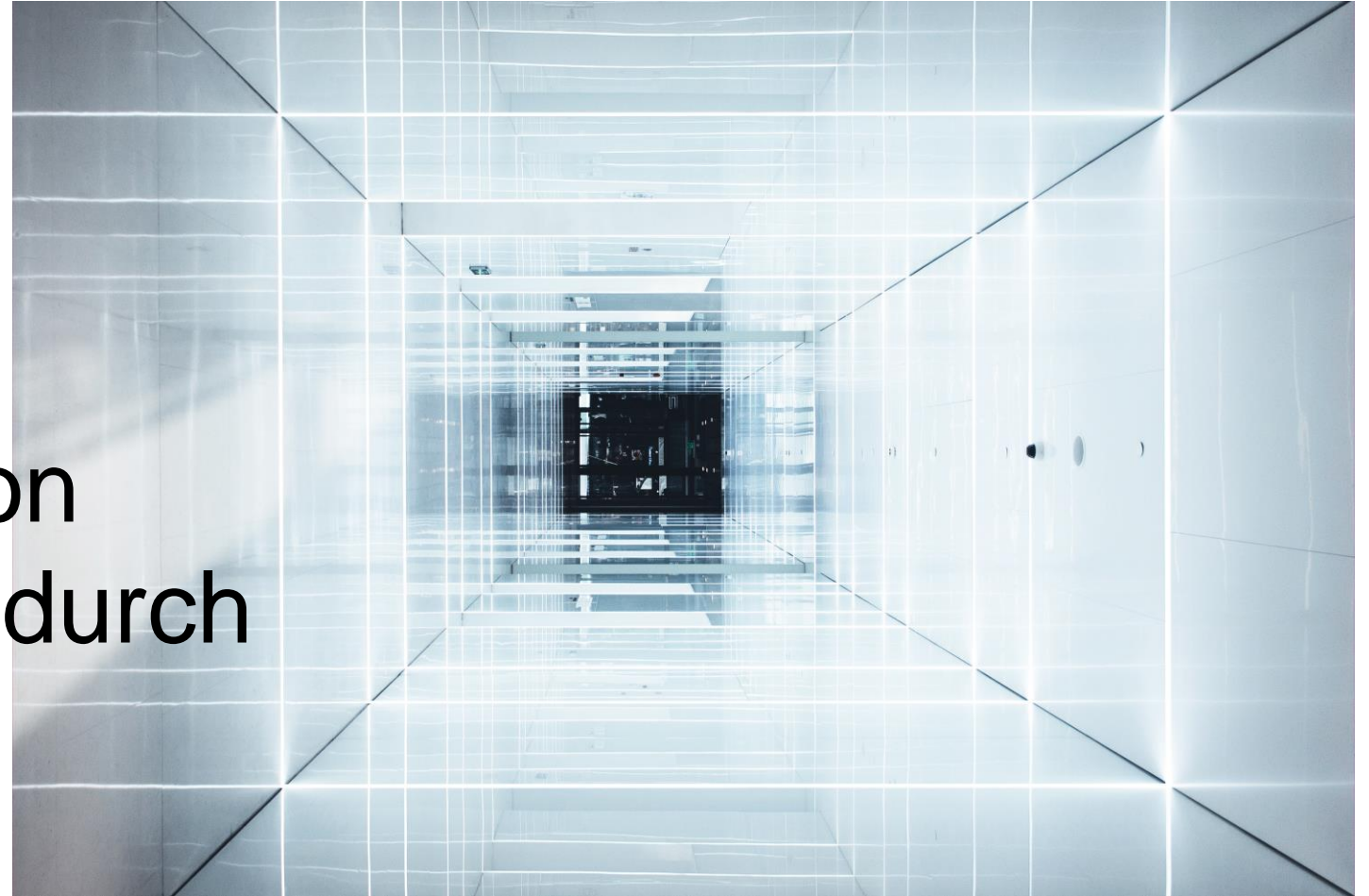


Semantische Anreicherung von Named Entities durch Taxonomien

Qurator 2021

spaCy, Syntex und TermStores und
graphische Modelle



Christian Fillies, Semtation GmbH
Berlin, 17. 02. 2021

Das Unternehmen

- Gründung 2001
- Standort Potsdam
- 10 Mitarbeiter
- Microsoft-Partner (ISV) seit 2001
- Eigenes Partnernetzwerk
- Büro im Denkmal

SEM TATION



Produktstrategie und Leistungsangebot

- Plattformhersteller:
 - Eigener Editor für Prozessmodelle: SemTalk Editor
 - Portal Templates und –Bausteine: SemTalk Services
 - Microservices zur Modellanreicherung
 - Metamodell-Konfigurator
- Einführungsunterstützung
- Prozess-basierte Implementierung auf der Microsoft-Plattform:
 - Strukturen in Microsoft Teams und SharePoint
 - Workflows



SemTalk®

SEM TATION

Das Qurator-Projekt

- Thema: Digitale Kuratierung
Bereitstellung elektronisch verfügbarer Inhalte
- 10 Partner aus Berlin und Brandenburg:
 - Forschungspartner: DFKI, Fraunhofer FOKUS, Wikimedia, Staatsbibliothek – Stiftung Preußischer Kulturbesitz
 - Unternehmenspartner: Semtation, ART+COM, Condat, Ubermetrics, 3pc, Ada Health
- Gefördert durch das BMBF im Rahmen des Programms „Wachstumskerne – Unternehmen Region“
- Struktur: je Projektpartner ein Teilprojekt, jeweils unterschiedliche Domänen, Zusammenarbeit bei ähnlichen Aufgabenstellungen, ggf. gemeinsame Nutzung von Microservices und Plattformen

Ziele dieses Vortrags

- Wie kann man mit nur wenig fachlichem Ontologie-Modellierungsaufwand deutlich mehr Nutzen aus der maschinellen Textverarbeitung erzielen
 - Automatisches „Sortieren“ von Dokumenten
 - Automatische Verschlagwortung und Extraktion von Metadaten

(Kombination von Verfahren des maschinellen Lernens mit semantischen Ansätzen, um den Aufwand für das maschinelle Lernen zu reduzieren)?

Inhalt

- Use Cases
- Named Entity Recognition
- spaCy: Natural Language Processing für NER
- Microsoft Cortex / Syntex mit Taxonomien
- Holmes: spaCy mit Ontologie
- Beispiel-Ontologie

Use Cases

- Filtern anhand von mehrdimensionalen Metadaten
- Facettierte Suche anhand von mehrdimensionalen Metadaten
- Dokumentenklassifikation (Rechnung, Stundenzettel,...)
- Extraktion von Metadaten
- Chat Bots (Verarbeitung natürlicher Sprache)

Use Case: Facettierte Suche

The screenshot shows the Microsoft Teams search interface. At the top, there is a search bar with the text 'Suche P' and a search icon. Below the search bar, there are several filters on the left side: 'FileType' (Set to 'Select value'), 'Sachbegriff' (with a list of categories like 'Café', 'Chausseehaus', etc.), 'Ort', 'Architekt', and 'Bauherr'. The main search results area shows 37 results for the query 'Gebäude'. The results are displayed in a grid format, with each result showing a document icon, a title, and the author's name and date. The results include documents like 'Denkmalschutz Potsdam', 'Microsoft Word - Die 10 häufig...malschutz', 'nowawes.pdf', 'weberhaeuser.pdf', 'ACDSee PDF Image', and '2012_ambrosius_horn_architect...rchia.pdf'.

- Im Team
- In Teams
- In Teams mit dem gleichen Prozessziel
- In Teams mit der gleichen Rolle

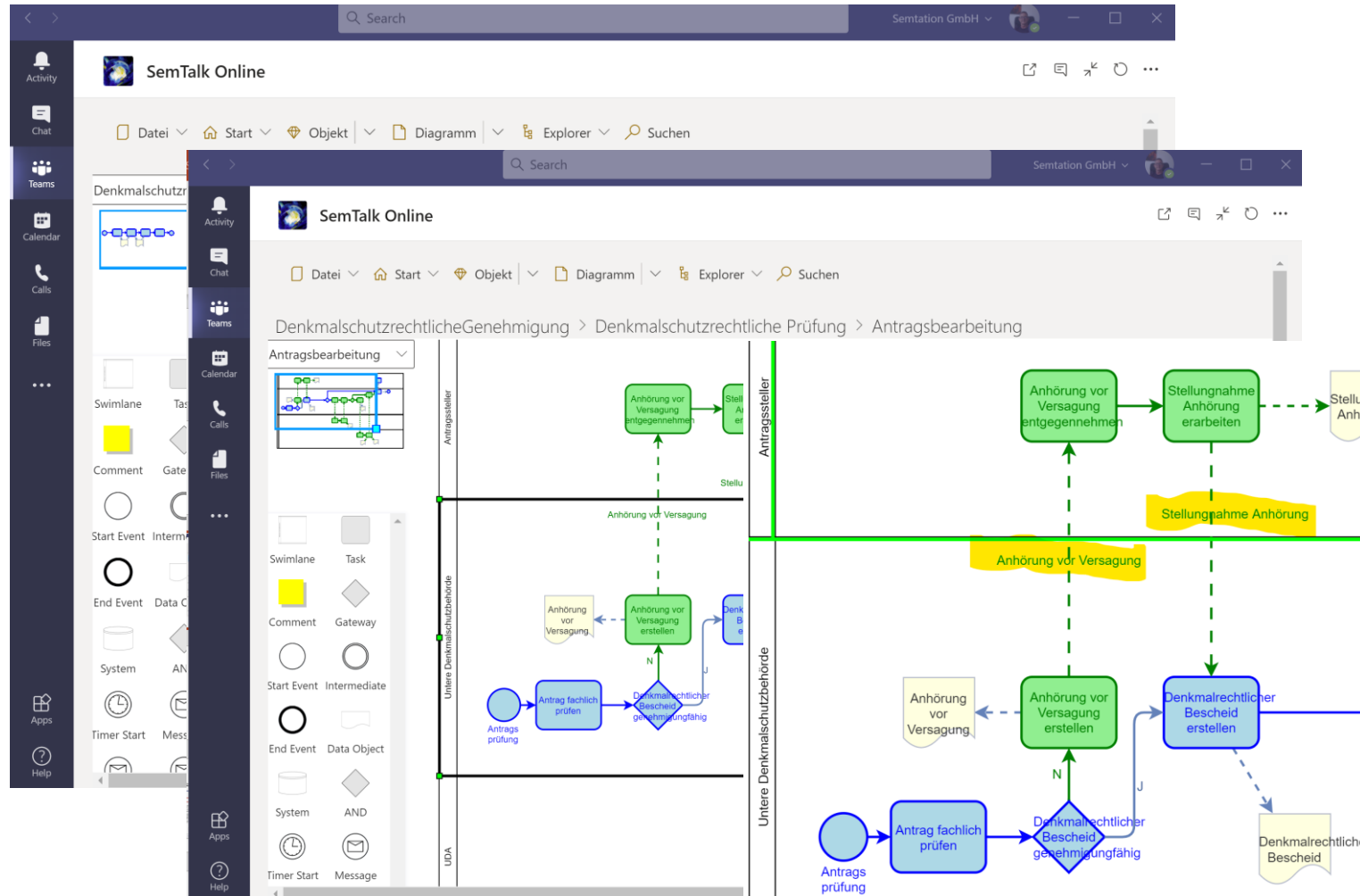
Beispiel Bauamt

- Genehmigungsprozess:
 - Dokumente / Objekte in der Denkmalschutzprüfung
 - Dokumente / Objekte in der Brandschutzprüfung
 - Dokumente / Objekte in der Statik Prüfung
 - Suche in vorherigen Bauanträgen nach diversen Kriterien

Standard-Geschäftsprozess

- Antragsprozess: Stellen, Prüfen, Bearbeiten, Überwachen
 - Formular zur Antragsstellung
 - cit Intelliform
 - Bot
 - Mitgeltende Dokumente, Vorlagen, Gesetze
- Bei jeder Ausführung des Prozesses entstehen Dokumente, Nachrichten, Chats, Protokolle, Meetings, erledigte ToDo Aufgaben, die wir gerne wiederfinden würden

Prozessmodell



- Im Prozessmodell ist beschrieben welche Daten / Dokumente von wem erzeugt werden
- Die Entscheidungen erfolgen aufgrund der Daten / Dokumente

Kann man aus Erfahrungen ähnlicher Projekte lernen und wie findet man die?

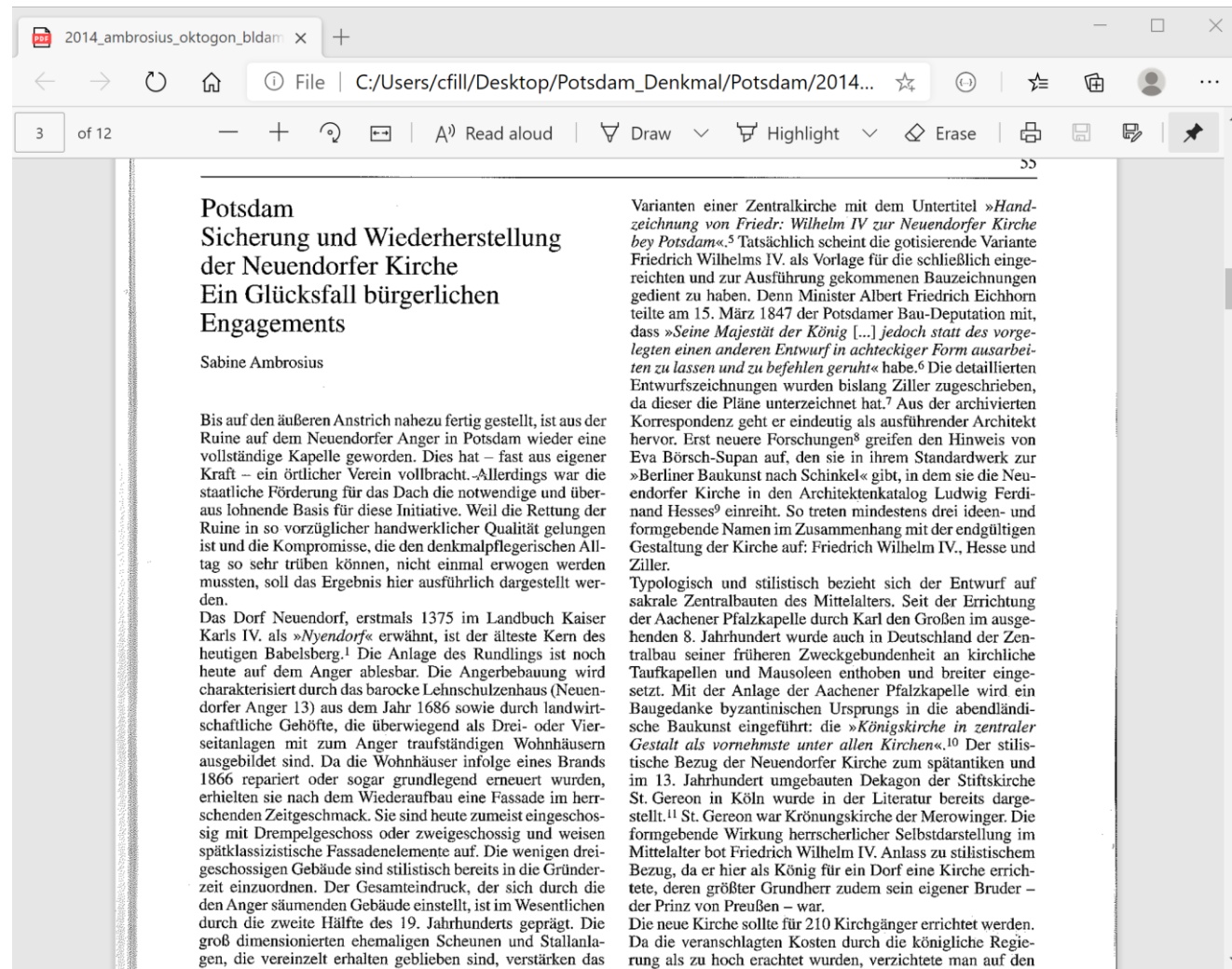
- Wilhelminenhof: Umzug einer Hochschule (HTW) in ein Industriedenkmal der AEG
- Siemensstadt 2.0: Innovations-Campus in einem Industriedenkmal der Siemens AG
- Flughafen Tegel: Umzug der Beuth-Hochschule in ein Industriedenkmal
- Zukunftsort Flughafen Berlin Tempelhof

Potentiale für Lösungen/Prozessverbesserungen mit KI

- Unterstützung für den Antragsteller
 - Beispiele zu „ähnlichen“ Anträgen
 - Hintergrundinformation, Merkblätter
 - Bot: Fachterminologie in „einfache Sprache“ umsetzen
- Unterstützung für den Antragsbearbeiter
 - Empfehlung an Entscheidungspunkten wie mit „ähnlichen“ Anträgen bisher verfahren wurde
 - Semantische Suche in Anträgen
 - Semantische Suche in Hintergrundinformationen

Named Entity Recognition

Beispieltext von www.potsdam.de



2014_ambrosius_oktogonal_bldam

File | C:/Users/cfill/Desktop/Potsdam_Denkmal/Potsdam/2014...

3 of 12

Read aloud | Draw | Highlight | Erase

Potsdam Sicherung und Wiederherstellung der Neuendorfer Kirche Ein Glücksfall bürgerlichen Engagements

Sabine Ambrosius

Bis auf den äußeren Anstrich nahezu fertig gestellt, ist aus der Ruine auf dem Neuendorfer Anger in Potsdam wieder eine vollständige Kapelle geworden. Dies hat – fast aus eigener Kraft – ein örtlicher Verein vollbracht. Allerdings war die staatliche Förderung für das Dach die notwendige und überaus lohnende Basis für diese Initiative. Weil die Rettung der Ruine in so vorzüglicher handwerklicher Qualität gelungen ist und die Kompromisse, die den denkmalpflegerischen Alltag so sehr trüben können, nicht einmal erwogen werden mussten, soll das Ergebnis hier ausführlich dargestellt werden.

Das Dorf Neuendorf, erstmals 1375 im Landbuch Kaiser Karls IV. als »*Nyendorf*« erwähnt, ist der älteste Kern des heutigen Babelsberg.¹ Die Anlage des Rundlings ist noch heute auf dem Anger ablesbar. Die Angerbebauung wird charakterisiert durch das barocke Lehnshulzenhaus (Neuendorfer Anger 13) aus dem Jahr 1686 sowie durch landwirtschaftliche Gehöfte, die überwiegend als Drei- oder Vierseitanlagen mit zum Anger traufständigen Wohnhäusern ausgebildet sind. Da die Wohnhäuser infolge eines Brands 1866 repariert oder sogar grundlegend erneuert wurden, erhielten sie nach dem Wiederaufbau eine Fassade im herrschenden Zeitgeschmack. Sie sind heute zumeist eingeschossig mit Drempelgeschoss oder zweigeschossig und weisen spätklassizistische Fassadenelemente auf. Die wenigen dreigeschossigen Gebäude sind stilistisch bereits in die Gründerzeit einzuordnen. Der Gesamteindruck, der sich durch die den Anger säumenden Gebäude einstellt, ist im Wesentlichen durch die zweite Hälfte des 19. Jahrhunderts geprägt. Die groß dimensionierten ehemaligen Scheunen und Stallanlagen, die vereinzelt erhalten geblieben sind, verstärken das

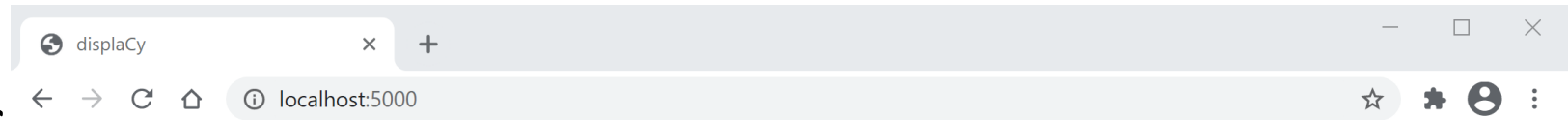
Varianten einer Zentralkirche mit dem Untertitel »*Handzeichnung von Friedr. Wilhelm IV zur Neuendorfer Kirche bey Potsdam*«. ⁵ Tatsächlich scheint die gotisierende Variante Friedrich Wilhelms IV. als Vorlage für die schließlich eingereichten und zur Ausführung gekommenen Bauzeichnungen gedient zu haben. Denn Minister Albert Friedrich Eichhorn teilte am 15. März 1847 der Potsdamer Bau-Deputation mit, dass »*Seine Majestät der König [...] jedoch statt des vorgelegten einen anderen Entwurf in achteckiger Form ausarbeiten zu lassen und zu befehlen geruht*« habe. ⁶ Die detaillierten Entwurfszeichnungen wurden bislang Ziller zugeschrieben, da dieser die Pläne unterzeichnet hat. ⁷ Aus der archivierten Korrespondenz geht er eindeutig als ausführender Architekt hervor. Erst neuere Forschungen ⁸ greifen den Hinweis von Eva Börsch-Supan auf, den sie in ihrem Standardwerk zur »*Berliner Baukunst nach Schinkel*« gibt, in dem sie die Neuendorfer Kirche in den Architektenkatalog Ludwig Ferdinand Hesses ⁹ einreicht. So treten mindestens drei ideen- und formgebende Namen im Zusammenhang mit der endgültigen Gestaltung der Kirche auf: Friedrich Wilhelm IV., Hesse und Ziller.

Typologisch und stilistisch bezieht sich der Entwurf auf sakrale Zentralbauten des Mittelalters. Seit der Errichtung der Aachener Pfalzkapelle durch Karl den Großen im ausgehenden 8. Jahrhundert wurde auch in Deutschland der Zentralbau seiner früheren Zweckgebundenheit an kirchliche Taufkapellen und Mausoleen entthoben und breiter eingesetzt. Mit der Anlage der Aachener Pfalzkapelle wird ein Baugedanke byzantinischen Ursprungs in die abendländische Baukunst eingeführt: die »*Königskirche in zentraler Gestalt als vornehmste unter allen Kirchen*«. ¹⁰ Der stilistische Bezug der Neuendorfer Kirche zum spätantiken und im 13. Jahrhundert umgebauten Dekagon der Stiftskirche St. Gereon in Köln wurde in der Literatur bereits dargestellt. ¹¹ St. Gereon war Krönungskirche der Merowinger. Die formgebende Wirkung herrscherlicher Selbstdarstellung im Mittelalter bot Friedrich Wilhelm IV. Anlass zu stilistischem Bezug, da er hier als König für ein Dorf eine Kirche errichtete, deren größter Grundherr zudem sein eigener Bruder – der Prinz von Preußen – war.

Die neue Kirche sollte für 210 Kirchgänger errichtet werden. Da die veranschlagten Kosten durch die königliche Regierung als zu hoch erachtet wurden, verzichtete man auf den

Key Phrases

```
C:\Users\cfill\Desktop\Potsdam_D
File Edit Search View Encoding Lan
2014_ambrosius_oktagon_bldam.pdf
1 {
2 "KeyPhrases": [
3 "Neuendorfer Kirch
4 "Neuendorfer Anger
5 "Potsdam Sicherung
6 "Friedrich Wilhelm
7 "Bau der Kirche",
8 "Französische Kirc
9 "Anger säumenden G
10 "Restaurierung",
11 "Anger traufständi
12 "gotisierende Vari
13 "Landbuch Kaiser K
14 "späteren Kaiser W
15 "Entwürfen Karl Fr
16 "Wiederherstellung
17 "Glücksfall bürger
18 "Fachwerkkirche",
19 "Sabine Ambrosius"
20 "Minister Albert F
21 "Rettung der Ruine
22 "Christian Nülken"
23 "Thomas Schubert",
24 "Hand des Königs",
25 "Mitte des Angers"
26 "Potsdamer Bau",
27 "Majestät der Köni
28 "Königs Wusterhaus
29 "Thomas Sander",
30 "Baufälligkeit ent
31 "Form des Neubaus"
32 "BRANDENBURGISCHE
33 "Carsten Schneider
34 "Ilona Schneider",
35 "dimensionierten e
36 "wenigen dreigesch
37 "Regierungsbauinsp
38 "erstmaß",
39 "Dach",
40 "Katja Müller",
41 "Untertitel",
42 "Handzeichnung",
43 "Vorlage",
44 "..."
```



Bis auf den äußeren Anstrich nahezu fertig gestellt, ist aus der Ruine auf dem **Neuendorfer Anger LOC** in **Potsdam LOC** wieder eine vollständige Kapelle geworden. Dies hat — fast aus eigener Kraft — ein örtlicher Verein vollbracht. -Allerdings war die staatliche Förderung für das Dach die notwendige und überaus lohnende Basis für diese Initiative. Weil die Rettung der Ruine in so vorzüglicher handwerklicher Qualität gelungen ist und die Kompromisse, die den denkmalpflegerischen Alltag so sehr trüben können, nicht einmal erwogen werden mussten, soll das Ergebnis hier ausführlich dargestellt werden. Das Dorf **Neuendorf LOC**, erstma³ß 1375 im Landbuch Kaiser **Karls IV. PER** als » **Nyendorfi < MISC** erwähnt, ist der älteste Kern des heutigen **Babelsberg LOC**. Die Anlage des **Rundlings LOC** ist noch heute auf dem Anger ablesbar. Die Angerbebauung wird charakterisiert durch das **barocke MISC** Lehnshulzenhaus (**Neuendorfer Anger 13 LOC**) aus dem **Jahr 1686 MISC** sowie durch landwirtschaftliche Gehöfte, die überwiegend als Drei- oder Vierseitenanlagen mit zum Anger traufständigen Wohnhäusern ausgebildet sind. Da die Wohnhäuser infolge eines **Brands LOC** 1866 repariert oder sogar grundlegend erneuert wurden, erhielten sie nach dem Wiederaufbau eine Fassade im herrschenden **Zeitgeschmack LOC**. Sie sind heute zumeist eingeschossig mit Drempeigeschoss oder zweigeschossig und weisen spätklassizistische Fassadenelemente auf. Die wenigen dreigeschossigen Gebäude sind stilistisch bereits in die Gründerzeit einzuordnen. Der Gesamteindruck, der sich durch die den Anger säumenden Gebäude einstellt, ist im Wesentlichen durch die zweite Hälfte des 19. Jahrhunderts geprägt. Die groß dimensionierten ehemaligen **Scheunen LOC** und Stallanlagen, die vereinzelt erhalten geblieben sind, verstärken das Bild des **Neuendorfer**

re

Named Entity Recognition (NER) Out-of-the-Box)

```
{
  "Name": "St. Nikolaikirche",
  "Type": "Organization",
  "SubType": null,
  "WikipediaLanguage": null,
  "WikipediaId": null,
  "WikipediaUrl": null,
  "BingId": null,
  "Matches": [
    {
      "Offset": 3189,
      "Length": 17,
      "EntityTypeScore": 0.95
    }
  ]
},
```

```
{
  "Name": "Hans Georg III",
  "Type": "Person",
  "SubType": null,
  "WikipediaLanguage": null,
  "WikipediaId": null,
  "WikipediaUrl": null,
  "BingId": null,
  "Matches": [
    {
      "Offset": 2812,
      "Length": 14,
      "EntityTypeScore": 1.0
    },
    {
      "Offset": 2892,
      "Length": 14,
      "EntityTypeScore": 1.0
    },
    {
      "Offset": 3266,
      "Length": 14,
      "EntityTypeScore": 1.0
    }
  ]
},
```

- 40 Beispieldokumente zum Denkmalschutz von www.potsdam.de
- Jedes wird einzeln betrachtet

Name
2010_ambrosius_drachenberg_das_ehemalige_untersuchungsgefaengnis_der.pdf.docx
2010_heft_2_brandenburgische_denkmalpflege_ambrosius_schneider_villa_henckel.pdf.docx
2012_ambrosius_horn_architectura_minora_-_chausseehaeuser_in_der_neumark_nowa_marchia.p...
2012_kulturstrassen_ambrosius-horn_chausseehaeuser_-_meilensteine.pdf.docx
20120209141544_0.pdf.docx
2014_ambrosius_oktogonal_bldam.pdf.docx
2018_ambrosius_konservierung_ohne_restaurierung.die_debatte_ueber_denkmale_mit_quellench...
51-kalesse.pdf_offene_kirchen_2020.pdf.docx
DNK-Denkmaeler-Privateigentum-Bd-59.pdf.docx
1bmalkoch.pdf.docx
1_aufsatz_gg.pdf.docx

Fazit: Ohne zusätzliches Training werden nur wenig aussagekräftige Ergebnisse erzielt, also man kann mit den Ergebnissen nicht viel anfangen (es gibt nur wenige Standardklassen von erkannten Objekten)

HiDA statt Wikipedia

The screenshot shows a web browser window displaying the search results for 'Potsdam' on the HiDAweb BLDAM Brandenburg website. The page title is 'Denkmale in Brandenburg' with a sub-header 'Stand: 09.10.2020'. The search criteria are 'Suche: Potsdam in Ort'. The results are sorted by 'Sachbegriff' and show 2320 documents. The left sidebar lists various filters: Landkreis (Potsdam, 2320), Ort (Potsdam, 2320), sorbisch, Gemeinde (Potsdam, 2320), Straße Hausnr. (Ahornstraße 28, 30, 32, Albert-Einstein-Straße 1, Albert-Einstein-Straße 1 a, mehr), Sachbegriff (Akzisemauer, Allee, Altersheim, Amtshaus, Apotheke, mehr), and Künstlername (Ackermann, Adolf; Adler, Friedrich; Aeppli, ?; Ahlins, Dieter). The main content area displays three detailed entries for monuments in Potsdam, including their location, street address, and historical description.

Denkmale in Brandenburg
Stand: 09.10.2020

Facetten-Suche Erweiterte Suche
Suche: **Potsdam** in Ort

Ergebnisse: 2320 Dokumente
Sortieren nach: Sachbegriff Go! Seite: 1 2 3 4 5 ... Nächste

Landkreis
• [Potsdam](#) (2320)

Ort
• [Potsdam](#) (2320)

sorbisch

Gemeinde
• [Potsdam](#) (2320)

Straße Hausnr.
• [Ahornstraße 28](#) (2)
• [Ahornstraße 30](#) (2)
• [Ahornstraße 32](#) (2)
• [Albert-Einstein-Straße 1](#) (2)
• [Albert-Einstein-Straße 1 a](#) (3)
[mehr](#)

Sachbegriff
• [Akzisemauer](#) (1)
• [Allee](#) (3)
• [Altersheim](#) (3)
• [Amtshaus](#) (2)
• [Apotheke](#) (1)
[mehr](#)

Künstlername
• [Ackermann, Adolf](#) (1)
• [Adler, Friedrich](#) (1)
• [Aeppli, ?](#) (1)
• [Ahlins, Dieter](#) (1)

Landkreis: Potsdam
Ort: **Potsdam**
Gemeinde: Potsdam
Sachbegriff: Allee
Ausweis-Art: ED
Denkmalbezeichnung: Lindenallee westlich der Commons des Neuen Palais bis zur Stadtgrenze
Dokumentnr: 09155823

Landkreis: Potsdam
Ort: **Potsdam**
Gemeinde: Potsdam
Straße Hausnr.: Leipziger Straße 9 | 9 a | 9 b | 9 c | 10 b | Friedrich-Wilhelm-Boelcke-Straße 2 | 4 | 5 | 6 | 7
Sachbegriff: Amtshaus | Verwaltungsgebäude
Baudaten: Datierung 1834-1836
Künstlername: Entwurf (vermutlich) Hampel, Carl
Ausweis-Art: ED
Denkmalbezeichnung: Heeres-Proviantamt mit Magazin 1, Magazin 3, Magazin 5, 7 sowie Dienst- und Amtswohngebäude des Proviantmeisters
Dokumentnr: 09156640
Beschreibung: Dehio Brandenburg, 2012: [PotsLeipz7_8ProviantamtMehlmagazin](#)

Landkreis: Potsdam
Ort: **Potsdam**
Gemeinde: Potsdam
Straße Hausnr.: Friedrich-Wilhelm-Boelcke-Straße 5 | 7
Sachbegriff: Speicher

- [HiDAweb BLDAM Brandenburg: Suche \(gis-bldam-brandenburg.de\)](https://gis-bldam-brandenburg.de)
- 2320 Einträge für Potsdam („Named Entities“)
- 300 Sachbegriffe („Entity Labels“)

Training

```
TRAIN_DATA = [  
    ('Who is Nishanth?', { 'entities': [(7, 15, 'PERSON')] }),  
    ('Who is Kamal Khumar?', { 'entities': [(7, 19, 'PERSON')] }),  
    ('I like London and Berlin.', { 'entities': [(7, 13, 'LOC'), (18, 24, 'LOC')] }) ]
```

Oder: Beispiellisten für
Trainingsdaten einspielen
sofern vorhanden

Bots und Spracherkennung (analog)

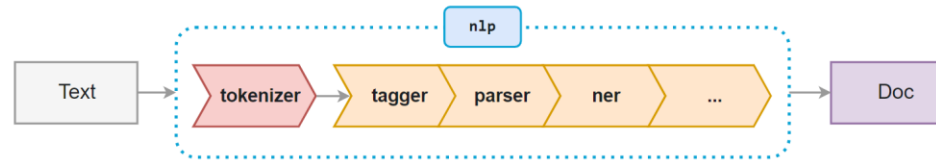
```
JSON
[
  {
    "text": "order a pizza",
    "intentName": "ModifyOrder",
    "entityLabels": [
      {
        "entityName": "Order",
        "startCharIndex": 6,
        "endCharIndex": 12
      }
    ]
  },
  {
    "text": "order a large pepperoni pizza",
    "intentName": "ModifyOrder",
    "entityLabels": [
      {
        "entityName": "Order",
        "startCharIndex": 6,
        "endCharIndex": 28
      },
      {
        "entityName": "FullPizzaWithModifiers",
        "startCharIndex": 6,
        "endCharIndex": 28
      },
      {
        "entityName": "PizzaType",
        "startCharIndex": 14,
        "endCharIndex": 28
      },
      {
        "entityName": "Size",
        "startCharIndex": 8,
        "endCharIndex": 12
      }
    ]
  }
]
```

- Named Entities in Texten und im Chat bzw. Bot
- LUIS:
<https://docs.microsoft.com/en-us/azure/cognitive-services/luis/get-started-get-model-rest-apis?pivots=programming-language-javascript>

Problemspezifische Entities / Tags würden mehr Nutzen stiften

- Listen von Orten
 - Listen von Personen
 - Arten von Denkmälern / Gebäuden
 - Zeitperioden
 - Baustile
-
- Prozesse bei denen die Dokumente erstellt / verwendet wurden
 - Rollen aus diesen Prozessen

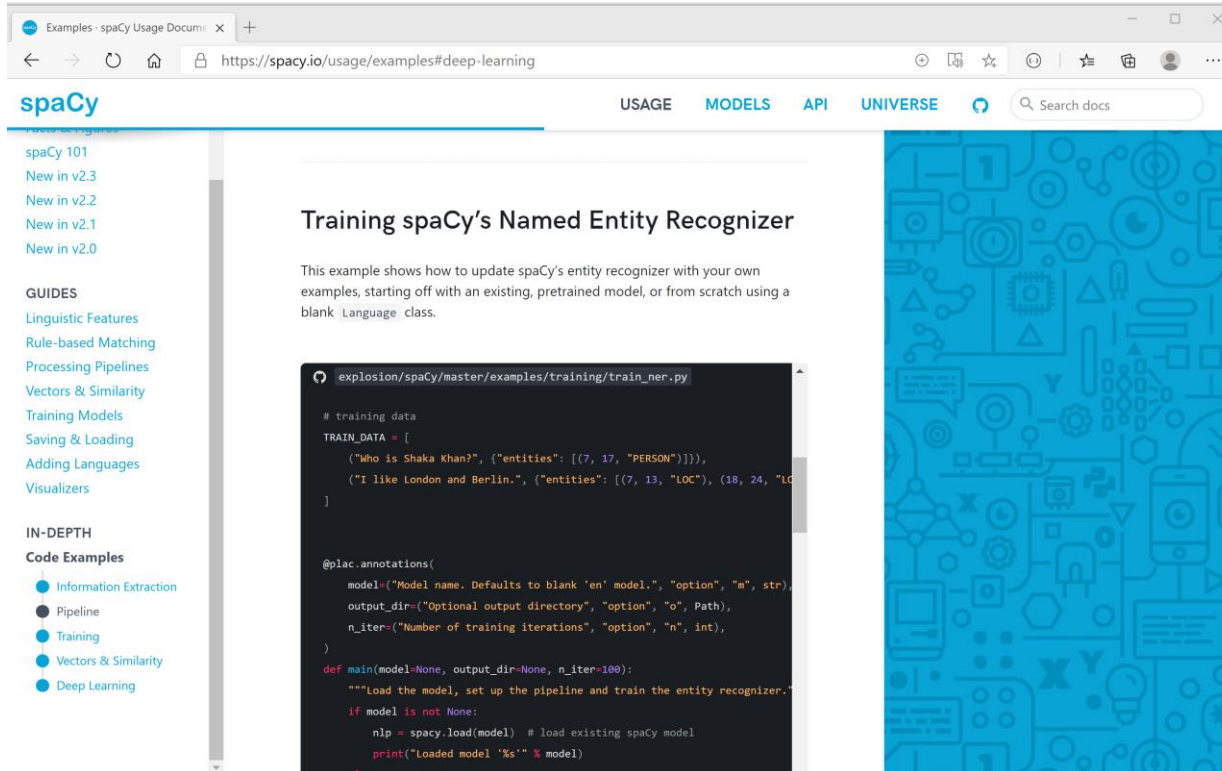
Prozess der NER (hier in der spaCy NLP Pipeline)



NAME	COMPONENT	CREATES	DESCRIPTION
tokenizer	Tokenizer	Doc	Segment text into tokens.
tagger	Tagger	Doc[i].tag	Assign part-of-speech tags.
parser	DependencyParser	Doc[i].head, Doc[i].dep, Doc.sents, Doc.noun_chunks	Assign dependency labels.
ner	EntityRecognizer	Doc.ents, Doc[i].ent_iob, Doc[i].ent_type	Detect and label named entities.
textcat	TextCategorizer	Doc.cats	Assign document labels.
...	custom components	Doc._.xxx, Token._.xxx, Span._.xxx	Assign custom attributes, methods or properties.

The processing pipeline always **depends on the statistical model** and its capabilities. For example, a pipeline can only include an entity recognizer component if the model includes data to make predictions of entity labels. This is why each model will specify the pipeline to use in its meta data, as a simple list containing the component names:

Custom Named Entities



The screenshot shows a web browser window displaying the spaCy documentation page for training a Named Entity Recognizer. The page title is "Training spaCy's Named Entity Recognizer". The main content area contains a code block for training a model. The code defines training data and a function to train the model.

```
explosion/spaCy/master/examples/training/train_ner.py

# training data
TRAIN_DATA = [
    ("Who is Shaka Khan?", {"entities": [(7, 17, "PERSON")]}),
    ("I like London and Berlin.", {"entities": [(7, 13, "LOC"), (18, 24, "LOC")]}),
]

@plac.annotations(
    model=("Model name. Defaults to blank 'en' model.", "option", "m", str),
    output_dir=("Optional output directory", "option", "o", Path),
    n_iter=("Number of training iterations", "option", "n", int),
)
def main(model=None, output_dir=None, n_iter=100):
    """Load the model, set up the pipeline and train the entity recognizer."""
    if model is not None:
        nlp = spacy.load(model) # load existing spaCy model
        print("Loaded model '%s'" % model)
```

- [Custom Named Entity Recognition Using spaCy | by Kaustumbh Jaiswal | Towards Data Science](#)
- [NER @ CLI: Custom-named entity recognition with spaCy in four lines \(codecentric.de\)](#)



displaCy Named Entity Visualizer

When Sebastian Thrun started working on self-driving cars at Google in 2007, few people outside of the company took him seriously. "I can tell you very senior CEOs of major American car companies would shake my hand and turn away because I wasn't worth talking to," said Thrun, now the co-

Model ⓘ
English - en_core_web_sm (v2.3.0)

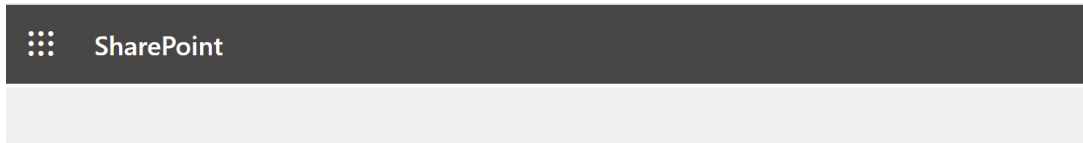
- Entity labels (select all)
- PERSON
 - NORP
 - ORG
 - GPE
 - LOC
 - PRODUCT
 - EVENT
 - WORK OF ART
 - LANGUAGE
 - DATE
 - TIME
 - PERCENT
 - MONEY
 - QUANTITY
 - ORDINAL
 - CARDINAL

When Sebastian Thrun PERSON started working on self-driving cars at Google ORG in 2007 DATE, few people outside of the company took him seriously. "I can tell you very senior CEOs of major American NORP car companies would shake my hand and turn away because I wasn't worth talking to," said Thrun ORG, now the co-founder and CEO of online higher education startup Udacity, in an interview with Recode earlier this week DATE.

A little less than a decade later DATE, dozens of self-driving startups have cropped up while

Syntex

Microsoft SharePoint/Teams - Grundlagen Informationsmanagement



ad.skype_wiki
19:7774edd526ba42818
5484179a7c6bbfd@thre
ad.skype_wiki

Potsdam

Site contents

EDIT LINKS

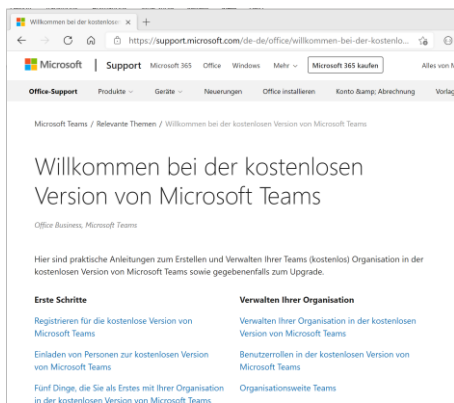
Columns

A column stores information about each document in the document library. The following table lists the columns in the document library:

Column (click to edit)	Type	Required
Title	Single line of text	
Sachbegriff	Managed Metadata	
Architekt	Managed Metadata	
Bauherr	Managed Metadata	
Baustil	Managed Metadata	
Ort	Managed Metadata	
Created	Date and Time	
Modified	Date and Time	
Created By	Person or Group	
Modified By	Person or Group	
Checked Out To	Person or Group	

- [Create column](#)
- [Add from existing site columns](#)
- [Column ordering](#)
- [Indexed columns](#)

- SharePoint
- Bibliotheken / Listen
- Spalten / Lookup Spalten / Metadaten
- Inhaltstypen
- Managed Metadata
- TermStore (Taxonomie-Datenbank zur gruppierten Speicherung von Termsets mit Synonymen, Unter-/Oberbegriffen, Attributen)
- Teams



Models > DenkmalArtikel > DenkmalArtikel Classifier

Label Train Test

Label by selecting text in at least 5 examples. Train after you've labeled.

1/1 of required negative example file

Labeled examples

+ Add example files - Remove

	Name	Label
<input checked="" type="checkbox"/>	2010_heft_2_brandenburgische_denkma...	Positive
<input type="checkbox"/>	2012_ambrosius_horn_architectura_mino...	
<input type="checkbox"/>	2012_kulturstrassen_ambrosius-horn_ch...	
<input type="checkbox"/>	2014_ambrosius_oktogonal_bldam.pdf.docx	Positive
<input type="checkbox"/>	2018_ambrosius_konservierung_ohne_re...	
<input type="checkbox"/>	2075_10_DNK-Denkmaeler-Privateigent...	Positive
<input type="checkbox"/>	Anhörung Versagung Bescheid.docx	Negative
<input type="checkbox"/>	AnhörungVersagung Bescheid.docx	Negative

View original 2010_heft_2_brandenburgische_denkmalpflege_ambrosi...

← Previous file File 1 of 40 Next file →

Is this file an example of DenkmalArtikel? Yes No

50 Sabine Ambrosius, Ilona Schneider und Ulrich Schneider
 Sabine Ambrosius, Ilona Schneider und Ulrich Schneider
 Potsdam Die Villa Henckel 53
 Potsdam Die Villa Henckel 53
 50 Sabine Ambrosius, Ilona Schneider und Ulrich Schneider
 BRANDENBURGISCHE
 PFLEGE
 BRANDENBURGISCHE DENKMALPFLEGE
 Jahrgang 19 2010 Heft 2
 Herausgeber:
 Brandenburgisches Landesamt für Denkmalpflege und
 Archäologisches Landesmuseum
 Verlag Willmuth Arenhövel Berlin
 Martin Petsch
 Georg Frank
 Andreas Kitschke und Bernd Redlich
 Hans Burger und Klaus Ricken

Microsoft Cortex / Syntex

- Classifier:
 - Automatische Erkennung von Inhaltstypen anhand weniger Beispieldokumente mit Hilfe von Beispielsätzen (Document Center)
 - Inhaltstypen definieren strukturierte Datenfelder zu Dokumenten (z.B. Rechnungsdatum, Rechnungsbetrag, Rechnungsart)
 - <https://www.sharepointnutsandbolts.com/2020/12/Project-Cortex-SharePoint-Syntex-Classifier.html>

- Extractor:
 - Automatisches Füllen dieser Datenfelder anhand von Managed Metadata (TermStore)

Out of the box ohne Programmieren, aber nur für "einfache" Szenarien geeignet

Models > DenkmalArtikel > Sachbegriff Extractor

40
Accuracy ⓘ

Label Train Test

Label at least 5 examples. Train after you've labeled.

Labeled examples

+ Add example files - Remove

	Name	Label
<input type="radio"/>	2010_heft_2_brandenburgische_denkmal...	Villa
<input type="radio"/>	2012_ambrosius_horn_architectura_mino...	Chausseehaus
<input type="radio"/>	2012_kulturstrassen_ambrosius-horn_ch...	
<input type="radio"/>	2014_ambrosius_oktogon_bldam.pdf.docx	Kirche
<input checked="" type="radio"/>	2018_ambrosius_konservierung_ohne_re...	Gefängnis
<input type="radio"/>	2075_10_DNK-Denkmaeler-Privateigent...	
<input type="radio"/>	Anhörung Versagung Bescheid.docx	No label
<input type="radio"/>	AnhörungVersagung Bescheid.docx	No label
<input type="radio"/>	Bescheid über Einfriedung.docx	No label
<input type="radio"/>	Bescheid Aufrechterhaltung d...	No label

View original 2018_ambrosius_konservierung_ohne_restaurierung.die_...

← Previous file File 5 of 40 Next file → Save

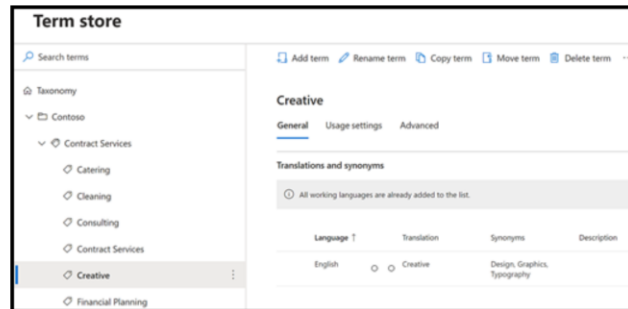
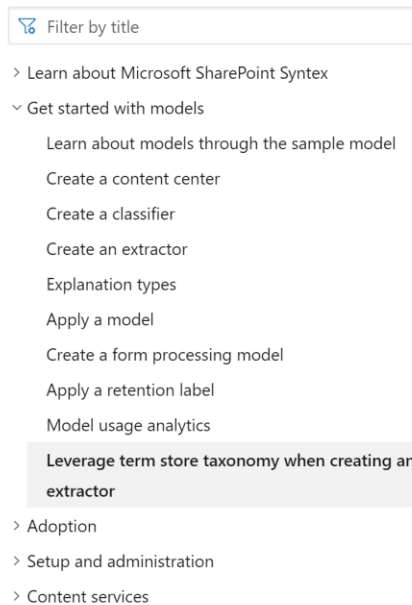
No label Accessibility Mode Off

ertorderliche Rechtsgrundlage zu bewahren. Das gelang allentalls, wenn sie sich etwa auf einem öffentlichen Grundstück befanden oder die Eigentümer selbst einen Wert darin erkannten.

Die ehemals repräsentativen Villen und Landhäuser der aufgelösten Geheimdienststadt wurden schnell wieder in ihr altes Erscheinungsbild überführt. Entstellungen wurden beseitigt und die Spuren der militärischen Nutzung gingen verloren. Heute schmückt das Gebiet mit seinen farbenfrohen Villen zwischen Pflingstberg und Neuem Garten die Berlin-Potsdamer Kulturlandschaft, wie sie bis 1945 gewachsen war und 1991 von der UNESCO zum Welterbe gekürt wurde. Erst seit 2013 markiert ein Geschichtspfad die wenigen Relikte aus der Zeit des Sperrgebiets.7

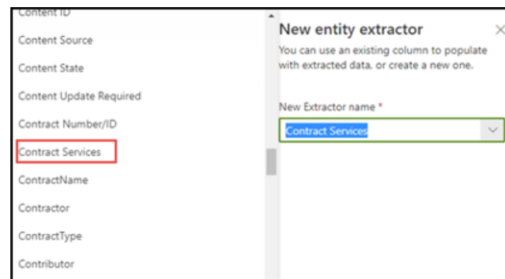
Worin unterschied sich 1994 die Leistikowstraße 1 von den umgebenden Villen? Ganz offensichtlich machte sich der Denkmalwert hier an einer Schicht fest, die an jedem anderen Haus als grobe Verunstaltung des Gebäudes gewertet worden wäre. Aufgrund seiner Nutzung und Deformation als Gefängnis war das Gebäude ein „Ort der Opfer und der Täter.“8 Das Maß der Entstellung unterschied es von den umgebenden Gebäuden und machte es zu einem herausragenden Denkmal mit

Syntex: Leverage term store taxonomy when creating an extractor



There could be a number of reasons why you might want to use a synonym in your term set: outdated terms, renamed terms, or variations between your organizations departments on

To make the managed metadata field available for you to select when you create your extractor, [add it as a managed-metadata site column](#). After you add the site column, it will be available for you to select when you create your extractor for your model.



After applying your model to the document library, when documents are uploaded to library, the extractor will display the preferred term (*Creative*) when the extractor finds any of the synonym values (*D*

- <https://docs.microsoft.com/en-us/microsoft-365/contentunderstanding/leverage-term-store-taxonomy>

- Home
- Sites
- Policies
- Settings
- Content services
 - Term store**
 - Content type gallery
- Migration
- Advanced
 - API access
 - More features
- OneDrive admin center
- Customize navigation

Term store

Search terms

- Denkmalschutz
 - Anhörung
 - Antrag
 - Antragstellung
 - Auflagen
- Bauvorhaben
 - bauliche Anlagen
 - bauliche Maßnahmen
 - Beseitigung eines Denkmals
 - Instandsetzung
 - Modernisierung
 - Sanierung
 - Veränderungen am Erscheinungsbild

Add term Rename term Copy term Move term ...

Bescheid

General Usage settings Navigation Advanced

Translations and synonyms

+ Add

Language ↑	Translation	Synonyms	Description
German	Bescheid		

Feedback

SKOS format reference for SharePoint taxonomy

The following table shows the [SKOS](#) equivalents for the [SharePoint taxonomy](#) vocabulary. SharePoint does not support [SKOS](#) values that have no SharePoint taxonomy equivalent.

SharePoint taxonomy	SKOS equivalent
sharepoint-taxonomy:Term	skos:Concept
sharepoint-taxonomy:TermSet	skos:ConceptScheme
sharepoint-taxonomy:inTermSet	skos:inScheme
sharepoint-taxonomy:hasTopLevelTerm	skos:hasTopConcept
sharepoint-taxonomy:topLevelTermOf	skos:topConceptOf
sharepoint-taxonomy:defaultLabel	skos:prefLabel
sharepoint-taxonomy:termSetName	skos:prefLabel
sharepoint-taxonomy:propertyName	skos:prefLabel
sharepoint-taxonomy:otherLabel	skos:altLabel
sharepoint-taxonomy:description	skos:definition

Holmes: spaCy mit Ontologie

- Alternative z.B. Elastic Search
- <https://medium.com/modern-nlp/semantic-search-fuck-yeah-e371c0f639d>

OVERVIEW

All Projects

PROJECTS

Pipeline

Training

Conversational

Research

Scientific

Visualizers

Containers & APIs

Non-Python

Standalone

Models

EDUCATION

Books

Courses

Videos

Podcasts

A modern named entity visualizer

An open-source virtual assistant for Ubuntu based Linux distributions

EpiTator



Extracts case counts, resolved location/species/disease names, date ranges and more

ExcelCy

Excel Integration with spaCy. Training NER using XLSX from PDF, DOCX, PPT, PNG or JPG.

explacy



A small tool that explains spaCy parse results

gobbi



Deep learning for text classification doesn't have to be scary

gracyql



A thin GraphQL wrapper around spacy

Graphbrain



Automated meaning extraction and text understanding

Holmes



Information extraction from English and German texts based on predicate logic

Kindred

Biomedical relation extraction using spaCy

languagecrunch

lemminflect



+ Connector zu Microsoft Search



Topic Matching Demonstration

This is a demonstration search engine based on the [Holmes](#) library (version 2.2). Holmes works by understanding each query and reading through the loaded search documents to look for places that look as though they deal with the same topic. Unlike other search engines, Holmes analyses the logical relationships between the words in a query and prioritises hits that share these relationships. This means it is important to **use full, grammatical expressions and to take care with spelling.**

Start by selecting example queries from the dropdown box, then go on to enter your own queries into the text box. Select *Info* from the menu or menu bar for advice on writing effective queries as well as for more technical information.

The current search language is German ▾

The loaded search documents are around 350 stories by the Gebrüder Grimm, Hans Christian Andersen, Johann Wilhelm Wolf and Ludwig Bechstein; a maximum of one result per story is returned.

Ein Bauer hat drei Pferde ▾

Ein Bauer hat drei Pferde 25/200 Search

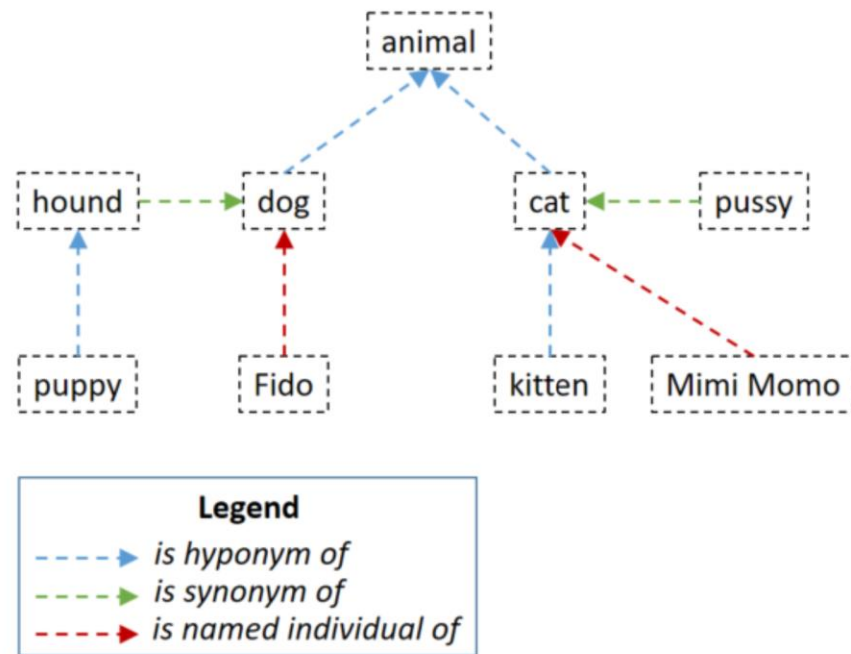
Texts show **single words**; **relations involving two words**; **relations involving three or more words**; and **relevance activation peaks**. All highlighted words have tooltips explaining the match at the individual word level.

Results for query **Ein Bauer hat drei Pferde**:

Rank	Document	Text	Score
1	Gebrüder Grimm - Die kluge Bauerntochter	Und wie sie so daher kam, sagte der König sie hätte das Räthsel getroffen, und es wäre alles erfüllt. Da ließ er ihren Vater los aus dem Gefängnis, und nahm sie bei sich als seine Gemahlin und befahl ihr das ganze königliche Gut an. Nun waren etliche Jahre herum, als der Herr König einmal auf die Parade zog, da trug es sich zu, daß Bauern mit ihren Wagen vor dem Schloß hielten, die hatten Holz verkauft; etliche hatten Ochsen vorgespannt, und etliche Pferde . Da war ein Bauer , der hatte drei Pferde , davon kriegte eins ein junges Füllchen, das lief weg und legte sich mitten zwischen zwei Ochsen, die vor dem Wagen waren. Als nun die Bauern zusammen kamen, fiengen sie an sich zu zanken, zu schmeißen und zu lärmern, und der Ochsenbauer wollte das Füllchen behalten und sagte die Ochsen hätten gehabt ; und der andere sagte nein, seine Pferde hätten gehabt , und es wäre sein. Der Zank kam vor den König, und der that den Ausspruch wo das Füllen gelegen hätte , da sollt es bleiben; und	87.8

[GitHub - msg-systems/holmes-extractor: Information extraction from English and German texts based on predicate logic](#)

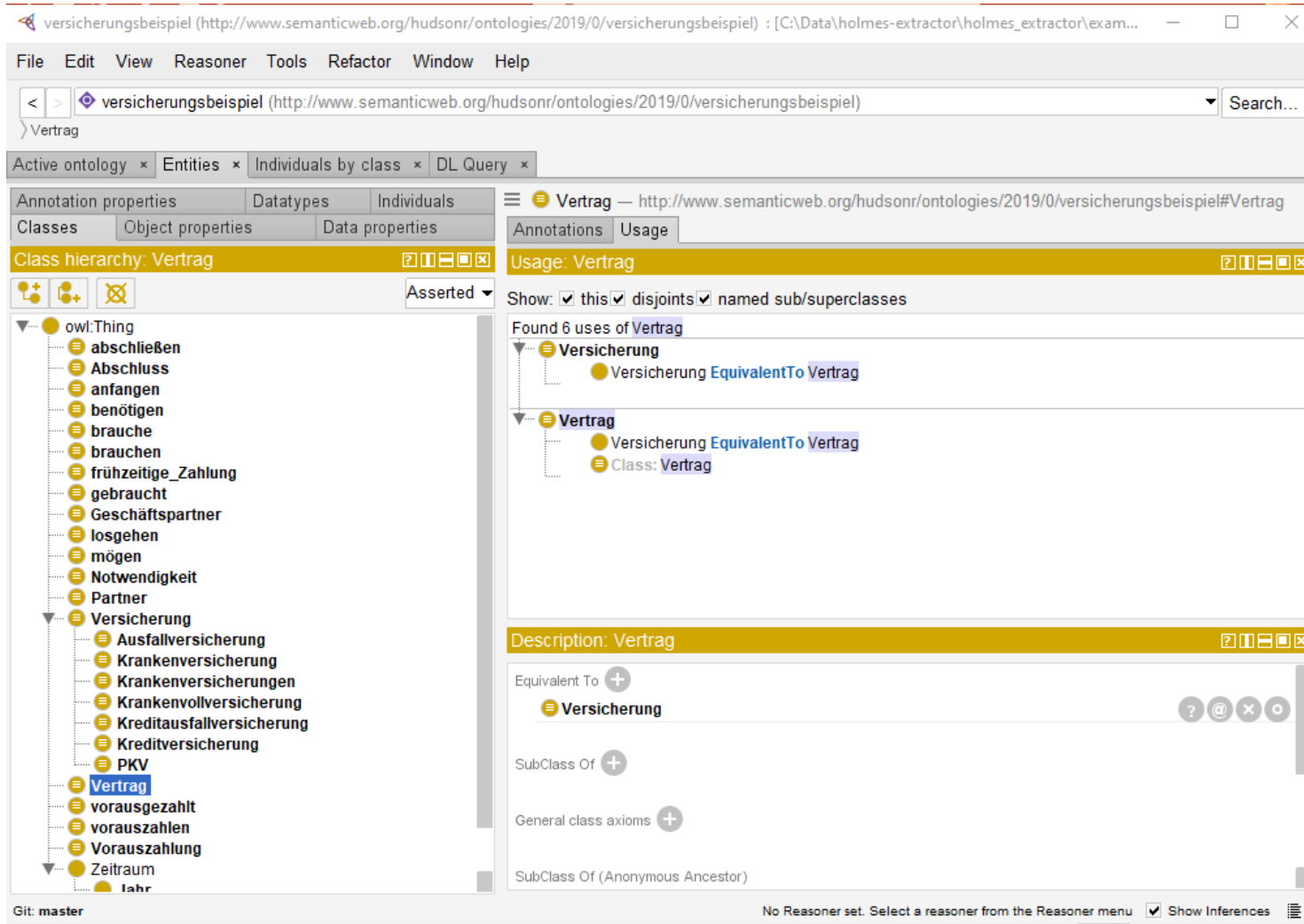
Figure 1: An English ontology example



Ontologies are defined to Holmes using the [OWL ontology standard](#) serialized using [RDF/XML](#). Such ontologies can be generated with a variety of tools. For the Holmes [examples](#) and [tests](#), the free tool [Protege](#) was used. It is recommended that you use Protege both to define your own ontologies and to browse the ontologies that ship with the examples and tests. When saving an ontology under Protege, please select *RDF/XML* as the format. Protege assigns standard labels for the hyponym, synonym and named-individual relationships that Holmes [understands as defaults](#) but that can also be overridden.

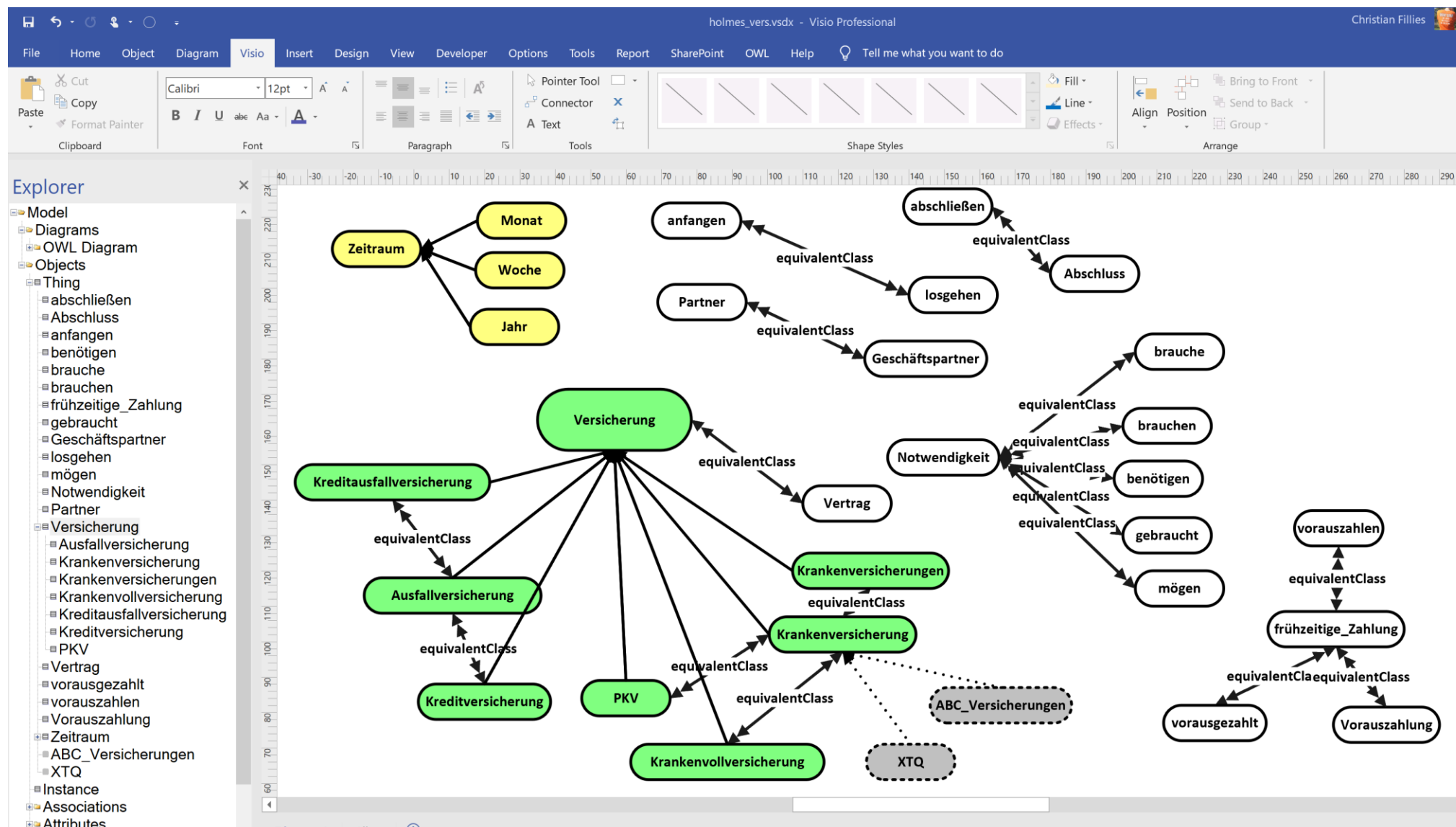
- **Holmes (Richard Hudson, .msg)**
- The **structural matching** ontology is used to analyse the content of both training and test documents. Each word from a document that is found in the ontology is replaced by its most general hypernym ancestor.
- It is important to realise that an ontology is only likely to work with structural matching for supervised document classification if it was built specifically for the purpose: such an ontology should consist of a number of separate trees representing the main classes of object in the documents to be classified.
- In the example ontology shown above, all words in the ontology would be replaced with *animal*; in an extreme case with a WordNet-style ontology, all nouns would end up being replaced with *thing*, which is clearly not a desirable outcome!
- The **classification** ontology is used to capture relationships between classification labels: that a document has a certain classification implies it also has any classifications to whose subtree that classification belongs.
- Synonyms should be used sparingly if at all in classification ontologies because they add to the complexity of the neural network without adding any tangible value; and although it is technically possible to set up a classification ontology to use symmetric matching, there is no sensible reason for doing so.

Holmes Demo-Ontologie in Protegé



- „Standard“-
Ontologie Editor
- Mehr für das
wissenschaftliche
Umfeld geeignet
- W3C Formate
OWL, RDF, SKOS

Holmes Ontologie in SemTalk / Visio



Beispiel: 40 Dokumente zu Denkmälern in
Potsdam

OWL Export

```
C:\Users\cfil\Desktop\Potsdam_Denkmal\potsdam_denkmal2.owl - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
potsdam_denkmal2.owl
558 <rdf:type rdf:resource="#Organization" />
559 <rdfs:label>Archäologisches Landesmuseum</rdfs:label>
560 <rdfs:label xml:lang="de">Archäologisches Landesmuseum</rdfs:label>
561 </owl:NamedIndividual>
562 <owl:NamedIndividual rdf:about="Brandenburgisches_Landesamt_für_Denkmalpflege">
563 <rdf:type rdf:resource="#Organization" />
564 <rdfs:label>Brandenburgisches Landesamt für Denkmalpflege</rdfs:label>
565 <rdfs:label xml:lang="de">Brandenburgisches Landesamt für Denkmalpflege</rdfs:label>
566 </owl:NamedIndividual>
567 <owl:NamedIndividual rdf:about="Ernst_Petzholtz">
568 <rdf:type rdf:resource="#Hofbaumeister" />
569 <rdfs:label>Ernst Petzholtz</rdfs:label>
570 <rdfs:label xml:lang="de">Ernst Petzholtz</rdfs:label>
571 <base:errichtet rdf:resource="#Villa_Henckel" xmlns:base="http://www.test.org#" />
572 </owl:NamedIndividual>
573 <owl:NamedIndividual rdf:about="Villa_Henckel">
574 <rdf:type rdf:resource="#Villa" />
575 <rdfs:label>Villa Henckel</rdfs:label>
576 <rdfs:label xml:lang="de">Villa Henckel</rdfs:label>
577 <base:has_Baustil rdf:resource="#spätklasszistisch" xmlns:base="http://www.test.org#" />
578 <base:steht_in rdf:resource="#Nauener_Vorstadt" xmlns:base="http://www.test.org#" />
579 <base:Adresse rdf:datatype="http://www.w3.org/2001/XMLSchema#string" xmlns:base="http://www.test.org#">Große Weinmeisterstraße 43
</base:Adresse>
</owl:NamedIndividual>
581 <owl:NamedIndividual rdf:about="Untere_Denkmalenschutzbehörde_der_Stadt_Potsdam">
582 <rdf:type rdf:resource="#Organization" />
583 <rdfs:label>Untere Denkmalschutzbehörde der Stadt Potsdam</rdfs:label>
584 <rdfs:label xml:lang="de">Untere Denkmalschutzbehörde der Stadt Potsdam</rdfs:label>
585 </owl:NamedIndividual>
586 <owl:NamedIndividual rdf:about="spätklasszistisch">
587 <rdf:type rdf:resource="#Baustil" />
588 <rdfs:label xml:lang="de">spätklasszistisch</rdfs:label>
589 </owl:NamedIndividual>
590 <owl:NamedIndividual rdf:about="Groß_Glienicke">
591 <rdf:type rdf:resource="#Ortsteil" />
592 <rdfs:label>Groß Glienicke</rdfs:label>
593 <rdfs:label xml:lang="de">Groß Glienicke</rdfs:label>
594 </owl:NamedIndividual>
595 <owl:NamedIndividual rdf:about="Klein_Glienicke">
596 <rdf:type rdf:resource="#Ortsteil" />
597 <rdfs:label>Klein Glienicke</rdfs:label>
598 <rdfs:label xml:lang="de">Klein Glienicke</rdfs:label>
599 </owl:NamedIndividual>
600 <owl:NamedIndividual rdf:about="Jesus_Christus">
601 <rdf:type rdf:resource="#Person" />
602 <rdfs:label>Jesus Christus</rdfs:label>
eXtensible Markup Language file length: 65.697 lines: 1.191 Ln: 1 Col: 1 Pos: 1 Windows (CR LF) UTF-8-BOM IN:
```

- Zum Beispiel für Holmes

TermStore Interface

SemTalk online

The screenshot shows the 'Taxonomy' interface in SemTalk online. On the left, there is a diagram with nodes and arrows, including labels like 'Hof', 'Kape', and 'Friedhof'. On the right, a 'TermSet' is defined as 'Person'. Below it, an 'Export' toggle is turned on. A tree view shows the following structure:

- Location
- Organization
- Person
 - Architekt
 - Hofbaumeister
 - Büring, Johann Gottfried
 - Schinkel, Karl Friedrich
 - Gontard, Carl Philipp Christian von
 - Bauherr
 - König
 - Hans Georg III
 - Hans Georg II
 - Gutsherr
 - Hans Georg III
 - Hans Georg II
 - Paulus

The screenshot shows the 'Term store' interface in the SharePoint admin center. The left sidebar contains navigation options like Home, Sites, Policies, Settings, Content services, Term store, Migration, Advanced, and More features. The main area displays a tree structure of terms:

- Organization
- Person
 - Architekt
 - Büring, Johann Gottfried
 - Gontard, Carl Philipp Ch...
 - Hofbaumeister
 - Schinkel, Karl Friedrich
 - Bauherr
 - Gutsherr
 - Hans Georg II
 - Hans Georg III
 - König
 - Friedrich II
 - Friedrich Wilhelm I
 - Friedrich Wilhelm II

Actions: Add term, Rename term, Copy term, Move term, Delete term

Bauherr

General | Usage settings | Advanced













Translations and synonyms

+ Add

Language ↑	Translation	Synonyms	Description
German	Bauherr		

Managed MetaData (SharePoint)

Potsdam

 Name ▾	Sachbegriff ▾	Architekt ▾	Bauherr ▾	Baustil ▾	Ort ▾
 2010_ambrosius_drachenberg_das_ehem...	Ehemaliges Wohnhaus un				
 2010_heft_2_brandenburgische_denkmal...	Dorfkrche Paretz Villa Henckel	Ernst Petzholtz	Friedrich Wilhelm IV	spätklasszistisch	Nauener Vorstadt
 2012_ambrosius_horn_architectura_mino...	Chausseehaus				
 2012_kulturstrassen_ambrosius-horn_cha...	Chausseehaus				
 20120209141544_0.pdf.docx	Dorfkirche Groß Glienicke		Hans Georg II		
 2014_ambrosius_oktogon_bldam.pdf.docx	Neuendorfer Kirche				
 2018_ambrosius_konservierung_ohne_re...	Ehemaliges Wohnhaus un				
 2020_48-51-kalesse.pdf_offene_kirchen_...	Dorfkirche Groß Glienicke		Hans Georg III	gotisch	Groß Glienicke
 2075_10_DNK-Denkmaeler-Privateigentu...					
 aufsatzgrabmalkoch.pdf.docx	Erbbegrabnis				
 bankmann_aufsatz_gg.pdf.docx					

Taxonomie-geführte Annotierung

.com/teams/denkmalerschutz/Potsdam/Forms/AllItems.aspx

Search this library

dit in grid view Open Share Copy link Download Delete Pin to top Rename

sdam

Name	Sachbegriff	Architekt	Bauherr
2010_ambrosius_drachenberg_das_ahem...	Ehemaliges Wohnhaus un...		
2010_heft_2_brandenburgische_denkmal...	Dorfkrche Paretz Villa Henckel	Ernst Petzholtz	Friedrich Wilhelm IV
2012_ambrosius_horn_architectura_mino...	Chausseehaus		
2012_kulturstrassen_ambrosius-horn_cha...	Chausseehaus		
20120209141544_0.pdf.docx	Dorfkirche Groß Glienicke		Hans Georg II
2014_ambrosius_oktogonal_bl...	Neuendorfer Kirche		
2018_ambrosius_konservierung_ohne_re...	Ehemaliges Wohnhaus un...		
2020_48-51-kalesse.pdf_offene_kirchen_...	Dorfkirche Groß Glienicke		Hans Georg III
2075_10_DNK-Denkmaeler-Privateigentu...			
aufsatzgrabmalkoch.pdf.docx	Erbbegräbnis		
bankmann_aufsatz_gg.pdf.docx			

Sachbegriff

- Haus
- Hof
- Keller
- Kirche**
 - Dorfkirche
 - Dorfkirche Groß Glienicke
 - Dorfkrche Paretz
 - Kapelle
 - Friedhofskapelle
 - Ribbeck-Kapelle
 - Neuendorfer Kirche**
 - Spandauer St. Nikolaikirche
- Remise
- Scheune
- Speicher

Apply **Cancel**

Return to classic

Filter (SharePoint)

SharePoint interface showing a list of documents in a library named "Potsdam". The list is filtered by the "Sachbegriff" (Subject) column. The filter pane on the right shows the selected filter: "Ehemaliges Wohnhaus und späteres KGB-Gefängnis, jetzt Gedenk- und B...".

Navigation bar: + New, Upload, Edit in grid view, Sync, Add shortcut to OneDrive, Export to Excel, Power Apps, Automate, All Documents, Filter (active), Info, Share.

Name	Sachbegriff	Architekt	Bauherr	Baustil
2010_ambrosius_drachenberg_das_ahem...	Ehemaliges Wohnhaus und			
2010_heft_2_brandenburgische_denkmal...	Dorfkrche Paretz Villa Henckel	Ernst Petzholtz	Friedrich Wilhelm IV	spätklassis
2012_ambrosius_horn_architectura_mino...	Chausseehaus			
2012_kulturstrassen_ambrosius-horn_cha...	Chausseehaus			
20120209141544_0.pdf.docx	Dorfkirche Groß Glienicke		Hans Georg II	
2014_ambrosius_oktogonal_bldam.pdf.docx	Neuendorfer Kirche			
2018_ambrosius_konservierung_ohne_re...	Ehemaliges Wohnhaus und			
2020_48-51-kalesse.pdf_offene_kirchen_...	Dorfkirche Groß Glienicke		Hans Georg III	gotisch
2075_10_DNK-Denkmaeler-Privateigentu...				
aufsatzgrabmalkoch.pdf.docx	Erbbegrabnis			
bankmann_aufsatz_gg.pdf.docx				

Sachbegriff

- (Empty)
- Ehemaliges Wohnhaus und späteres KGB-Gefängnis, jetzt Gedenk- und B...
- Dorfkirche Groß Glienicke
- Chausseehaus
- Nedlitzer Fährgut
- [See All](#)

Bauherr

- (Empty)
- Hans Georg II
- Friedrich Wilhelm IV
- Hans Georg III

Ort

- (Empty)
- Neu Fahrland
- Nauener Vorstadt

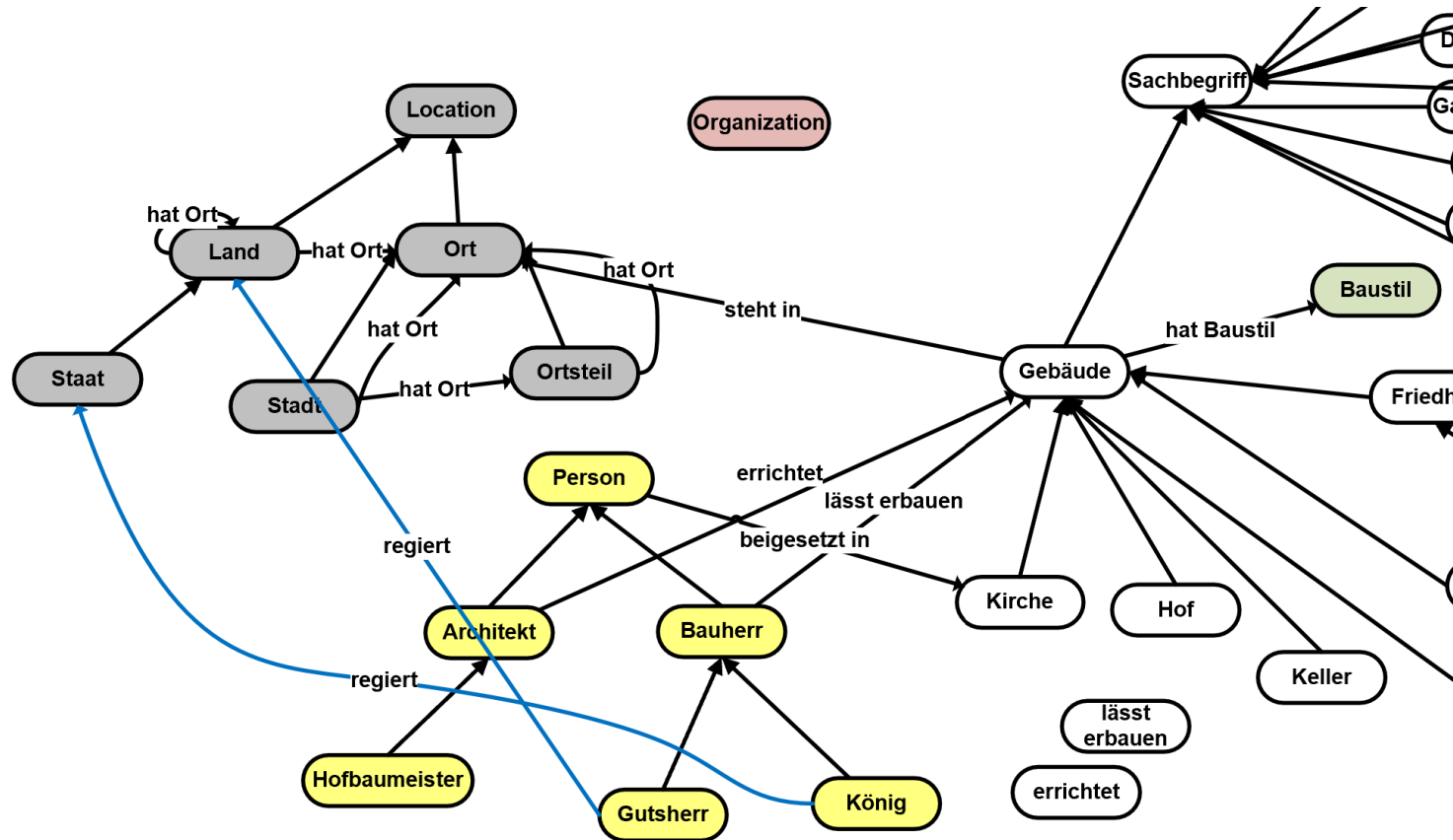
Wiederholung: Facettierte Suche

The screenshot shows a Microsoft Teams search interface. At the top, there is a search bar with the text "Suche P" and a search icon. Below the search bar, there are several filters on the left side:

- FileType**: A dropdown menu with "Select value" selected.
- Sachbegriff**: A list of checkboxes with the following items: Café, Chausseehaus, Dorfkirche Groß Glienicke, Dorfkirche Paretz, Ehemaliges Wohnhaus und späteres KG..., Erbbegräbnis, Nedlitzer Fährgut, Neuendorfer Kirche, and Villa Henckel.
- Ort**: A dropdown menu.
- Architekt**: A dropdown menu.
- Bauherr**: A list of checkboxes with the following items: Friedrich Wilhelm IV, Hans Georg II, and Hans Georg III.

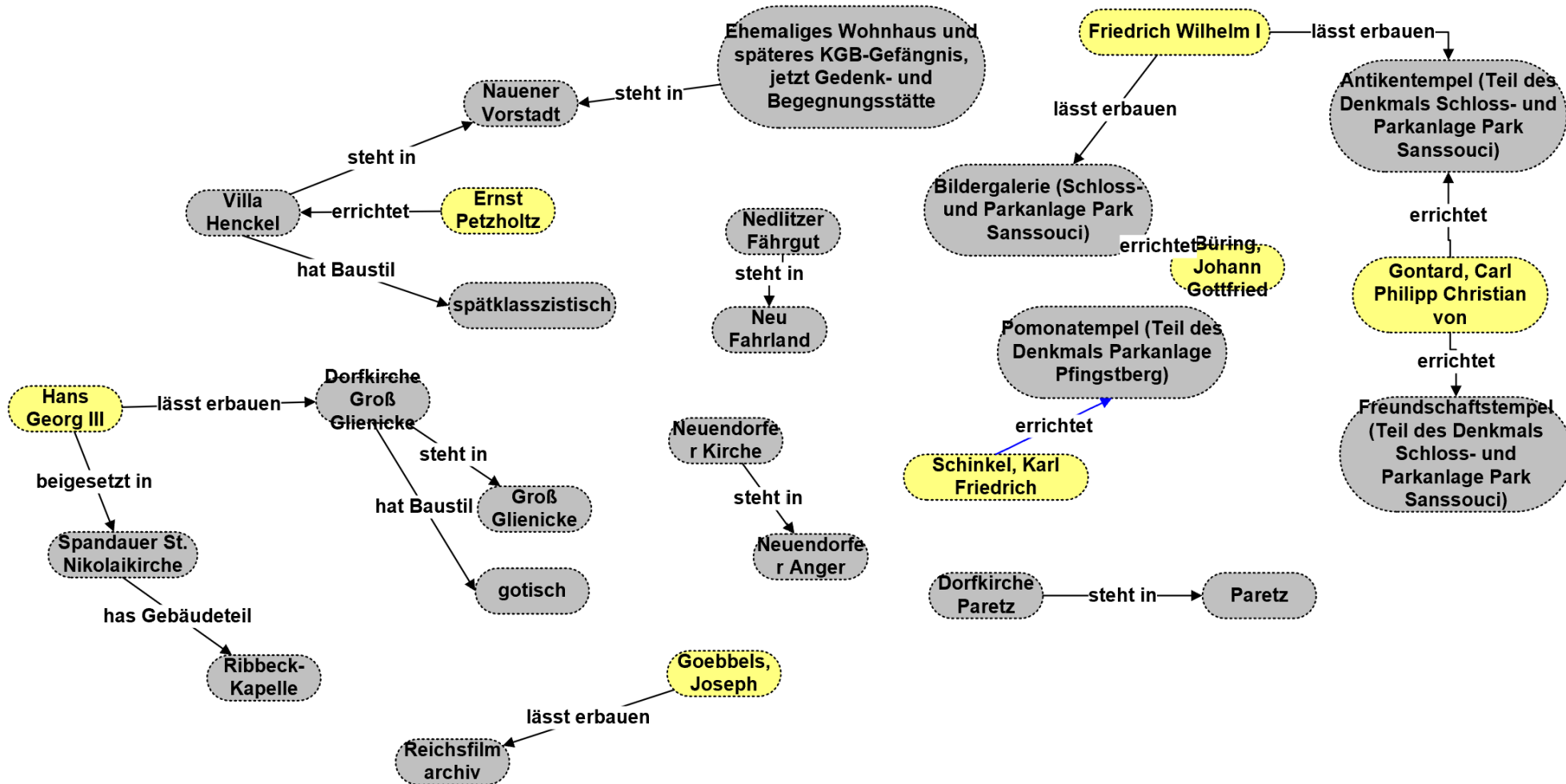
The search results are displayed in a grid of 37 items. Each item is a document thumbnail with a Word icon and the title "Denkmalschutz". The thumbnails are arranged in a 3x3 grid, with the last cell empty. The first row shows results for "Potsdam", "Microsoft Word - Die 10 häufig...malschutz", and "nowawes.pdf". The second row shows results for "weberhaeuser.pdf", "ACDSee PDF Image.", and "2012_ambrosius_horn_architect...rchia.pdf". The third row shows a result for "weberhaeuser.pdf".

Ontologie vs. Taxonomie



- Modellierung von Beziehungen statt Ober- Unterbegriffen
- Vererbungen von Beziehungen
- Objekt-orientierung statt Description Logik

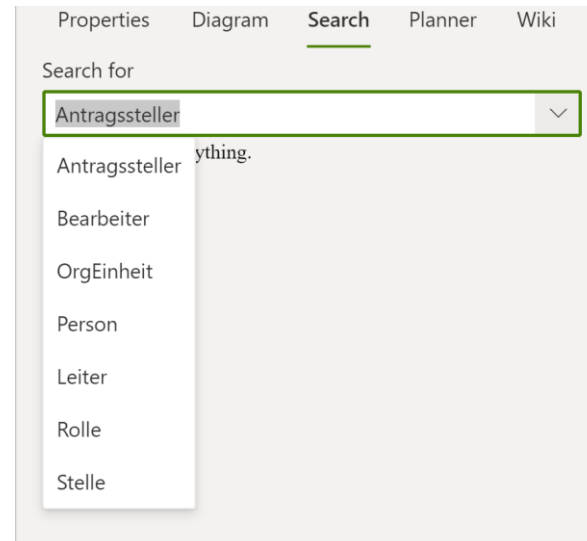
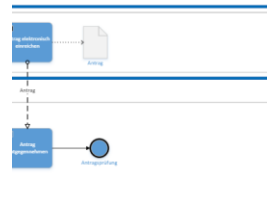
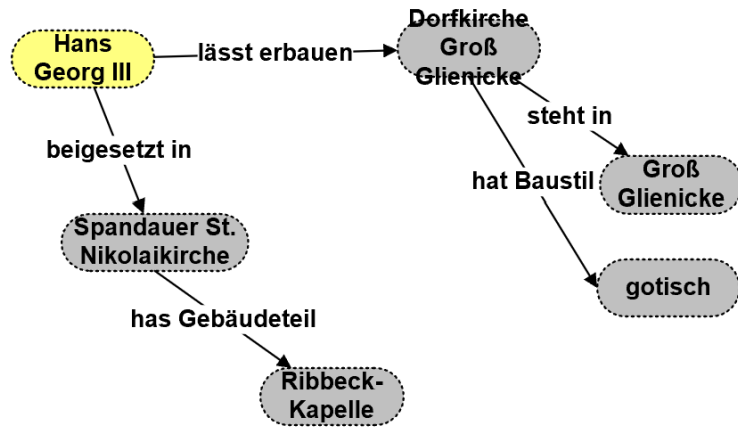
Optionale „Graph-Datenbank“



Auswertungen/
Filterung jenseits des
SharePoint
TermStores

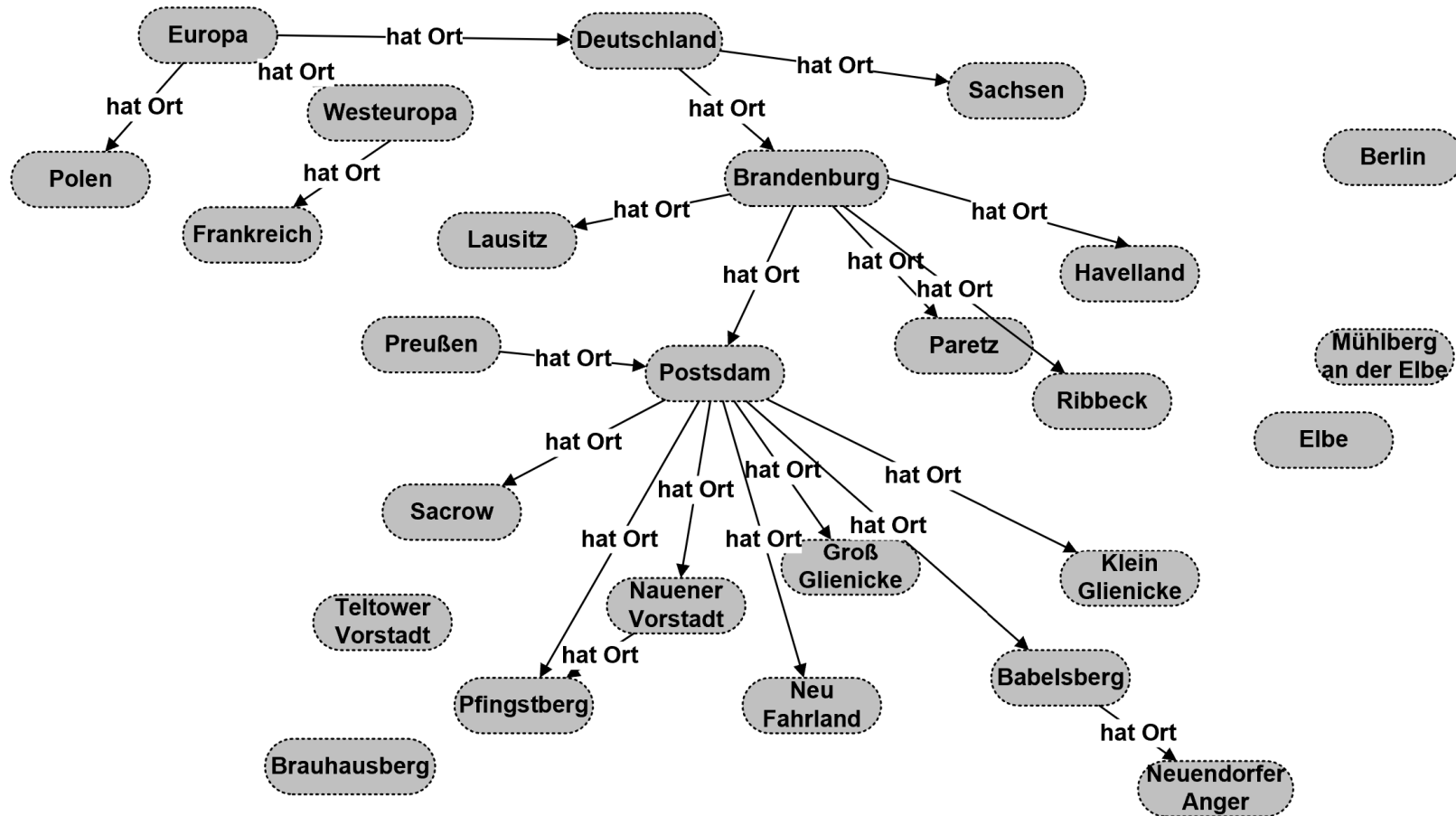
Schnittstelle zu
Graph-
Datenbank
(NoSQL, Neo4j
etc) ermöglicht
Auswertungen

SemTalk Search



- Vorschlagsliste anhand verbundener Objekte aus der Ontologie
- Explizite und implizite Beziehungen
- Query Expansion (Suche „Auto“ dann finde auch „PKW“) oder umgekehrt

„Teil von“ - Hierarchien



Taxonomy

← Import → Export

TermSet

Location

Associate As: *

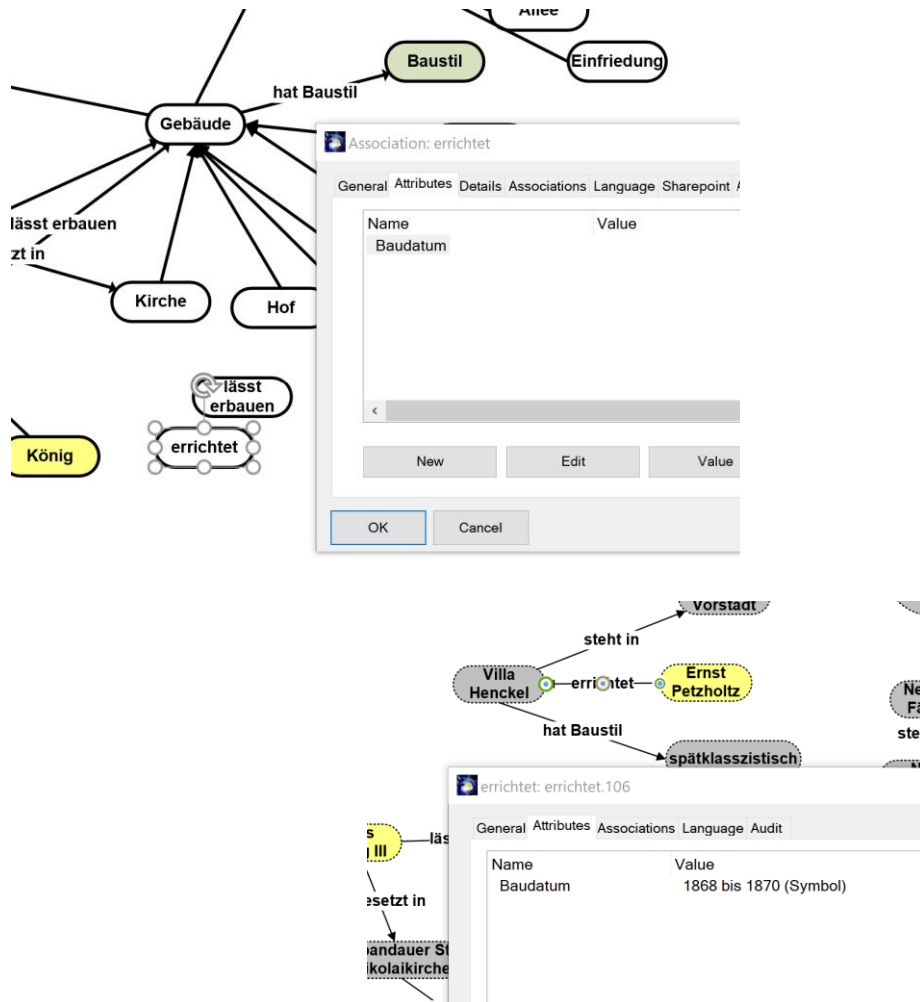
hat Ort

Export



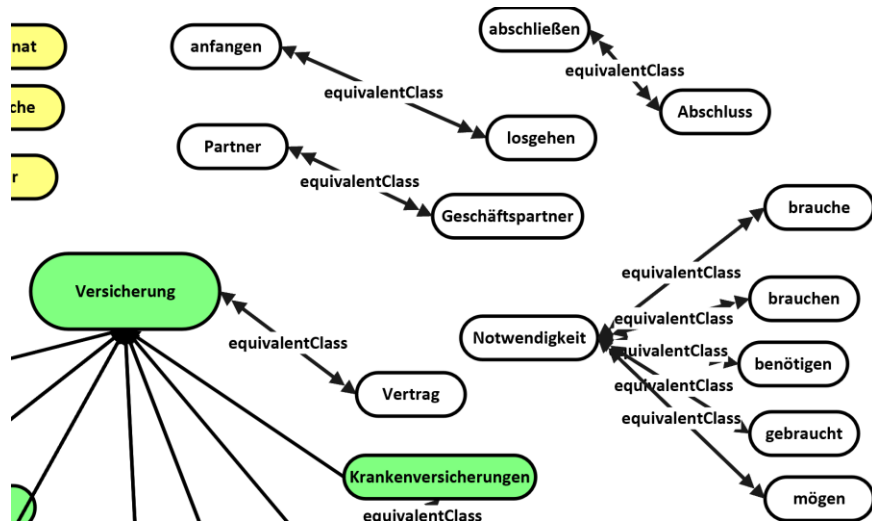
- > Preußen :
- ✓ Europa :
- > Westeuropa :
- ✓ Deutschland :
- ✓ Brandenburg :
- ✓ Potsdam :
- Sacrow
- ✓ Nauener Vorstadt :
- Potsdamer Pfingstberg
- Potsdamer Pfingstberg
- Groß Glienicke
- Klein Glienicke
- Neu Fahrland
- > Babelsberg :
- Paretz

Aspekt-orientierte Ontologien



- Beziehungen oder Eigenschaften sind oft nur aus einer bestimmten Perspektive relevant („Baustil“ vs. „Ladestation verfügbar“)
- Beziehungen oder Eigenschaften sind oft nur in einer bestimmten Zeitspanne gültig („Potsdam liegt in Preußen“)
- Gültigkeitsinformation an Kanten
- Konsistenzchecks: Nur eine Person kann zum Zeitpunkt X König von Y sein. Ohne Zeitfilter ist die Ontologie inkonsistent!
- Schwer im Training von NLP Modellen einsetzbar, da man ja für jede Kombination aktiver Aspekte neu trainieren müsste. TermStore ist genauso schwer denkbar
- Sehr interessant für dynamische Abfragen und Filterszenarien

Gedanken zur Modellierung von Ontologien



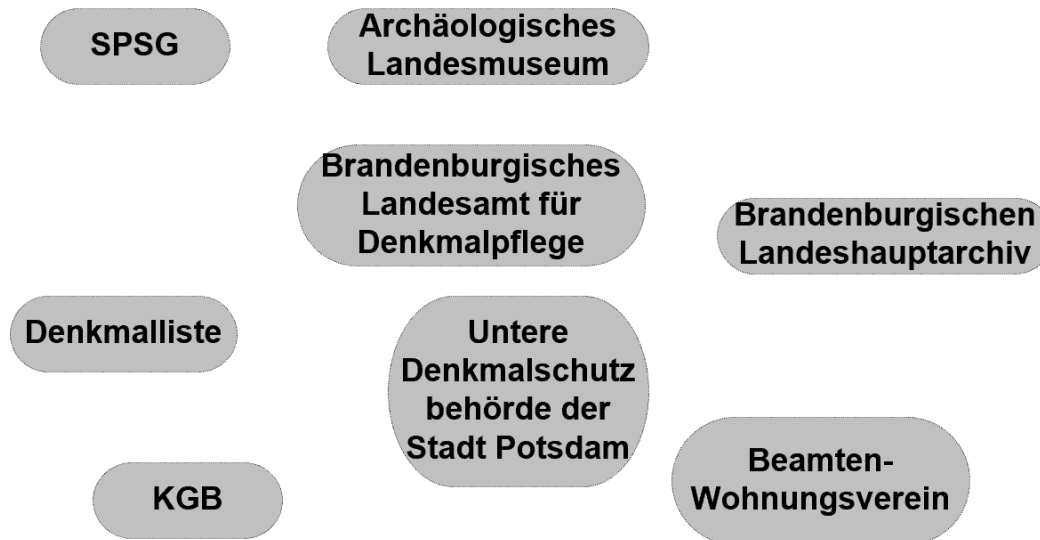
- > Antragstellung
- ✓ Auflagen
 - Aufforderung zur Nachbesserung
 - Auflagen Nachbesserung
 - überwachen
- > Bauvorhaben
- ✓ Bescheid
 - denkmalschutzrechtlicher Bescheid
 - Widerspruch
- > Datenschutz

- Verschiedene orthogonale Dimensionen, nach denen die Dokumente klassifiziert und gefiltert werden sollen
- Vorsicht mit Synonymen und Äquivalenzen. Hier ist die Anwendung sehr wichtig. Im beschränkten Kontext eines Bots kann es richtig sein, dass Vertrag == Versicherung gilt.
- Eindeutige Semantik der „Child Term“ Beziehung:
 - „Widerspruch“ „ist ein“ „Bescheid“?
 - "Auflagen“ „besteht aus“ "überwachen"?
- Geltungsbereich der Taxonomie: Tenant/Teams Template/Team
- Owner / Administrator der Taxonomie

Takeaways

- Entity-Listen und Ontologien verbessern und vereinfachen das Training von Spracherkennungsmodellen im Machine Learning
- Ontologien helfen Konsistenz zu erhalten bei der Annotierung von Dokumenten
- Ontologien sind die Grundlage für die automatische Annotierung von Dokumenten
- Ontologien helfen beim Filtern der Suchergebnisse und bei der Auswahl von Suchbegriffen
- Auch relativ kleine Ontologien lohnen sich für einige Domänen bzw. Prozesse
- Jeder, der Teams bzw. SharePoint einsetzt, kann direkt anfangen

Praxisteil



- Entwickeln Sie selber eine Ontologie zum Thema „Organisation“ die die bisherigen Instanzen besser klassifiziert
- (Mit, Excel, Protegé, SemTalk Online...)

- **Prozessportal**
 - Primäres Medium: (Grafisches) Prozessmodell
 - Publikation und Dokumentation von Vorgehensweisen (Prozessen)
 - Festlegung von Verantwortlichkeiten, Reihenfolgen und Kommunikationswegen
 - (IT-) Werkzeuge, Schemata z.B. für SharePoint Sites, Vorlagen für Workflows, empfohlene Apps usw.
 - Es enthält domänenspezifisches Wissen wie z.B. Taxonomien für SharePoint
 - Wenn das Prozessmodell in Teams veröffentlicht wird, kann man über das Prozessmodell chatten
 - Verbesserungsvorschläge und Feedback zum Prozess selbst oder zum Prozessmodell verwalten
 - Bot informiert über der Prozessmodell
 - MIS und Reporting zur Übersicht über Prozessinstanzen/Projekte/Workflows
- **Prozessmodellierungs- Team**
 - Primäres Medium: Prozessmodellierungstool
 - Unterstützung der Modellierung / Modellkoordination
 - Freigabe
 - Repositories, Vokabulare
 - Domänenspezifische Teams
 - Modellierungsprojekt-Vorgehensweisen / Projektmanagement / Sprints im Modellierungsprojekt
 - Listen, Formulare usw. z.B. für Interviews
 - Sammlung von Information zum Thema
- **Prozessoptimierungs- Team**
 - Auswertung, Simulation, Prozessänderungen („Digitalisierung“, Reifegrade)
 - Umsetzung
- **Prozess/Projekt Team**
 - Primäres Medium: ToDo Liste, Projektplan o.ä.
 - Plattform zu kollaborativen Ausführung eines Projektes/Prozesses
 - Stakeholder werden zu konkreten internen und externen Personen zugeordnet
 - Konkrete Objekte des Prozesses werden im Chat diskutiert und in der Teams-Site in Bibliotheken und Listen gesammelt und bearbeitet
 - Die Teamsite kann teilweise aus dem Prozessmodell provisioniert werden
 - Apps, Listen, Workflows, Managed Meta Data usw. werden in der Teamsite möglichst automatisch installiert
 - Komplexere Projekte können pro Sprint einen Channel mit eigenem Board provisionieren. Eine Seite im Prozessmodell entspricht einem Channel. Ein Board enthält initial die Aufgaben einer Prozessmodellseite
 - Bots und Workflows unterstützen bei der Ausführung des Prozesses
 - Links zum Prozessportal zur Information. Das Prozessmodell wird meist als Goldstandard gesehen von dem aber flexibel abgewichen werden kann
 - Prozess-Wissensmanagement kann Hilfe bei konkreten Fragen bieten aus ähnlichen Projekten bieten
- **Prozess-Wissensmangement**
 - Primäres Medium: Suchcenter
 - Die Menge aller Prozess/Projekt Team Sites bildet den Erfahrungs- bzw. Wissenspool der Organisation in dem die bisherigen Projekte abgelegt sind
 - Durch die Verwendung einer ähnlichen Vorgehensweise, Struktur, Sprache sowie durch die Zuordnung zum Prozessmodell werden die Projekte vergleichbar und auswertbar