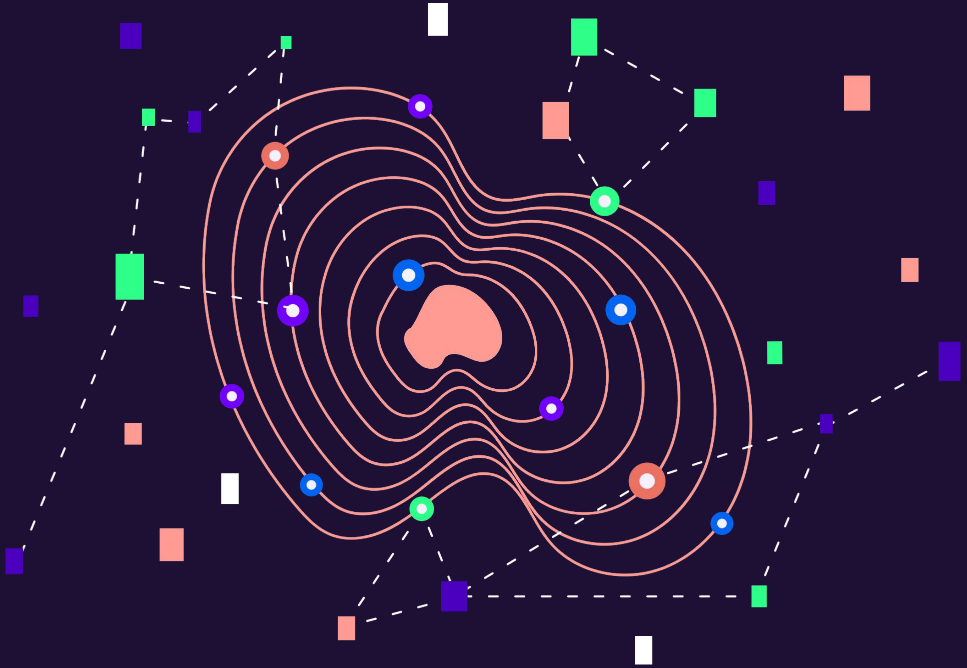
KNSI Golem, Warsaw University of Technology, 21.04.2022

Applica:
Natural Language Processing for Document Understanding

Julita Ołtusek
Research Scientist at Applica

Presentation plan

1. About Applica
2. Information Extraction
3. Language Models
4. TILT live show



A few words about me

- Warsaw University of Technology
  - EiTI, ISI
  - KNSI “Golem”
- Applica
  - Internship and Master’s Thesis: *“Entity labeling in business documents based on contextual information”*
  - Atlas team (Research team, but more engineering stuff)
  - Baldur team (Research, checkboxes project)

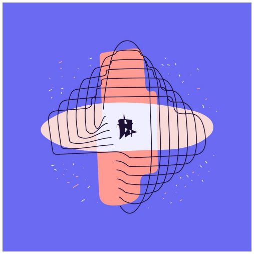
Applica

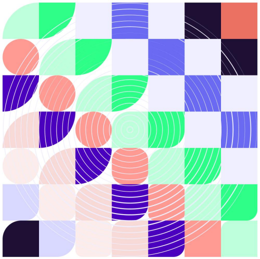
APPLICA

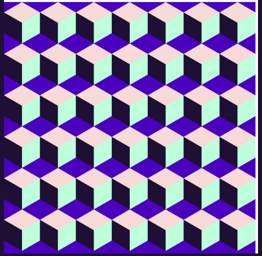
- Not quite a startup but certainly not a corporation (100+ employees)
- Automation of Information Extraction from Business Documents
- Deep Learning, NLP, Data Science

Our use cases

- Automation of previous manual processes
- Datasets: business documents: structured, semi-structured or unstructured, e.g.: invoices, loss runs, contracts, lab requests
- Customers: banks, insurance companies, medical sector
- Our target market: mainly USA but also Europe and Poland





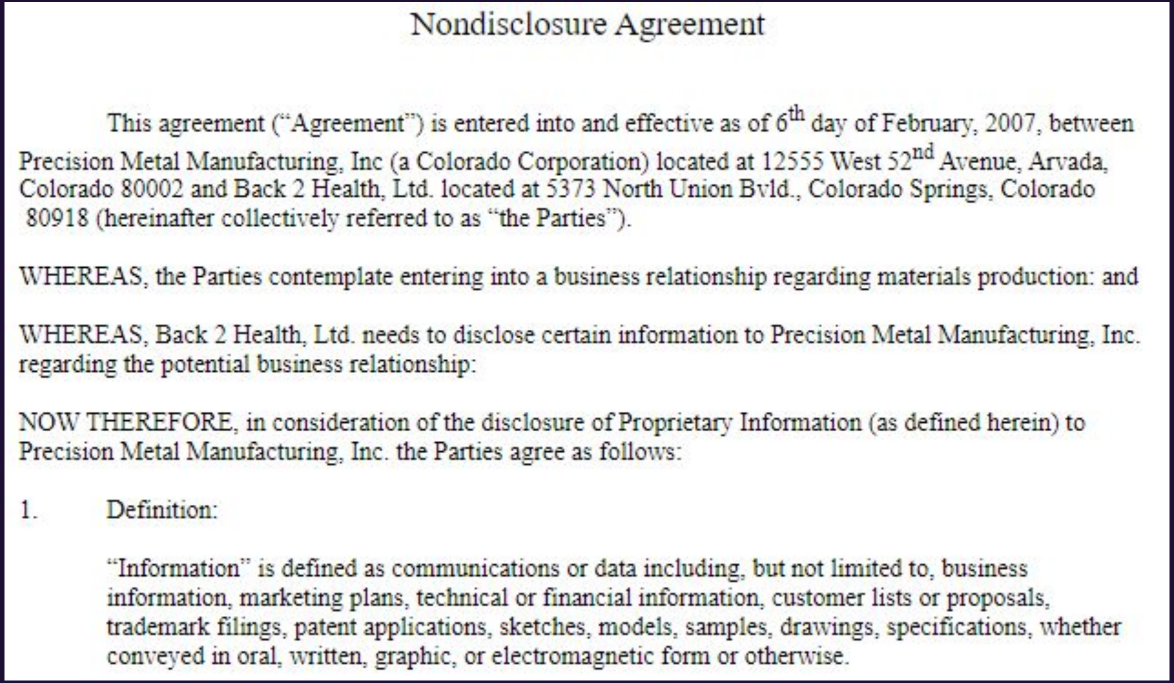


Infrastructure

- NVIDIA DGX A100 (inception program)
  - GPUs for research purposes in total:
    - 8x 80GB
    - 16x 40GB
    - 16x 32GB
  - And more for test and prod envs
- Cloud computing
  - AWS + S3 for storage
  - Azure
  - Google Cloud

Information Extraction from business documents

Information Extraction



Nondisclosure Agreement

This agreement (“Agreement”) is entered into and effective as of 6th day of February, 2007, between Precision Metal Manufacturing, Inc (a Colorado Corporation) located at 12555 West 52nd Avenue, Arvada, Colorado 80002 and Back 2 Health, Ltd. located at 5373 North Union Bvld., Colorado Springs, Colorado 80918 (hereinafter collectively referred to as “the Parties”).

WHEREAS, the Parties contemplate entering into a business relationship regarding materials production: and

WHEREAS, Back 2 Health, Ltd. needs to disclose certain information to Precision Metal Manufacturing, Inc. regarding the potential business relationship:

NOW THEREFORE, in consideration of the disclosure of Proprietary Information (as defined herein) to Precision Metal Manufacturing, Inc. the Parties agree as follows:

1. Definition:

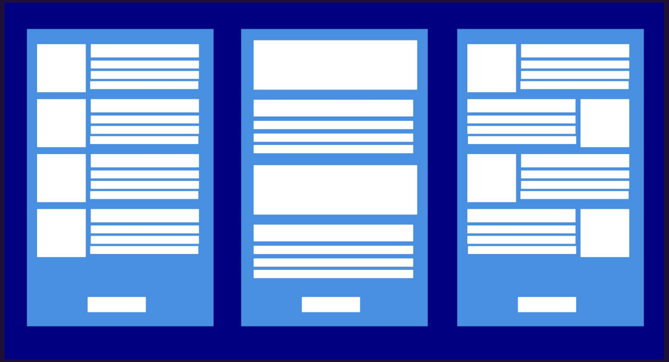
“Information” is defined as communications or data including, but not limited to, business information, marketing plans, technical or financial information, customer lists or proposals, trademark filings, patent applications, sketches, models, samples, drawings, specifications, whether conveyed in oral, written, graphic, or electromagnetic form or otherwise.

Key	Value
effective date	2007-02-06
party	Precision Metal Manufacturing, Inc
party	Back 2 Health, Ltd.
…	…

Layout awareness

Handling different layout elements:

- tables,
- forms,
- continuous text,
- paragraphs,
- lines,
- headers,
- footers,
- other graphical information…

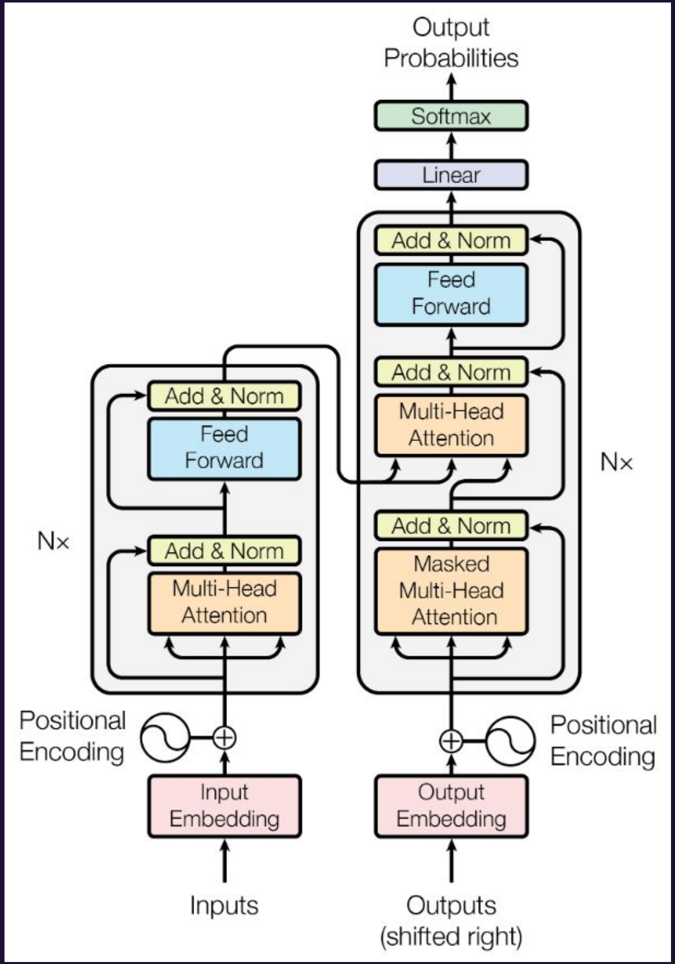


Language models: overview of approaches

Transformer

- Bidirectional encoder maps an input sequence to a sequence of continuous representations, which is then fed into a decoder.
- Decoder receives the output of the encoder together with the decoder output at the previous time step, to generate an output sequence.

- Based entirely on attention mechanism instead of recurrent units



Source: “Attention is all you need” [4].

Language Models based on Transformer

- **June 2018**: GPT, the first pretrained Transformer model, used for fine-tuning on various NLP tasks and obtained state-of-the-art results
- **October 2018**: BERT, another large pretrained model, this one designed to produce better summaries of sentences
- **February 2019**: GPT-2, an improved (and bigger) version of GPT that was not immediately publicly released due to ethical concerns
- **July 2019:** RoBERTa, based on BERT with modified hyperparameters, removed the next-sentence pretraining objective and trained with much larger mini-batches and learning rates
- **October 2019**: DistilBERT, a distilled version of BERT that is 60% faster, 40% lighter in memory, and still retains 97% of BERT's performance
- **October 2019**: BART and T5, two large pretrained models using the same architecture as the original Transformer model (the first to do so)
- **May 2020**, GPT-3, an even bigger version of GPT-2 that is able to perform well on a variety of tasks without the need for fine-tuning (called *zero-shot learning*)

Source: Hugging Face [3].

Categories of Transformer models

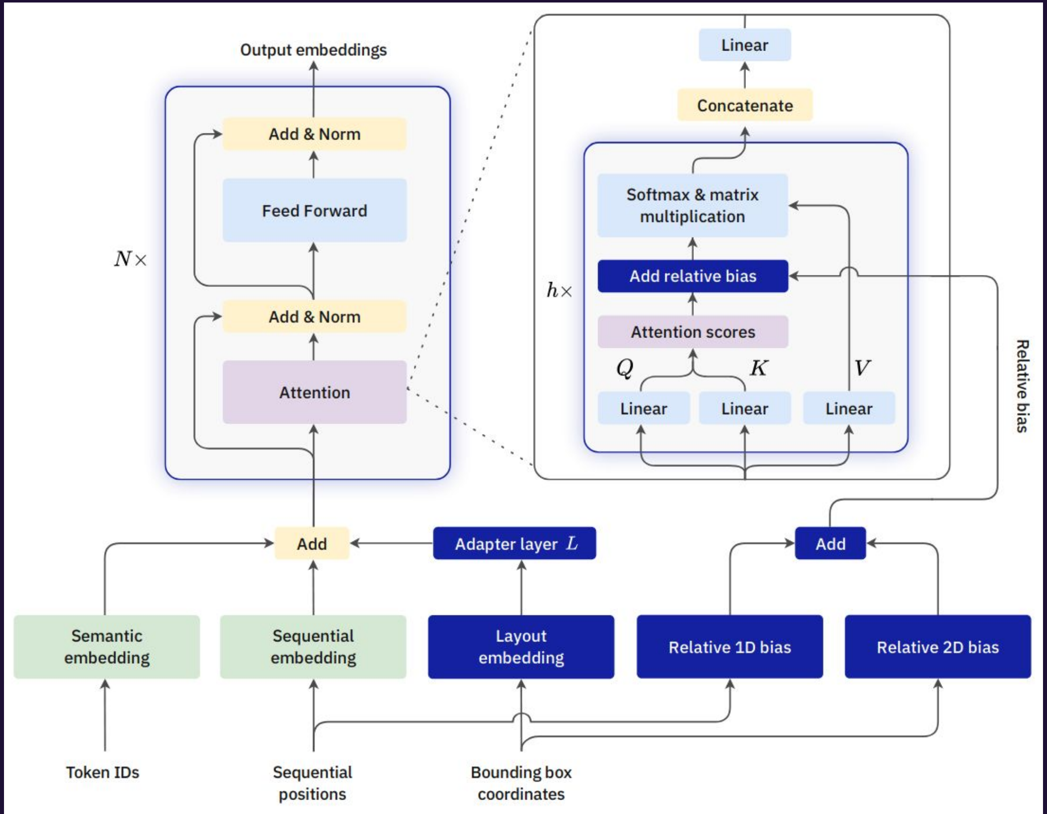
- *auto-regressive* or *decoder-only* (GPT)
- *auto-encoding* or *encoder-only* (BERT)
- *sequence-to-sequence* or *encoder-decoder* (BART, T5)

https://huggingface.co/docs/transformers/index

Models created in Applica

LAMBERT: Layout-Aware Language Modeling for Information Extraction [paper]

Text viewed not simply
as a sequence of words,
but as a collection of tokens
on a two-dimensional page.



Dark blue elements are introduced in LAMBERT.

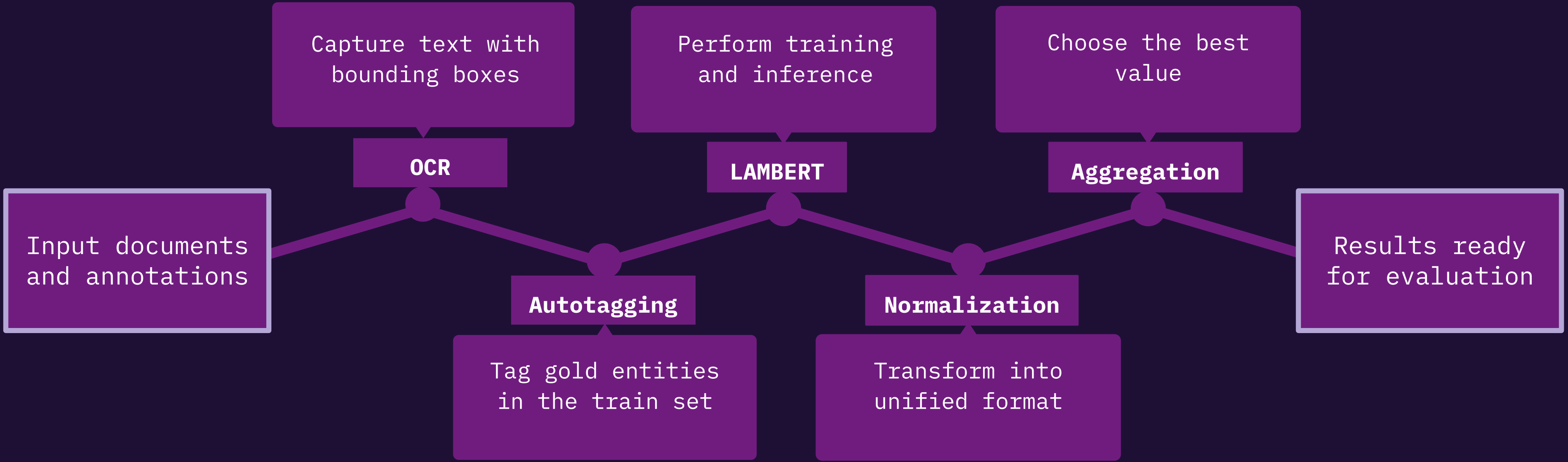
Source: LAMBERT paper [1].

LAMBERT - pretraining and fine-tuning

- Pretraining
  - Initialised with weights of RoBERTa implemented in *transformers* library
  - Self-supervised fashion
  - Masked language modeling objective
  - Collection of around 315k documents (3.12M pages) PDFs extracted from Common Crawl, filtered by an SVM binary classifier to obtain business documents with non-trivial layout

- Fine-tuning
  - Supervised learning (using labeled data)
  - Sequence labeling - classification of tokens
  - Multiple downstream information extraction tasks

Fine-tuning LAMBERT - the whole pipeline



Input documents and annotations

OCR

Capture text with bounding boxes

Autotagging

Tag gold entities in the train set

LAMBERT

Perform training and inference

Normalization

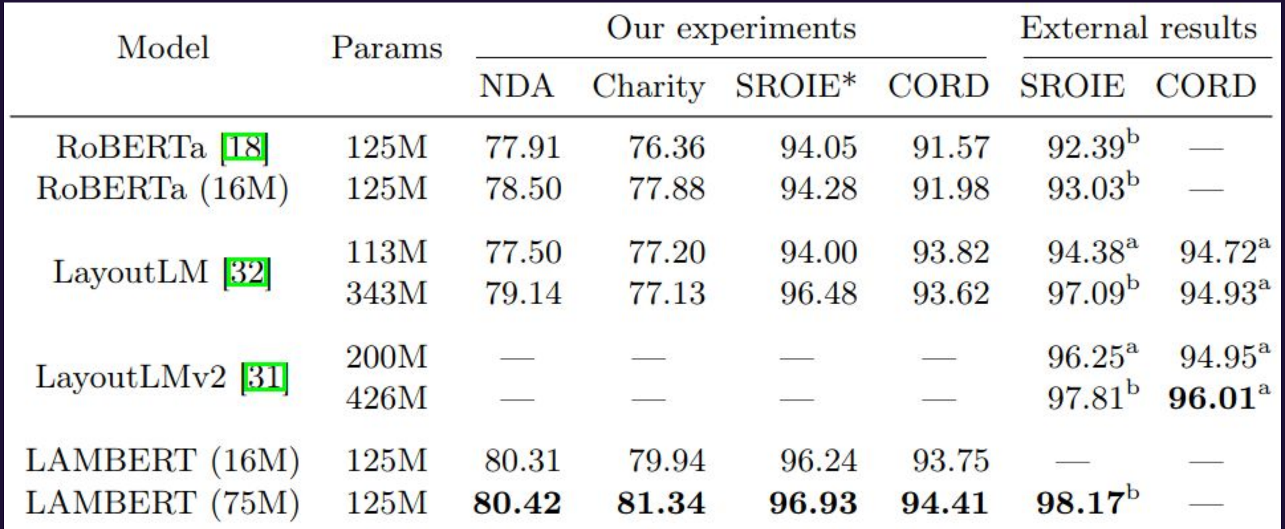
Transform into unified format

Aggregation

Choose the best value

Results ready for evaluation

LAMBERT - results



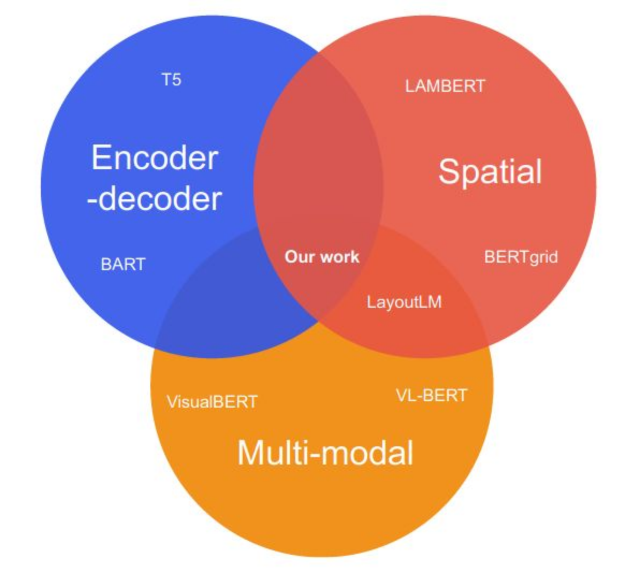
Model	Params	Our experiments				External results	
		NDA	Charity	SROIE*	CORD	SROIE	CORD
RoBERTa [18]	125M	77.91	76.36	94.05	91.57	92.39[b]	—
RoBERTa (16M)	125M	78.50	77.88	94.28	91.98	93.03[b]	—
LayoutLM [32]	113M	77.50	77.20	94.00	93.82	94.38[a]	94.72[a]
	343M	79.14	77.13	96.48	93.62	97.09[b]	94.93[a]
LayoutLMv2 [31]	200M	—	—	—	—	96.25[a]	94.95[a]
	426M	—	—	—	—	97.81[b]	**96.01**[a]
LAMBERT (16M)	125M	80.31	79.94	96.24	93.75	—	—
LAMBERT (75M)	125M	**80.42**	**81.34**	**96.93**	**94.41**	**98.17**[b]	—

Source: LAMBERT paper [1].

TILT - Text-Image-Layout transformer [paper]

- Generative model - extraction performed in a question answering manner

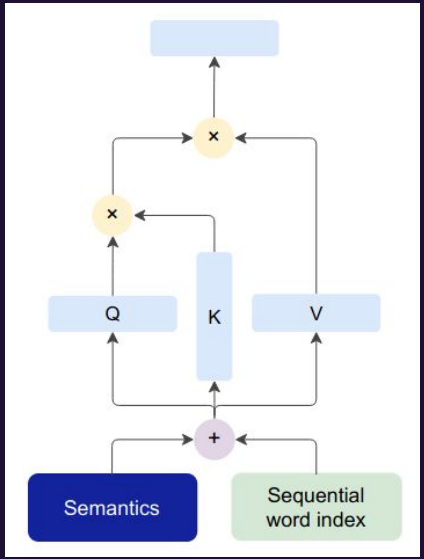
- T5 architecture (sequence to sequence)
- + 2d relative bias
- + U-Net as a backbone visual encoder network



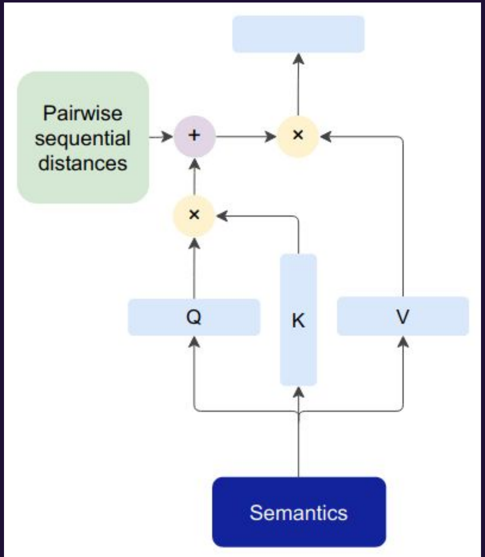
TILT in comparison to other works.

Source: TILT paper [2].

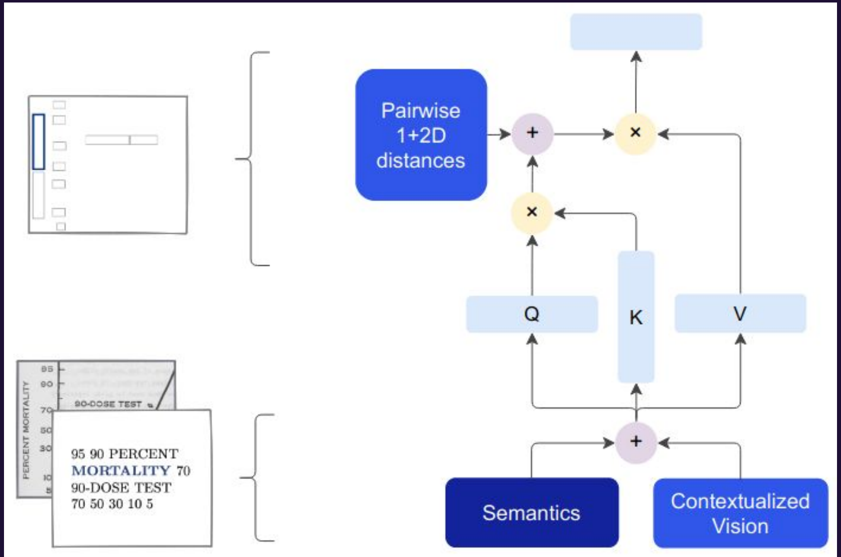
TILT - architecture



Vanilla Transformer



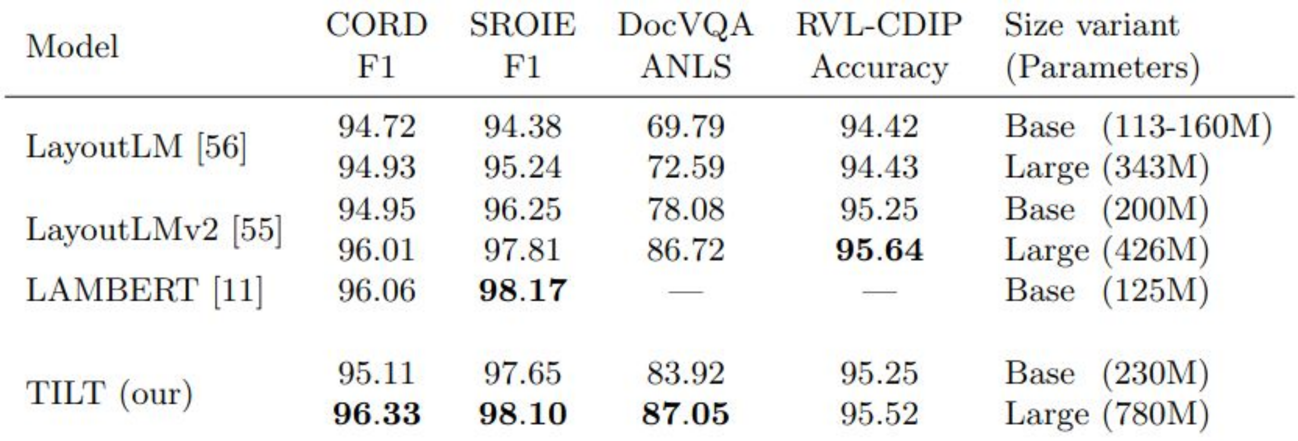
T5 with sequential bias separated from semantics



TILT with additional spatial and graphical information

Source: TILT paper [2]

TILT - results



Model	CORD F1	SROIE F1	DocVQA ANLS	RVL-CDIP Accuracy	Size variant (Parameters)
LayoutLM [56]	94.72	94.38	69.79	94.42	Base (113-160M)
	94.93	95.24	72.59	94.43	Large (343M)
LayoutLMv2 [55]	94.95	96.25	78.08	95.25	Base (200M)
	96.01	97.81	86.72	**95.64**	Large (426M)
LAMBERT [11]	96.06	**98.17**	—	—	Base (125M)
TILT (our)	95.11	97.65	83.92	95.25	Base (230M)
	96.33	**98.10**	**87.05**	95.52	Large (780M)

Source: TILT paper [2].

Awards at ICDAR 2021 - 16th International Conference on Document Analysis and Recognition

- ★ **Best Industry Related Paper Award** for LAMBERT: Layout-Aware Language Modeling for Information Extraction
- ★ Top of the leaderboard of **Infographics Visual Question Answering Challenge** and **Single Document Visual Question Answering Challenge** for TILT

TILT demo!

We're hiring!

- Senior/Mid ML Python Developer
- Junior ML Python Developer
- Intern Research Engineer/Scientist

RESEARCH

- DevOps / SRE Engineer
- Support Manager

DELIVERY

- Senior Frontend Developer
- Senior QA Automation Engineer
- Senior Product Manager

PRODUCT

https://www.applica.ai/about/careers

References

1. LAMBERT: Layout-Aware Language Modeling for Information Extraction
2. Going Full-TILT Boogie on Document Understanding with Text-Image-Layout Transformer
3. Hugging Face course
4. Attention is all you need

Thank you for attention!