



**LIB-IT DMS GmbH**

# **Nutzung von Vector Search in FILERÖ über Spracheingabe**

Dagmar Causley (Geschäftsführende Gesellschafterin)

Jash Jack Inthavong (IU Duales Studium Informatik / Computer Science)



# LIB-IT DMS GmbH – kurz und knapp

- Gegründet 2003
- Sitz in Pleidelsheim (zwischen Stuttgart und Heilbronn)
- Eignerin/Geschäftsführerin: Dagmar Causley
- 22 Mitarbeiter
- Ausbildungsbetrieb
- InterSystems Partner seit 22 Jahren
- **Produkt:** FILERO (Enterprise Information Management System)
- **Buzz words:** BigData, Digitale Transformation + Prozesse, EIMS, DMS, Revisionssichere Archivierung, GoBD, DSGVO, Compliance, Performance



# Was ist FILERO

- FILERO ist ein webbasiertes Enterprise Information Management System (EIMS), mit dem Daten aller Art verwaltet, indexiert und recherchiert werden können.
- Alle in FILERO verwalteten Daten und Dateien können mit verschiedenen Suchmethoden wieder aufgefunden werden, z.B. über die Volltextsuche, Categoriesuche, Feldsuche oder Schnellsuche.
- Es können Daten angelegt, geändert, gespeichert und gelöscht werden. Dateien, die mit Office Programmen erstellt wurden, können Datensätzen hinzugefügt und mitverwaltet werden. Beim Einladen dieser Dateien werden diese ebenfalls komplett indexiert und suchbar gemacht.
- Daten können über Schnittstellen (SOAP API oder RESTful API) oder manuell angelegt werden.



# Was ist FILERO

- Alle in FILERO verwalteten Funktionen sind berechtigungsabhängig, d.h. ein Benutzer kann z.B. nur die Daten suchen, ändern oder löschen, für die er die entsprechende Berechtigung hat. Daten, für die ein Benutzer keine Berechtigung hat, werden diesem Benutzer auch nicht angezeigt.
- Jede Aktion in FILERO, die die Neuanlage oder Änderung eines Datensatzes zur Folge hat, wird in der Versionshistorie des Datensatzes protokolliert.



# Was ist FILERO

**Bewährtes Standard UI:**  
**Links:** Strukturanzeige  
**Rechts:** Datensatzanzeige

Es können unterschiedliche Strukturen angelegt werden, z.B. :

- Kunden-/Lieferantenakte
- Personalakte
- Produktakte
- Messeakte
- Inventarverzeichnis
- Handbücher
- uvm.

2024 © LIB-IT DMS GmbH



# Problem: Suchsyntax (1)

In FILERÖ gibt es verschiedene (und auch komplexe) Suchmöglichkeiten.

Die drei wichtigsten sind:

- Globale Suche (Google<sup>®</sup>-Like-Suche)
- Menügesteuerte Kategoriensuche (=Klassensuche)
- Menügesteuerte Feldsuche

Verwendet wird am häufigsten die Globale Suche.

Für Standardanwender ist eine Suche über Suchbefehle (z.B. Boolesche Operatoren) fast immer zu schwierig:



# Problem: Suchsyntax (2)

z.B. Suche

- Alle GPs, die unter 2024 einen Ordner Eingangsrechnungen haben
  - ❖ `Kunden() contains AdminYear2024(contains Ordner:Ordnerart(Eingangsrechnungen))`
- Alle GPs, bei denen in 2024 eine Person hinzugefügt wurde
  - ❖ `Kunden() contains Person(creationdate(2024))`
- Alle Datensätze anzeigen, in denen DC verlinkt ist
  - ❖ `linksto(RefLIBITMitarbeiter(fid(640002)))`



# Kundenwunsch

Einfach suchen  
über Spracheingabe  
und „natürliche“ Suchanfragen



# Suchbeispiele

- Geschäftspartner Taunusstein
- Geschäftspartner aus Bayern
- Geschäftspartner hat keinen Mandanten Status
- Suche Personen, die in München wohnen
- Suche nach Patienten
- Weiblich
- Patienten München



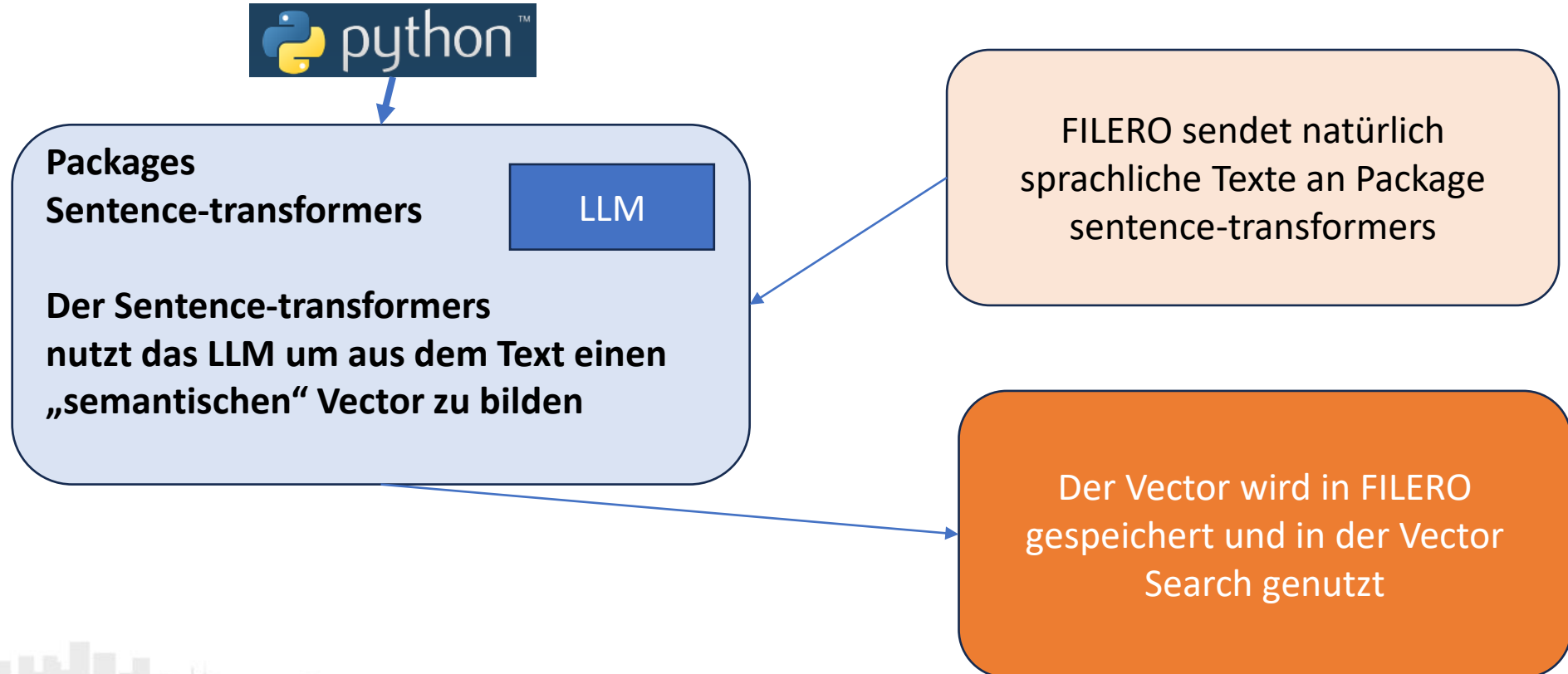
# Wie ist das technisch gelöst? (1)

- **IRIS for Windows (x86-64) 2024.3.0 VECANN**
- **FILERO 2024.10**
- **Hinweis:**  
mit IRIS Version 2024.1 hatten wir erhebliche Performanceprobleme



# Wie ist das technisch gelöst? (2)

## Architecture





# Wie ist das technisch gelöst? (3)

1. Folgende Klassen wurden für diesen Test vektorisiert
  - GP (3.789 Datensätze)
  - Adresse (3.334 Datensätze)
  - Person (2.504 Datensätze)
- Zuwachs der Datenbank von 3,39 GB auf 22,2 GB
- Es wurden über 800.000 Vektoren gebildet



# Wie ist das technisch gelöst? (3a - GP)

**Geschäftspartner (GP)**

GP ID*	G1018	Versand-Mail	info@bsp-6.de	Korrespondenzsprache*	Deutsch		
Zahlart	<input checked="" type="radio"/> Überweisung <input type="radio"/> Lastschrift <input type="radio"/> Kreditkarte	Rechnungsversand	<input type="radio"/> Post <input checked="" type="radio"/> Email <input type="radio"/> Portal	Debitor	1000001	Kreditor	7000001
Name kurz*	LIB-IT DMS GmbH						
Vollständiger Name	LIB-IT DMS GmbH						
Abkürz./Exnamen/Suchbegriffe							
Mandantenstatus*	Mandant			FN Name 1 ⓘ	LIB-IT DMS GmbH		
Mandatsführer Recht				FN Name 2 ⓘ			
Mandatsführer Steuer	Mitarbeiter, Peter			FN Name 3 ⓘ			
Kooperationsstatus				Lieferantenstatus	Lieferant		
Hauptadresse	SuperArkaden, blablubb Straße 15, 80331 München, Bayern, Deutschland						
Rechnungsadresse	SuperArkaden, blablubb Straße 15, 80331 München, Bayern, Deutschland						
Korrespondenzadresse	SuperArkaden, blablubb Straße 15, 80331 München, Bayern, Deutschland						
Telefon-Zentrale				Telefon ohne Format			
Fax-Zentrale				Fax ohne Format			
Zentrale E-Mail	info@bsp6.de			Homepage			
Notiz							



# Wie ist das technisch gelöst? (3b - Adresse)

**A Adresse**

GP Link	G1018 • LIB-IT DMS GmbH (München) • Mandant • Lieferant • Mitarbeiter, Peter		Parent ID	G1018
Adressart*	Geschäftsadresse	Hauptadresse	<input checked="" type="checkbox"/>	
Versandart		Korrespondenzadresse	<input checked="" type="checkbox"/>	
Zusatzfeld	SuperArkaden			
Straße Hausnr.	blablubb Straße 15			
PLZ Ort, Land*	80331 München, Bayern, Deutschland			
Google Maps	<a href="http://www.google.com/maps?output=classic&amp;ie=UTF8&amp;hl=de&amp;saddr=blablubb%20Stra%C3%9Fe%2015%2C%2080331%20M%C3%BCnchen%2C%20Deutschland">www.google.com/maps?output=classic&amp;ie=UTF8&amp;hl=de&amp;saddr=blablubb%20Stra%C3%9Fe%2015%2C%2080331%20M%C3%BCnchen%2C%20Deutschland</a>			
Google Route	<a href="http://www.google.com/maps?output=classic&amp;ie=UTF8&amp;hl=de&amp;daddr=blablubb%20Stra%C3%9Fe%2015%2C%2080331%20M%C3%BCnchen%2C%20Deutschland&amp;saddr=Sonnenstra%C3%9F">www.google.com/maps?output=classic&amp;ie=UTF8&amp;hl=de&amp;daddr=blablubb%20Stra%C3%9Fe%2015%2C%2080331%20M%C3%BCnchen%2C%20Deutschland&amp;saddr=Sonnenstra%C3%9F</a>			
Notizen				



# Wie ist das technisch gelöst? (3c - Person)

**Person**

GP Link: G1018 • LIB-IT DMS GmbH (München) • Mandant • Lieferant • Mitarbeiter, Peter

Personen-ID\*: P1007

Anrede\*: Frau

Nachname\*: Chef

Anrede im Brief: Sehr geehrte Frau Chef

Position:

Bild:

Tel.-Zentrale:

Telefon mit DW: -

Fax-Zentrale:

Mobil 1:

E-Mail 1:

Büroadresse: SuperArkaden, blablubb Straße 15, 80331 München, Bayern, Deutschland

Beziehung zu Person:

Geburtsdatum:

Mitarbeiter Link:

Unterschrift Zeile 1:

Unterschrift Zeile 2:

Unterschrift Zeile 3:

Adresse privat Link:

Strasse Nr. privat:

Tel. privat:

E-Mail 1 privat:

Notizen:

Persönl. Steuer-ID:

Titel:

Vorname(n): Ina

zu Händen: Frau Chef

Abteilung:

Portalzugriff:

Durchwahl: -

Fax Durchwahl:

Mobil 2:

E-Mail 2:

Beziehung:

Geburtsmonat:

Ansprechpartner Briefvorlage:

Durchwahl Brief:

PLZ Ort privat:

Mobil privat:

E-Mail 2 privat:



1. Aufbereitung der FILERO-Datenstruktur, sodass wir für das LLM Vektoren bilden können
2. Im FILERO-Backend wurden dann die gewünschten Daten (Klassen) vektorisiert

2024 © LIB-IT DMS GmbH



# Wie ist das technisch gelöst? (5)


- Javascript: Spracheingabe über freie **Web Speech API** (SpeechRecognition)
  - [https://developer.mozilla.org/en-US/docs/Web/API/Web\\_Speech\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API)

- Hinweis:
  - Das API-Modul funktioniert nicht mit FireFox:

												
	 Chrome	 Edge	 Firefox	 Opera	 Safari	 Chrome Android	 Firefox for Android	 Opera Android	 Safari on iOS	 Samsung Internet	 WebView Android	 WebView on iOS
<div>SpeechRecognition</div>	<div>✓</div> <div>33</div> <div>**</div>	<div>✓</div> <div>79</div> <div>**</div>	<div>✗</div> <div>No</div> <div></div>	<div>✓</div> <div>20</div> <div>**</div>	<div>✓</div> <div>14.1</div> <div>*</div>	<div>✓</div> <div>33</div> <div>**</div>	<div>✗</div> <div>No</div> <div></div>	<div>✓</div> <div>20</div> <div>**</div>	<div>✓</div> <div>14.5</div> <div>*</div>	<div>✓</div> <div>2.0</div> <div>**</div>	<div>✓</div> <div>4.4.3</div> <div>**</div>	<div>✓</div> <div>14.5</div> <div>*</div>



# Wie ist das technisch gelöst? (6)

-  als Programmiersprache
- LLM (Large Language Model)  
als Package in Python:  
**sentence-transformers/all-MiniLM-L6-v2**
  - <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>



# Erkenntnisse

- Chunking-Strategie ausschlaggebend für die Präzision der Vector Search
- Vektor-Index in größeren Datenbanken nötig
  - Hierarchical Small World Index (HNSW)-Index



# Anforderungen / Zukunft (1)

- Alle Klassen und Attribute inkl. ihrer Relationen und Hierarchien sollen in FILERÖ vektorisiert werden
- Aktualisierung der Vektoren bei Änderung.  
Neue Datensätze automatisch vektorisieren.
- Diskussion: Historische Versionen vektorisieren?



# Anforderungen / Zukunft (2)

- **Besondere Herausforderung:**

FILERO ist ein generisches System,  
d.h. der Kunde generiert Klassen, Attribute, Relationen  
und Hierarchien selbst.



# Anforderungen / Zukunft (3)

- In Zukunft soll eine KI die semantischen Zusammenhänge von Klassen und Attributen erkennen und dann automatisch vektorisieren.





# Dankeschön → Fragen?

