EMILiA

Entwicklung einer Software für die Archivierung und Nutzbarmachung von E-Mails

Nico Beyer, Felix Gericke, Alexander Hinze-Hüttl

EMILiA

IBB Business Team

Freie Universität Berlin

ARCHIV DER MAX-PLANCK-GESELLSCHAFT

Themen

1 Einleitung | Nico Beyer

2 Übernahme | Felix Gericke

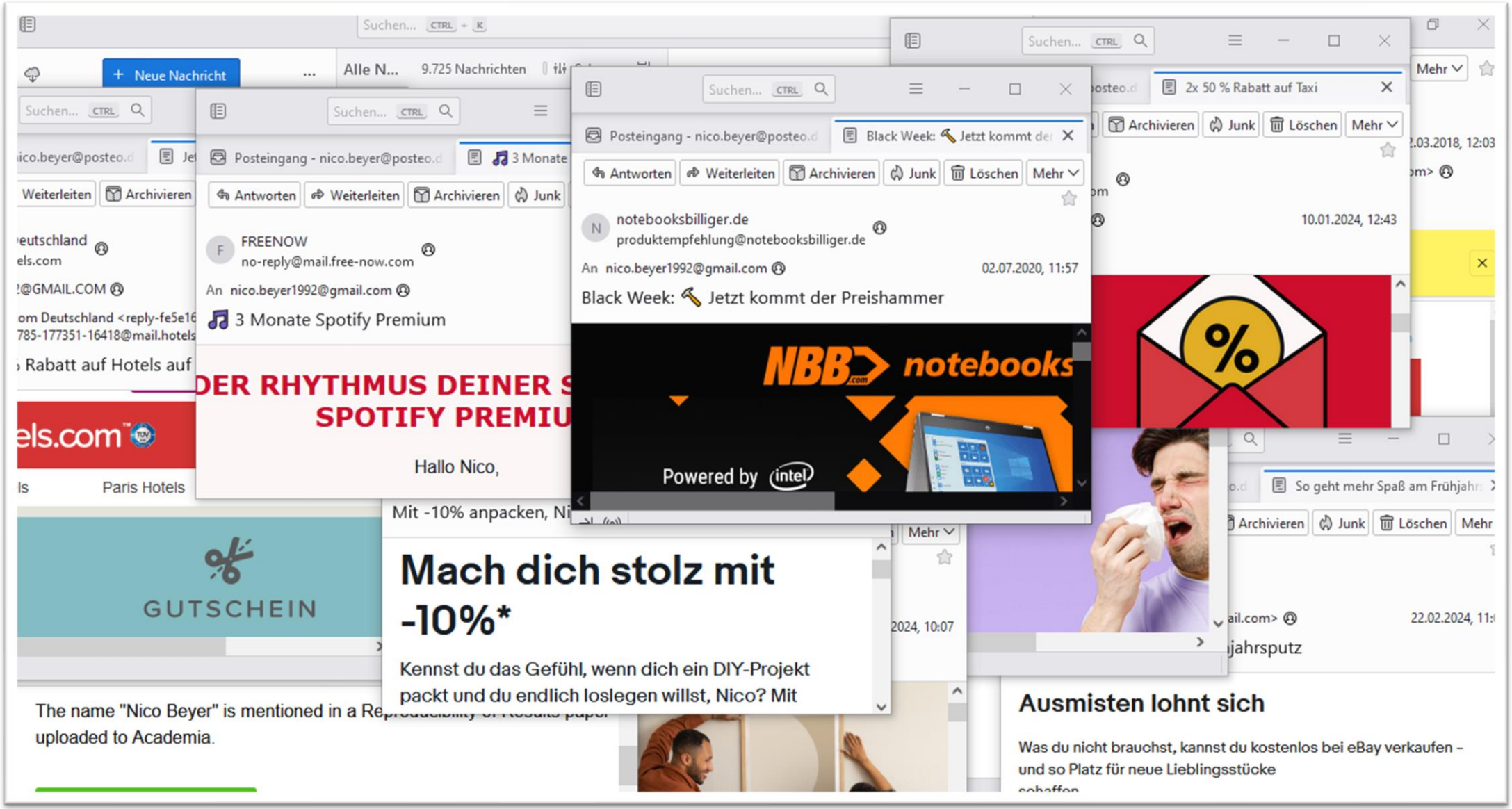
3 Personenbezogene Daten | Alexander Hinze-Hüttl

4 Ausblick | Nico Beyer

5 Zeit für Fragen

EMILiA

1 Einleitung


Neue Nachricht
Alle N...
9.725 Nachrichten
Posteingang - nico.beyer@posteo.d
Black Week: Jetzt kommt der
Antworten
Weiterleiten
Archivieren
Junk
Löschen
Mehr
FREENOW
no-reply@mail.free-now.com
An nico.beyer1992@gmail.com
3 Monate Spotify Premium
DER RHYTHMUS DEINER S
SPOTIFY PREMIU
Hallo Nico,
notebooksbilliger.de
produktempfehlung@notebooksbilliger.de
An nico.beyer1992@gmail.com
02.07.2020, 11:57
Black Week: Jetzt kommt der Preishammer
NBB notebooks
Powered by intel
2x 50 % Rabatt auf Taxi
10.01.2024, 12:43
Paris Hotels
GUTSCHEIN
Mit -10% anpacken, Ni
Mach dich stolz mit
-10%*
Kennst du das Gefühl, wenn dich ein DIY-Projekt
packt und du endlich loslegen willst, Nico? Mit
So geht mehr Spaß am Frühjahr
22.02.2024, 11:
jahrsputz
Ausmisten lohnt sich
Was du nicht brauchst, kannst du kostenlos bei eBay verkaufen –
und so Platz für neue Lieblingsstücke
The name "Nico Beyer" is mentioned in a Re
uploaded to Academia.

EMILiA

Informationsflut

Herausforderungen

Spam, Werbung, E-Mails ohne Informationswert

Händische Bewertung und Erschließung ist aufgrund von riesigen Datenmengen nicht zu bewerkstelligen

Schadhafte Anhänge

E-Mails können Viren enthalten

Personenbezogene Daten

Viele E-Mails müssen anonymisiert werden

EMILiA

Herausforderungen

Schwacher E-Mail-Standard

Eingehende Metadaten können sehr unterschiedlich aussehen

Anomalien und Defekte

Unterschiedliche E-Mail-Clients, proprietäre Dateiformate und Alter der Daten erschweren Verarbeitung

Verschlüsselung

Einige E-Mails sind verschlüsselt

EMILiA

Die Nadel im Heuhaufen

EMILiA

Automatisierung als Chance

- Fortschritte im Bereich der Automatisierung können dabei helfen, Prozesse zu vereinfachen

- Wichtige Entscheidungen sollen aber nach wie vor von Archivfachkräften getroffen werden

Projektrahmen

Projektrahmen

- 2015: Beginn der Konzeption im Archiv der MPG in Kooperation mit dem Fachbereich Informatik der FU Berlin mit
- Entwicklung eines Prototyps
- 2024: Förderprogramm „ProValid“ der Investitionsbank Berlin

EMILiA

Aktueller Stand

Archivischer Lebenszyklus

Erfassung, Management, Indizierung, Limitierung, intelligente Analyse

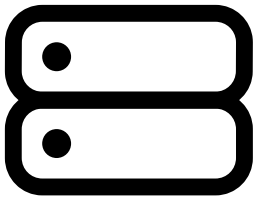
Abgabe an das Archiv

Übernahme

- Import
- Virenprüfung
- Formaterkennung
- Authentizität
- Integrität
- Erstellung von SIPs

Bewertung & Erschließung

- Erkennung von Spam und Dubletten
- Erkennung von Themen, Personen und Orten
- Identifikation personenbezogener Daten



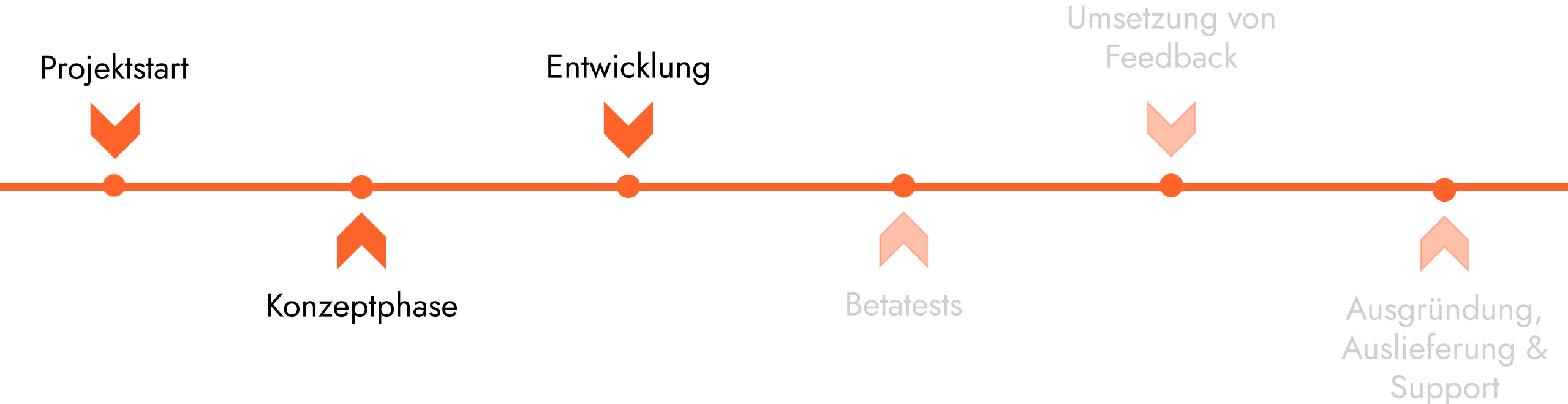
Übergabe an digitales Langzeitarchiv

Nutzung

- Recherchedatenbank
- Anonymisierung
- Erstellung von DIPs
- Darstellung

EMILiA

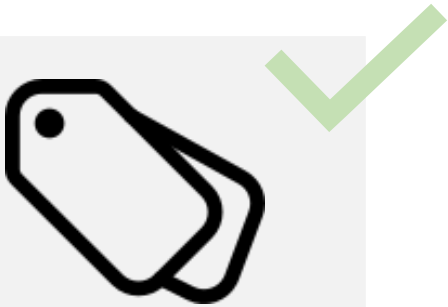
Meilensteine



EMILiA

Entwicklungsstand

Übernahme



Bewertung &
Erschließung

Nutzung

Live-Demonstrationen

2

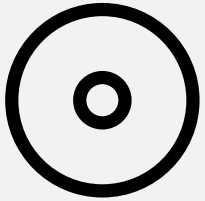
Übernahme

Optionen für die Übernahme

1

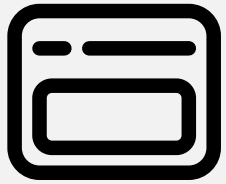
Abgabe über Mailserver

2



Physische Abgabe

3



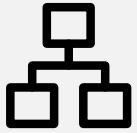
Abgabetool

Übernahme-Workflow

Inhalt der SIPs

BagIt-Format

Angaben zum SIP (.xml)



Strukturdatei (.xml)

PREMIS-Metadaten (.xml)

E-Mails (.txt) und Anhänge (Original)

Prüfsummen (.txt)

EMILiA

3

Personenbezogene Daten

Potentiale

- Anreicherung einer Recherchedatenbank
- Anonymisierung durch Schwärzung

Herausforderung

- Informationen können in unterschiedlicher Komplexität auftreten
- Einige Informationen wirken erst in Kombination personenbezogen
- Kontext bestimmt die Brisanz der Informationen

EMILiA

Herausforderung

Simple Strukturen:

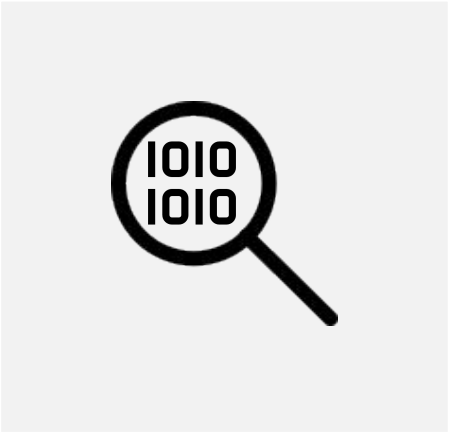
- E-Mail-Adressen
- Telefonnummern
- Kontonummern
- IP-Adressen
- KFZ-Kennzeichen

Komplexe Strukturen:

- Namen
- Adressen/Ortsangaben
- Firmen/Organisationen

EMILiA

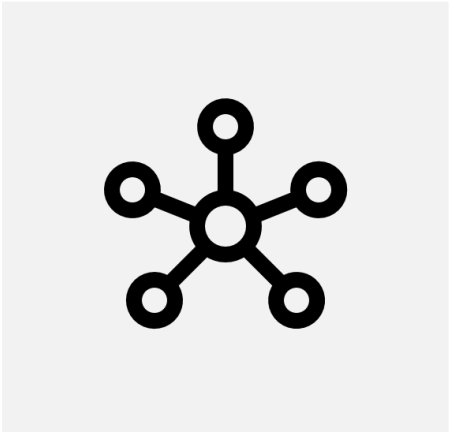
Werkzeuge



Pattern-Matching:

erkennt simple Strukturen anhand von Regeln.

leichtgewichtig



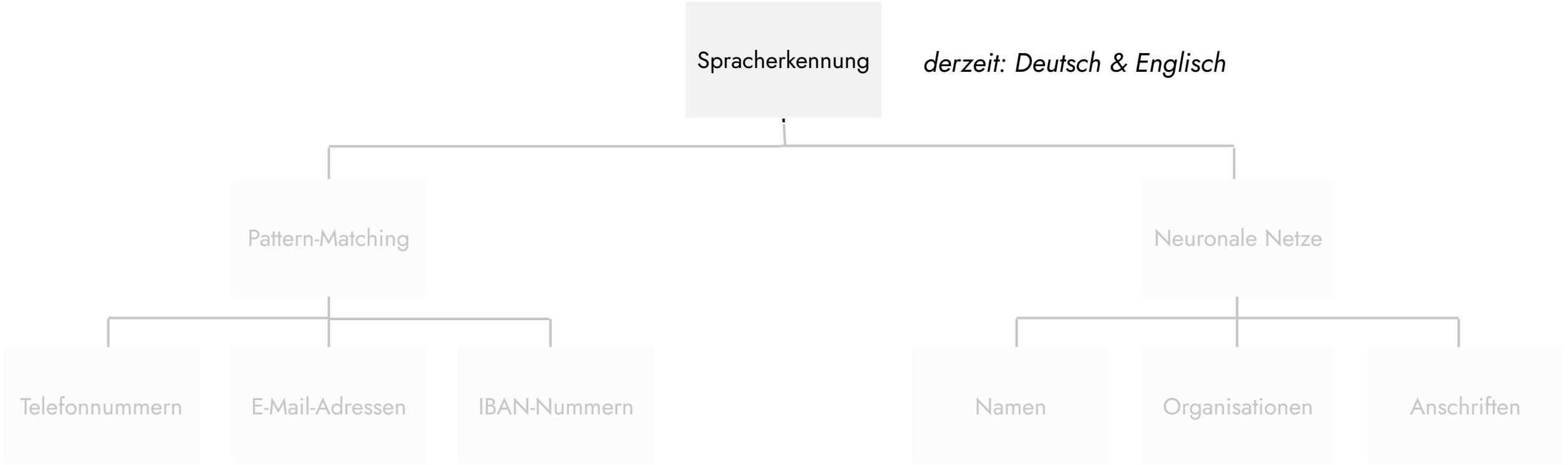
Neuronale Netze:

Klassifiziert Wörter anhand des Kontextes.

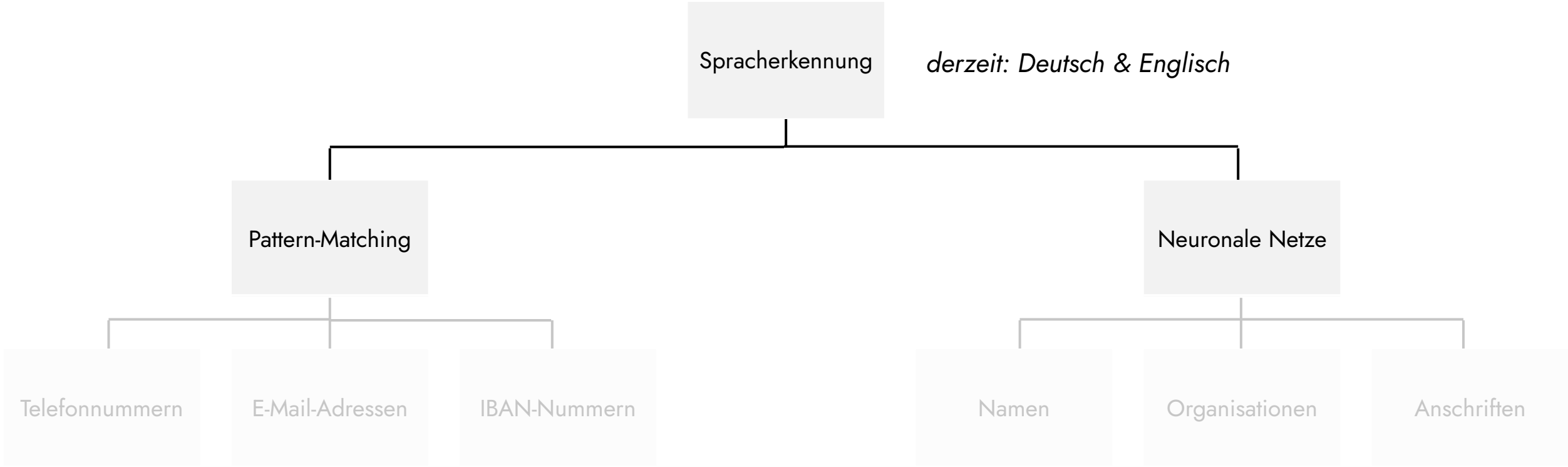
rechenintensiv

EMILiA

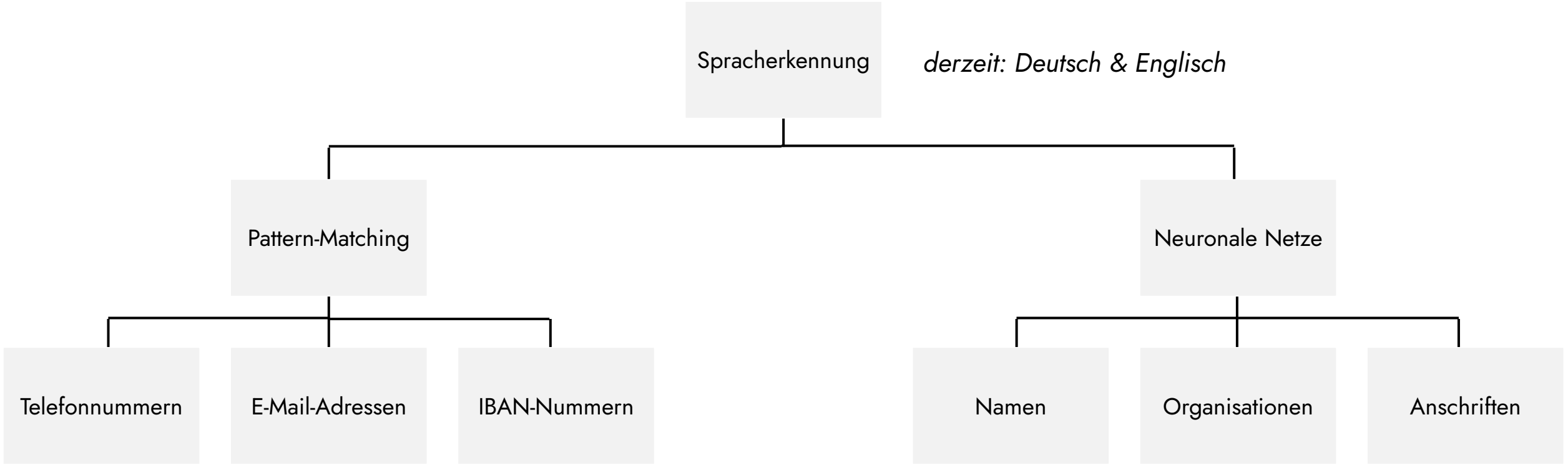
Aktuelle Pipeline



Aktuelle Pipeline



Aktuelle Pipeline



Live-Demonstration

Frau Anna Schmidt, Humboldtstraße 123, 12345 Berlin.

Für die Zahlungen bitten wir Sie, die Rechnungsbeträge auf folgende IBAN-Konten zu überweisen:

IBAN DE59290501010001149590 (Kontoinhaber: Firma XYZ GmbH)

IBAN AT505001051761522741 53 (Kontoinhaber: Herr Frank Mayer)

Für Rückfragen können Sie mich unter meiner Handynummer +49 15115349572 erreichen.

Für nicht dringende Angelegenheiten stehe ich Ihnen auch gerne unter der Festnetznummer 030 34567890 zur Verfügung.

Bitte senden Sie alle Dokumente an meine E-Mail-Adresse anna.schmidt@xyz.com.

TOOL STARTEN

EMILiA

Live-Demonstration

Probieren Sie es gerne aus!

http://164.92.130.147:8501/

! Obwohl wir keine Informationen speichern, empfehlen wir Ihnen, keine sensiblen Daten für die Tests zu verwenden.

EMILiA

4

Ausblick

Ausblick

- Wir arbeiten aktiv an der Implementierung weiterer Funktionen. Um auf dem Laufenden zu bleiben, können Sie unseren Newsletter abonnieren und weitere Webinare besuchen.

Zeit für Fragen

5

Vielen Dank für Ihre Aufmerksamkeit!

Fragen und Vorschläge sind jederzeit willkommen.

E-Mail: info@emilia-archiv.de
Telefon: +49 30 841 337 15

Archiv der Max-Planck-Gesellschaft
EMILiA-Projekt
Boltzmannstraße 14
14195 Berlin-Dahlem

www.emilia-archiv.de

EMILiA