Introducing Kairntech: AI-powered Text Analysis

Stefan Geißler, Kairntech

Oct, 2019

About Kairntech

kairntech
AI-powered Text Applications for Enterprise

❖ Kairntech: The company

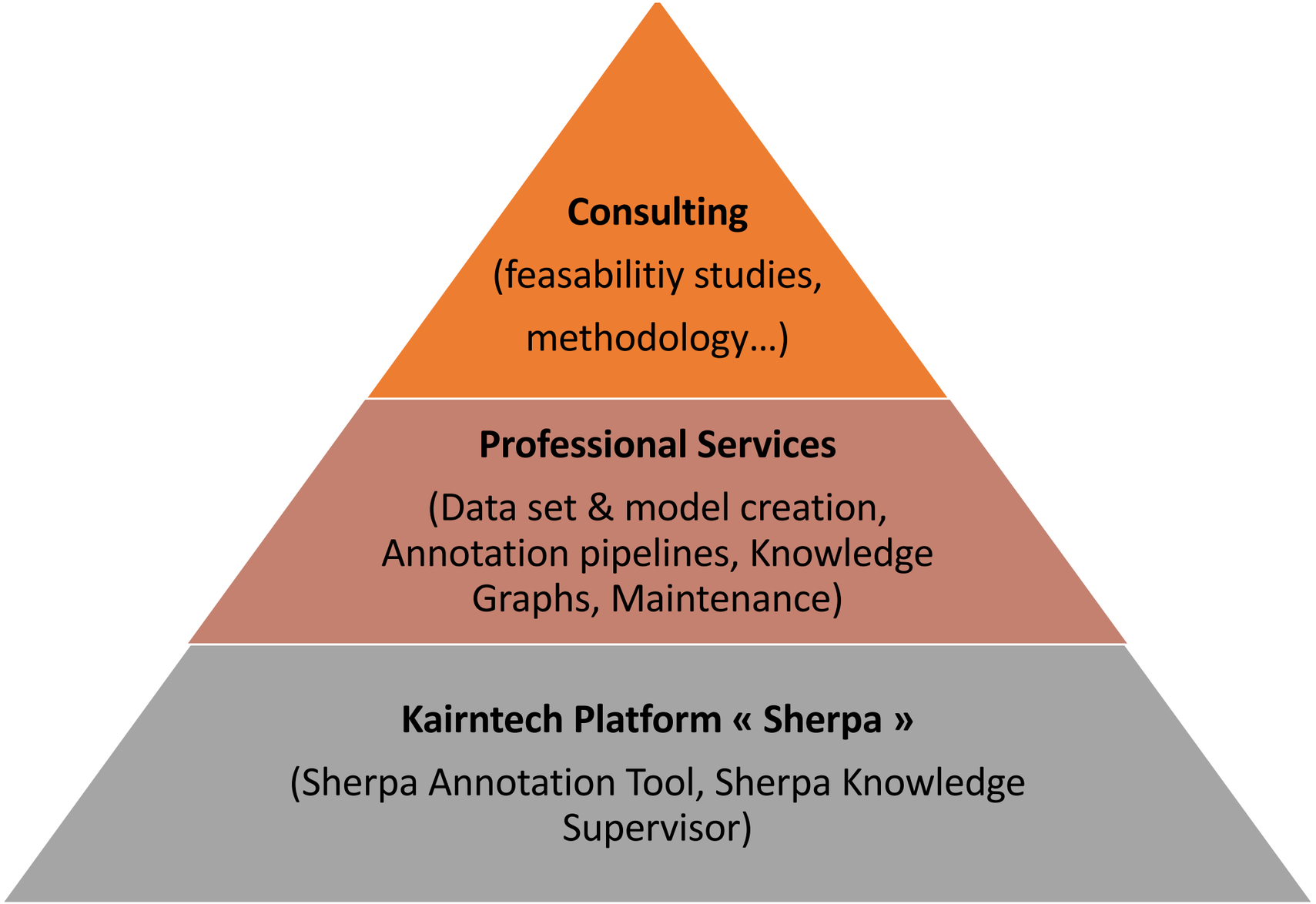
- ❖ Created in dec 2018, 10 partners
- ❖ France (Paris & Grenoble/Meylan), Germany (Heidelberg)

❖ Kairntech: The team

- ❖ Background in Software engineering, Machine Learning, Sales, Management
- ❖ +15 years of experience in NLP development and deployment from Xerox, IBM, TEMIS. Development of components currently in production at CERN, NASA, EPO…

Kairntech: Our offering

kairntech
AI-powered Text Applications for Enterprise



Machine Learning & NLP: Pain points

kairntech
AI-powered Text Applications for Enterprise

- Off-the-shelf NLP models often **don’t work for specific needs**
- Implementation is **slowed down** by the need of building specific training dataset
- AI/NLP services are often require **integration of business glossaries & knowledge graph**
- Absence of maintenance leads to **quality deviations**

So: We need data, not only algorithms

kairntech
AI-powered Text Applications for Enterprise

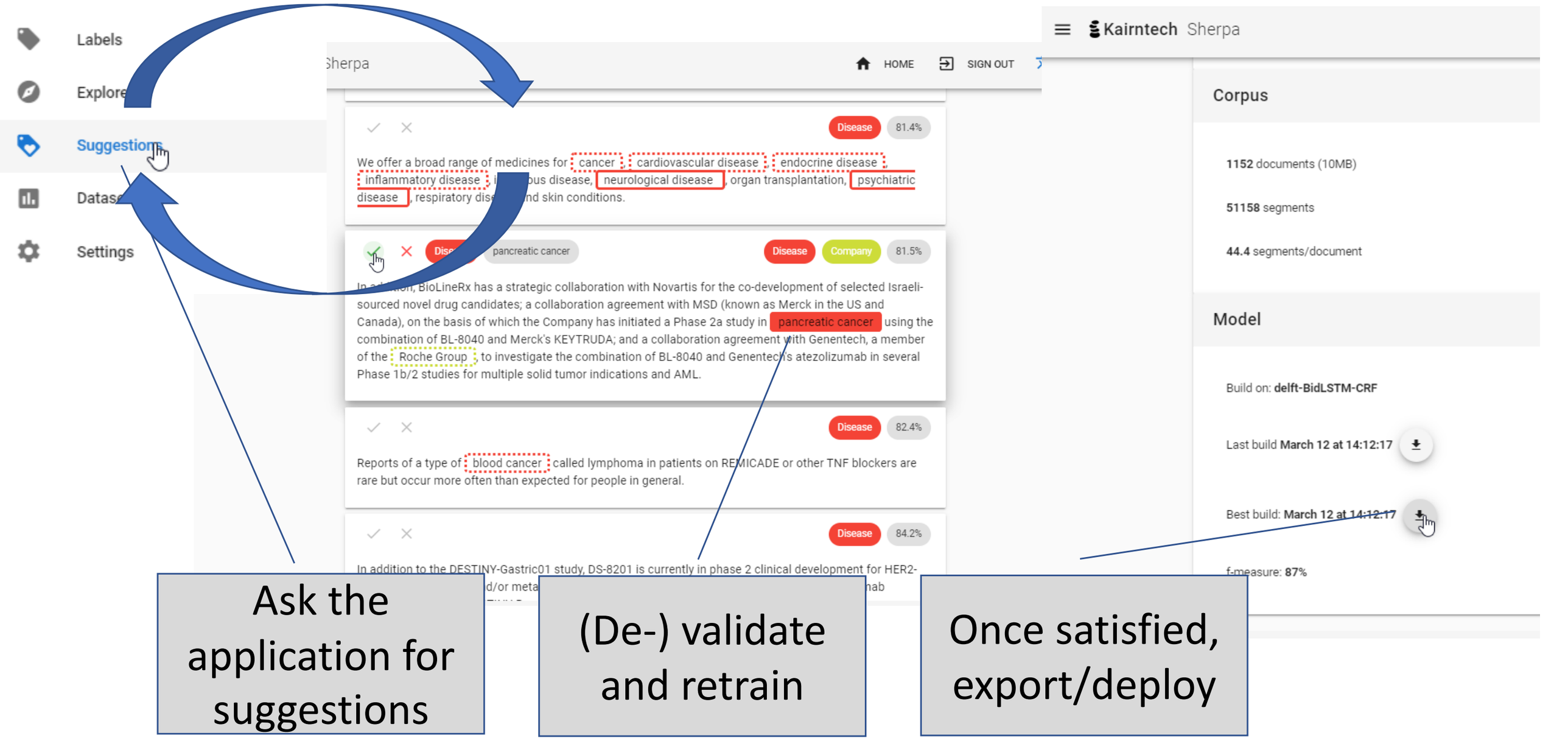
What people think AI is about

The reality

Charts copied from https://hackernoon.com/%EF%B8%8F-big-challenge-in-deep-learning-training-data-31a88b97b282

The Sherpa: A non-expert workflow for dataset creation

kairntech
AI-powered Text Applications for Enterprise



Ask the application for suggestions

(De-) validate and retrain

Once satisfied, export/deploy

On dataset preparation: Requirements

kairntech
AI-powered Text Applications for Enterprise

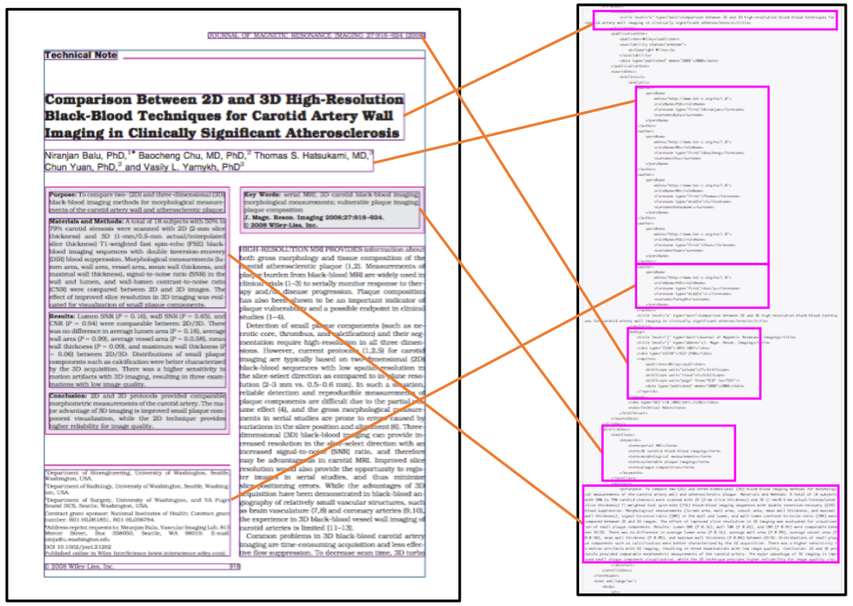
- **Web-based** (no install), intuitive GUI, usable by domain experts
- Limit manual annotation efforts: ***Active Learning***
- **Collaboration** (work in teams, measure inter-annotator agreement)
- **Not just NER** annotation: Entity typing, document categorization, …
- Facilitate **deployment-to-production**

Kairntech beyond the Sherpa

kairntech
AI-powered Text Applications for Enterprise

Activities that are not part of a product:

- ❖ Annotation directly into PDF documents (respecting document structure)
- ❖ Annotation (normalization, disambiguation, scoring, linking) according to existing vocabularies and thesauri
- ❖ Integration with Graph Databases



Conclusions

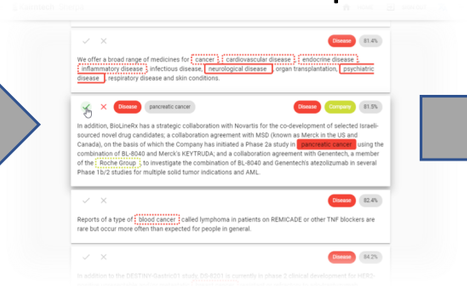
kairntech
AI-powered Text Applications for Enterprise

- **So much data!**
  - But very little of it labelled and useful for superised learning
- **So many pretrained models!**
  - But most of the time they do not quite do what you need in your project
- **So many algorithms!**
  - But a library alone will not allow you to implement the solution you need

Kairntech & The European Language Grid

kairntech
AI-powered Text Applications for Enterprise

3rd party component
Kairntech Sherpa



3rd party component X

3rd party component Y

3rd party component Z

Language corpora:
Social Media, News,
Scientific Documents,
Contracts, Laws,

Document analysis modules
(ML models)

Other data …

Other data …

European Language Grid: language data, analysis modules, ….

Issues, requirements

kairntech
AI-powered Text Applications for Enterprise

- **Documented and uniform format for Document corpora**
- **API to allow third party components to access and process data**
- **API to feed resulting modules back into the ELG**
- **License issues:**
  - Limitations on document corpora and resulting modules
  - Definition of business model for 3rd party components