

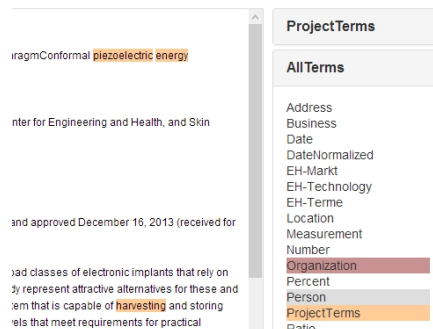
RECHERCHE- UND MONITORING-TOOL ZUR INFORMATIONSEXTRAKTION

EFFIZIENTE FILTERUNG RELEVANTER DATEN UND INFORMATIONEN

Die Suche nach relevanten Informationen in einer Vielzahl von Datenquellen und die Extraktion wichtiger Informationen daraus ist heutzutage eine weit verbreitete Aufgabe in vielen Bereichen eines Unternehmens wie z.B. in der Produkt- oder Vor-Entwicklung.

Diese sich in der Regel wiederholende Aufgabenstellung besteht aus der Suche und dem Zusammentragen von neuen Informationen zu einzelnen Themenfeldern aus unterschiedlichen Quellen.

Die besondere Anforderung hierbei ist, aus einer ständig wachsenden Menge an Informationen möglichst effizient die relevanten Daten und Informationen zu identifizieren. Oft sind solche Informationen über mehrere Dokumente verteilt, was die manuelle Extraktion von Informationen erschwert, da viel Zeit aufgewendet werden muss, um alle Dokumente zu durchsuchen.



Den Ausgangspunkt für die Technologiefrüherkennung bildet der generelle Informationsbedarf:

- + *Wie kann die Menge an zur Verfügung stehenden Informationen über neue Technologien (bspw. Fachdatenbanken, Patente und wissenschaftliche Veröffentlichungen) effizient und schnell bewältigt werden?*
- + *Wie kann mein Unternehmen frühzeitig die relevanten technologischen Möglichkeiten identifizieren?*
- + *Welche neuen Technologien zeichnen sich als Substitutionsgefahr ab oder ermöglichen neue Funktionalitäten für die eigenen Produkte?*
- + *Wie ist deren Leistungsfähigkeit in Bezug auf das eigene Anwendungsgebiet einzuschätzen?*
Wie ist der Reifegrad zu beurteilen?

Technologische Entwicklungen & Trends frühzeitig erkennen

In technologie- und forschungstreibenden Unternehmen ist die Technologiefrüherkennung ein wichtiger Teil des strategischen Innovations- und Technologie-Managements.

Ziel der Technologiefrüherkennung ist, aussichtsreiche Ansätze oder Trends zu ermitteln und deren Entwicklungspotential zu beurteilen.

Technologische Entwicklungen und Trends müssen identifiziert, technologische Möglichkeiten rechtzeitig erkannt und deren Auswirkungen und Bedeutung für das Unternehmen bewertet werden. Dies stellt Unternehmen vor enorme Herausforderungen in Form zeitlicher und personeller Ressourcen.

Semantische Informations- und Analysewerkzeuge

Da eine manuelle Analyse gefundener Dokumente sehr zeitintensiv ist, macht es Sinn, auch für diese Suchprozesse spezialisierte, software-basierte **Recherche-** und auch **Monitoring-Werkzeuge** einzusetzen, die eine schnelle und zielgerichtete Extraktion relevanter Informationen aus digitalen Texten ermöglichen. Hierzu erfolgt eine semantische Anreicherung der Texte (*Semantic Content Enrichment*) um Metadaten wie Personen, Orte, Firmen, spezifische Projekt- und Firmen-Terme und andere Annotationen.

Informationsbedarf / Suchprozesse

Ein Suchprozess nach neuen Informationen geht in der Regel zunächst immer vom eigenen Anwendungsfeld aus. Sinnvoll ist aber auch im Sinne einer Cross-Industry-Innovation Technologien aufzufinden, die in anderen Anwendungsfeldern bereits eingesetzt werden.

Bevor eine Suche konfiguriert werden kann, muss der individuell der **situationspezifische Informationsbedarf** bestimmt werden. Auf dieser Grundlage können die zu analysierenden Informationsquellen ausgesucht sowie die „key words“ für die Suche festgelegt werden. Diese „key words“ können unter Umständen je nach Informationsquelle auch unterschiedlich sein (z.B. bei Suchen in Fachdatenbanken, Presseportalen, Patentdatenbanken innerhalb des gleichen Themenfeldes).

Um relevante Dokumente in online Medien zu finden, die wichtige Inhalte enthalten könnten, beeinflussen zwei Punkte wesentlich die Qualität und Relevanz der zu analysierenden Dokumente:

1. Die verfügbaren **Datenquellen**: Fachdatenbanken, Fachzeitschriften und –Journale, Patentdatenbanken, offene Suchmaschinen (indizierte Dokumente und Websites im Internet), RSS-Feeds, Presseportale.
2. Die Bestimmung und logischen Verknüpfungen der „**key words**“, mit der eine Suche in der gewählten Datenquelle gestartet wird.

Weiterhin gibt es für die Suchprozesse zwei generelle Zielstellungen:

1. Die Identifikation **relevanter Dokumente**:
sind prinzipiell relevante Informationen in Bezug auf meine Suchstrategie in einem Dokument enthalten?
2. Das Auffinden **relevanter Textstellen** in den gefundenen Dokumenten:
welches sind im Dokument die interessantesten, relevantesten Textpassagen in Bezug auf das Themenfeld der Suche (Begriffe, Sätze, Absätze)?



REMARK: Ein Software-Tool zur Recherche, Analyse und Auswertung von digitalen Texten (im Internet)

Eines der Projektergebnisse des vom BMBF geförderten Projektes syncTech ist das Recherche– und Monitoringtool **REMARK**. Ziel bei der Entwicklung dieser Anwendung war die Realisierung eines interaktiven Werkzeuges zur Ausführung von Annotationsprozessen, sowie zur Analyse und Beurteilung von Ergebnissen der Informationsextraktion.

Für die grundlegenden Aufgaben der **Informationsextraktion** wird in REMARK das Software-System GATE (s. u.) eingesetzt, eine freie Open Source Software-Plattform zur Extraktion von Informationen aus Texten.

Eine der Hauptaufgaben der Informationsextraktion ist die Identifikation und Annotierung von Begriffen in Texten (Named-Entity-Recognition). Diese basiert zum einen auf **Begriffs– bzw. Gazetteer-Listen**. Zum anderen gibt es **regelbasierte Formalismen**, die Annotationen aufgrund von Wortarten und –stellungen und bereits in Sätzen vorhandenen Annotationen erzeugen. Dies erfordert eine zuvor durchgeführte linguistische (sprachabhängige) Analyse des Textes (*Tokenizer, SentenceSplitter, Part-of-Speech-Tagger*).

GATE (General Architecture for Text Engineering) ist ein text-technologisches System und Software- Framework zur Informationsextraktion und linguistischen Verarbeitung von Texten. Es wird seit mehr als 15 Jahren an der University of Sheffield (GB) als frei erhältliche Open Source Software entwickelt. Seine Architektur ermöglicht die Entwicklung und den Einsatz aufgabenorientierter Plugins ebenso wie die interaktive Nutzung über ein Graphical User Interface (GUI) und auch das Einbinden in bestehende Software über ein Application Programming Interface (API) .

REMARK — HIGHLIGHTS

1. Angebundene Datenquellen:

Suchmaschinen (**Google / Bing**), Publikations-/Fachdatenbanken (**WTI**), Journale (**JournalTOCs**), Patentdatenbanken (**EPO — European Patent Office, USPTO—United States Patent and Trademark Office**), **RSS-/News-Feeds**, im eigenen **File-System** abgelegte Dateien.

2. Einrichtung von **Projekträumen**: Themenfelder, in denen inhaltlich zusammengefasst Suchen in verschiedenen Datenquellen durchgeführt werden können (z.B. Technologien, Forschungsthemen, Marktanalysen (bezogen auf ein Themenfeld), Konkurrenzanalysen, eigenes Unternehmen).

3. Suche nach relevanten Dokumenten über **key words / Zeitraum**.

4. Verarbeitung von PDF-, Text- und HTML-Dateien.

5. Verwendung von systemseitigen Gazetteer-Listen und Regeln zur Annotierung/Text-Identifikation:

GATE Standard-Annotationen: Personen, Organisationen, Datum, Zahlen, Währungen und andere Named Entities (atomare Elemente im Text / Satz).

6. Suche in den Dokumenten über firmen-/instanz- und projektspezifische **Begriffslisten**:

A) allgemeine **Firmen/Instanz-Terme**, B) themen-spezifische **Projekt-Terme**.

7. **Extraktion / Annotation** von Begriffen und auch Sätzen mit relevanten Informationen.

8. Berechnung der absoluten und relativen (Term frequency) Anzahl von Firmen- und ProjektTermen zur Beurteilung der Relevanz eines Dokumentes.

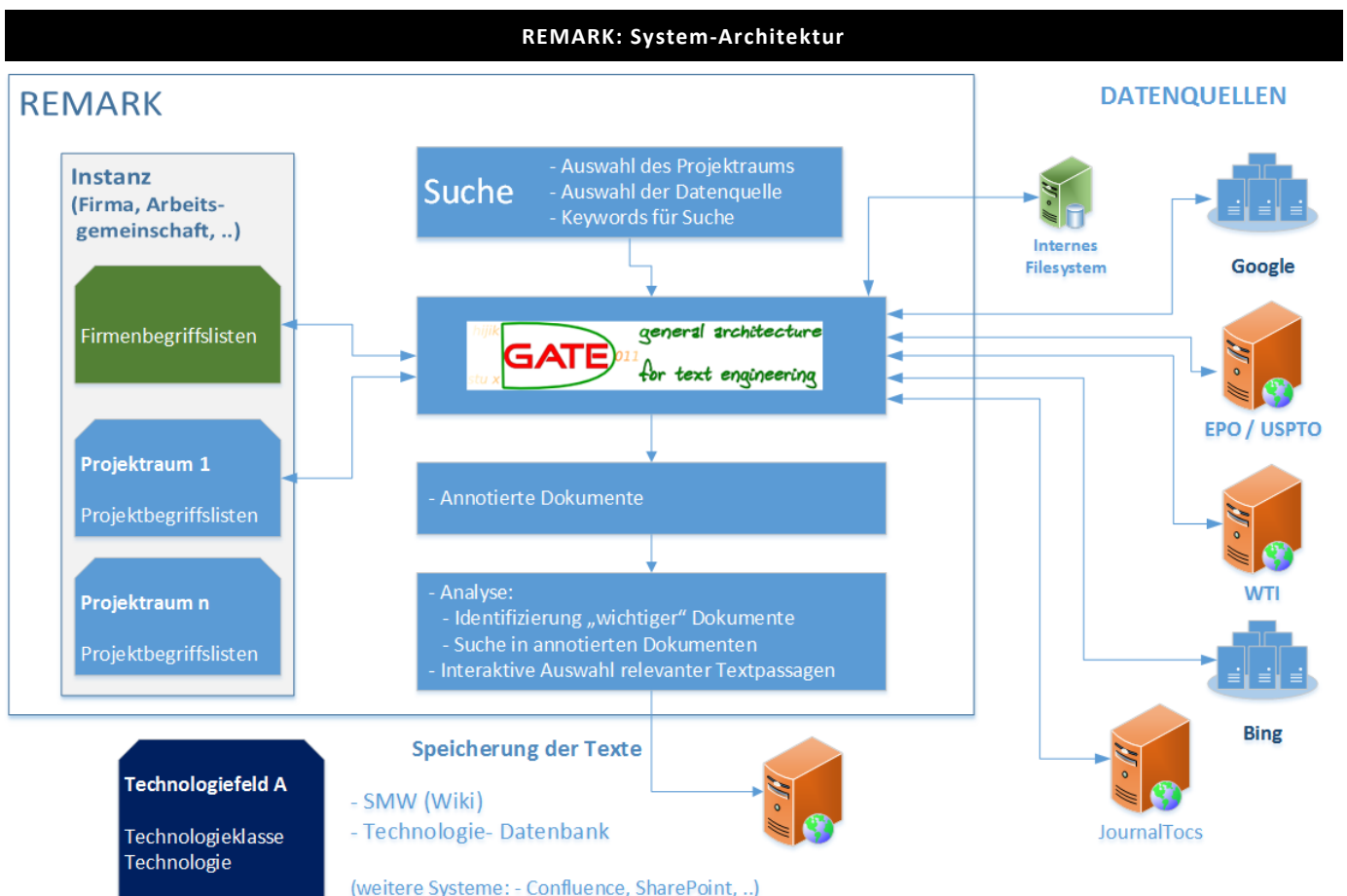
9. Analyse / Auswertung von **Einzeldokumenten** und auch **Dokumenten-Corpora**.

10. **Markierung** relevanter Dokumente (interessante Dokumente, Literaturrecherche).

11. Einrichtung von **ad-hoc-** und **zeitgesteuerten Suchen** (Monitoring).

12. Automatische **Erkennung** von Dokumenten, die bereits im Projektraum annotiert sind (z.B. durch ähnliche Suchanfragen) → keine doppelte Annotierung.

13. Interaktive Übernahme von relevanten Textpassagen in Wissens-Datenbanken.



**Kontaktieren Sie uns für weitere Informationen:
EINS GmbH**

Entwicklung Interaktiver Software
Unterweingartenfeld 6,
D-76135 Karlsruhe

Tel: +49-721-955 3976
Fax: +49-721-53 103 55
info@eins-gmbh.de
www.eins-gmbh.de

In Kooperation mit

Fraunhofer-Institut für Arbeitswirt-
schaft und Organisation (IAO)
Nobelstraße 12
70569 Stuttgart
Herr Markus Korell
Markus.Korell@iao.fraunhofer.de
www.iao.fraunhofer.de

REMARK: Such- und Annotations-Ergebnisse

294821	Key words: Energy harvesting Osmose Document Type: ALL Results: 6	Documents	All docs	Start: 2015/02/12 / 12:20:39 End: 2015/02/12 / 12:20:46	FINISHED	ADHOC
294012	Key words: Energy harvesting Thermoelektrisch Document Type: ALL Results: 6 scheduled for: 12.02.2015 12:17:00	Documents	All docs	Start: 2015/02/12 / 12:17:00 End: 2015/02/12 / 12:17:11	FINISHED	ADHOCTIMER

Bild 1: Übersicht über durchgeführte ad-hoc Suchen

before: match: after:

open	link	Extern Link	interested
▼	Result ID:259865	http://www.degruyter.com/view/j/ehs.2014.1.issue-3-4/ehs-2014-0001/ehs-2014-0001.xml	<input type="checkbox"/>
before	match	after	
J., M. Sanghadasa, and S. Priya. 2014. "Vibration Modeling of Arc-Based Cantilevers for Energy Harvesting Applications." Energy Harvesting and Systems 1:57. Bedekar, V., J. Oliver, and S. Priya. 2010. "	Design and Fabrication of Bimorph Transducer for Optimal Vibration Energy Harvesting.	" IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control 57(7):1513–23. [Web of Science] [CrossRef] Burghardt, F. L., A. C. Waterbury, I. Paprotny, L. M. Miller, P. Minor, R. Send, Q. Xu,	
Y.-D. Kim, et al. 2014. "Magneto-Mechano-Electric (MME) Energy Harvesting Properties of Piezoelectric Macro-Fiber Composite/Ni Magnetoelectric Generator." Energy Harvesting and Systems 1:3.	Kinetic Energy Harvesting Using Piezoelectric and Electromagnetic Technologies-State of the Art	" IEEE Transactions on Industrial Electronics 57(3):850–60. [Web of Science] [CrossRef] Kong, N., D. S. Ha, A. Erturk, and D. J. Inman. 2010. "Resistive Impedance Matching Circuit for Piezoelectric Energy Harvesting"	

Bild 2: Ergebnisse einer (Annotationen-)Suchanfrage an einen Dokumenten-Corpus (Datenquelle: JournalTOCs):
Ausgabe aller Sätze, die genau drei Projektterme enthalten (siehe Spalte "match").
Filter auf die Begriffe in den Spalten *match*: "Energy", *after*: "IEE"

Bekanntmachung - Forschung - BMBF vom 15. Dezember 2014

Die Nachbarschaftspolitik der EU mit der Region des südlichen und östlichen Mittelmeers gewinnt zunehmend an Bedeutung. Deshalb hat die Europäische Kommission mit dem 7. Forschungsrahmenprogramm die Förderung eines ERA-Netzes mit dieser Region beschlossen. Das ERANETMED hat zum Ziel, die forschungspolitische Zusammenarbeit im Mittelmeerraum zwischen den beteiligten Partnern aus den EU-Mitgliedstaaten, den Assoziierten Staaten und den mediterranen Partnerländern zu intensivieren und politische Dialogprozesse zu stärken. Es leistet damit einen Beitrag zur Umsetzung und Begleitung des EUROMED-Prozesses.

Diese gemeinsame ERANETMED-Förderbekanntmachung soll es Wissenschaftlerinnen und Wissenschaftlern aus Europa sowie den südlichen und östlichen Mittelmeeranrainern ermöglichen, gemeinsam an einem Forschungsthema im Bereich „Erneuerbare Energien und Energieeffizienz“ zu arbeiten.

Folgende Förderorganisationen fördern die multilateralen Verbundvorhaben für „Erneuerbare Energien und Energieeffizienz“ im Rahmen der ERANETMED-Förderbekanntmachung:

Algerien
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique – Direction Générale de la Recherche Scientifique et du Développement Technologique (DG-RSDT)

Ägypten
Deutschland
Frankreich
Frankreich

ProjectTerms

AllTerms

Address
Business
Date
EH-Terme
Jobtitle
Location
Money
Number
Opinion
Organization
Percent
Person
ProjectTerms
Ratio
Sentence
Sentence1ProjectTerm
Sentence2ProjectTerms
Sentence Business

Bild 3: Analyse eines Einzel-Dokumentes