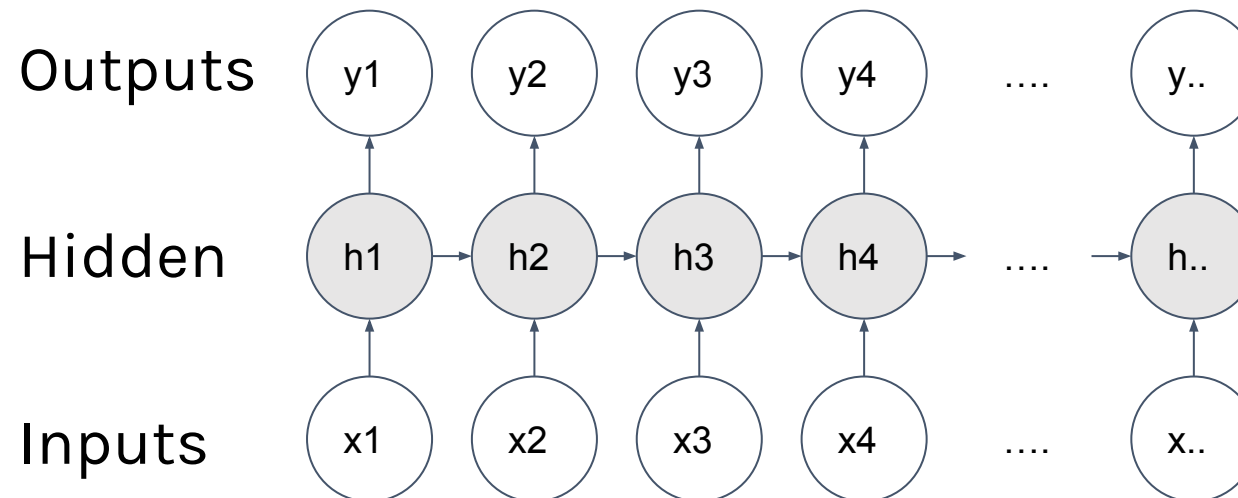
- 01-11-2000 - heden  
Office Manager bij **CTB Flex Software & Services BV** te Ede.  
Functieomschrijving: o.a. beantwoorden van inkomende telefoongesprekken met behulp van telefooncentrale; notuleren bij (interne) vergaderingen; het bestellen kantoorartikelen; bijhouden van agenda's; archiveren; koffie/thee en lunches verzorgen ten behoeve van (interne) besprekingen; inkomende en uitgaande post inschrijven en verwerken; controleren van de inkomende facturen; distributie van de uitgaande offertes.
- 01-08-1997 - 01-11-2000  
Secretaresse bij **Allure Uitzendbureau BV** te Utrecht.  
Functieomschrijving: o.a. telefoon beantwoorden en zelfstandig correspondentie voeren in de vorm van brieven of e-mail; informatie verstrekken aan uitzendkrachten naar aanleiding van zeer diverse en specifieke vragen Wet Flexibiliteit en Zekerheid; het adviseren van de directie bij het in dienst nemen van werknemers met betrekking tot contracten en salarissen; notuleren en verslagen schrijven; archiveren; mede organiseren van promotionele activiteiten voor klanten en uitzendkrachten.
- 14-10-1996 - 01-08-1997  
Boekhoudkundig medewerkster **Autobedrijf P.P. Langebroek** te Wageningen.  
Functieomschrijving: o.a. postboek bijhouden; coderen en invoeren in administratie van inkomende facturen; kas-, bank-, girostukken; registratie klantgegevens; betaalopdrachten doen met telebanking; debiteuren bewaken; advertenties maken voor lokale klanten; vitrine inrichten met autoaccessoires; zelfstandig correspondentie voeren; agenda's bijhouden; archiveren; telefoon beantwoorden middels kleine telefooncentrale; informatie verstrekken aan klanten.

# Phrase extraction



# Sequence labeling

**Named Entity Recognition (NER)** and **POS tagging** are the most common sequence labeling problem.



# Our conversion

Our problem is very similar to **NER**.

So we developed RNN based methods specific to our domain.

Going from HMM + CRF to RNNs:

- ~10-30% error reduction.
- Much more flexible.
- Much less feature engineering.



# Our conversion

Less partial extraction:

Aircraft Engine/Airframe **power plant**  
**fitter**

Aircraft Engine/Airframe power plant  
fitter

Better extraction from running text:

I was a **customer services manager** for the Motor Industry

Less extraction from bad context:

I was reporting to the **General manager** of the Sales department

## This talk

- Introduction.
- LSTM sequence labeling model.
- Serving your model.
- What next?

## This talk

- Introduction.
- LSTM sequence labeling model.
- Serving your model.
- What next?

# Bidirectional LSTM sequence labeling models

[Huang et al. \(2015\) “Bidirectional LSTM-CRF Models for Sequence Tagging”](#)

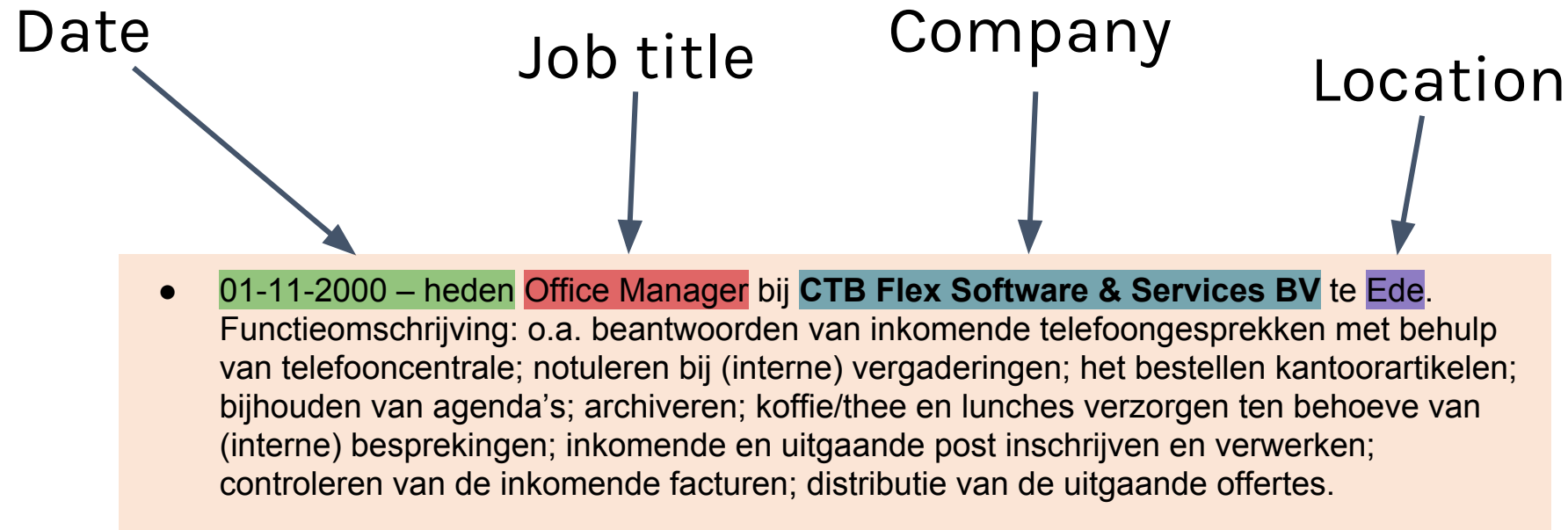
[Ma & Hovy \(2016\) “End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF”](#)

[Lample et al. \(2016\) “Neural Architectures for Named Entity Recognition”](#)

[Chiu & Nichols \(2015\) “Named Entity Recognition with Bidirectional LSTM-CNNs”](#)

This talk will show how to implement a simple BLSTM model.

# Anatomy of a CV



# Input data

Tokens	Labels
## - ## - ####	B-date
-	I-date
heden	I-date
Office	B-jobtitle
Manager	I-jobtitle
bij	0
CTB	B-company
Flex	I-company
Software	I-company



# Notebook Demo

## This talk

- Introduction.
- LSTM sequence labeling model.
- **Serving your model.**
- What next?

```
from flask import Flask
app = Flask(__name__)

predictor = tf.contrib.predictor.from_saved_model(export_dir)

@app.route("/inference", methods=['POST'])
def inference():
    json = request.get_json()
    input_data = DATA_READER.read_lines(json['input'])
    input_dict = {
        'x': input_data,
        'seqlen': [len(input_data)]
    }
    output = predictor(input_dict)
    return jsonify(output)
```

## This talk

- Introduction.
- LSTM sequence labeling model.
- Serving your model.
- What next?

# What next?

Tuning!

Handling unknown tokens.

Multilinguality.

CRF readout.

Line embeddings.

# textkernel

Machine Intelligence for People and Jobs

## **RNN sequence labeling for document parsing in Tensorflow**

Carsten van Weelden  
cvanweelden@textkernel.com

**<https://textkernel.careers/>**