Building Radar: Generating new business made simple

Agenda

- What we do & use case
- NLP @ Building Radar
- Next Challenges
- Q&A Session

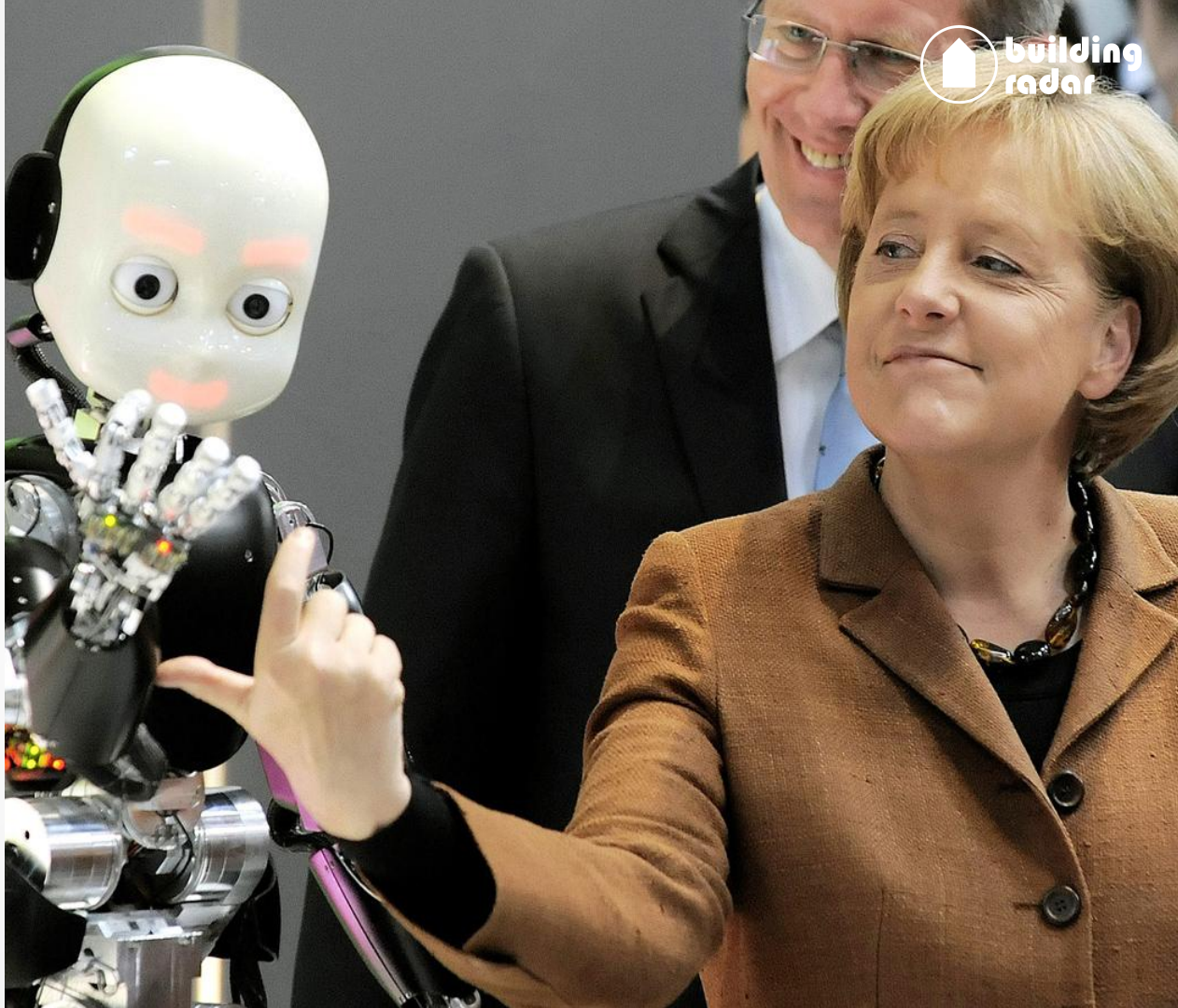
What we do

Building Radar generates a measurable sales pipeline for companies in the construction industry by providing customers with business opportunities.

We deliver leads for their target market and support them through the sales process.

Growth of AI

- Other industries had success
- New Applications
- Early adopters and visionaries
- Pandemic



Use Case

- Find construction projects early
- Spend more time selling and not researching
- Getting in contact

How do we identify relevant construction leads?

Every hour, we are searching through **50.000+ global sources** to find signs of construction projects.

Neural networks are evaluating more than **1.000.000 articles daily** and identify over **5.000 brand-new construction projects** worldwide.

Based on your criteria, we are providing you with **highly relevant construction leads** in real-time.

NLP @ Building Radar

Building Radar: Generating new business made simple

Customer's expectations & BR techniques

Scraping of news websites: Not all articles are related to construction industry
⇒ Filtering

Customers are interested in enriched data
⇒ Labeling, Tagging

Phases
- Under construction
- Planned
- Finished

Categories
- Agriculture
- Industrial
- Residential
- ...

Entities
- Construction's location
- Construction's value
- ...

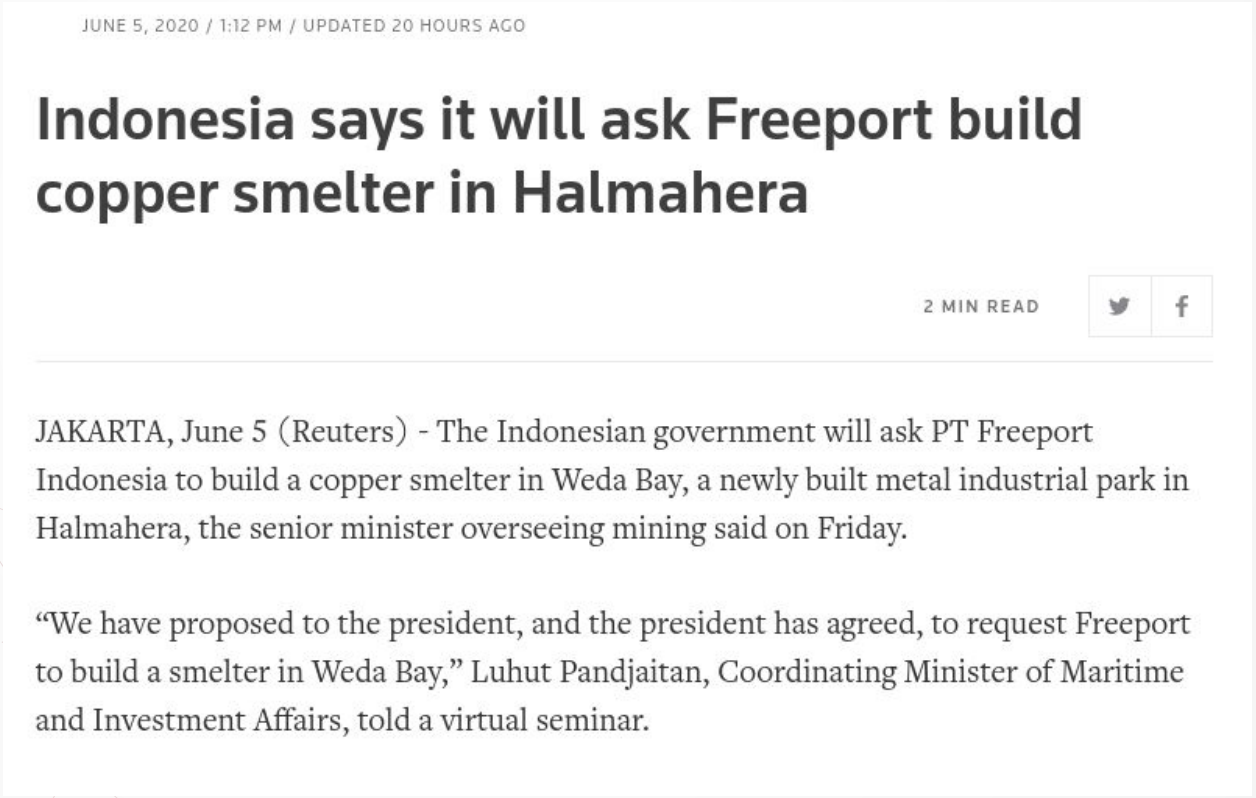
Why use NLP?

Different sources more or less (un)structured:

1. **Tenders**
2. **News Articles**
3. **Architect Websites**
4. **Images**

Different sources more or less (un)structured:

News articles



JUNE 5, 2020 / 1:12 PM / UPDATED 20 HOURS AGO

Indonesia says it will ask Freeport build copper smelter in Halmahera

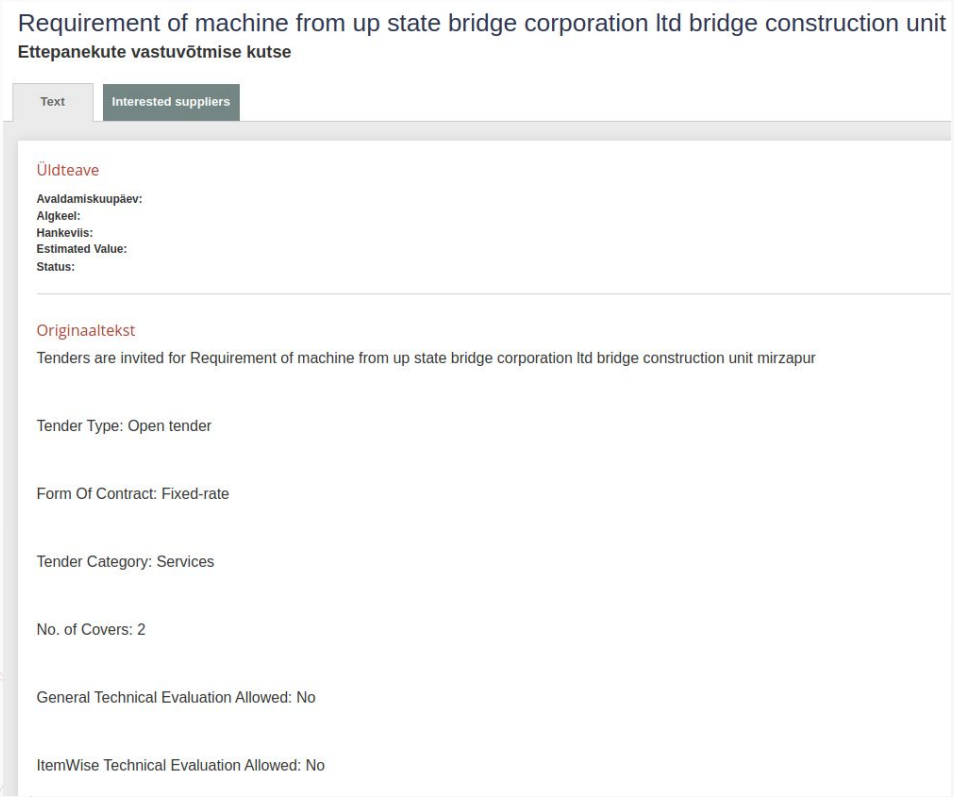
2 MIN READ

JAKARTA, June 5 (Reuters) - The Indonesian government will ask PT Freeport Indonesia to build a copper smelter in Weda Bay, a newly built metal industrial park in Halmahera, the senior minister overseeing mining said on Friday.

"We have proposed to the president, and the president has agreed, to request Freeport to build a smelter in Weda Bay," Luhut Pandjaitan, Coordinating Minister of Maritime and Investment Affairs, told a virtual seminar.

Different sources more or less (un)structured:

Tenders



Different sources more or less (un)structured:

Architect Websites

Siemensstadt 2.0

The master plan opens the Siemensstadt 2.0 and creates a new multifunctional center for culture, research and production – a new attractor for Berlin and beyond.
Through the hybrid program and its versatile uses, the Siemensstadt is creating a new community and home for over 30,000 people. Daycare and school, affordable housing, sports fields, hotels, as well as a public plaza and green spaces upgrade the location and make it a place for everyone to experience. It is no longer just work and production that take place here; numerous new residents from several generations are revitalizing the Siemensstadt and shaping its new community identity.
The careful handling of Siemens' history in Berlin and the corresponding listed buildings are another special feature of the project. In the Schaltwerk hall complex, diverse uses and a wide range of cultural activities can be experienced in a completely new way. Rooms for experimentation and discovery are available in the form of Living Labs and offer places for innovation and inspiration. The existing ten-storey factory and office building are linked to the newly designed Schaltwerk halls by connecting buildings and activated by a mix of coworking and future workshops. The neighboring administration building can be accessed on the northern side by a new wave-like entrance and allows to use the potential of the inner courtyards. A connecting conference and workshop zone is created by an overarching canopy.
The urban structure enables change and growth through a scalable pattern. The systematic area arrangement defines the adaptable and optimized master plan, which can develop in a self-similar way over time as needed.
In the Innovation Campus Siemensstadt 2.0, work and leisure spheres merge into one another for a more intensive exchange and productive entanglement. The historic 70 hectare site is being renovated, modernized and re-functionalized in order to create a forward-looking Smart City district for Berlin.

Location
Berlin, DE

Client
Siemens

Area GFA
1 175 405 m²

Status
Competition

Year
2020

Different sources more or less (un)structured:

Images

building radar

How do we deliver structured data?

Various NLP tasks

1. **Classification**
2. **NER (Named Entity Recognition)**
3. **Q&A (Question & Answering)**

1 Classification

Multi-Class Multi-Label classification

> “Jetzt haben es Stadtverwaltung und Rat schriftlich: Der Misserfolg in der Beethovenhalle hat viele Mütter und Väter. Nachdem die städtischen Rechnungsprüfer das millionenschwere Desaster gewohnt nüchtern analysiert haben, formt sich ein Bild.”

- Infrastructure
- Office
- Sport
- Agriculture
- Etc...

1 Classification - Solution

BERT model: bert-base-german-cased with FARM

```
# 4. Create an AdaptiveModel
# a) which consists of a pretrained language model as a basis

lang_model = "bert-base-german-cased"
language_model = LanguageModel.load(lang_model, language="german")

# b) and a prediction head on top that is suited for our task => Text classification

prediction_head = MultiLabelTextClassificationHead(num_labels=len(label_list))

model = AdaptiveModel(language_model=language_model,
                      prediction_heads=[prediction_head],
                      embeds_dropout_prob=0.1,
                      lm_output_types=["per_sequence"],
                      device=device)
```

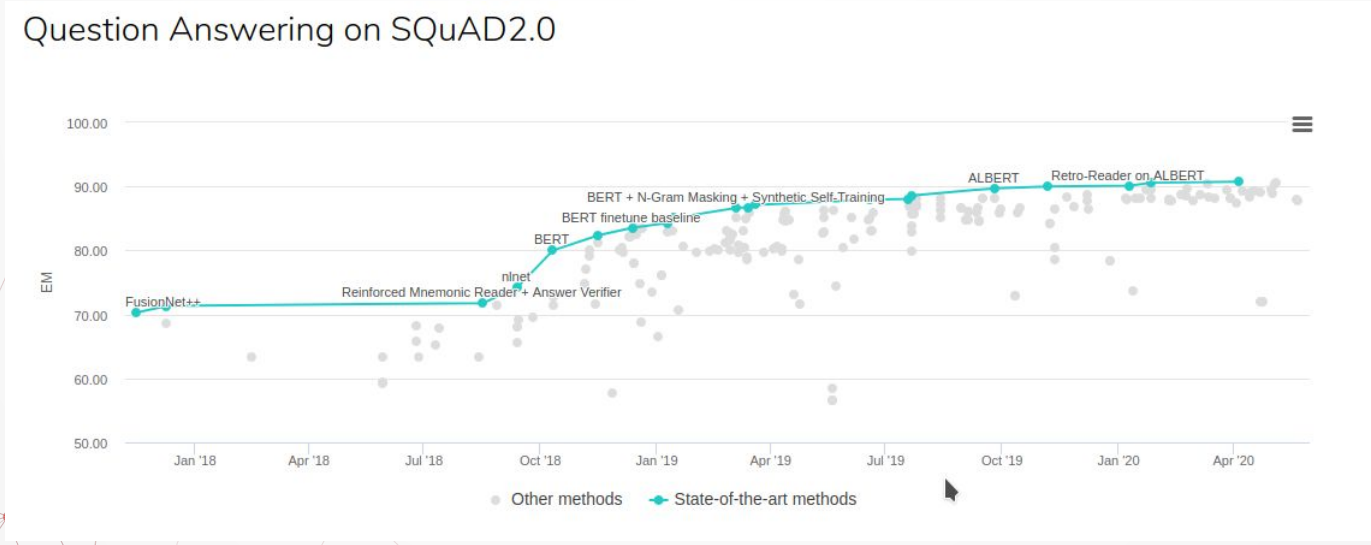
Excursus BERT

Building Radar: Generating new business made simple

BERT (Bidirectional Encoder Representations for Transformers)

Released by Google end of 2018

Application of Transformer (attention model) in order to create a language model.
⇒ Outperforms existing language models in contextual understanding of language



BERT - Transformer

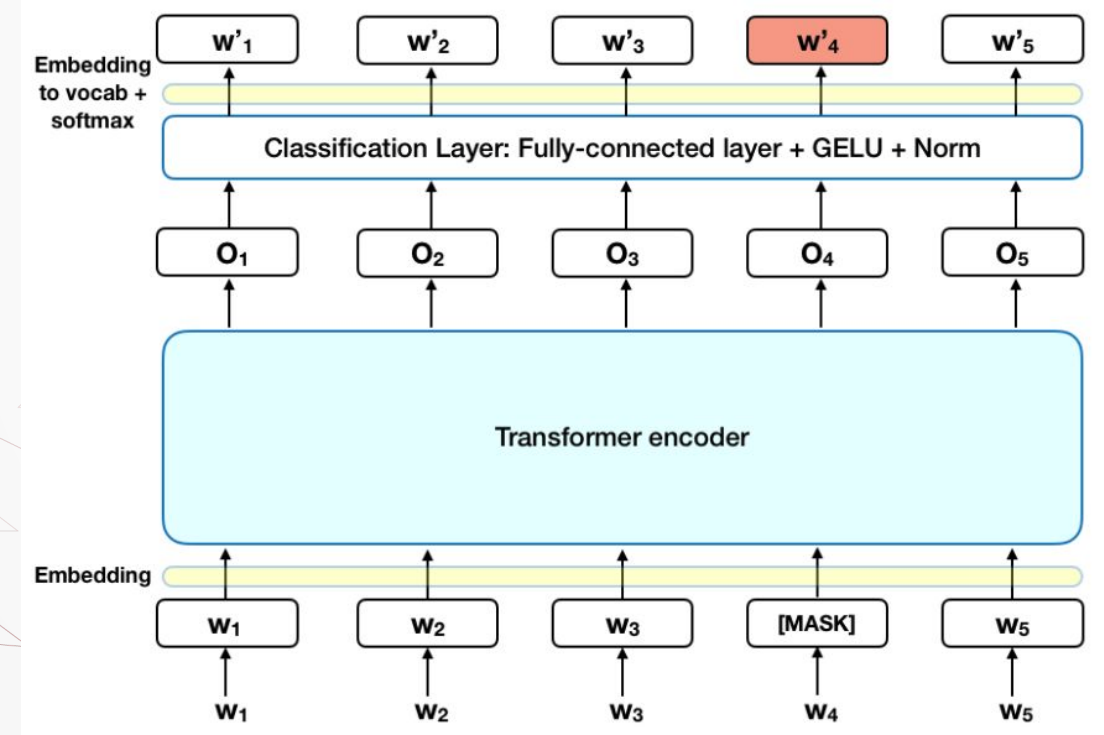
Transformer, an attention mechanism that learns contextual relations between words.

Transformer has two mechanisms:

- Encoder reading the text input
- Decoder producing a prediction for the task (not used to create the language model)

BERT - Transformer Encoder

In pre-processing 15% of the words are replaced with a [MASK] token. The model tries to predict the original token.



BERT - Why use BERT?

Ease of implementation and use of a BERT language model:

- Text classification: Add a classification layer
- NER: Feed the output vector of each token into a classification layer

Frameworks allow us to implement these models easily (FARM, Hugging Face)

2 NER

Customers are also interested in more than the phases and categories of a construction.

They also need to know:

- Where the construction will happen
- When it will start
- What the size of the construction is
- ...

2 NER - Example

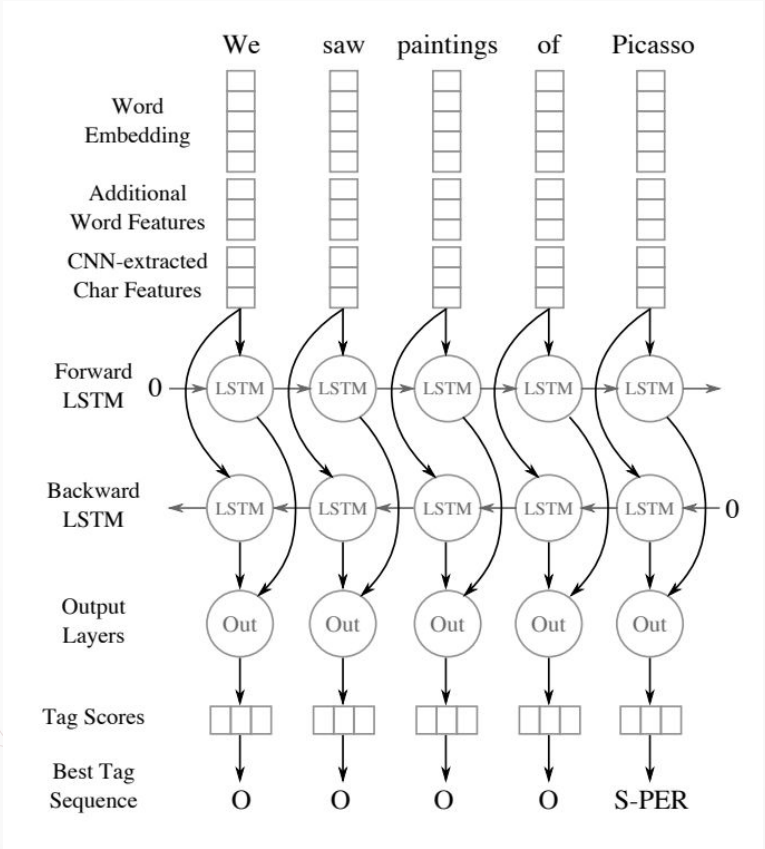
Address location

“**Haßloch.** Nach der Bekanntgabe der Lockerungen der Corona-Maßnahmen des Landes **Rheinland-Pfalz** wird auch der Holiday Park ab dem 10. Juni wieder seine Tore öffnen. Das Konzept zur Wiedereröffnung des Holiday Parks beinhaltet unter anderem die Beschränkung der maximalen Besucherzahl, die verpflichtende Online Anmeldung für die Besucher und zahlreiche Hygiene- und Sicherheitsmaßnahmen innerhalb des Parks. Mit der Wiedereröffnung präsentiert der Holiday Park, mit „DinoSplash“ auch dieses Jahr wieder ein neues Abenteuer für die ganze Familie.”

Where will the construction happen?

“Rheinland-Pfalz”

2 NER - Theory


We
saw
paintings
of
Picasso
Word Embedding
Additional Word Features
CNN-extracted Char Features
Forward LSTM
0
LSTM
Backward LSTM
0
Output Layers
Out
Tag Scores
Best Tag Sequence
O
O
O
O
S-PER

3 Q&A - Cons of sequence tagging

The difference between a location and construction location is small.
It is hard for a NER model to make this difference.

Within a text a model can predict more than one construction location.
Challenge: The correct entity must be chosen

3 Q&A - Theory

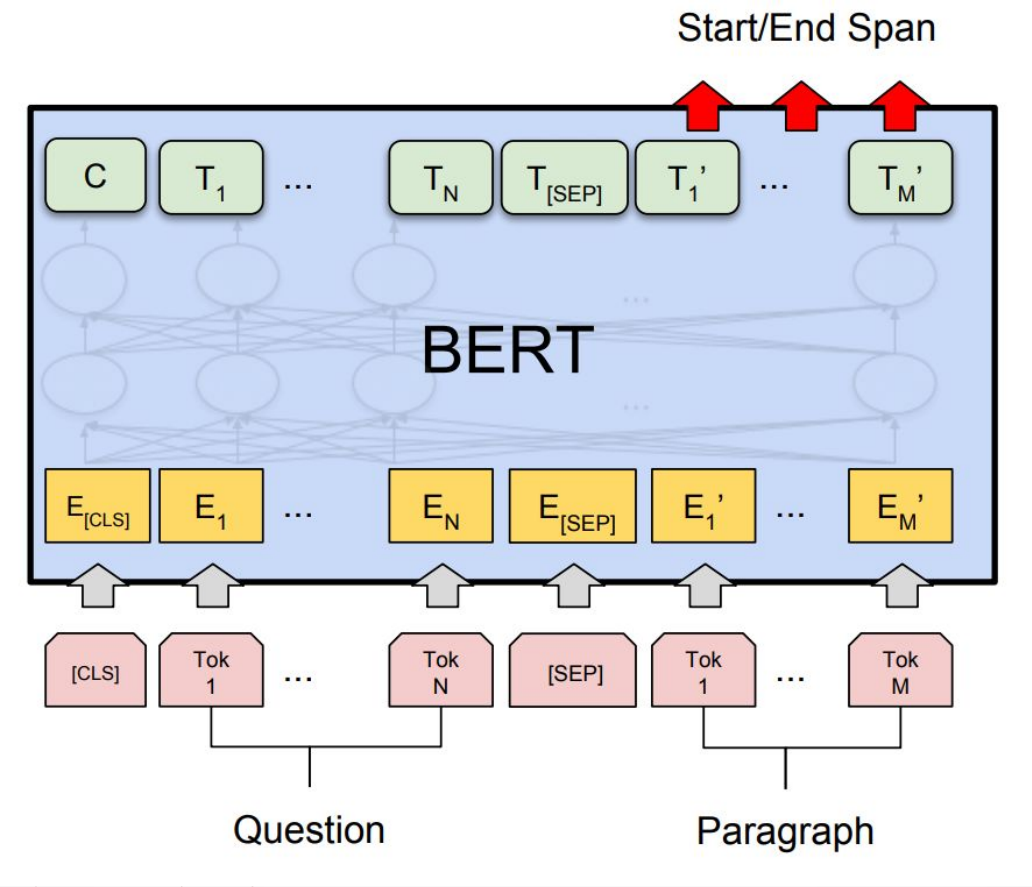
Match a question with a text's span

"Immediately behind the basilica is the Grotto, a Marian place of prayer and reflection. It is a replica of the grotto at Lourdes, France where the Virgin Mary reputedly appeared to Saint Bernadette Soubirous in 1858. At the end of the main drive (and in a direct line that connects through 3 statues and the Gold Dome), is a simple, modern stone statue of Mary"

To whom did the Virgin Mary allegedly appear in 1858 in Lourdes France?

"answer_start": 515, "text": "Saint Bernadette Soubirous"

3 Q&A - Theory


Start/End Span
C
T1
...
TN
T[SEP]
T1'
...
TM'
BERT
E[CLS]
E1
...
EN
E[SEP]
E1'
...
EM'
[CLS]
Tok 1
...
Tok N
[SEP]
Tok 1
...
Tok M
Question
Paragraph

3 Q&A - Example

Address location

“Ampeln an der B1: CDU Dortmund registriert neue Pläne der Stadtverwaltung mit Kopfschütteln - und macht Gegenvorschlag”

“In welcher Stadt ist das Projekt ?”
“Dortmund”

“Wo ist das Projekt ?”
“an der B1”

What is the result?

- Relevant construction projects can be easily found
- Information is readily available, less research time
- Market intelligence

Article

Interesting | Done | Not interesting | Snooze | Export | Report

Rathaus-Umbau: Baubeginn verzögert sich

Translate content? TRANSLATE NO

German Articles | EAB | News Imexan | TÜV Süd: News | Rieder Bau: Eink | TüV N Architekten | N: Querfeld Design (ohne Schlagwörter)

Key facts

Address	Category	Phase	Construction costs
Hallbergmoos Germany	Public Service and Administration	Planned	1M - 5M €

Dates

Construction start in 4 months

Description

Seit fünf Jahren bereitet der geplante Rathaus-Umbau dem Gemeinderat arges Kopfzerbrechen: Von 1,7 auf 4,7 Millionen Euro sind die Kosten während der ...

Associated companies

meuer – planen beraten Architekten GmbH
Architect

Munich
Germany

Landline +49 89 75201330
Web www.m-pb.de

Gemeinde Hallbergmoos

Hallbergmoos
Germany

Landline +49 811 55220
Web www.hallbergmoos.de

strictly confidential - please do not share

Next challenges

Building Radar: Generating new business made simple

NLP Topics

- Document Level NER
- Single Model for Classification and NER tasks
- Contextual Document Similarity
- Document summarization / paraphrasing
- Auto ML

Come and Join the Team!

1. Working Student, Thesis and Full-time positions

2. **68+** employees from **13** nations based in Munich

3. **1.000.000+** newspaper articles read. Number of employees manually researching construction projects: **0**

Marco Both
m.both@buildingradar.com

Aurélien Levecq
a.levecq@buildingradar.com

Thank you!

Building Radar: Generating new business made simple