silk
data

machine learning
excellence delivered

AI-based search and navigation for book collections

silk
data

Demo: https://edubase.silkdata.ai/
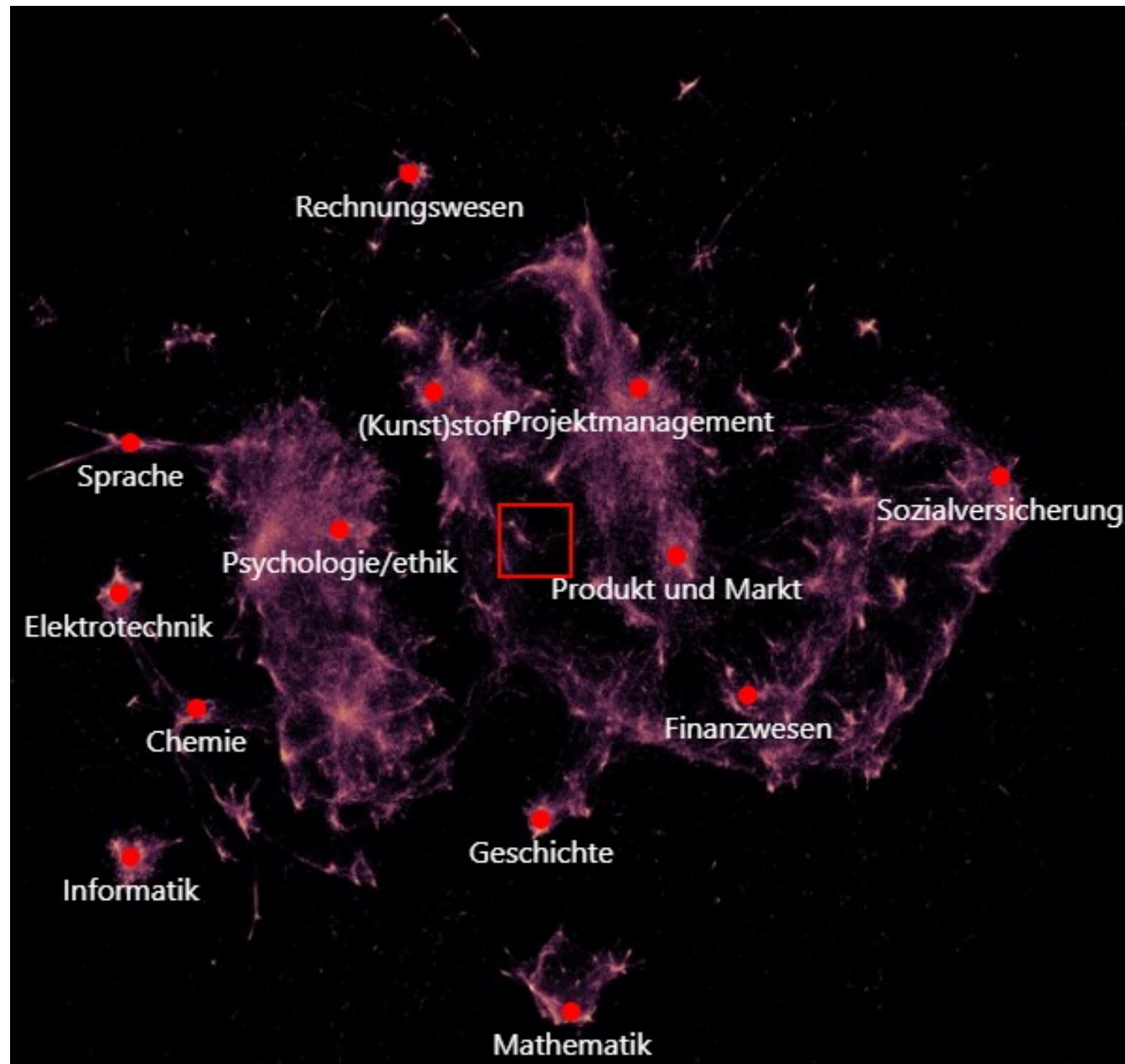
Outline

- Motivation
- Implementation principle
    - Word map: keywords and topics
    - Document map
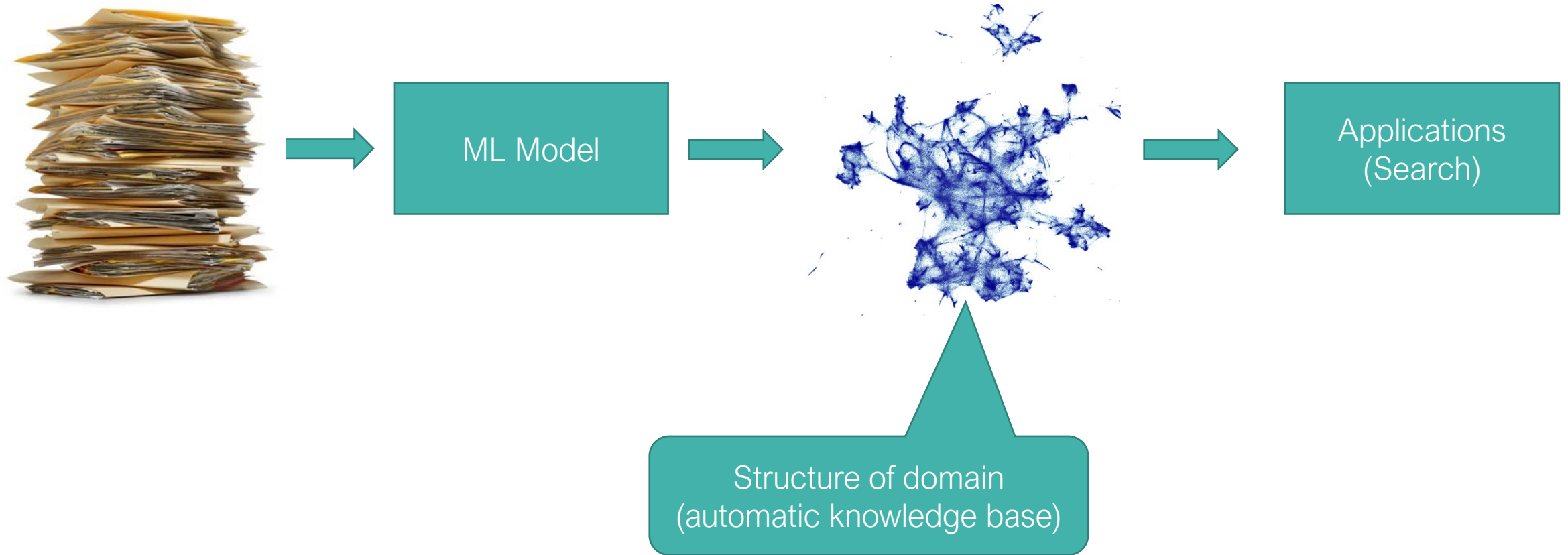   - Search clarification
- Search queries
- Applications

Motivation

- Knowledge workers spent up to **2.5 hour per day*** to find the information
  - Equals to **€25 million per year** for 1000-workers corporation
  - Extra cost of **not finding** important data

- Issues:
  - Varying/special vocabulary
  - Duplicate documents
  - Large number of results
 - Lack of user data ('clickstream') in corporate and special search systems

* Source: http://intranetfocus.com/2-5-hours-a-day-spent-searching-really/

Implementation: semantic map

ML Model

Applications
(Search)

Structure of domain
(automatic knowledge base)

Data for demos

- Books from Edubase platform
  - https://app.edubase.ch/
  - 2694 German-language books
  - 430217 pages in total
  - Search demo: https://edubase.silkdata.ai/
- Scientific papers related to COVID-19 crisis
  - Source of papers: https://www.semanticscholar.org/cord19
  - Free access
  - 53000 papers in our demo
  - Search demo: https://covid19.silkdata.ai/
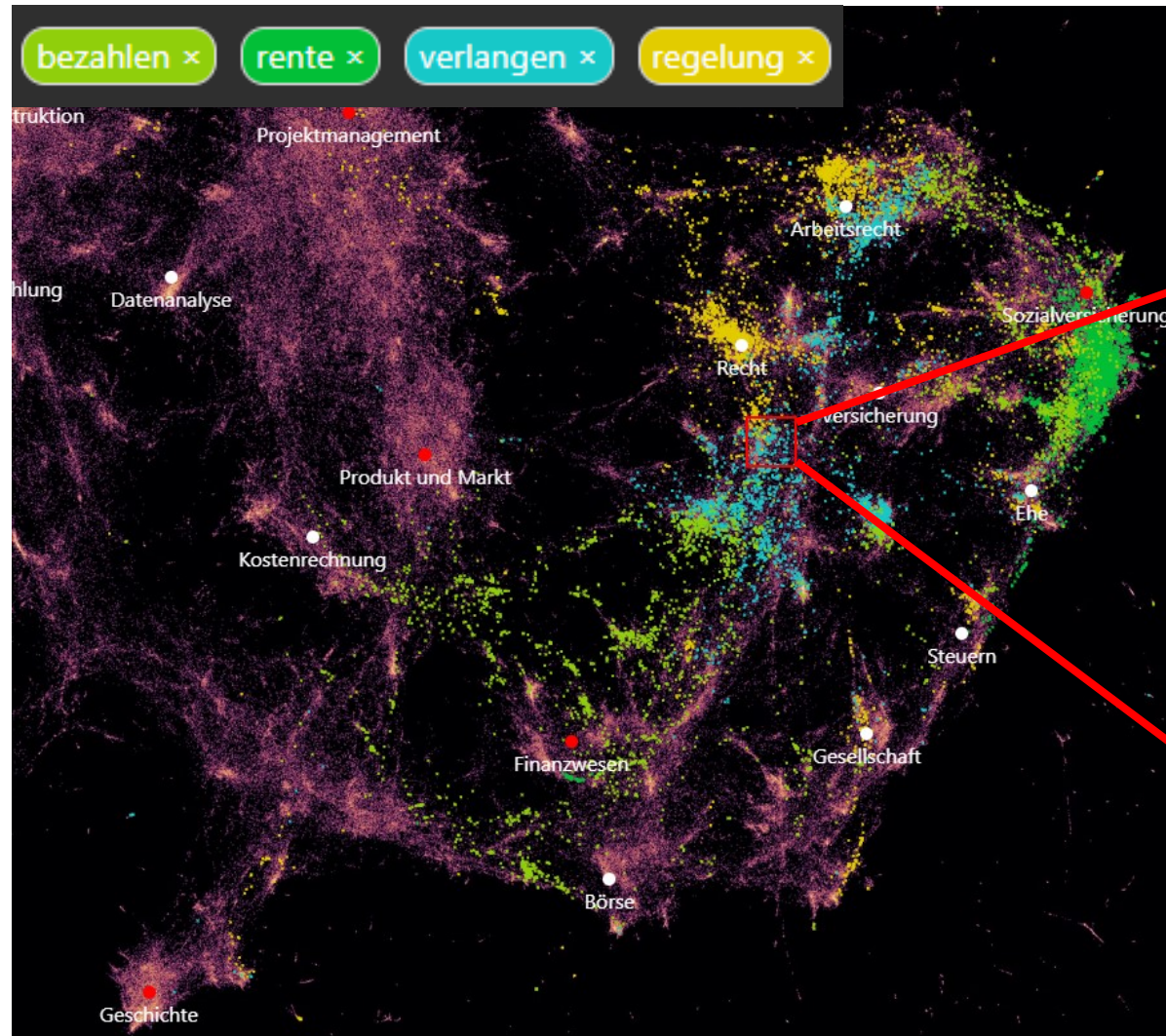
Word map

Connection between keywords, topics and documents (snippets)

NLP pipeline
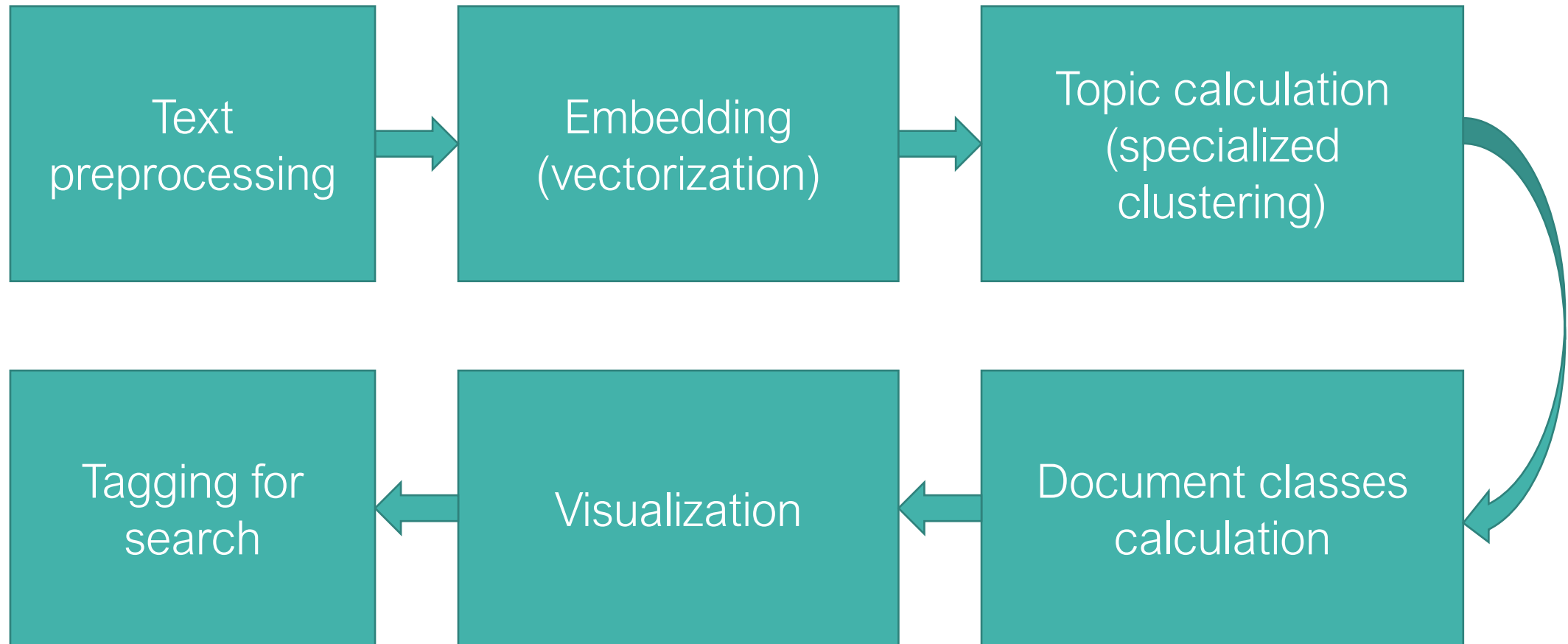
Document map
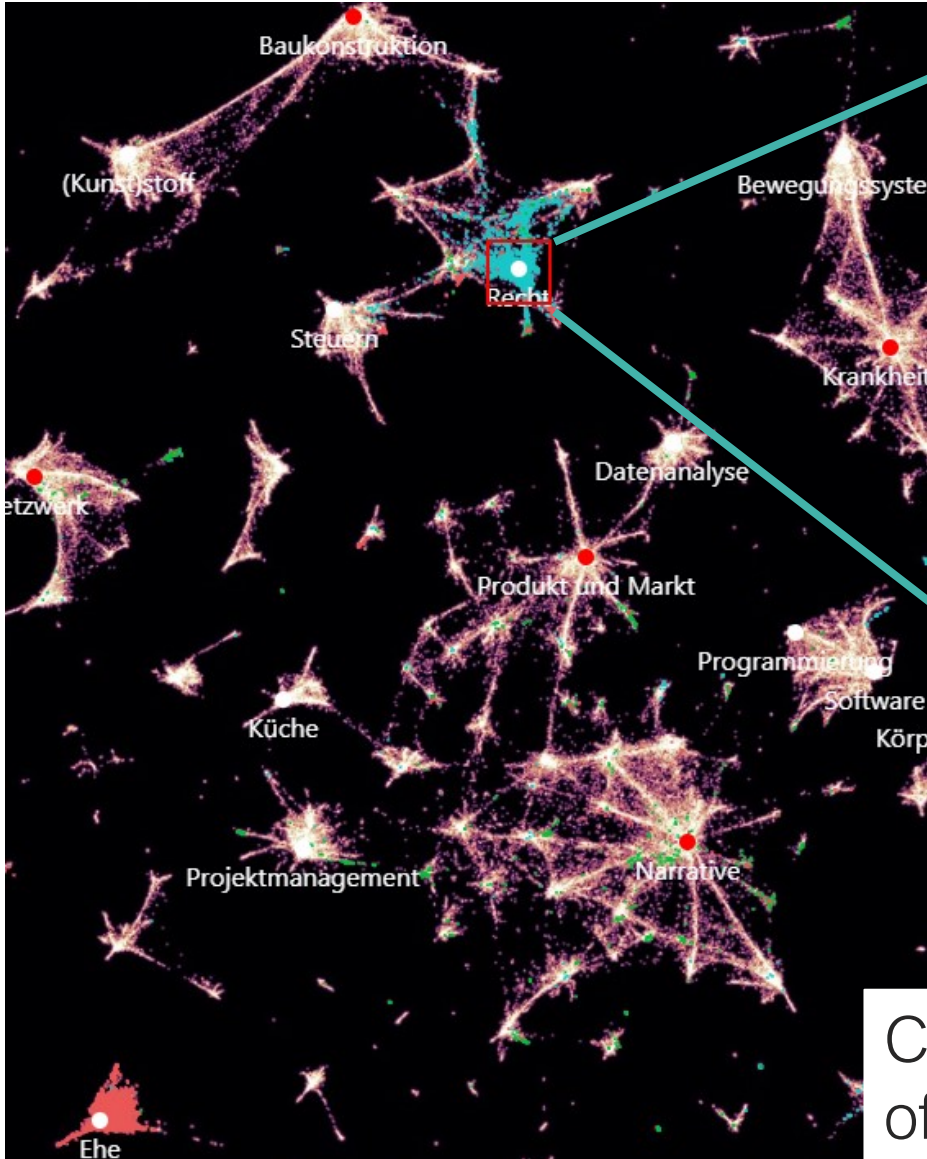
ARTICLES ⓘ

Internationales Zivilprozessrecht, p.465
https://app.edubase.ch/#doc/35910/465

Schweizerisches Obligationenrecht - Allgemeiner Teil,
p.323
https://app.edubase.ch/#doc/31654/323

Internationales Zivilprozessrecht, p.352
https://app.edubase.ch/#doc/35910/352

Allgemeines Verwaltungsrecht, p.271
https://app.edubase.ch/#doc/35914/271

Schweizerisches Obligationenrecht. Allgemeiner Teil I & II,
p.323
https://app.edubase.ch/#doc/31974/323

Nachbarschaft - was gilt im Konfliktfall, p.44
https://app.edubase.ch/#doc/40266/44

ZGB. Das Schweizerische Zivilgesetzbuch, p.748
https://app.edubase.ch/#doc/31986/748

ZGB. Das Schweizerische Zivilgesetzbuch, p.101
https://app.edubase.ch/#doc/31986/101

Allgemeines Verwaltungsrecht, p.615
https://app.edubase.ch/#doc/35914/615

Connection between document classes (groups
of topics) and documents (book pages)

Document map: principle

Search

'Find similar':
looks for semantically similar documents (in topic space)

Document classes (contexts) for search clarification

Principle at a glance

Exploration

Word map	→	Document map	→	Search

Word map

Document snippets (150-200 words)

Keywords

Topics (themes)

Relations between concepts

Document map

Book pages

Tags*

Relations between pages

Search

Book pages

Tags

Search clarification
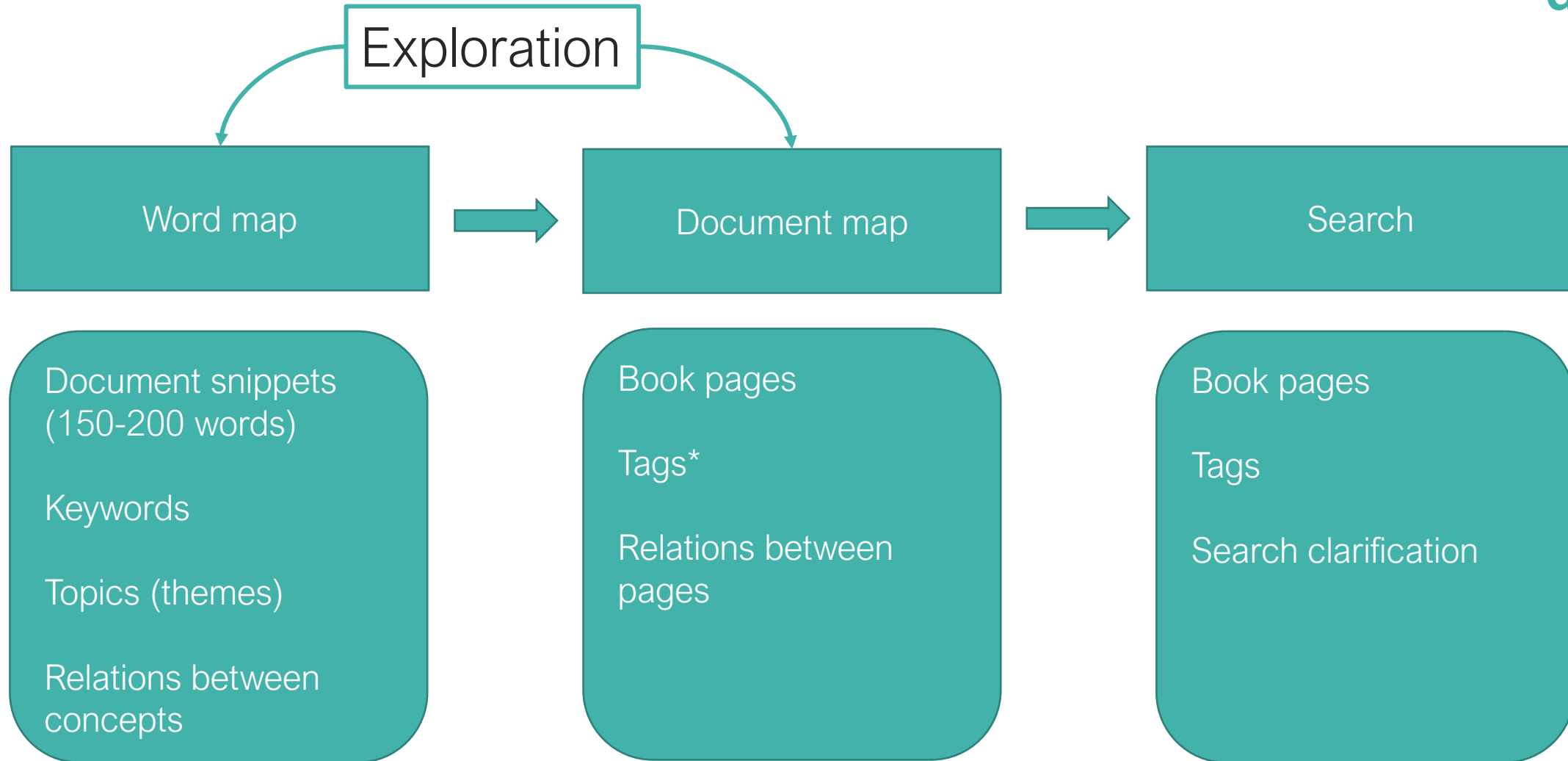
* Document classes, combinations of topics; **single tag** per document (book page)

Search queries

- Edubase platform

- Collected from Nov 2019 till Aug 2020

- 310537 queries in total

- 1650 per day

- 58889 unique queries

- 1 to 31 words (99% - 8 words or less)

- 2 to 234 symbols (99% - up to 63 symbols)

- Up to 300,000 results per query
  (25% queries yields more than 230 results)

- Up to 104 document classes

Number of results per query
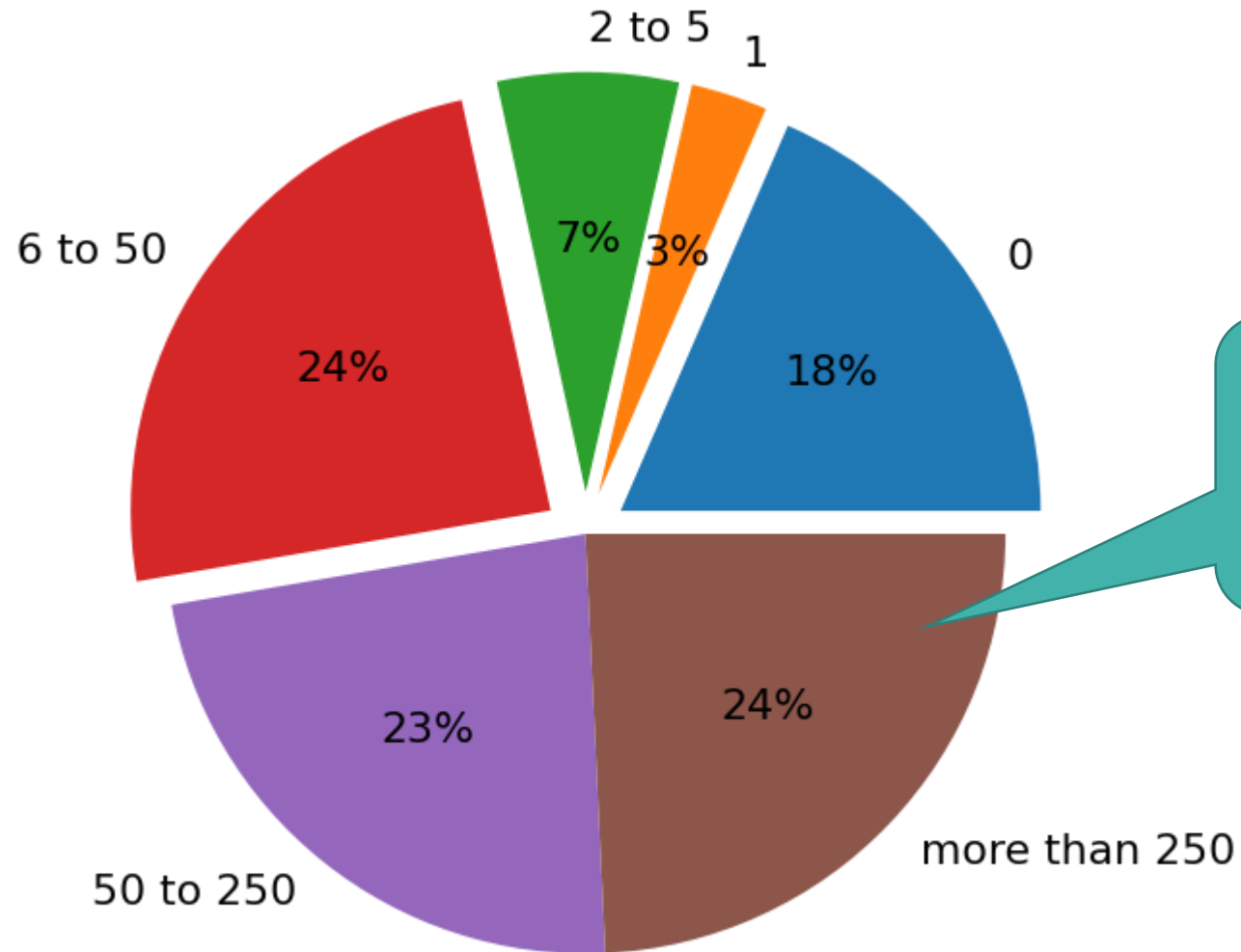
Almost half of (unique) queries have more than 50 results

Number of tags per query

Usual solution

Add more keywords to search query

Tag histogram

Search query: 'Regelung'
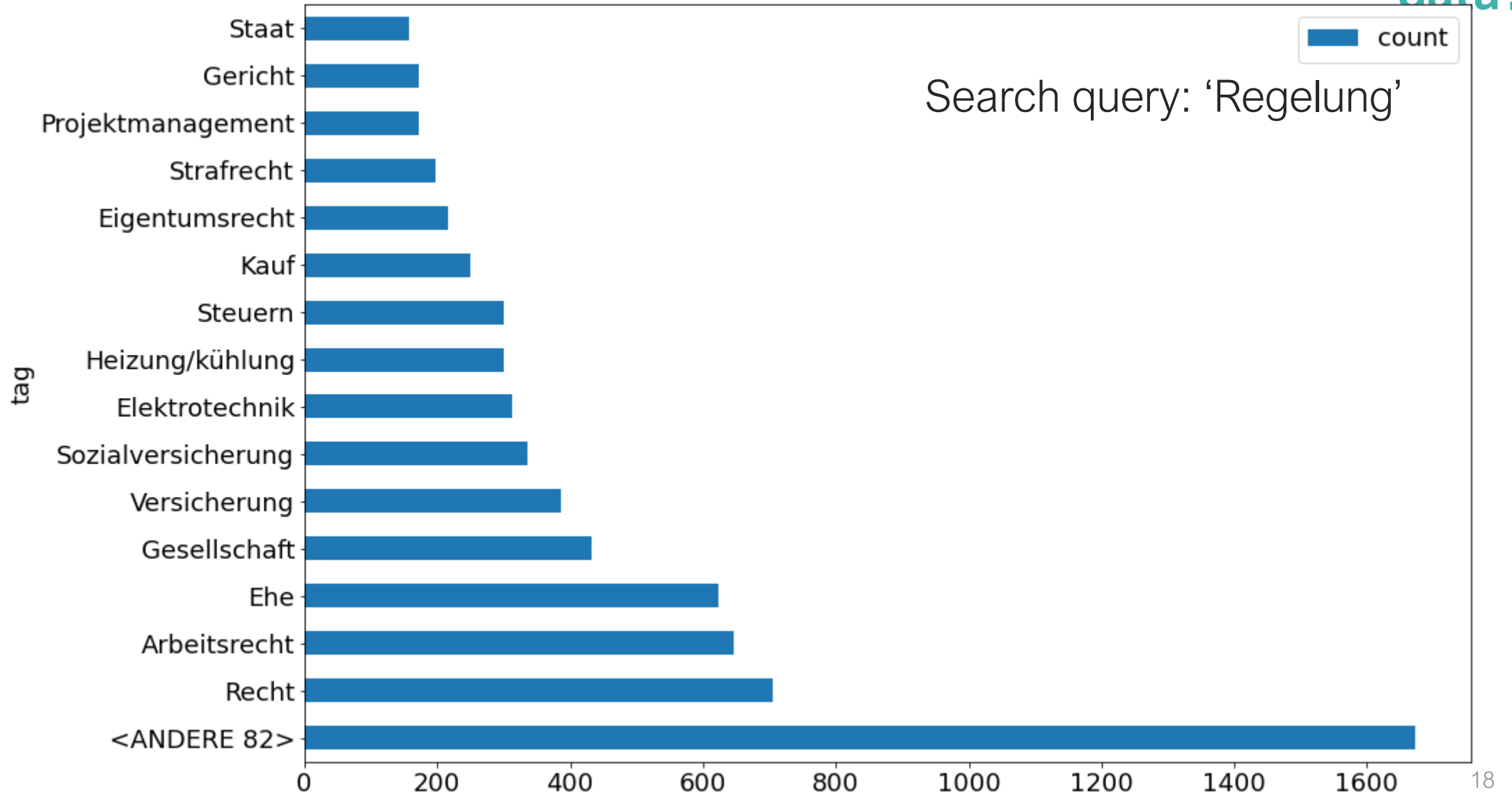
Our proposal

Provide automatic suggestions to narrow down search

Search clarification

Full: 688 pages
Recht: 71 pages
Ehe: 63 pages

Similar documents

Document vectors is based on topic distribution

Applications

- Search improvement

- Search on specialized data (where Google and Co. cannot be used)

- Exploratory search

- Ontology engineering

- 'Visual contents' of a book

- Visual topic model

Possible data for search and navigation:

- Books/journals (publishers)

- Internal document collections (research, corporate, …)

- Emails

- Legal data (laws, regulations, contracts, standards)

Our demo: https://edubase.silkdata.ai/

Thank you for attention!

visit: silkdata.ai   mail to: hello@silkdata.ai

www.silkdata.ai
hello@silkdata.ai
+49 (2151) 387 3531

Headquarters:
Silk AI lines LLC

Hikaly 3-7
220005 Minsk
Belarus

Partner in EU:
SilkCode GmbH

Luisenstraße 62
D-47799 Krefeld
Germany