

# Email-Topic Analysis

Kevin Bönisch  
NLP | Data Science | Software Development  
[k.boenisch@outlook.com](mailto:k.boenisch@outlook.com)

# ING DiBa

---

# Task



- **Input:** Dataset: 627 German mails
- **Output:** NLP-pipeline for processing, focusing on Topic Modelling



# Topic Modelling with BERTopic

Training a small **BERTopic** model on the given emails

- Create embeddings for each email (they aren't very long)
- Train a BERTopic model on these
- Upload the model to Hugging Face
- Use the topic model in the NLP pipeline



# Topic Modelling with BERTopic

## Naive baseline:

- Bilingual embedding model *cross-encoder/msmarco-MiniLM-L6-en-de-v1*
- No Hyperparameter tuning
- Train BERTopic, plot the results

# Topic Modelling with BERTopic





# Topic Modelling with BERTopic

- Embedding model isn't capturing clusters well
  - Use *deepset/gbert-large*
- Hyperparameters are off:
  - *min cluster size probably too large by default?*
    - Adjust to smaller value = 30
  - *Word filtering?*
    - Filter very rare terms
    - Drop ubiquitous terms

# Topic Modelling with BERTopic





# Topic Modelling with BERTopic

## Topics:

- -1\_specificationsheet\_der xsl\_specs\_xsl
- 0\_mac\_die installation\_wenden sie sich\_wenden sie
- 1\_auf den\_sie bitte\_die bestellnummer\_anzeigen
- 2\_fuer\_moechte\_ueber\_freundlichen gruessen
- 3\_aktivierungscode\_fehlerhaften\_27\_ticket
- 4\_intel\_core\_ethernet\_ghz
- 5\_build\_exception\_build\_project build\_exception\_disc project build

**=> Not very readable**





# Topic Modelling with BERTopic

**Use LLM to generate custom labels: *meta-llama/Llama-3.1-8B-Instruct***

*[INST]*

*Ich habe ein Thema, das die folgenden Dokumente enthält:*

*[DOCUMENTS]*

*Das Thema wird durch die folgenden Schlüsselwörter beschrieben: '[KEYWORDS]'.*

*Basierend auf den obigen Informationen zum Thema erstelle bitte ein kurzes Label für dieses Thema. Stelle sicher, dass du nur das Label zurückgibst und nichts weiter.*

*[/INST]*



# Topic Modelling with BERTopic

## Topics:

- -1\_specificationsheet\_der\_xsl\_specs\_xsl => Fehlermeldung "Fehler beim Parsen der XSL Daten"
- 0\_mac\_die\_installation\_wenden\_sie\_sich\_wenden\_sie => Interbase SPARQL Server - Installation fehlerhaft
- 1\_auf\_den\_sie\_bitte\_die\_bestellnummer\_anzeigen => Warehouses Prozessverwaltung und -übertragung
- 2\_fuer\_moechte\_ueber\_freundlichen\_gruessen => Softwareprobleme und Support
- 3\_aktivierungscode\_fehlerhaften\_27\_ticket => Softwarefehler bei Sales First Class
- 4\_intel\_core\_ethernet\_ghz => Hardwareprobleme bei der Installation von Sales First Class
- 5\_build\_exception\_build\_project\_build\_exception\_disc\_project\_build => Brennvorgänge mit Build-Exception-Fehlern



# Topic Modelling with BERTopic





# Topic Modelling with BERTopic

## Evaluation with Silhouette Score

- How well each document fits its assigned topic compared to other topics.
- Score ranges from -1 to 1
  - *Values close to 1 indicate that documents are well matched to their topic and distinct from other topics.*
  - *Values near 0 suggest that documents lie between topics.*
  - *Negative values imply that documents may have been assigned to the wrong topic.*
- Silhouette Score (all docs): 0.45396703481674194

# NLP-Pipeline



UIMA

- Transform Emails into the Unstructured Information Management Architecture format

DUUI

- Define the Docker Unified UIMA Interface NLP pipeline (Abrami et al., 2025)

UCE

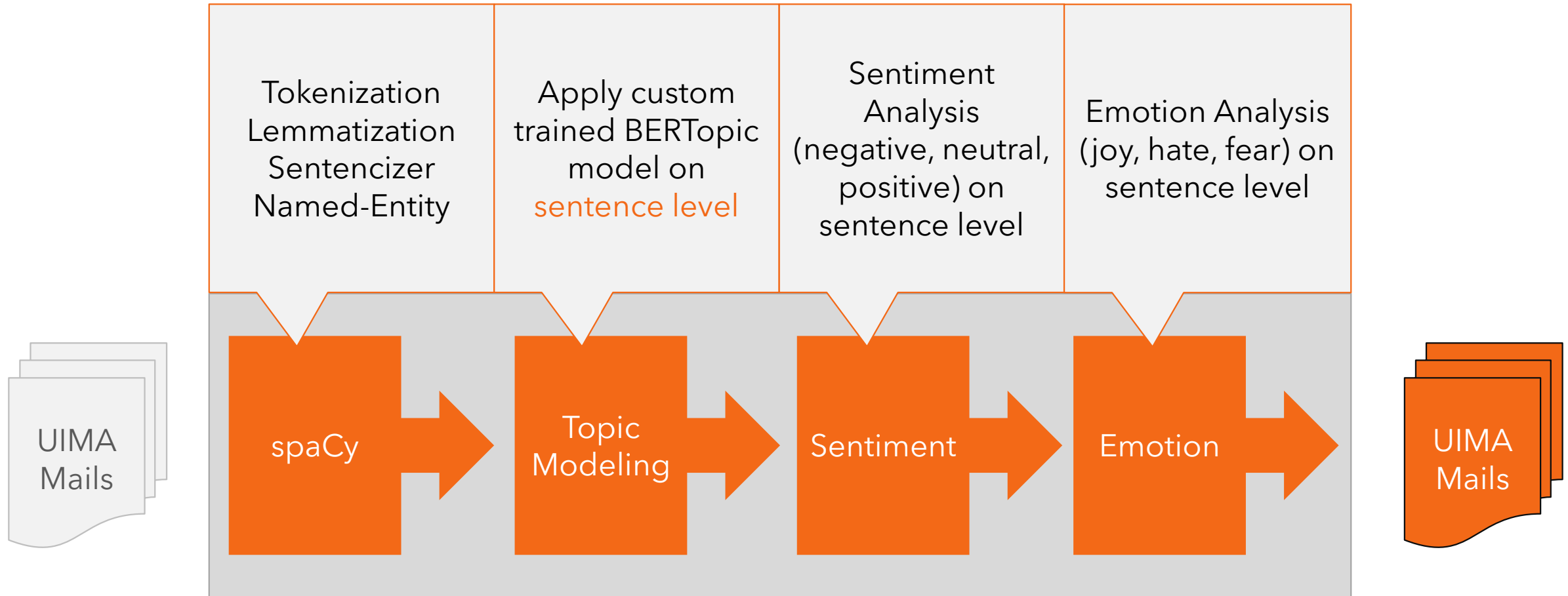
- Setup the Unified Corpus Explorer web portal



```
<xmi:XMI xmlns:xmi="http://www.omg.org/XMI"
  xmlns:pp1="http://org.texttechnologylab/type/morphosyn/tag/pp1.ecore"
  xmlns:convant1="http://org.texttechnologylab/type/morphosyn/tag/conv/convant1.ecore"
  xmlns:convcon="http://org.texttechnologylab/type/morphosyn/tag/conv/convcon.ecore"
  xmlns:cnjtemp1="http://org.texttechnologylab/type/morphosyn/tag/cnj/cnjtemp1.ecore"
  xmlns:twitter="http://org.texttechnologylab/annotation/twitter.ecore"
  xmlns:IsoSpatial="http://org.texttechnologylab/annotation/semaf/IsoSpatial.ecore"
  xmlns:negmod="http://org.texttechnologylab/type/morphosyn/tag/neg/negmod.ecore"
  xmlns:semantics="http://org.hucompute/textimager/uima/type/semantics.ecore"
  xmlns:convpost="http://org.texttechnologylab/type/morphosyn/tag/conv/convpost.ecore"
  xmlns:type16="http://org.texttechnologylab/annotation/temporal/type.ecore"
  xmlns:conv="http://org.texttechnologylab/type/morphosyn/tag/conv.ecore"
  xmlns:gnfinder="http://org.texttechnologylab/annotation/biofid/gnfinder.ecore"
  xmlns:cnjcoord1="http://org.texttechnologylab/type/morphosyn/tag/cnj/cnjcoord1.ecore"
  xmlns:morphosyn="http://org.texttechnologylab/type/morphosyn.ecore" xmi:version="2.0">
  <cas:NULL xmi:id="0" />
  <tcas:DocumentAnnotation xmi:id="2" sofa="1" begin="0" end="157" language="de" />
  <type2:DocumentMetaData xmi:id="3" sofa="1" begin="0" end="156" documentTitle="Mail_1"
    documentId="1"
    documentUri="file:///C:/kevin/projects/duui/ing-diba-mail-topics/data/xmi-raw/mail_1.xmi"
    documentBaseUri="file:///C:/kevin/projects/duui/ing-diba-mail-topics/data/xmi-raw/"
    isLastSegment="false" />
  <cas:Sofa xmi:id="1" sofaNum="1" sofaID="_InitialView" mimeType="text"
    sofaString="Hallo, ich habe seit 2 Tagen in meiner Warehouse Software unter Exportieren-File Optionen ein fast nicht lesbare Schrift. Wi
  <cas:View sofa="1" members="2 3" />
</xmi:XMI>
```



# DUUI NLP-Pipeline





dependency:ROOT xmi:id="1415" sofa="4" begin="524" end="531" Governor="822" Dependent="822" DependencyType="--" flavor="basic"/>  
dependency:ROOT xmi:id="1416" sofa="4" begin="573" end="580" Governor="830" Dependent="830" DependencyType="--" flavor="basic"/>  
dependency:ROOT xmi:id="1417" sofa="4" begin="685" end="688" Governor="850" Dependent="850" DependencyType="--" flavor="basic"/>  
type3:NamedEntity xmi:id="1418" sofa="4" begin="0" end="9" value="MISC"/>  
type3:NamedEntity xmi:id="1419" sofa="4" begin="128" end="155" value="LOC"/>  
type3:NamedEntity xmi:id="1420" sofa="4" begin="176" end="201" value="ORG"/>  
type3:NamedEntity xmi:id="1421" sofa="4" begin="233" end="236" value="PER"/>  
type3:NamedEntity xmi:id="1422" sofa="4" begin="264" end="267" value="PER"/>  
type3:NamedEntity xmi:id="1423" sofa="4" begin="622" end="633" value="MISC"/>  
duui:ReproducibleAnnotation xmi:id="3" timestamp="9394634963113" pipelineName="wiki" description="/Td6WfOAAATm1rRGAgAhARYAAAB0L+Wj4AFUARbAD2IiedENISNkF8NimONoMb/oRVKAmfISW6d30imJPRG1qE7KB0zpitrGdyF493Y9qVrOX15wO2+IhI/zM/w1No7nR0Mw0NVGediNcwFtikiyYtCfuj/12  
duui:ReproducibleAnnotation xmi:id="1610" timestamp="25308142105600" pipelineName="ing-diba-pipeline" description="/Td6WfOAAATm1rRGAgAhARYAAAB0L+Wj4AHGAURdAD2IiedENISNkF8NimONoMw4kdGQnyu3CZ9LySb6x0wvaNLv1s7zupnYyIhbXuG6GSbxqPKhwV6UUDG32lmq0ZdxwK/i58RMu1bcw  
duui:ReproducibleAnnotation xmi:id="2" timestamp="9702770686707" pipelineName="wiki" description="/Td6WfOAAATm1rRGAgAhARYAAAB0L+Wj4AFTAPJdAD2IiedENISNkF8NimONroAYnfDRTp2aAVJtS2R3f6fyTCetAyT/nFXJQ870ZE5irjm/D1T2Vqxvb21qt07hySdreiu5XqBU30MynkT1teA4zKj88Dp7BJ  
duui:ReproducibleAnnotation xmi:id="1" timestamp="9429821123352" pipelineName="wiki" description="/Td6WfOAAATm1rRGAgAhARYAAAB0L+Wj4AFXAP9dAD2IiedENISNkF8NimONoMb/oRVKAmfISW6d30imJPRG1qE7KB0zpitrGdyF493Y9qVrOX15wO2+IhI/zM/w1No7nR0Mw0NVGediNcwFtikiyYtCfuj/12  
duui:ReproducibleAnnotation xmi:id="1511" timestamp="25307695132100" pipelineName="ing-diba-pipeline" description="/Td6WfOAAATm1rRGAgAhARYAAAB0L+Wj4AHHAUNDAD2IiedENISNkF8NimONoMw4kdGQnyu3CZ9LySb6x0wvaNLv1s7zupnYyIhbXuG6GSbxqPKhwV6UUDG32lmq0ZdxwK/i57gQM0Mwx  
model:MetaData xmi:id="1424" sofa="4" begin="0" end="0" Lang="EN" Source="https://huggingface.co/TheItCrOw/bertopic-german-mails-small" ModelVersion="1bc72fb28573f4c6c839faa331089fa836ed53d4" ModelName="TheItCrOw/bertopic-german-mails-small"/>  
model:MetaData xmi:id="1498" sofa="4" begin="0" end="0" Lang="Multi" Source="https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment" ModelVersion="f2f1202b1bdeb07342385c3f807f9c07cd8f5cf8" ModelName="cardiffnlp/twitter-xlm-roberta-base-sentim  
model:MetaData xmi:id="1513" sofa="4" begin="0" end="0" Lang="Multi" Source="https://huggingface.co/02shanky/finetuned-twitter-xlm-roberta-base-emotion" ModelVersion="28e6d080e9f73171b574dd88ac768da9e6622c36" ModelName="02shanky/finetuned-twitter-xlm-rober  
annotation2:TopicValue xmi:id="1425" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" probability="0.9522950053215027"/>  
annotation2:TopicValue xmi:id="1426" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" probability="0.9583427309989929"/>  
annotation2:TopicValue xmi:id="1427" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" probability="0.8584617972373962"/>  
annotation2:TopicValue xmi:id="1430" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" probability="0.9596068263053894"/>  
annotation2:TopicValue xmi:id="1431" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" probability="0.9442148804664612"/>  
annotation2:TopicValue xmi:id="1429" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" probability="0.887855589389801"/>  
annotation2:TopicValue xmi:id="1436" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" probability="0.975091278553009"/>  
annotation2:TopicValue xmi:id="1435" sofa="4" value="Interbase SPARQL Server - Installation fehlerhaft" probability="0.967727541923523"/>  
annotation2:TopicValue xmi:id="1434" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" probability="0.9724609851837158"/>  
annotation2:TopicValue xmi:id="1432" sofa="4" value="Interbase SPARQL Server - Installation fehlerhaft" probability="0.9665926098823547"/>  
annotation2:TopicValue xmi:id="1428" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" probability="0.8495633602142334"/>  
annotation2:TopicValue xmi:id="1433" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" probability="0.8755331635475159"/>  
annotation2:BertTopic xmi:id="1437" sofa="4" begin="0" end="72" Topics="1425" model="1424"/>  
annotation2:BertTopic xmi:id="1439" sofa="4" begin="72" end="76" Topics="1433" model="1424"/>  
annotation2:BertTopic xmi:id="1441" sofa="4" begin="77" end="86" Topics="1429" model="1424"/>  
annotation2:BertTopic xmi:id="1443" sofa="4" begin="87" end="263" Topics="1436" model="1424"/>  
annotation2:BertTopic xmi:id="1445" sofa="4" begin="264" end="267" Topics="1427" model="1424"/>  
annotation2:BertTopic xmi:id="1447" sofa="4" begin="268" end="271" Topics="1428" model="1424"/>  
annotation2:BertTopic xmi:id="1449" sofa="4" begin="272" end="337" Topics="1426" model="1424"/>  
annotation2:BertTopic xmi:id="1451" sofa="4" begin="338" end="438" Topics="1434" model="1424"/>  
annotation2:BertTopic xmi:id="1453" sofa="4" begin="439" end="520" Topics="1431" model="1424"/>  
annotation2:BertTopic xmi:id="1455" sofa="4" begin="521" end="566" Topics="1430" model="1424"/>  
annotation2:BertTopic xmi:id="1457" sofa="4" begin="567" end="671" Topics="1435" model="1424"/>  
annotation2:BertTopic xmi:id="1459" sofa="4" begin="672" end="768" Topics="1432" model="1424"/>  
annotation2:TopicValueBaseWithScore xmi:id="1465" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" score="0.8495633602142334"/>  
annotation2:TopicValueBaseWithScore xmi:id="1467" sofa="4" value="Interbase SPARQL Server - Installation fehlerhaft" score="0.967727541923523"/>  
annotation2:TopicValueBaseWithScore xmi:id="1464" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" score="0.8584617972373962"/>  
annotation2:TopicValueBaseWithScore xmi:id="1461" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" score="0.8755331635475159"/>  
annotation2:TopicValueBaseWithScore xmi:id="1471" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" score="0.9442148804664612"/>  
annotation2:TopicValueBaseWithScore xmi:id="1466" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" score="0.975091278553009"/>  
annotation2:TopicValueBaseWithScore xmi:id="1470" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" score="0.9522950053215027"/>  
annotation2:TopicValueBaseWithScore xmi:id="1472" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" score="0.9596068263053894"/>  
annotation2:TopicValueBaseWithScore xmi:id="1463" sofa="4" value="Interbase SPARQL Server - Installation fehlerhaft" score="0.9665926098823547"/>  
annotation2:TopicValueBaseWithScore xmi:id="1468" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" score="0.9724609851837158"/>  
annotation2:TopicValueBaseWithScore xmi:id="1462" sofa="4" value="Fehlermeldung &quot;Fehler beim Parsen der XSL Daten&quot; in der Gewinnkalkulation" score="0.9583427309989929"/>  
annotation2:TopicValueBaseWithScore xmi:id="1469" sofa="4" value="Brennvorgänge mit Build-Exception-Fehlern" score="0.887855589389801"/>  
annotation2:UnifiedTopic xmi:id="1473" sofa="4" begin="0" end="72" Topics="1470"/>  
annotation2:UnifiedTopic xmi:id="1475" sofa="4" begin="72" end="76" Topics="1461"/>  
annotation2:UnifiedTopic xmi:id="1477" sofa="4" begin="77" end="86" Topics="1469"/>  
annotation2:UnifiedTopic xmi:id="1479" sofa="4" begin="87" end="263" Topics="1466"/>  
annotation2:UnifiedTopic xmi:id="1481" sofa="4" begin="264" end="267" Topics="1464"/>



# Visualize Results with UCE



## Unified Corpus Explorer (UCE)

(Boenisch et al., 2025)



- Import annotated UIMA objects
- Visualize their annotations & build a fully-fledged webportal around the corpus



Best Paper Award  
@ NAACL 2025  
System Demonstrations Track

Sparql &amp; Problem



Pro



Search results ≥ 5

## Keyword-in-Context

Search has been started without the Keyword-in-context feature.

Fulltext

Published

Title

Relevancy



Mail\_582 (582)

- 1 0,013



(Unbekannt)



Brennvorgänge mit Build-Exception-Fehlern



Softwareprobleme und Support



Interbase SPARQL Server - Installation fehlerhaft



Fehlermeldung bei Installation (Programmpunkt: Interbase **SPARQL** Server): ...bei Umwandlung ihres Produktdatenspeichers ist ein **Problem** aufgetreten.. Warehouse laesst sich dann nicht starten

1.



Mail\_584 (584)

- 1 0,011



(Unbekannt)



Interbase SPARQL Server - Installation fehlerhaft



Fehlermeldung "Fehler beim Parsen der XSL Daten" in der Gewinnkalkulation



Brennvorgänge mit Build-Exception-Fehlern



Sehr geehrte Frau Holpert, hier die Fehlermeldung beim Aufspielen der CD in den Rechner: Es wird die Komponente 4 übertragen: Interbase **SPARQL** Server. Fehlermeldung: "Bei Umwandlung Ihres Produktdatenspeichers ist ein **Problem** aufgetreten. Sie werden Ihren Datenspeicher deswegen mit dieser Software nicht verwenden können. Bitte wenden Sie sich an den Helpdesk." Nachdem der Fehler bei Interbase auftritt, ist es wegen

1.



Mail\_617 (617)

- 1 0,008



(Unbekannt)



Brennvorgänge mit Build-Exception-Fehlern



Interbase SPARQL Server - Installation fehlerhaft



Fehlermeldung: Interbase **SPARQL** Server-Warehouse Edition Setup Bei der Umwandlung Ihres Produktdatenspeichers ist ein **Problem** aufgetreten. Sie werden Ihre Datenbank mit dieser Software nicht verwenden können. Bitte wenden Sie sich an den Helpdesk

1.

Sparql &amp; Problem

Locations

4

People

5

Organisations

3

Misc

0

Taxonomy

0

Times

0

Navigation

&lt; 1 &gt;

W

Search

Wiki

ING DiBa Mails / Mail\_233 / Interbase SPARQL Server - Installation fehlerhaft

Topic - Annotation DTR-149: Interbase SPARQL Server - Installation fehlerhaft

Annotated within document: D-151

Document Distribution of the topic

Document ID	Topic Weight
288	1.0
117	1.0
254	1.0
327	1.0
261	1.0
247	1.0
267	1.0
608	1.0
301	1.0
241	1.0
252	0.88
612	0.86
263	0.85
61	0.83
253	0.83
527	0.83
599	0.81
248	0.80
214	0.80
49	0.79

W Mail\_233 (233)

Softwareprobleme und Support

Interbase SPARQL Server - Installation fehlerhaft

Brennvorgänge mit Build-Exception-Fehlern

Rechner habe, laesst sich der produktcreator 11 nicht mehr installieren. Was kann ich als ziemlicher Computer-Laie tun?...

W Mail\_437 (437)

1

0

Keyword-in-Context

Search has been started without the Keyword-in-context feature.

Locations 9

People 6

Organisations 5

Misc 9

Taxonomy 0

Times 0

Navigation

< 1 2 3 >

←

ING DiBa Mails

Corpus Overview

Meta

Name

ING DiBa Mails

Source

ING DiBa

Language

de-DE

Imported

2025-08-27T18:29:49

Documents

627

Pages

628

Annotations

OCRPages

☐

OCRParagraphs

☐

OCRBloc

☐

OCRLines

☐

Taxa

☐

BIOfid O

☐

SRLinks

☐

Named-Entities

☒

GeoNam

☐

Sentences

☒

Lemma

☒

Times

☐

Embeddings

☐

RAGBot

☐

WikiLink

☐

Negations

☐

Unified Topics

☒

W

Corpus Documents

Wiki

ING DiBa Mails

Corpus - Annotation C-1: ING DiBa Mails

corpusConfig.json

Annotations and Metadata

Description

Ein kleiner Korpus bestehend aus Helpdesk-Mails.

Topic Distribution in Corpus

Topic	Value
Softwareprobleme und Support	0.15
Softwarefehler bei Sales First Class	0.10
Fehlermeldung "Fehler beim Parsen der XSL Daten" in der Gewinnkalkulation	0.15
Warehouses Prozessverwaltung und -übertragung	0.10
Brennvorgänge mit Build-Exception-Fehlern	0.10
Interbase SPARQL Server - Installation fehlerhaft	0.10

Softwareprobleme und Support

Softwarefehler bei Sales First Class

Fehlermeldung "Fehler beim Parsen der XSL Daten" in der Gewinnkalkulation

Warehouses Prozessverwaltung und -übertragung

Brennvorgänge mit Build-Exception-Fehlern

Hardwareprobleme bei der Installation von Sales First Class

Interbase SPARQL Server - Installation fehlerhaft

You want to search? Use the UCE-Portal!

der XSL Daten" in der Gewinnkalkulation

1

beim Parsen der XSL Daten" in der Gewinnkalkulation

1

stzt man den Startzeiger an eine beliebige

1

Daten" Brennvorgänge mit Build-Exception-Fehlern

1

damit ich an diesem Punkt auch das element

1

Interbase SPARQL Server - Installation fehlerhaft

1

icht mehr geladen bzw.nur der ladevorgang

1

en mußte aber feststellen das ein bearbeiten

1

## Lexicon 14.388

Search



A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

unifiedtopic 68	PC 23
Windows 7 18	Problem 15
Zeit 14	XSL Daten 14
Mac 13	OFFICE 10
Sales 10	Internes Problem 10
Guten Tag 9	Windows 9
Gewinn-Kalkulation 9	MfG 9
Link 8	Datenscanner 8
IT 7	Device 7
.PRC 7	USB-Prozessverwalter 7
Vorschau 7	Sie werden Ihren Datenspeicher deswegen mit dieser Software nicht verwenden können. 7

## Occurrence



Select a lexical entry.

Load more

## Filter


NamedEntity ☒

GeoName ☒

Lemma ☐

Time ☒

GazetteerTaxon ☒

GnFinderTaxon ☒

CompleteNegation ☒

Focus ☒

Cue ☒

Scope ☒

XScope ☒

UnifiedTopic ☒

W

Internet

namedentity

1 2 3 4

Occurrence



Alphabet





Mail\_584

(Unbekannt)

Sehr geehrte Frau Holpert, hier die Fehlermeldung beim Aufspielen der CD in den Rechner: Es wird die Komponente 4 übertragen: joy t Interbase SPARQL joy t Server. joy t Fehlermeldung: "Bei Umwandlung Ihres Produktdatenspeichers ist ein Problem aufgetreten. fear t Sie werden Ihren Datenspeicher deswegen mit dieser Software nicht verwenden können. anger t Bitte wenden Sie sich an den Helpdesk." anger t Nachdem der Fehler bei Interbase auftritt, ist es wegen meiner Antivirensoftware Avira nicht einverstanden? anger t Geben Sie bitte mein Problem an einen Techniker weiter mit der Bitte, daß er mir hilft. joy t Meine Telefon-Nummer 41737-6183 Er soll mir bitte auch seine Durchwahlnummer geben, daß ich ihn ggfls. joy t zurückrufen kann. joy t Besten Dank joy t

- 1 -

## Control Panel

## Visualization

584

### Settings

Tt

### Functions

Toggle Focus

Toggle Highlighting

### Search Tokens

### Key Topics



Interbase SPARQL  
Server - Installation  
fehlerhaft

Fehlermeldung  
"Fehler beim Parsen  
der XSL Daten" in der  
Gewinnkalkulation

Brennvorgänge mit  
Build-Exception-  
Fehlern

Softwarefehler bei  
Sales First Class

W



Mail\_584

(Unbekannt)

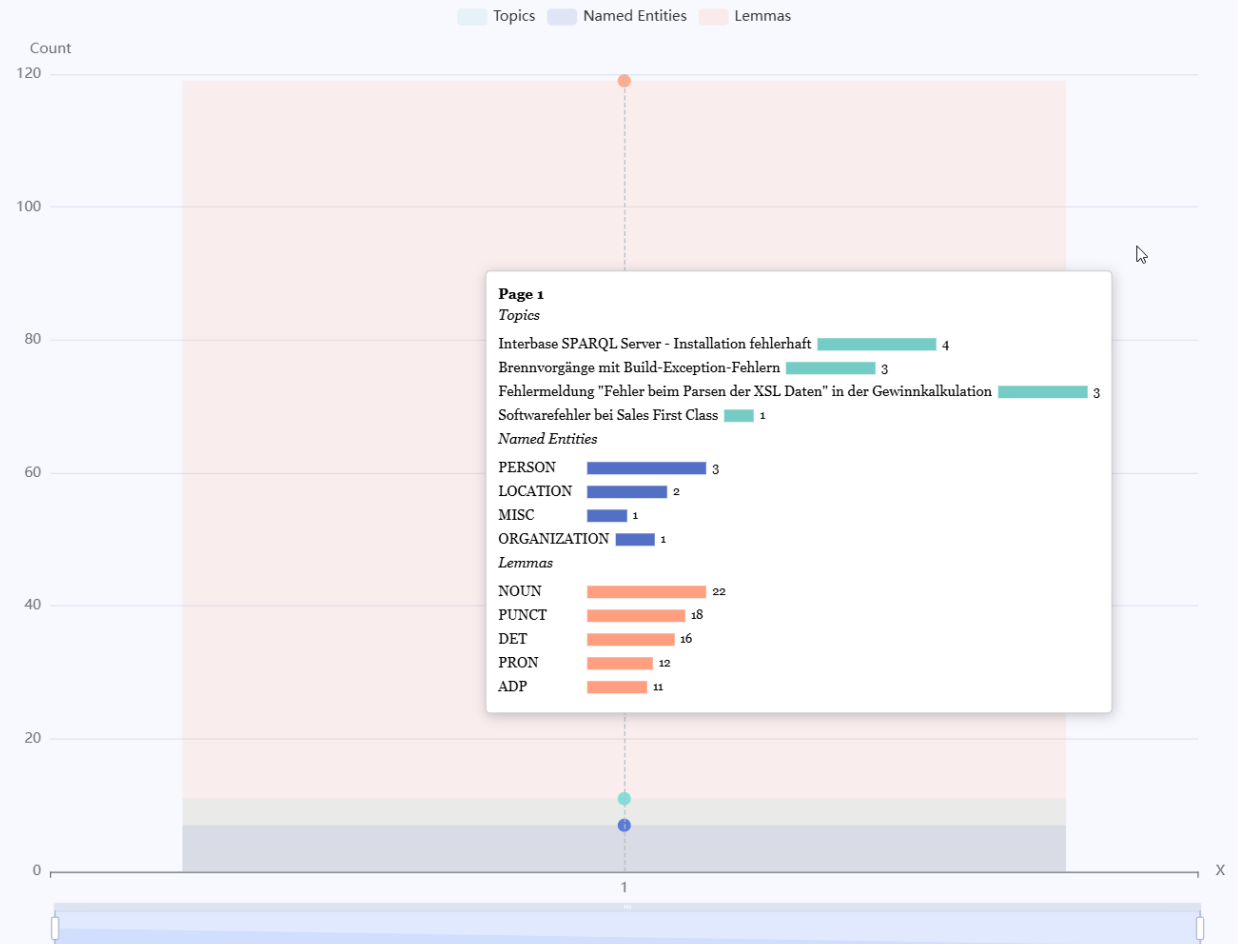
Sehr geehrte Frau Holpert, hier die Fehlermeldung beim Aufspielen der CD in den Rechner: Es wird die Komponente 4 übertragen: joy  
Interbase SPARQL joy Server. joy Fehlermeldung: "Bei Umwandlung Ihres Produktdatenspeichers ist ein Problem aufgetreten. fear  
Sie werden Ihren Datenspeicher deswegen mit dieser Software nicht verwenden können. anger Bitte wenden Sie sich an den Helpdesk." anger  
Nachdem der Fehler bei Interbase auftritt, ist es wegen meiner Antivirensoftware Avira nicht einverstanden? anger  
Geben Sie bitte mein Problem an einen Techniker weiter mit der Bitte, daß er mir hilft. joy Meine Telefon-Nummer 41737-6183 Er soll mir bitte auch seine Durchwahlnummer geben, daß ich ihn ggfls. joy zurückrufen kann. joy Besten Dank joy

- 1 -

W

## Control Panel

## Visualization





Semantic Density Topic-Entity Topic Landscape Topic Similarity Sentence Topic Flow





## Bericht zur Analyse - Seite 1

11/18/2021 → 9/2/2025

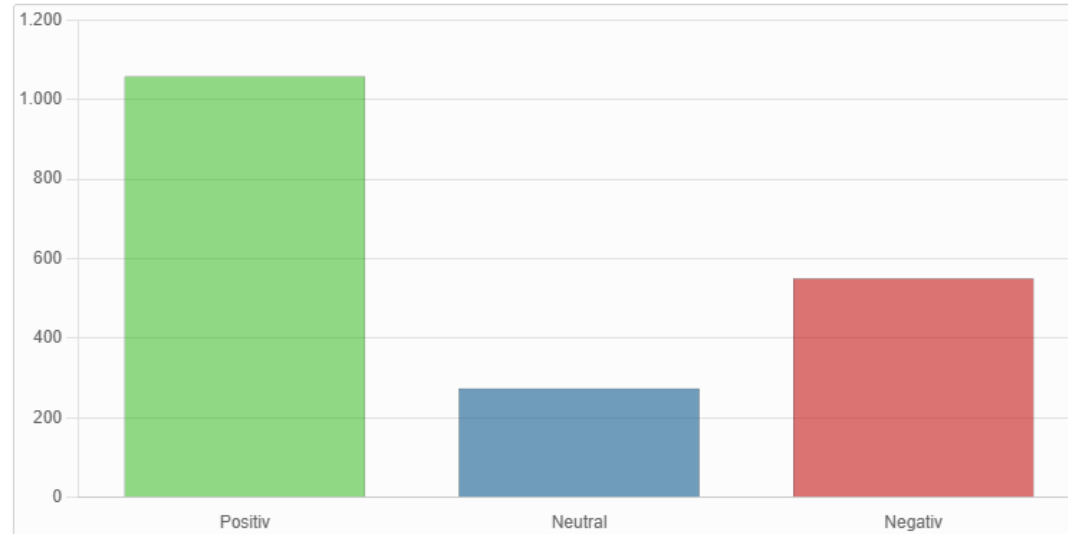
Auf dieser Seite wird die gewählte Entität auf das Thema "Ukraine" untersucht.

Alle Auswertungen, die unter Berücksichtigung des gewählten Zeitraums erfasst wurden, sind mit einem  Icon markiert.  
Das gleiche gilt parallel für das Berücksichtigen des Themas und dem  Icon.

## Überblick zur gewählten Entität

SPD		Gesamt	Zeitraum
	 Anzahl Reden	9392	4991
	 Anzahl Reden im Thema	806	763
	 Reden pro Sitzung	21,35	25,08
	 Reden pro Sitzung im Thema	1,83	3,83

## Sentimentverteilung (Auf Satzebene)



Reden zum Thema

Florian Hahn  
(CDU/CSU)

Besonders hoch ist dabei der Erwartungsdruck auf Deutschland. Dass die Bundesregierung glaubt, mit einer Lieferung von 5000 Helmen an die Ukraine dieses Signal der Unterstützung zu senden und dem Erwartungsdruck an uns Herr zu werden, ist ein großer Irrtum.



Spärlicher Applaus!

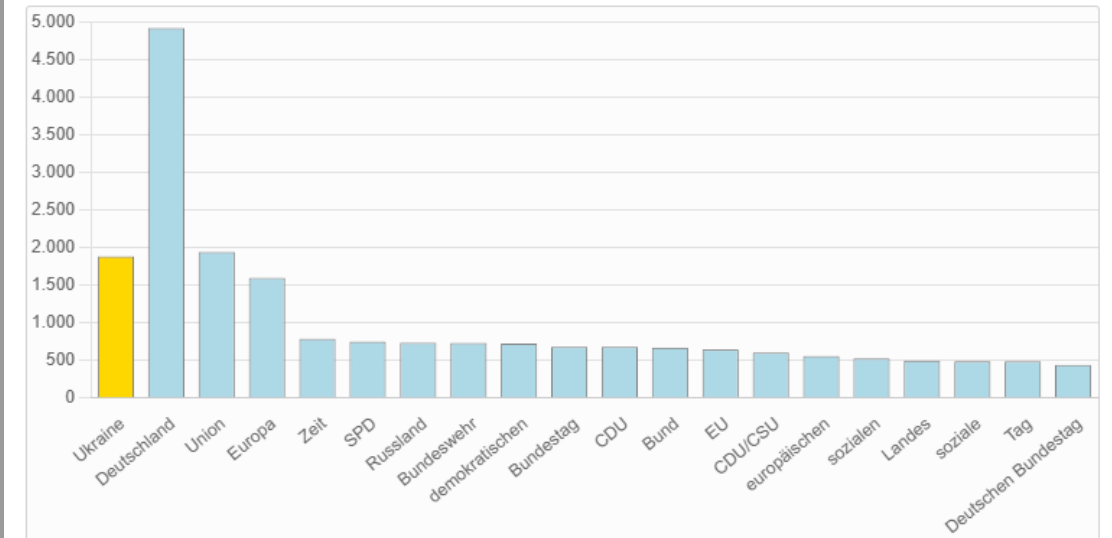
Johannes Schrap (SPD)

Roderich Kiesewetter  
(CDU/CSU)

Es ist Geschichtsklitterung, wenn Gregor Gysi hier Russland das Wort redet und Beifall von der AfD bekommt. Denn erinnern wir uns an die Auswirkungen, was die Ukraine durchzumachen hat: Im Jahre 1994 vereinbarte man im Budapester Memorandum und in vielen Verhandlungen davor, dass die Ukraine ihren Status als drittstärkste Nuklearmacht der Welt abgibt, die Atomwaffen übergibt, um selbst Integrität und Souveränität zu erhalten. Und 2014 ist dieses Abkommen gebrochen worden. Sämtliche Nuklearverhandlungen weltweit liegen auf Eis, weil alle das Beispiel Ukraine fürchten. Das ist die Wahrheit. Herr Klingbeil, ich hätte mir gewünscht, dass Sie so klare Worte gefunden hätten wie Alexander Graf Lambsdorff oder Friedrich Merz.

## Thema im Prioritäts-Vergleich

Es wird das gewählte Thema mit den Top 20 Themen der gewählten Entität bzgl. derer Erwähnungen verglichen.







# Future Work

- How does your solution work when multiple topics are addressed in one email?
  - *Sentence-based (or segment-based) Topic Modelling solves this issue*
- How does your solution cope with new topics that emerge over time?
  - *DUUI is able to redo annotations; BERTopic can be continuously trained (partial\_fit)*
- How could your solution be improved if you had access to more metadata?
  - *UCE offers dynamic filtering based on metadata (e.g. by customerId, by gender, by [GENERIC\_KEY])*

# Thank you



**GitHub**



**BERTopic Model**



**UCE**



**DUUI**



---

**KEVIN BÖNISCH**

**NLP | DATA SCIENCE | SOFTWARE DEVELOPMENT | KAGGLE MASTER**