

Alexandria

Die kollaborative Wissensmaschine

Florian Kuhlmann

neofonie*

Gefördert durch:



Bundesministerium
für Wirtschaft
und Technologie

aufgrund eines Beschlusses
des Deutschen Bundestages

Zur Person



- Dipl. Wirt.-Inf. **Florian Kuhlmann**
- Senior Project Manager F&E **Neofonie**
 - Neofonie: Berliner KMU mit 160 Mitarbeitern
 - Davon ca. 20 Mitarbeiter in F&E
 - Hauptkunden: Verlage und Marktplätze
- Geschäftsführender Gesellschafter **Leverton GmbH**

Alexandria - Überblick

- **THESEUS** Use Case
- Schwerpunkt **Social Media + Semantische Technologien**
- Ende 2011 abgeschlossen
- Showcase unter <http://alexandria.neofonie.de>



Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Alexandria – die wichtigsten Bausteine

- **Wissensbasis** (Ontologie)
- **Text-Analyse**
- **Question Answering**
- **Portal** „Alexandria“



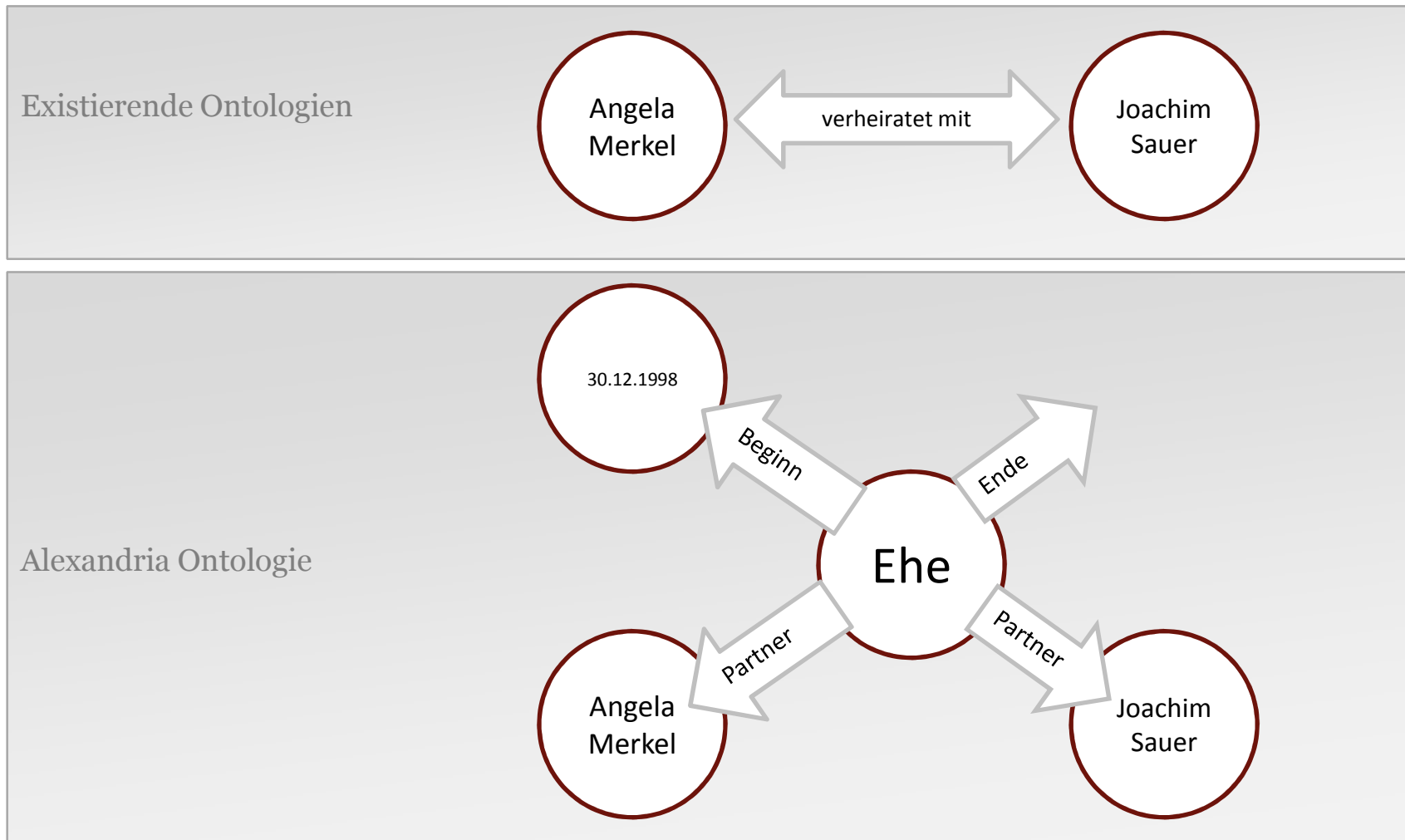


Die Wissensbasis

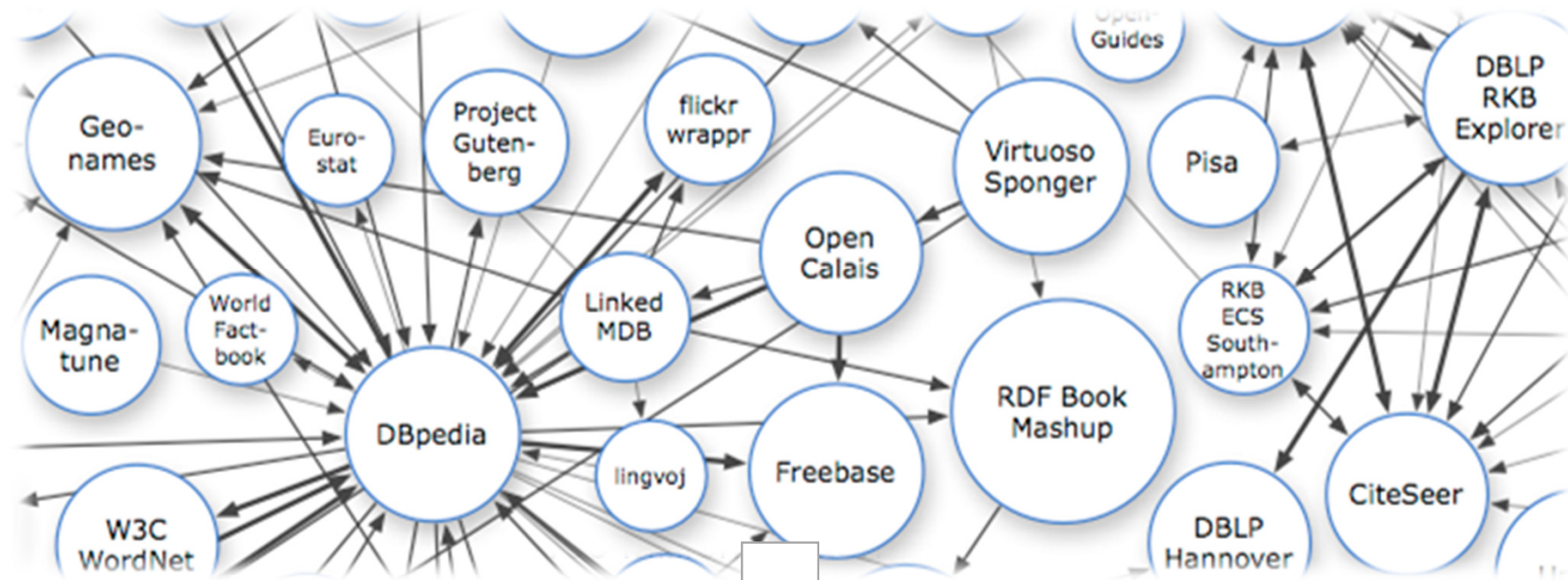
Die Wissensbasis: Ein Netzwerk von Entitäten



Komplexe Beziehungen



Datenquelle: Linked Open Data



Wissensbasis

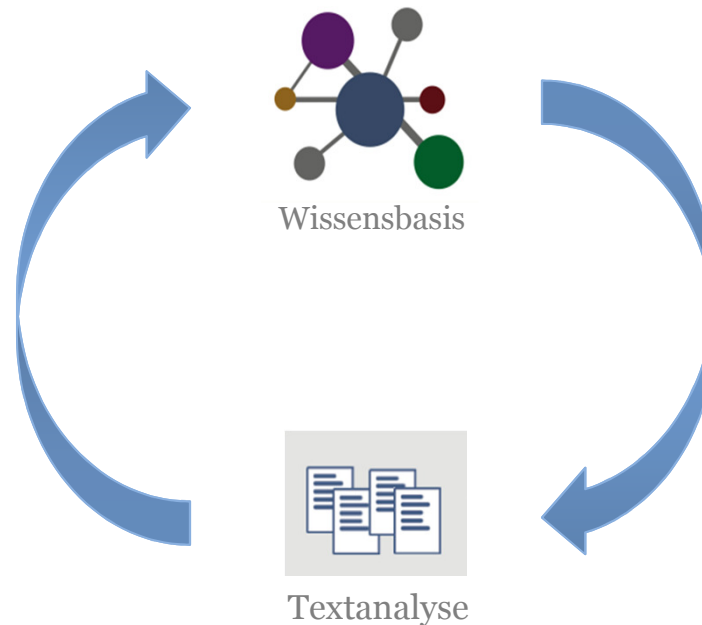
Datenquelle: Community



Wissensbasis

Grundlage für Textanalyse

- Alexandria **Wissensbasis**
- Wissen wird genutzt, um Analyse zu verbessern
- Analyse verbessert Wissensbasis



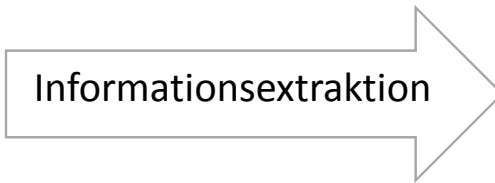


Textanalyse

Datenquelle: Textanalyse



Textdokumente

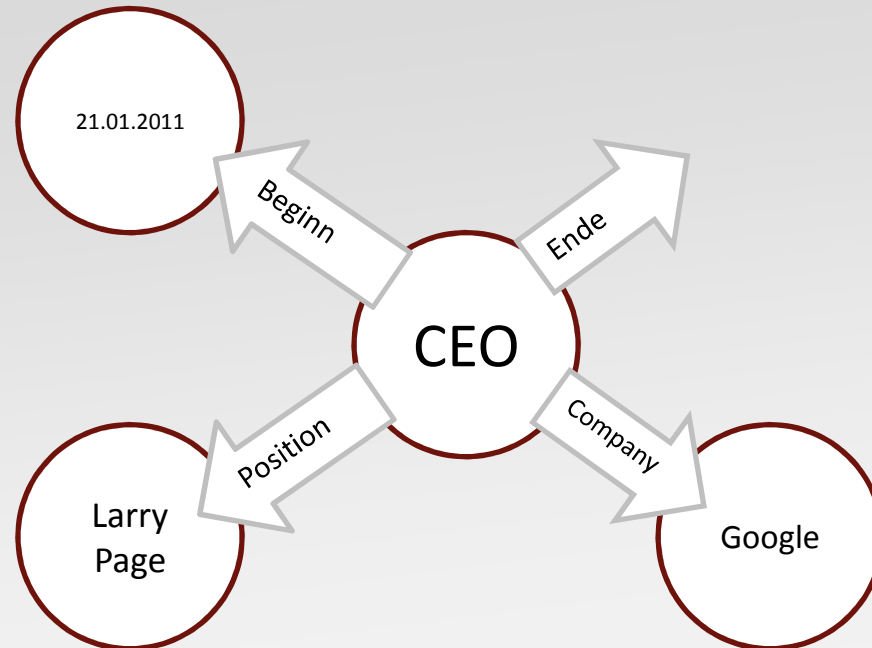


Informationsextraktion



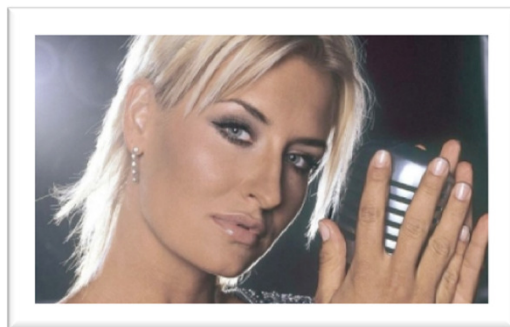
Wissensbas
is

*„Larry Page ist
neuer CEO bei
Google“*



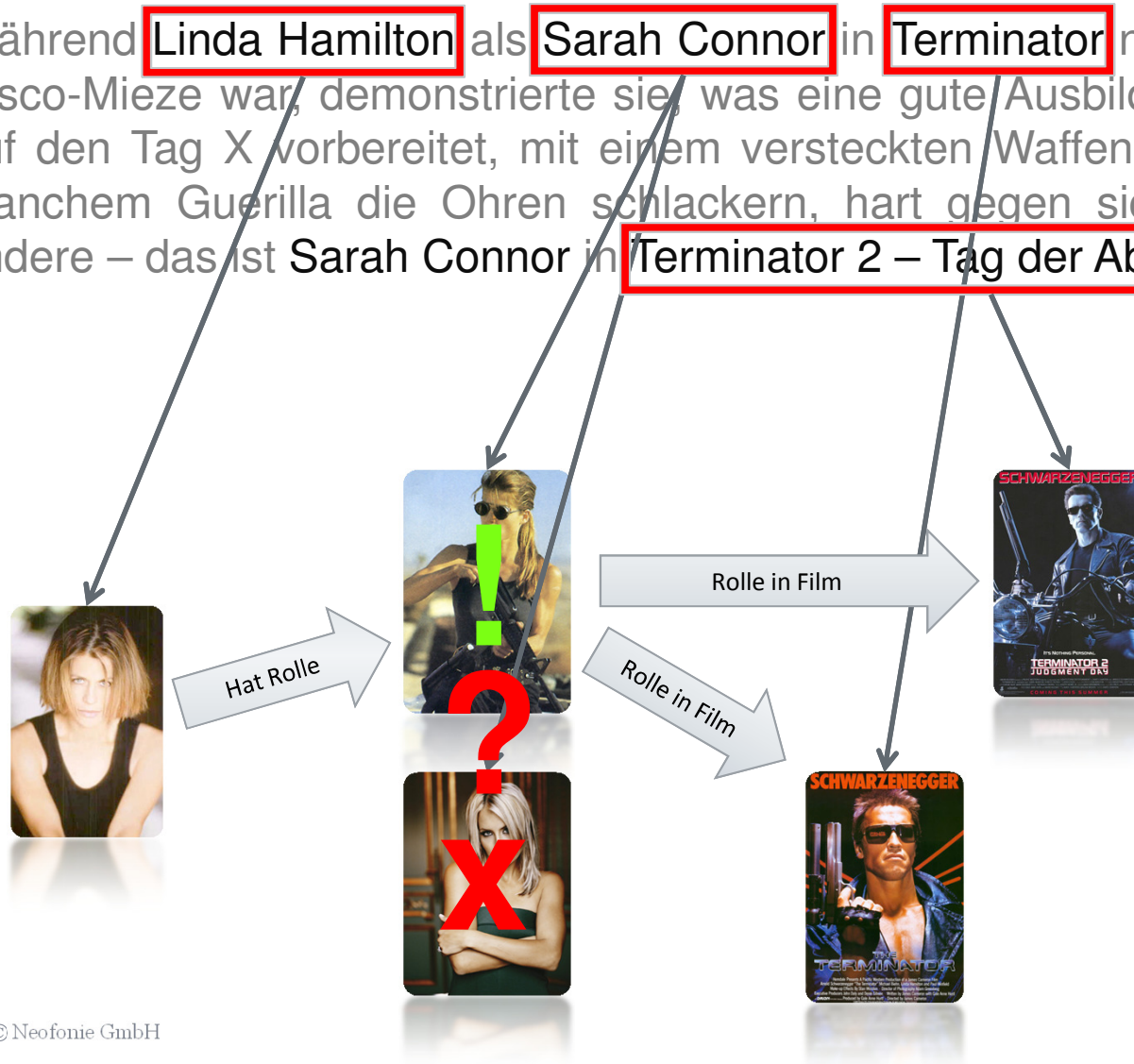
Wie hilft die Wissensbasis bei Textanalyse?

- **Erkannte Namen sind nicht eindeutig**
- Alexandria kennt Namen „*Peter Müller*“ in 18 Ausprägungen
- Prominenteres Beispiel:
 - Sarah Connor (Sängerin)
 - Sarah Connor (fiktive Figur im Terminator-Universum)



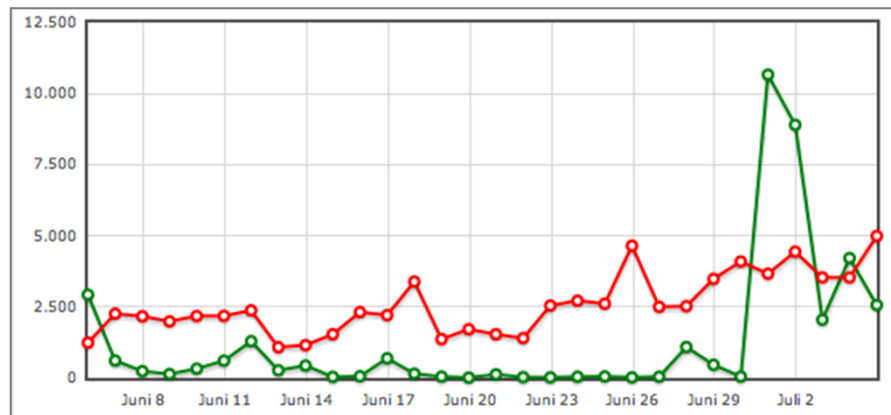
Identity Resolution - Beispiel

Während **Linda Hamilton** als **Sarah Connor** in **Terminator** noch eine 80er Jahre Disco-Mieze war, demonstrierte sie, was eine gute Ausbildung wert ist. Immer auf den Tag X vorbereitet, mit einem versteckten Waffenarsenal, bei dem so manchem Guerilla die Ohren schlackern, hart gegen sich und auch gegen andere – das ist Sarah Connor in **Terminator 2 – Tag der Abrechnung**.



Aufbauende Werkzeuge

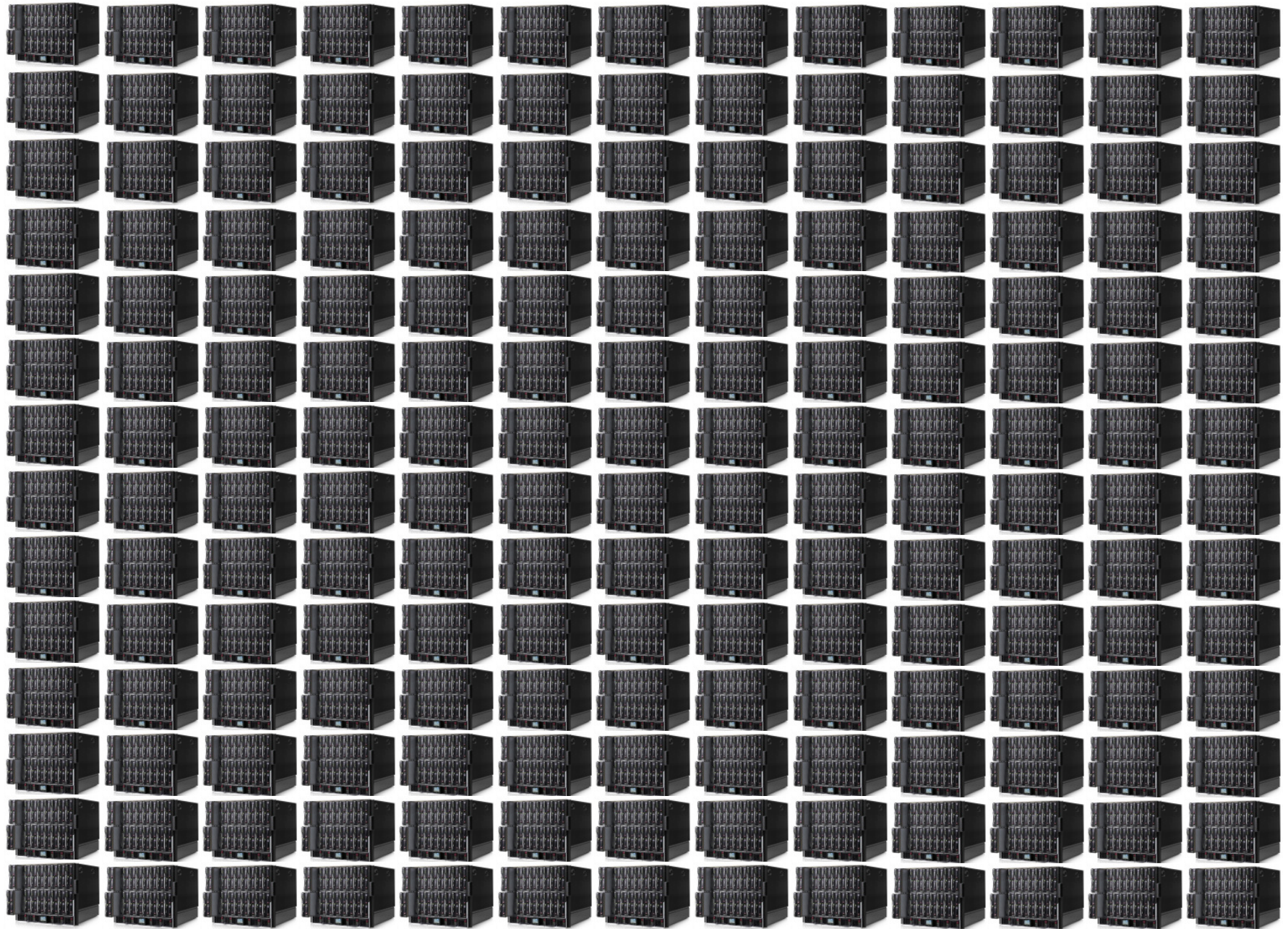
- Erkennung von
 - **Beziehungen zwischen Entitäten**
 - **direkter und indirekter Rede**
 - **Meinungen**
 - **Ereignisse** (inklusive Zeit und Ort)
- **Statistiken**



Skalierungs-Problematik

- Bisher nur Analyse **überschaubarer Datenmengen**
- Durchsatz: 200.000 Dokumente / Tag / Server
 - Ca. **3.000 Rechner** notwendig, um in 2 Wochen alle Dokumente des dt. Internets zu prozessieren





Skalierungs-Problematik

- Alle Rechner müssen konfiguriert werden
 - Behandlung von Ausfällen aufwändig
 - Kosten für KMUs kaum tragbar
- Viele **Geschäftsmodelle** auf Basis **tiefer Analyse von Big Data** nicht möglich

MIA



- Ein **cloud-basierter Marktplatz für Informationen und Analysen** auf dem **deutschsprachigem Web**
- Forschungsschwerpunkt:
 - Semantische Analysen auf „Big Data“
 - Abfragesprache zur Online-Analyse der Daten
 - Faire Abrechnungsmodelle
- Laufzeit: Anfang 2012 – Ende 2014
- Leitung: **TU Berlin / DIMA**
- Weitere Partner: Fraunhofer FIRST, Temis, VICO Research, Empulse

Erste Experimente



- **Verteiltes System (Hadoop)** anstatt viele Einzel-Systeme
- **Flache Strukturen** anstatt tiefe Graphen
- **Erste Experimente** mit einer verteilten Infrastruktur:
 - Ca. 300 Rechner anstatt 3000 Rechner, um das gesamte dt. Internet in 2 Wochen zu prozessieren
 - Automatische Skalierung + Failover-Behandlung
 - Aber: Bisher nur 70 % der Präzision

Fazit und Ausblick

- Semantische Technologien aus Alexandria heute im Einsatz
- Nur für **überschaubare Datenmengen (Private Content)**
- MIA adressiert **Skalierungsproblematik (Big Data)**
- ... und ermöglicht damit **neue Geschäftsmodelle**
- **Qualität darf nicht vernachlässigt werden**
 - Wichtig für eine breite Markt-Akzeptanz
- **Qualitätssteigerung durch**
 - Weitere Evolution von **Technologien zur Textanalyse (NLP)**
 - **skalierbare Wissensbasis**

neofonie*

Florian Kuhlmann
Research Manager

Neofonie GmbH
Robert-Koch-Platz 4
10115 Berlin
Germany

T +49.30 24627 525
F +49.30 24627 120
florian.kuhlmann@neofonie.de
www.neofonie.de