



SIN C

SYSTEMS INTEGRATION.
NETWORK.
CONSULTING.

RIAD	Akronym:	Rechtsinformatische Analyse von Dokumenten
RIAD	Ort:	Hauptstadt von Saudi-Arabien
RIAD	Gegend:	Provinz in Saudi-Arabien
RIAD	Arabisch:	Die Gärten
RIAD	Name:	Familienname z.B. in Ägypten
RIAD	Bauweise:	Traditioneller marokkanischer Atrium-Palast
RIAD	Tippfehler:	RAID = Spezialeinheit der frz. Polizei
...		

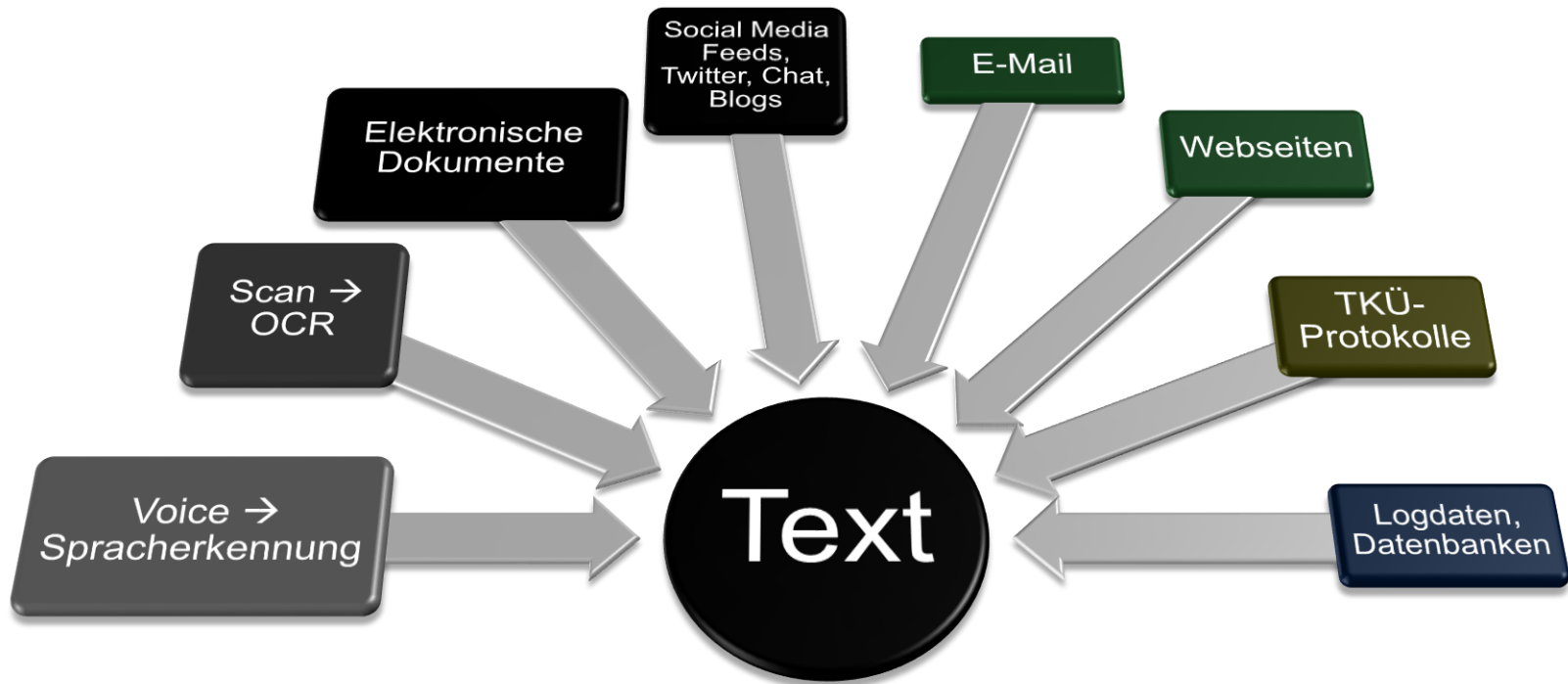
RIAD:

Massendaten und **Textanalyse**
für Entscheider und Ermittler

22. Europäischer Polizeikongress - Berlin
19.02.2019 – Till Elborg

Text als wichtigste Repräsentanz für Informationen in Massendaten

Makrotrend → rapide Zunahme der Datenmengen



Wo spielt Textanalyse bei der Verarbeitung von Massendaten im polizeilichen Umfeld eine Rolle?

**Prävention und
Kriminologie**

Strafverfolgung

z.B. Sammlung
von Informationen,
Gewinnung von
Erkenntnissen

z.B. Ermittlungen,
Beschlagnahme,
IT-Forensik

Modus der Textanalyse bei der Verarbeitung von Massendaten

Maschinelle
Dunkelverarbeitung
(**automatisiert**)

Intellektuelle
Durchdringung
(**interaktiv**)

z.B. Newsfeed, Twitter,
Live-Protokolldaten

z.B. **Dokumente**
umfangreicher
elektronischer Akten,
gescannte Ordnerregale,
beschlagnahmter Mail-
Server

Wann beginnt die „Big Data“ Komplikation bei Text?

„Big Data“ Komplikation → Systeme und herkömmliche Anwendungsarchitekturen stoßen an Grenzen in Bezug auf schnellen Speicherzugriff und insbesondere Antwortzeiten

Selten bei automatisierter Dunkelverarbeitung

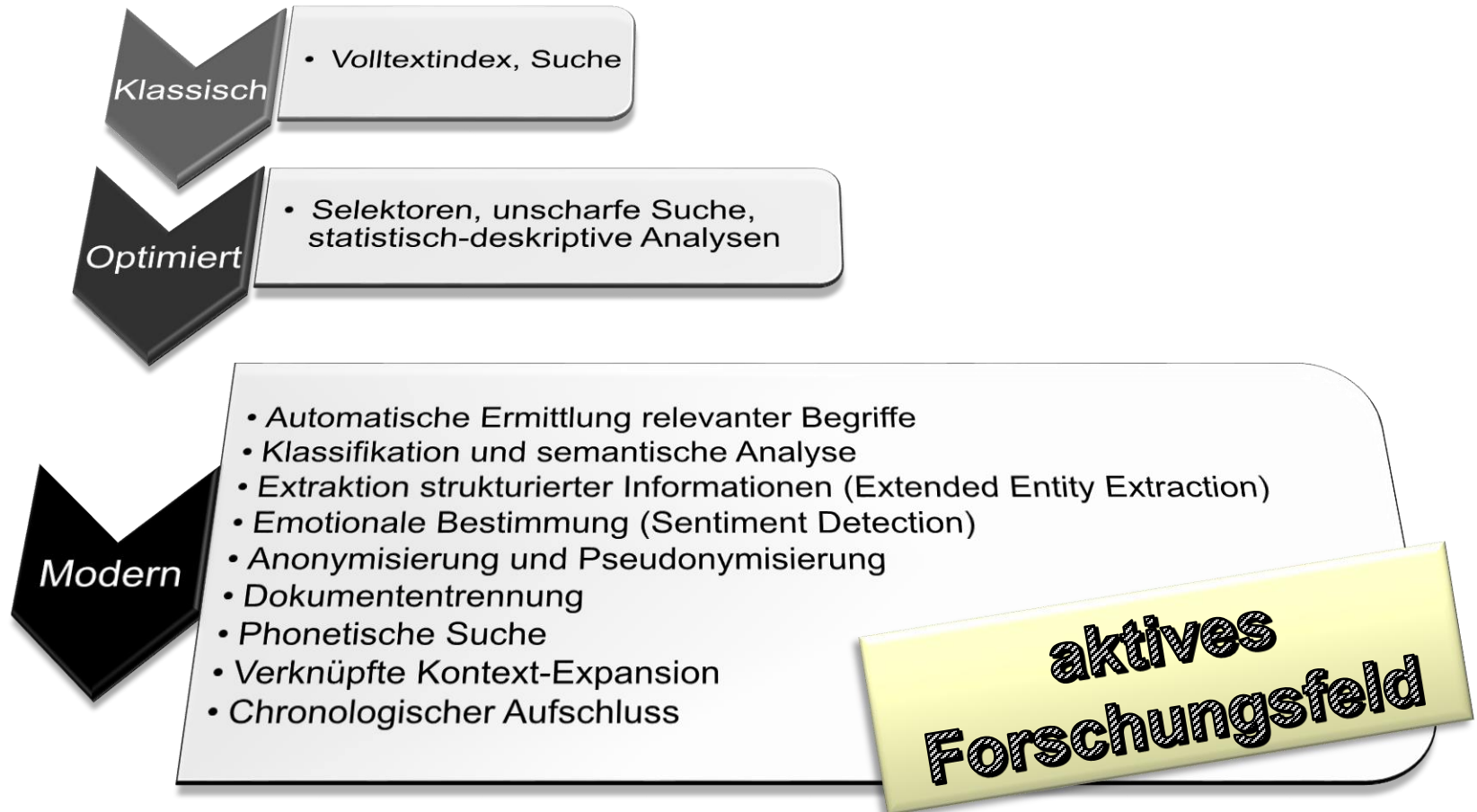
Verfahren oft schon seit Jahrzehnten etabliert, im Vordergrund steht Verarbeitungseffizienz; gute Skalierbarkeit mit klassischen Methoden

Bei interaktiver Unterstützung intellektueller Durchdringung beginnen die Herausforderungen aus heutiger Sicht schon ab wenigen Tausend Seiten

Fundstellen von Worten und Entitäten, dynamischen Indexdaten und Strukturinformationen vervielfachen den Bedarf an Speicher, auf den schnell zugegriffen werden muss

Exponentieller Zuwachs der Datenmenge Text in Umfangsverfahren

Evolution von Ansätzen und Methoden der Textanalyse



Künstliche Intelligenz in der Textanalyse

Statistisch-derivative Algorithmen

- Klassifikation, semantische Analyse
- Erkennung von Extremismus / Drohungen

Algorithmen zum Clustering von Text

- Dokumententrennung, Gruppierung zusammengehöriger Textabschnitte
- Ähnlichkeitssuche

Neuronale Netzwerke

- Neueste OCR-Generation
- Erkennung von Namen
- Erkennung komplexer Verweisketten
- Maschinelle Inhaltszusammenfassung

Erfahrungen aus der Praxis der Textanalyse „Lessons learned“ – kleine Auswahl

„General-purpose“-Algorithmen sind selten unmittelbar erfolgreich

→ Wenige Ausnahmen, z.B. Erkennung der Landessprache

Kombination mit Fachwissen (Domain Knowledge), sprachlichem und lexikalischem Wissen notwendig für praktischen Nutzen

→ Erfordernis einer abwägenden Vermittlungsinstanz (Super-Korrektiv)

Moderater Aufwand bei Portierung in andere Fach-Domäne; hoher Aufwand bei Portierung in andere Sprachräume

→ Beispiel: Phonetische Suche

Steigerung der Präzision durch den ergänzenden Einsatz rekurrenter neuronaler Netzwerke in hybriden Algorithmen immer häufiger erfolgreich

→ Beispiel: Spezialisierte OCR zur Erkennung von Aktenzeichen in gescannten Dokumenten

Zunehmende Bedeutung von Anonymisierung / Pseudonymisierung

Gestaltung von Zielen bei der Textanalyse im polizeilichen Umfeld

Sorgfältige Integration von Textanalyse-Funktionen in fachliche Prozesse / Vermeidung disruptiver Kontext-Wechsel

- Beispiel: Innovatives Plattform-Projekt für IT-Forensik
- Beispiel: Juristische Textanalyse in der Justiz mit Integration in Posteingang und E-Akte

Smarte algorithmische Assistenz statt automatisierter Suggestion

- Die Maschine dient dem Menschen, nicht umgekehrt

Ergonomische Nutzarmachung der Ergebnisse der Textanalyse

- Höchste Anforderungen bei interaktiver Durchdringung großer Datenvolumina mit Kontext-Expansion (virtualisierte Benutzeroberflächen, leistungsfähige „in memory“-Datenbanktechnologie zwingend)

Präzision und technische Exzellenz

- Entwicklung, Evaluation und Adaption neuer algorithmischer Ansätze am Puls der Forschung
- Förderung inländischer Forschung (starke Sprachabhängigkeit der Textanalyse)



SINC

SYSTEMS INTEGRATION.
NETWORK.
CONSULTING.

Vielen Dank !

SINC GmbH
Rheingastr. 182
65203 Wiesbaden

<http://www.sinc.de>

Telefon: +49 (0) 611 504518 – 00
Fax: +49 (0) 611 504518 – 28

info@sinc.de

Die Übersicht zum Schluss

