

Stammtisch

Luzern

Stammtisch
Luzern

SAP Stammtisch Luzern

Herzlich Willkommen!

Einleitung und Regeln – SAP Stammtisch Luzern

Beim SAP Stammtisch Luzern treffen sich Architekten, Berater, Entwickler, Studenten u.v.m. zum Erfahrungs- sowie Know-How Austausch

Unsere Regeln:

- Wir sind unabhängig
- Wir sind per «Du»
- Jeder ist willkommen
- Jeder zahlt sein Essen/Trinken selbst

Weitere Informationen

GitHub:

<https://sapstammtisch.github.io/Luzern>

XING:

<https://www.xing.com/communities/groups/sap-stammtisch-luzern-ed9c-1110610/about>

SAP Inside Track

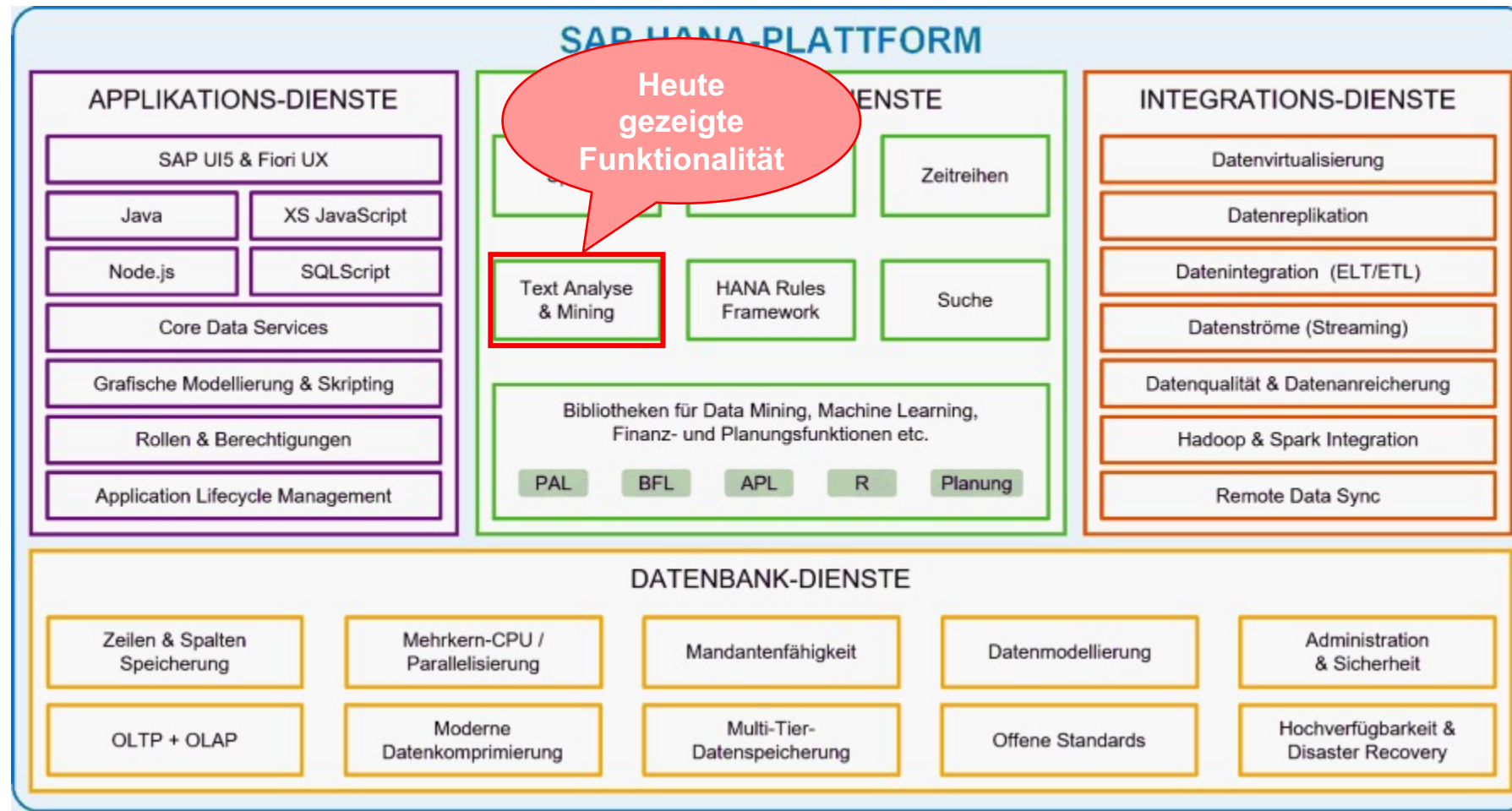
<https://sapstammtisch.github.io/Luzern/>

Google: ***SAP Stammtisch Luzern***

Nächster Termin: September - Mittwoch, 22.09.2021

Text-Analysis mit SAP HANA

HANA, S/4 HANA, HANA Plattform



Was wird gezeigt

1. Differenzierung zu S4/HANA und NetWeaver AS
2. SAP HANA Machine Learning
3. SAP HANA Machine Learning Erweiterungen
4. SAP HANA arbeiten mit Text
 - Fulltext und Fuzzy Search
 - Text analysis
 - Entity extraction
 - Fact Extraction (Sematische Textanalyse)
 - Text Mining
 - Dokumentklassifizierung, ähnliche Dokumente finden

S4/HANA und NetWeaver AS - Abgrenzung

1. SAP HANA ist der Datenbank unterbau von S4/HANA und SAP for HANA.
2. Läuft eigenständig
3. Unterschiedliche Konnektoren (ODBC, JDBC, Python- NodeJS- R-Integration)
4. Unabhängiges Lizenzmodel (Express sogar kostenlos)
5. Text-Analytics und Text-Mining verwendet bereits fertige Algorithmen und Modelle der SAP. Die Erstellung, Verwendung und Pflege eigener Modelle wird nicht gezeigt.

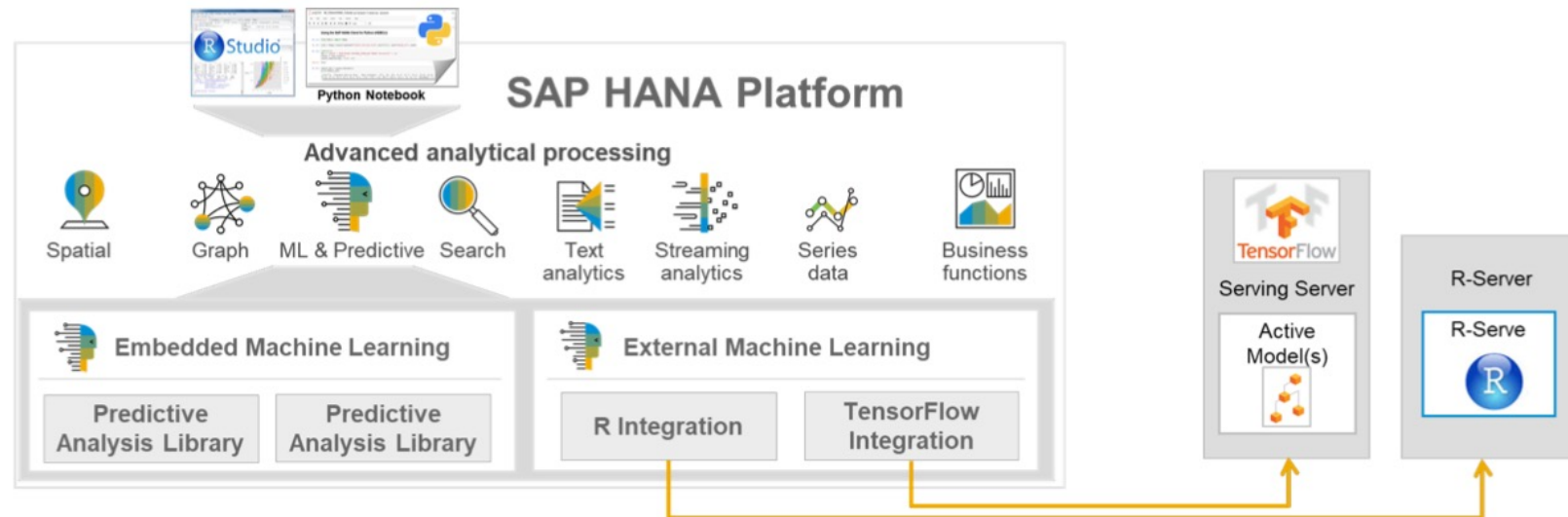
SAP HANA Machine Learning

SAP HANA machine learning

In-database and external machine learning capabilities

Richest multimodal in-database applications

- Trending ML algorithms for HANA embedded use and processing with in-memory performance
- Interfaces for Data Scientists in R and Python
- Combine and enrich spatial, text analysis, graph processing with machine learning



© 2019 SAP SE or an SAP affiliate company. All rights reserved. | PUBLIC

Arbeiten mit Text

Suche

Volltextsuche auf Tabellen und Views und Dokumenten. Fuzzy und konfigurierbar.

Textanalyse

Zerlegung von Texten in Token und Wordstämme, bis hin zu komplexeren Analysen wie Entitäts- und Fact-Extraction.

Text-Mining

Erkennung von relativen Inhalten, Klassifizierung und ähnliche Inhalte erkennen.

Beispiel Suche (Fuzzy und Dictionary)



Textanalyse

Sehr geehrter Versicherungs-AG,
bei meinem **Schadenfall 23.12231.23.6** habe ich eine
Taggeldzahlung über 300,00 CHF erhalten. Dies ist aus
meiner Sicht nicht korrekt, da eine **Arbeitsunfähigkeit**
von 40% angenommen wurde. Meine Arbeitsunfähigkeit
beträgt jedoch 80%.

Können Sie mich bitte kontaktieren um diesen
Sachverhalt zu klären.

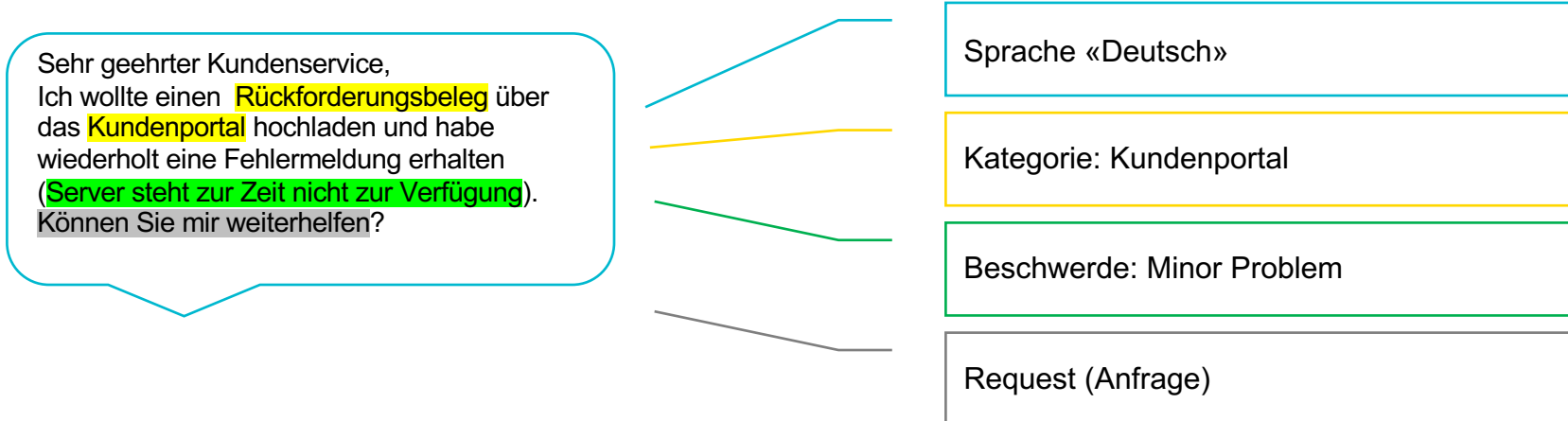
Mit freundlichen Grüßen

Alexander

Text	Typ
Versicherungs-AG	Organisation
Schadenfall	Topic
23.12231.23.6	CaseID
Taggeldzahlung	Topic
Arbeitsunfähigkeit	Topic
Können Sie... kontaktieren	Request
Alexander	Person

Anwendungsbeispiele

Beschwerde Management – automatische Klassifizierung basierend auf Text



Entity extraction

1. Entities:

<PERSON>**John Lennon**</PERSON>

war ein Mitglied der <ORGANIZATION>**Beatles**</ORGANIZATION>

2. Facts:

Ich

<STRONGPOSTIVESSENTIMENT>**liebe**</STRONGPOSTIVESSENTIMENT>

<TOPIC>**Ihre Versicherung**</TOPIC>

3. Entitätsextraktion ist keine Keyword suche:

– Can we **bill** you?

– **Bill** Smith was president.

Von HANA build-in unterstützte Extraktion

Who

People, job title, and national identification numbers

What

Companies, organizations, financial indexes, and products

When

Dates, days, holidays, months, years, times, and time periods

Where

Addresses, cities, states, countries, facilities, Internet addresses, and phone numbers

How much

Currencies and units of measure

Generic concepts

text data, global piracy, and so on

Sprachen

Arabic, English, Dutch, Farsi, French, German, Italian, Japanese, Korean, Portuguese, Russian, Simplified Chinese, Spanish, Traditional Chinese

Mitgelieferte Konfigurationen

Voice of customer

Sentiments: strong positive, weak positive, neutral, weak negative, strong negative, and problems

Requests: general and contact info

Emoticons: strong positive, weak positive, weak negative, strong negative

Customizing

1. Dictionaries
 - Eigene Entitäten
 - Wordstämme
 - Auch Semantik: (MajorProblem, Verb, Noun,...)

2. Regeln CGUL
Custom Grouper User Language
 - Regelbasierend
 - Reguläre Ausdrücke, Linguistische Attribute

Text Mining

Text-Mining hilft bei:

- Identifizierung ähnlicher Dokumente
- Identifizierung der Schlüsselwörter eines Dokumentes
- Identifizierung ähnlicher Begriffe
- Kategorisierung von Dokumenten

Beispiel:

- Kennzeichnung wichtiger Schlüsselwörter in einem Dokument
- Ähnliche Problemfälle und der Lösungen erkennen
- Automatiche Kategorisierung von Dokumenten

Text Mining Beispiel anhand von TripAdvisor-Daten



Zusammenfassung

1. SAP HANA bietet «out-of-the-box»:

- Parametrierbare Volltextsuche und Fuzzysuche
- Spracherkennung
- Text-Analytics, Sematische Texterkennung (gut, schlecht, Beleidigungen)
- Entitätsextraction z.B. für Produkt/Partnererkennung/Fall und Vertragserkennung
- Speicherung der Daten in Tabellen – somit wiederverwendbar für weitere Analysen, ML und Applikationen
- Text-Mining zur Klassifizierung und Einteilung von Texten/Dokumenten