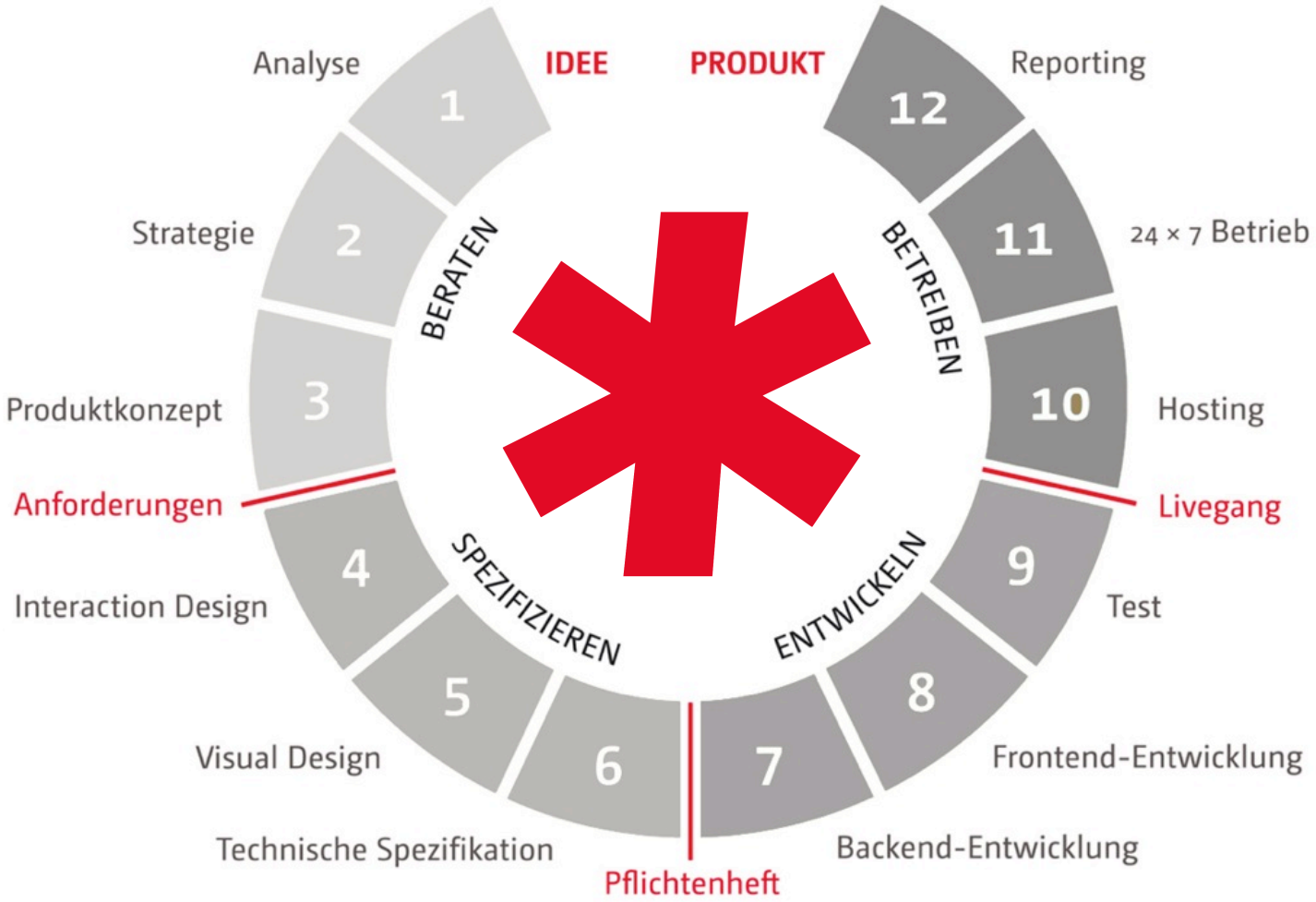
neofonie*

Resources! Resources! Resources!

Heiko Ehrig (Head of Research)

Berlin, 1998, 1st german search engine, 180 pl, 2 companys

What we offer



BMW
BAUER
WAZ MEDIEN GRUPPE
WORT & BILD VERLAG
PMG · Presse-Monitor
KGSt
T
MINI
HEROLD.at
SCOUT 24
e·plus
O2
Bundesverwaltungsamt
TOMORROW FOCUS
V.Ö. TRAVEL
axel springer
BBC
ebay
Bundesverband Deutscher Stiftungen
gavago
VERLAGSGESELLSCHAFT MADSACK
AO Foundation
Ringier
wall
kalaydo.de
das regionale Findernet
DEUTSCHE BUNDESBANK EUROSYSTEM
myToys
Springer
n-tv
nachrichten.de
Wir sind für Ihre Seite!
SIEMENS
FRIEND SCOUT 24
mobile.de
WHITE WALL
Berlin.de
JOB SCOUT 24
Frankfurter Allgemeine
ZEITUNG FÜR DEUTSCHLAND
KLICK
OLIMAR
Besonders reisen
Hubert Burda Media
HSE24
ICH SEH SHOPPEN
BERLINER PHILHARMONIKER
Die Bundesregierung
Creditreform
G+J
Deutsche Post
nexnet
billing, customer services & client management
Audi
YAPITAL
PARSHIP.de
Die Online-Partneragentur
ifolor.
FOTOS FÜRS LEBEN
Cornelsen
1&1
M. DuMont Schauberg
OLYMPUS
HAMBURG Tourismus

Research Department

- 12 Computer Scientists, Linguists, Mathematicians
- Text Mining and Analytics, Search
  - Text Classification
  - Named Entities and Concept Tagging
  - Topic Detection and Tracking
  - Sentiment Analysis (Customer‘s Voice)
- Data Analytics & Consulting
- Individual Projects

txt werk - a Textmining API

- Works On German Texts
  - Department Classification
  - Keyword Detection
  - Dates Detection
  - Entity Detection (person, location, organisation)
  - Concepts with Links to Freebase
  - Named Entities with Links to Freebase
  - Quotes
- Get Your API Key : http://bit.ly/txtwerk

Resources! Resources! Resources!

- German Resources are rare!
- Example Named Entity Linking
  - We did not find a Gold Standard
  - Manual Labeling
- ERD Challenge 2014 (SIGIR'14 workshop)
  - Googlers manually reannotated some hundred texts from ClueWeb (data set not public)

Roadmap

- More Entity Types (companies, products, brands)
- Individual Customer Lexica
- Sentiment Detection
- English and more languages

Wishes to the community

* Share your resources!
* Corporate-friendly licensing!
* If you leave the academia, share your resources!
* Lobby for resources @EC!
* Lobby for maintaining resources servers (like meta-share, datahub.io)
* Don‘t forget the Non-English Speaking World!