



ifis

Institut für Informationssysteme
Technische Universität Braunschweig

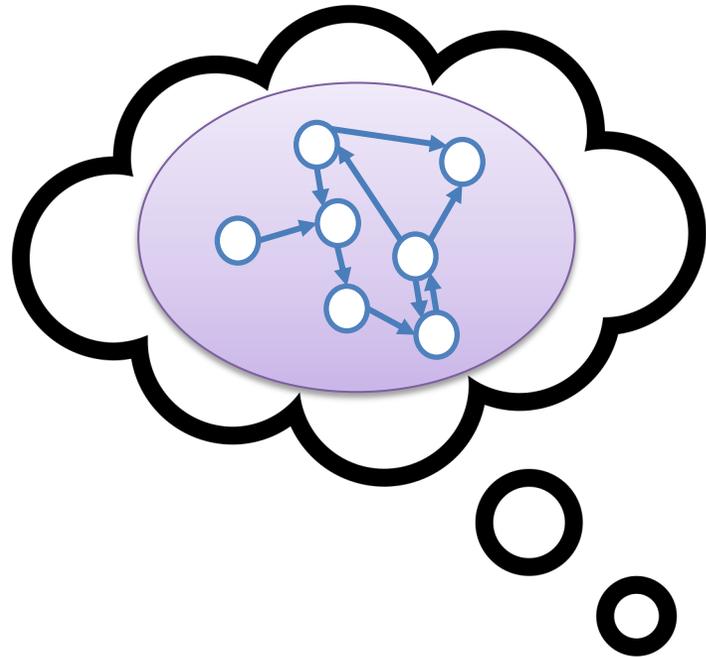
Enriching Simple Keyword Queries for Domain-Aware Narrative Retrieval

**Hermann Kroll & Christin Kreutz &
Pascal Sackhoff & Wolf-Tilo Balke**

Institute for Information Systems
TU Braunschweig, Germany



Motivation



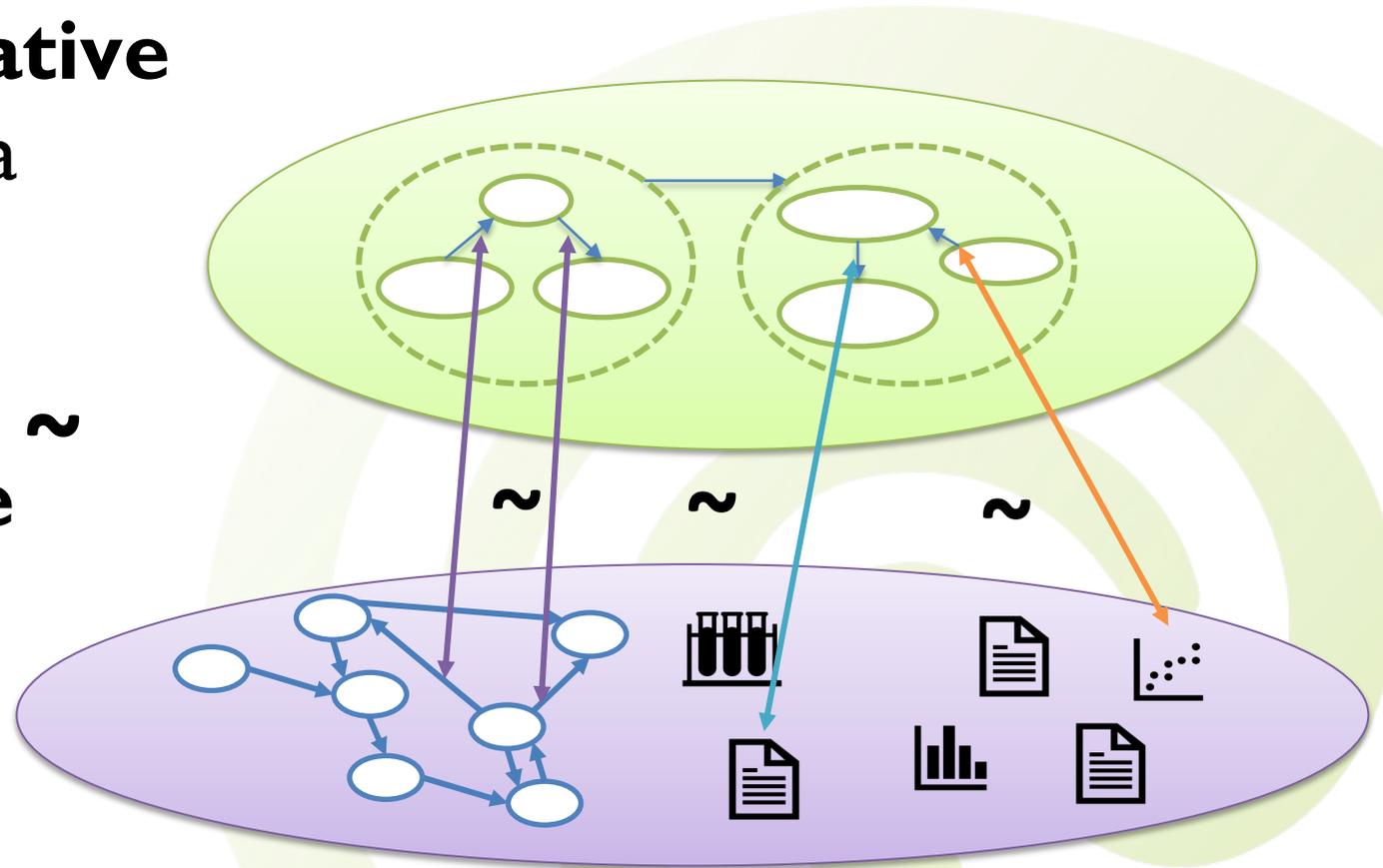
Did you know?





Narrative Information Access

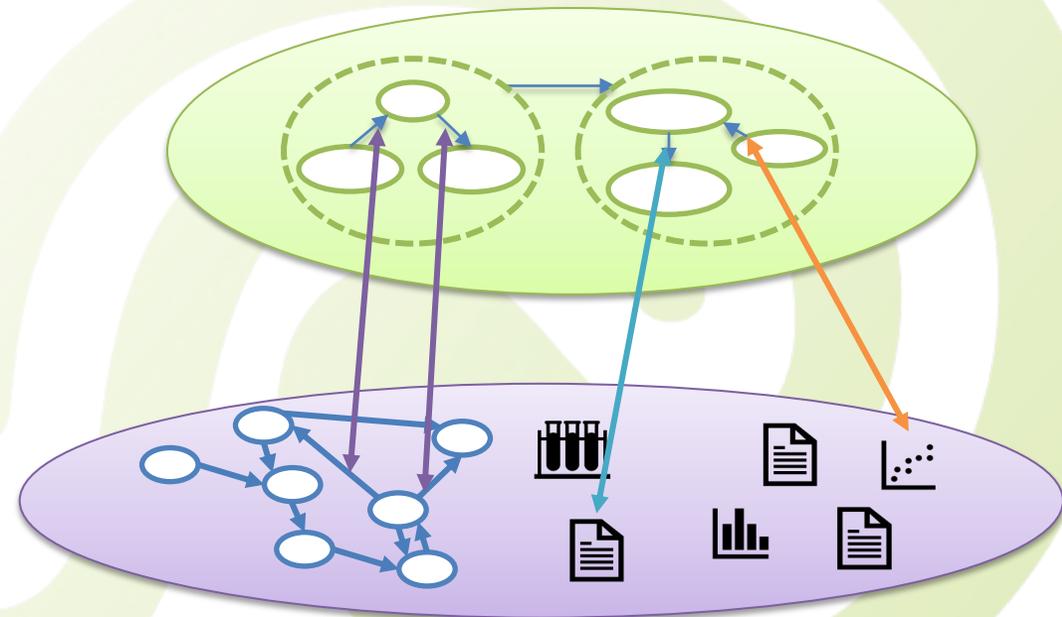
1. Bind the **whole narrative** against real-world data
2. All bindings must be **context-compatible** \approx





Last JCDL's Conclusion

- Narrative information access allows precise and structured searches in digital libraries
 - It is an **extension** to **knowledge base querying**
 - **Contexts** are **vital** to determine a statement's validity



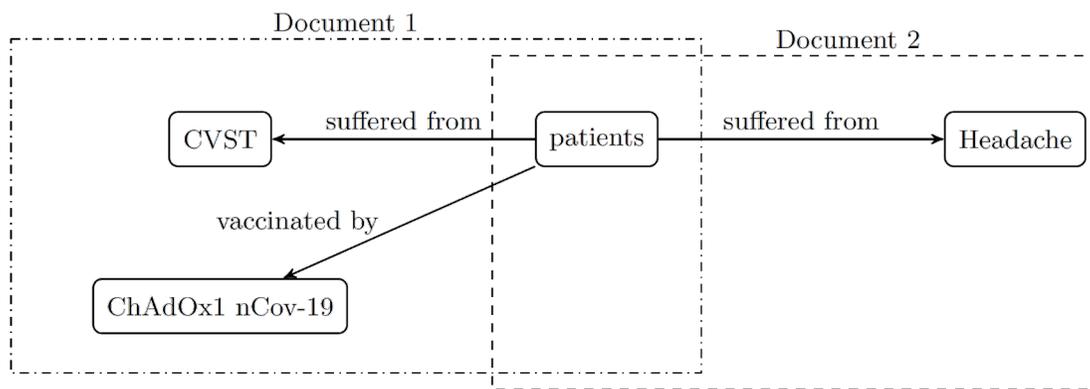


What we Did

- Implemented Narrative information access in pharmacy:
 - www.narrative.pubpharm.de



Implicit Context Model



Narrative Service Last updated 10.02.2023 Search Drug Overviews Long COVID Overview Help Impressum

metformin Browse treats Disease Browse Add Search

How to Search: ?

Example Queries

Data Source:
PubMed (Help)
LitCovid (Help)
Long Covid (Help)
Covid 19 Pre-Prints via ZBMED (Help)

Results by year:

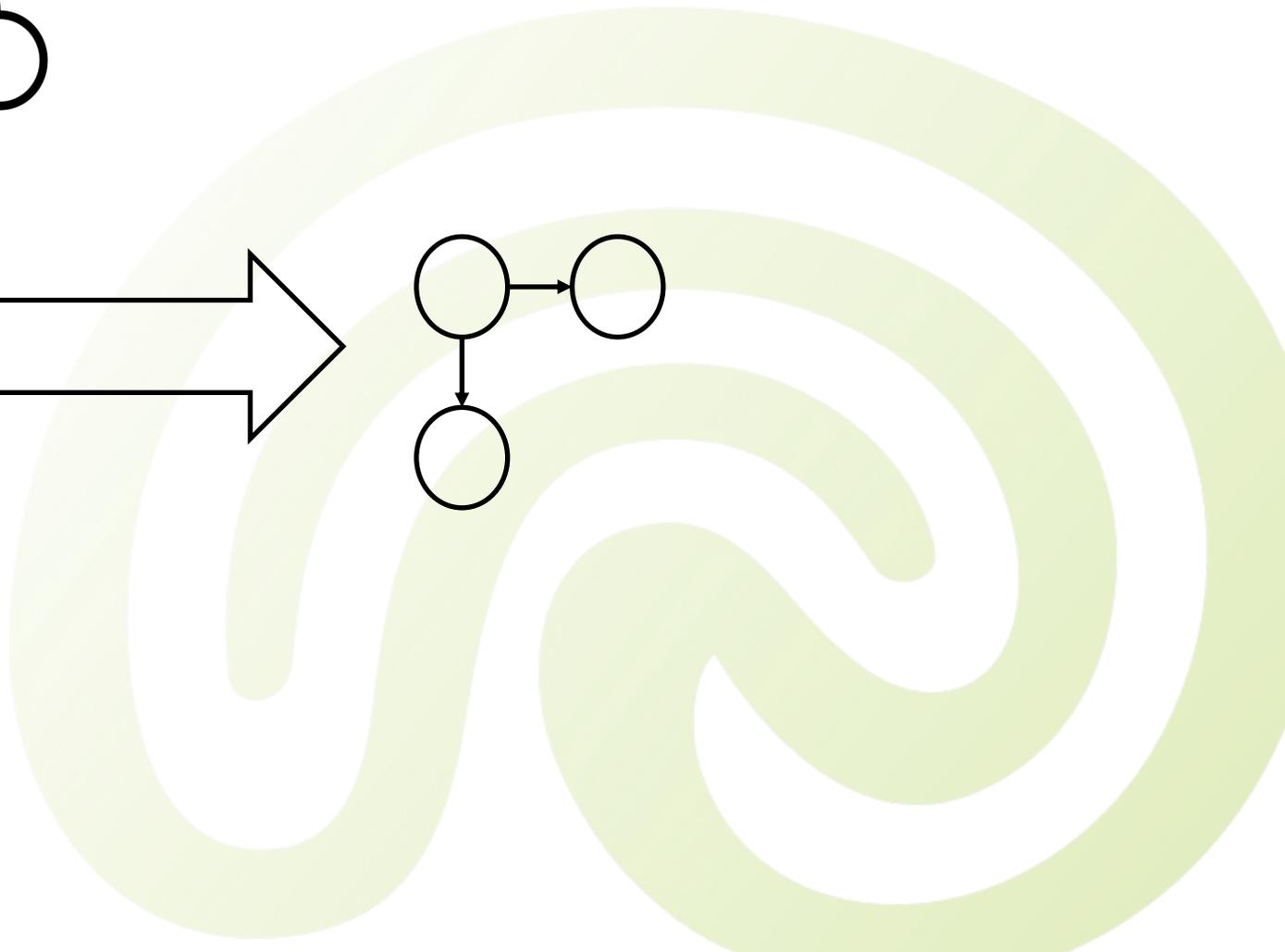
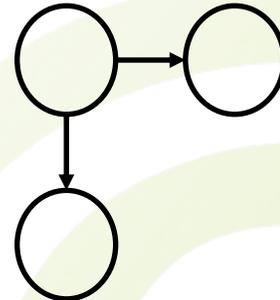
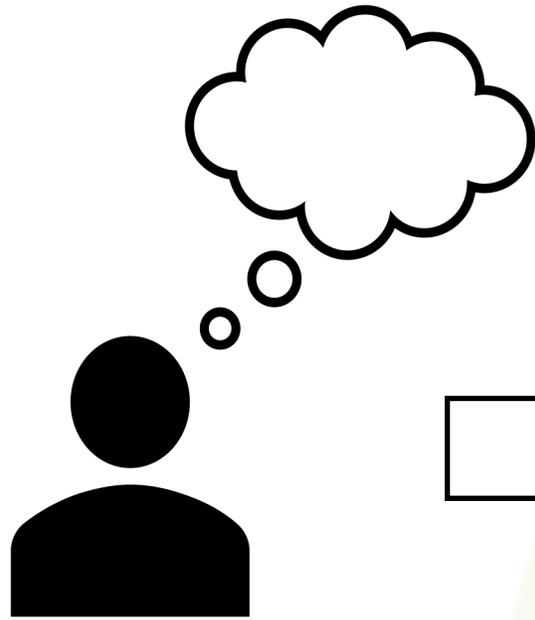
Visualization by:
Substitution
MeSH-Taxonomy

Results: ? Page: 1 of 3 Latest Publications First Most Frequent Substitutions First

Search in result titles:
child

36 Documents [?Disease= Insulin Resistance (Disease MESH:D007333 Q.)]	✓	✗
31 Documents [?Disease= Diabetes Mellitus (Disease MESH:D003926 Q.)]	✓	✗
29 Documents [?Disease= Obesity (Disease MESH:D009765 Q.)]	✓	✗
29 Documents [?Disease= Diabetes Mellitus, Type 2 (Disease MESH:D003924 Q.)]	✓	✗
10 Documents [?Disease= Weight Loss (Disease MESH:D015431 Q.)]	✓	✗
7 Documents [?Disease= Polycystic Ovary Syndrome (Disease MESH:D011085 Q.)]	✓	✗

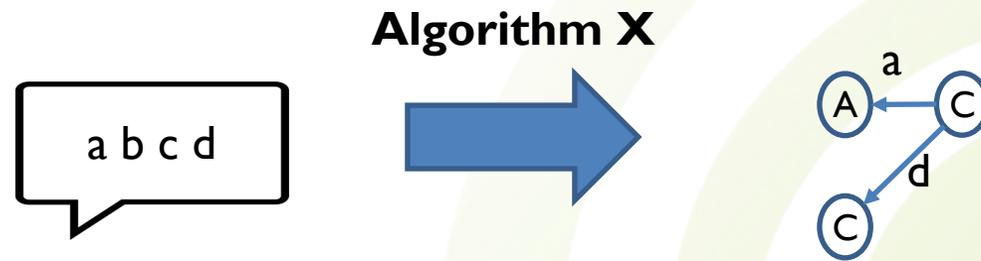
Omnicron period were 22.8% (95% CI -6.4-44.0) and 66.5% (95% CI 56.0-73.0). In **children** aged 12-15 years, **VE** against symptomatic **COVID-19** during the Delta period was 65.4% (95% CI 61.5-69.8) at 2-5 weeks after the first dose, with no observed cases at 2-5 weeks after the second dose. The corresponding **VE** against symptomatic **COVID-19** during the Omicron period were 30.2% (95% CI 18.4-40.3) and 81.2% (95% CI 77.7-84.2). The waning of the protective effect against the symptomatic disease began after five weeks post-first and post-second dose. Interpretation: During the study period, uptake of **BNP16242** in Scotland has covered more than two-thirds of CYP aged 12-17 years with the first dose and about 40% with the second dose. We found no increased likelihood of admission to hospital with a range of health outcomes in the period after **accelerated vaccination** with both doses was associated with a substantial reduction in the risk of **COVID-19** symptomatic disease during both the Delta and Omicron periods, but this protection began to wane after five weeks. Funding: UK Research and Innovation (Medical Research Council); Research and Innovation Industrial Strategy Challenge Fund; Chief Scientist's Office of the Scottish Government; Health Data Research UK; National Core Studies - Data and Connectivity.





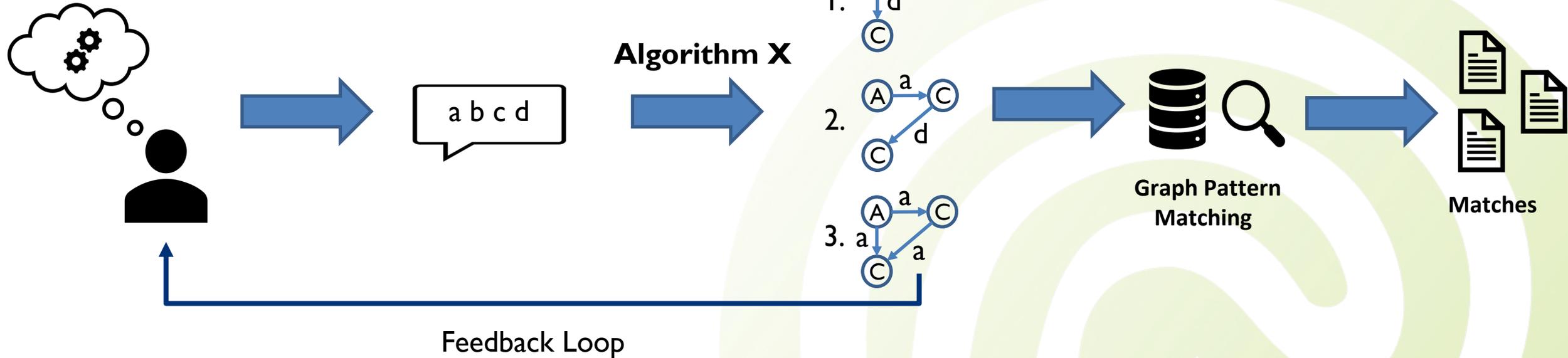
Research Question

- Can a user's **search intent** be deduced from a **keyword query**?





Schematic Overview

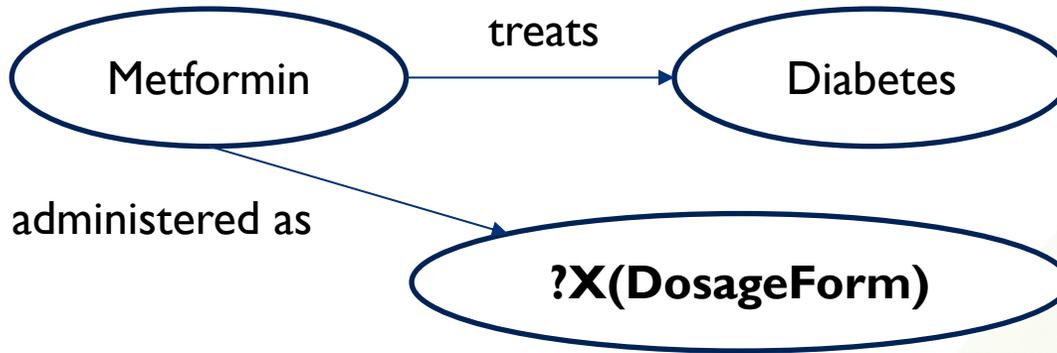


Narrative Queries





Narrative Queries



- A **narrative query** consists of:
 - **concepts as nodes**
 - **interactions as edges**
 - **terms** to support out-of-vocabulary concepts



Hard to connect to the pattern

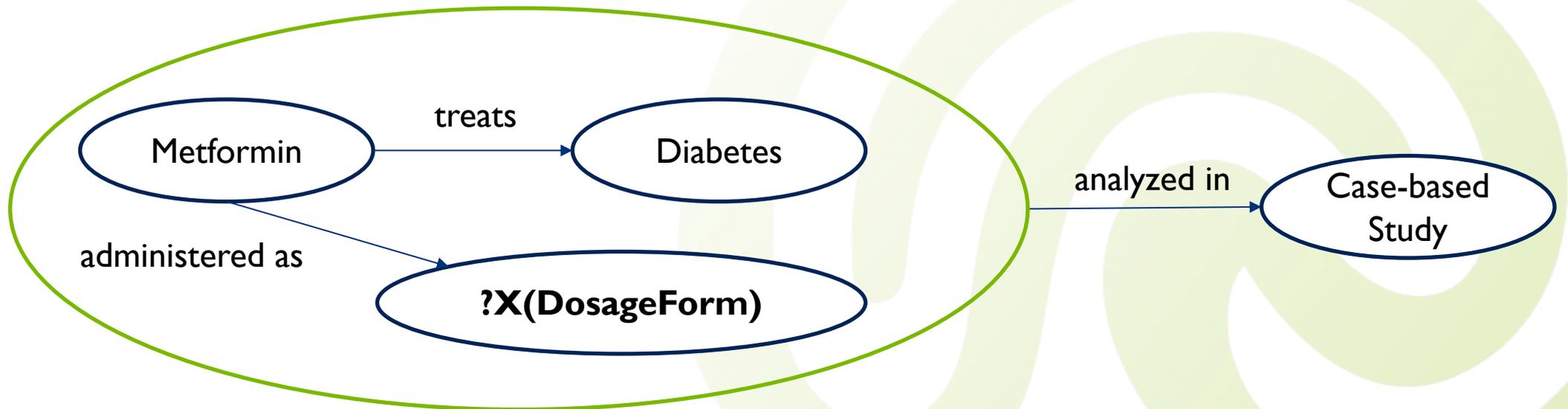
Young Men

Some searched terms might not be known as concepts (out-of-vocabulary)



Narrative Queries

- Nested structure would allow a higher **expressiveness**
 - However, hard to realize/implement
 - That is why we kept queries **flat** at the moment



Data Model

New Paper

Open in a new tab

- All
- Disease
- DosageFo...
- Drug
- LabMethod
- Method
- Species

Pharmacokinetic analysis of two different doses of simvastatin following oral administration in dogs.

Kim, M | Baek, I

Journal of veterinary pharmacology and therapeutics, Vol. 44 No. 3 (May 2021)

5/2021

Simvastatin, used orally to treat **hyperlipidemia**, exhibits highly variable pharmacokinetics (PKs) in **humans**. The aim of this study was to investigate **simvastatin** PKs using noncompartmental analysis and population PK models following a single **oral administration** of two doses (20 and 80 mg) in **dogs**. Forty **beagle dogs** were randomly divided into two groups corresponding to the two doses. Blood samples were collected from each group according to the assigned schedule after **oral administration**. The plasma concentration of **simvastatin** was determined using liquid chromatography-tandem **mass spectrometry**. The area under the curve and maximum concentration of **simvastatin** increased in a dose-dependent manner with high variability. A two-compartment model with first-order absorption ($K_a = 1.83 \text{ hr}^{-1}$) and first-order elimination (clearance $[CL/F] = 292 \text{ L/h}$; volume of distribution in the central compartment $[V_c / F] = 1506 \text{ L}$) well described the PKs of **simvastatin** in **dogs**. Large variability in the PKs of **simvastatin** was quantitated via modeling approaches, allowing the differentiation of between-subject variability (144.8 CV% for K_a ; 94.7 CV% for CL/F ; 97.5 CV% for V_c / F) and residual variability (62.7%). These findings will help facilitate the development of an optimal dose regimen of **simvastatin** in **canines** with **hypercholesterolemia** and may be useful in developing novel formulations.

Classifications

PharmaceuticalTechnology:

SVM

Pharmaceutical:

pharm*:Pharmacokinetic(0,15), pharm*:pharmacokinetics(177,193)

pharm*:Pharmacokinetic(0,15), pharm*:pharmacokinetics(177,193)

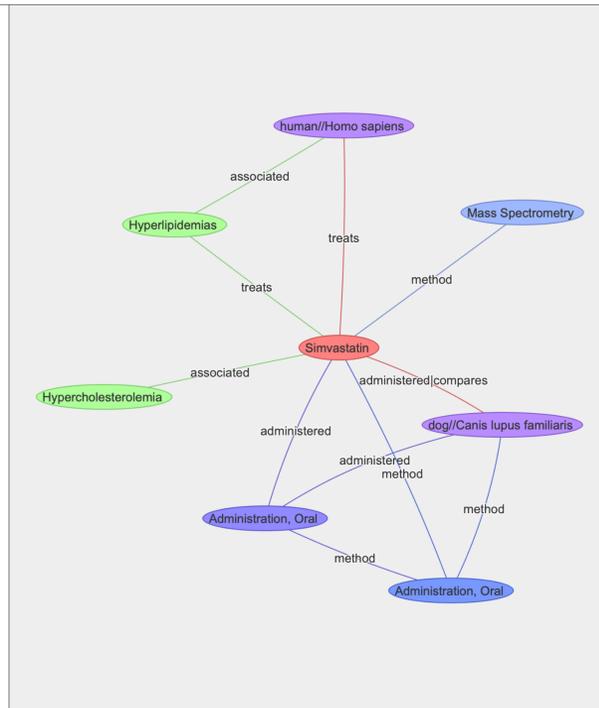
pharmaceutical:

SVM

pharmaceuticalTechnology:

Classifications

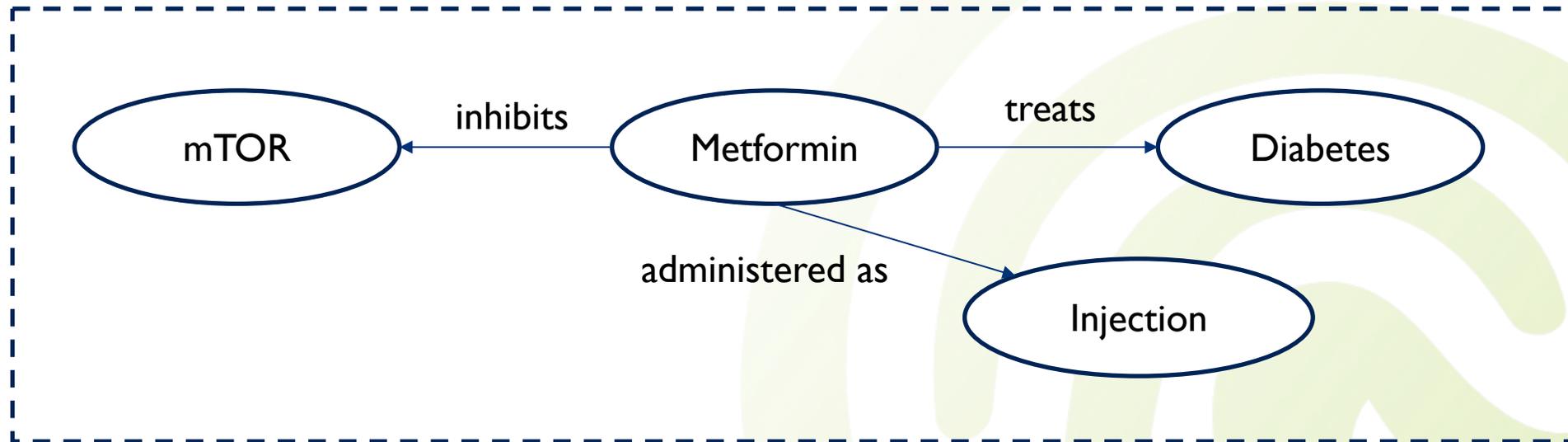
hypercholesterolemia and may be useful in developing novel formulations





Document Graphs

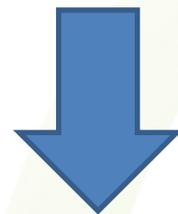
Document I23



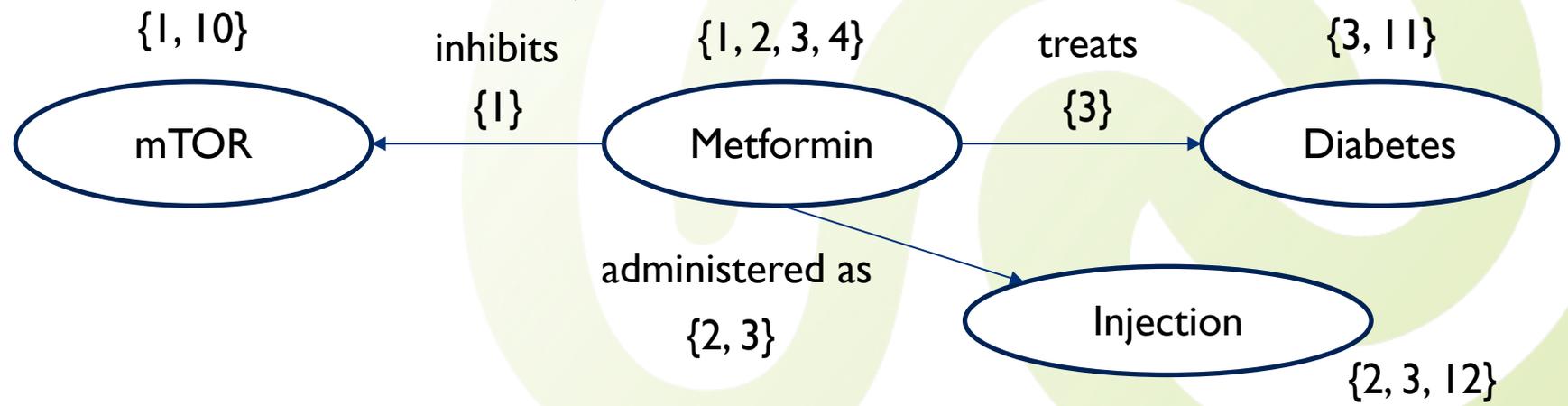


Document Collection Index

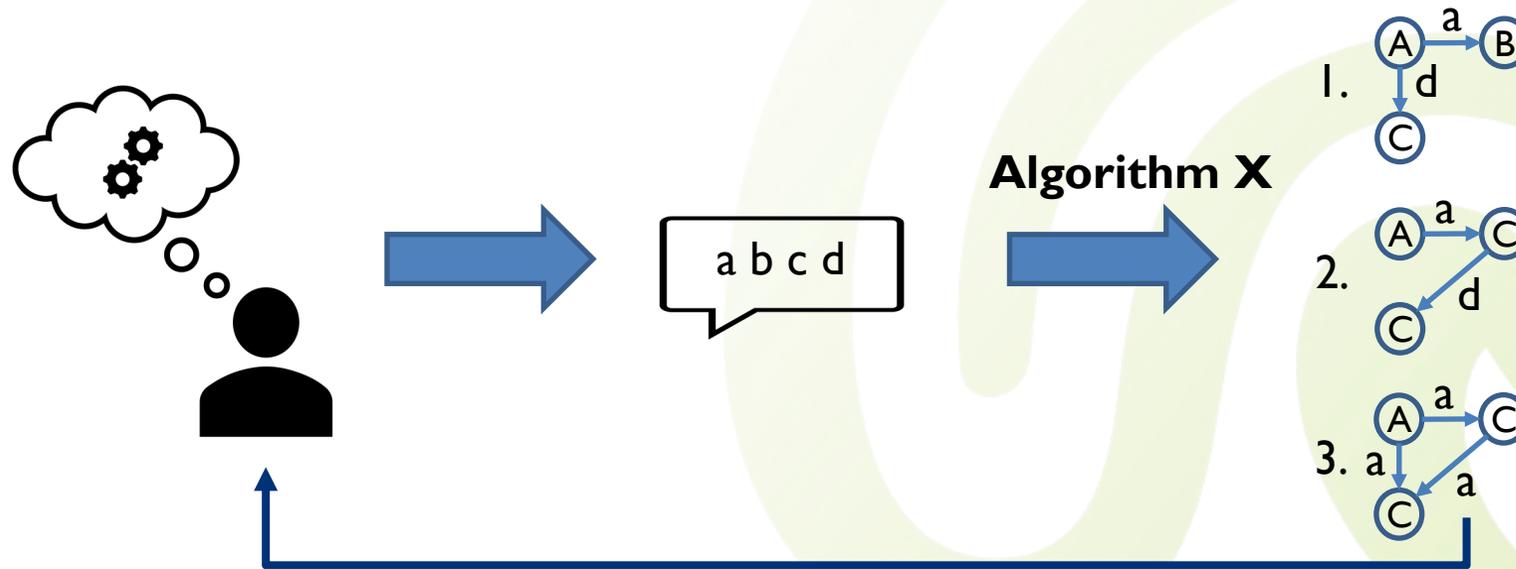
Documents



Document
Collection
Index



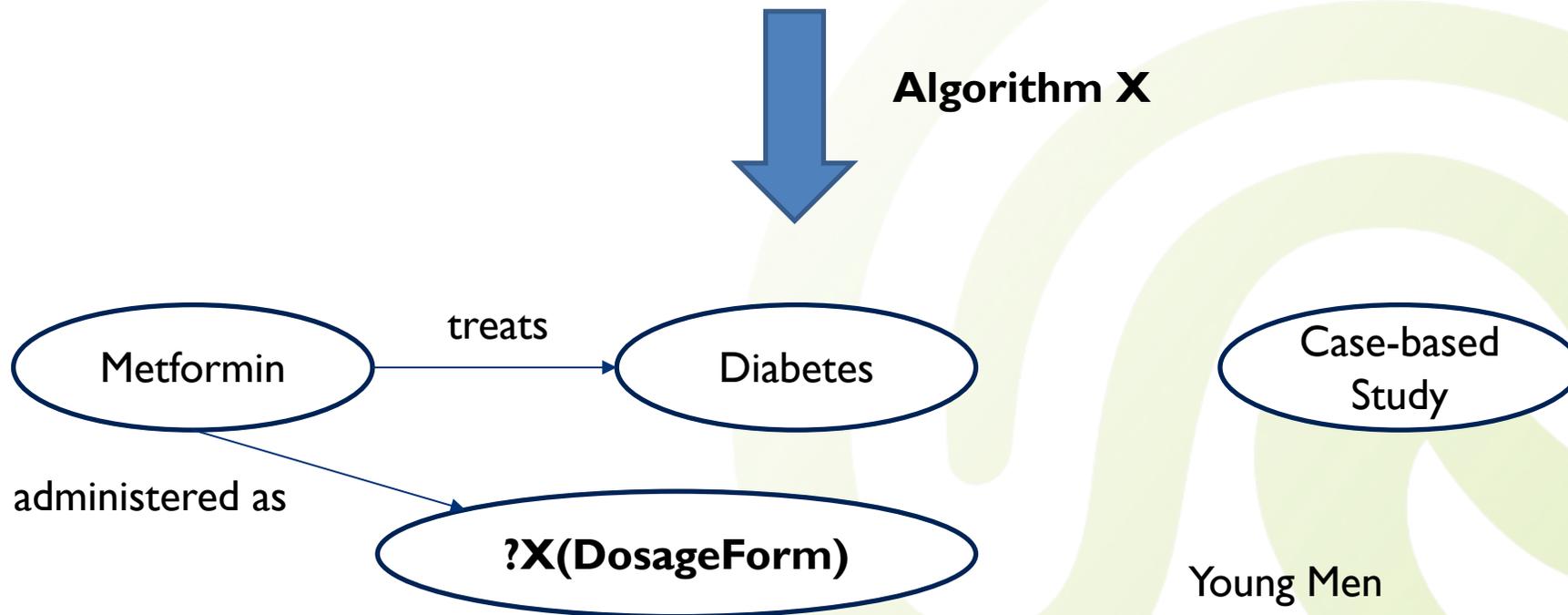
Translation





Translation Goal

Metformin Diabetes Dosage Form
Case-based Study Young Men





Query Generation Strategy

1. Map keywords (or combinations) to...
 - A concept / a term / a predicate
2. Generate all possible narrative queries
 - Keep queries that yield at **least a single document** (via the document collection index) to ensure **hits for the users**
3. Select queries based on different criteria:
 - **Most-specific / Most-general / Most-supported** (based on hits)

Evaluation



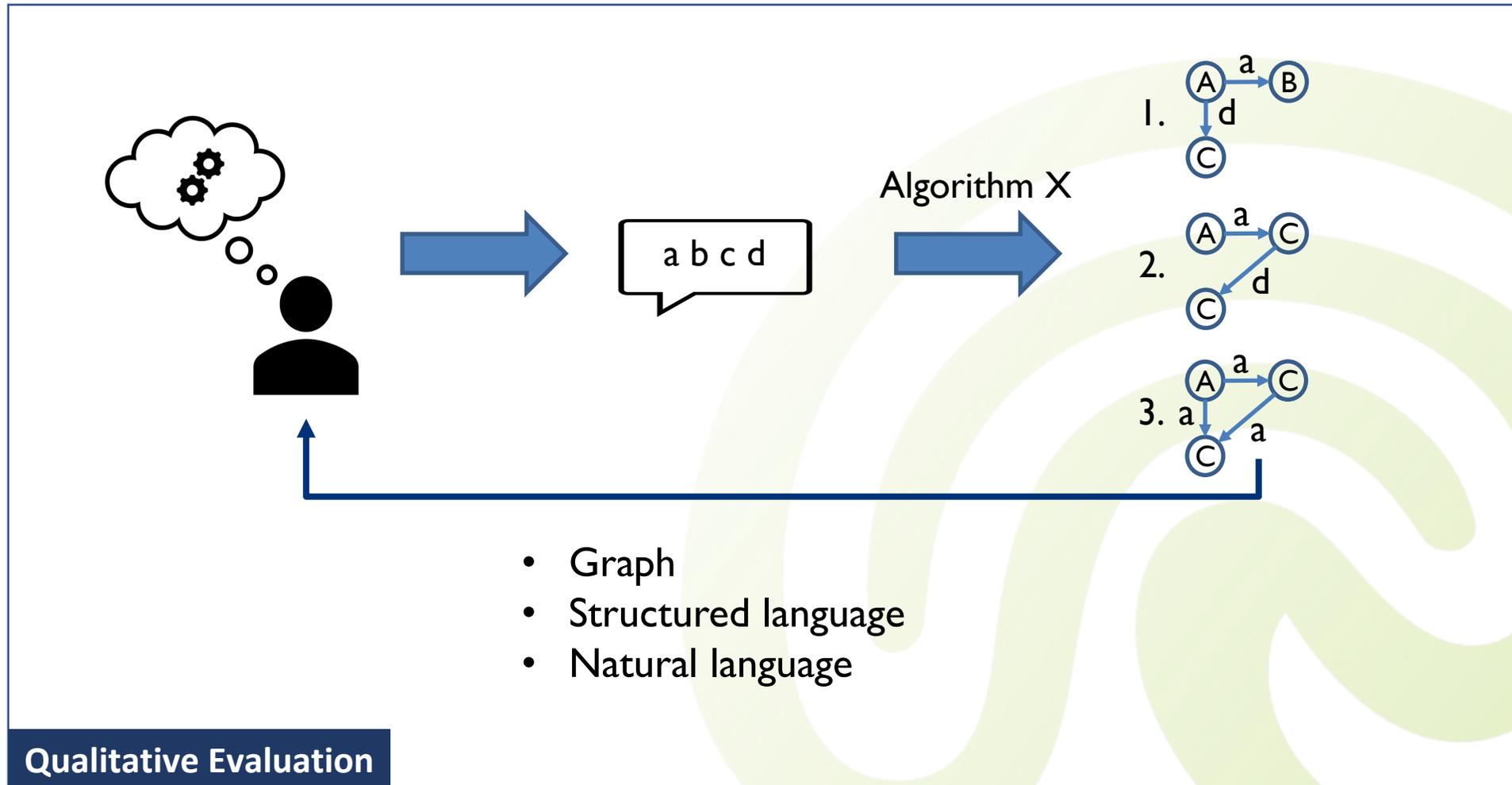


Evaluation Aspects

- Can a user's search intent be deduced from a keyword query?
 - **Query Representation:** How should generated patterns be presented to the users, i.e., which query representation is suitable for our users?
 - **System Satisfaction:** How useful is the end-to-end system?
 - **Translation Effectiveness:** How effectively does our method translate keyword-based queries to narrative queries for users?

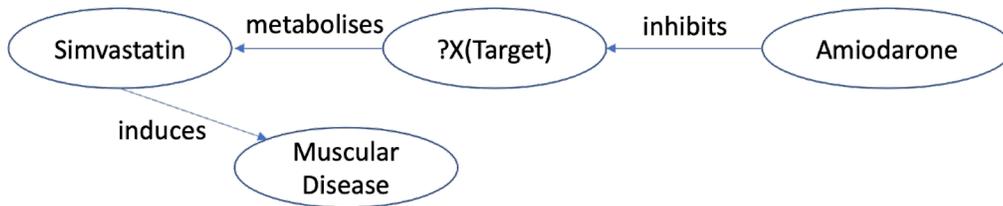


Query Representation





Query Representation



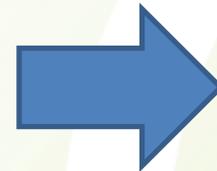
(a) Graph query of *'Through which target do Simvastatin and Amiodarone interact so that Simvastatin may induce a Muscular Disease?'*

Metformin administered ?X(DosageForm)
?X(DosageForm) administered Patients
Patients associated Diabetes

(b) Structured query of *'As which dosage forms can Metformin be administered to diabetic patients?'*

Some vaccine is associated with COVID 19.
The same vaccine is administered to patients.
These patients suffer from CVST.

(c) Natural language query of *'Which COVID 19 vaccines may make patients suffer from cerebral venous sinus thrombosis (CVST)?'*



IU	++	+	+/-	-	--
graph	3	3	2	1	0
structured	1	4	2	2	0
natural language	3	4	2	0	0



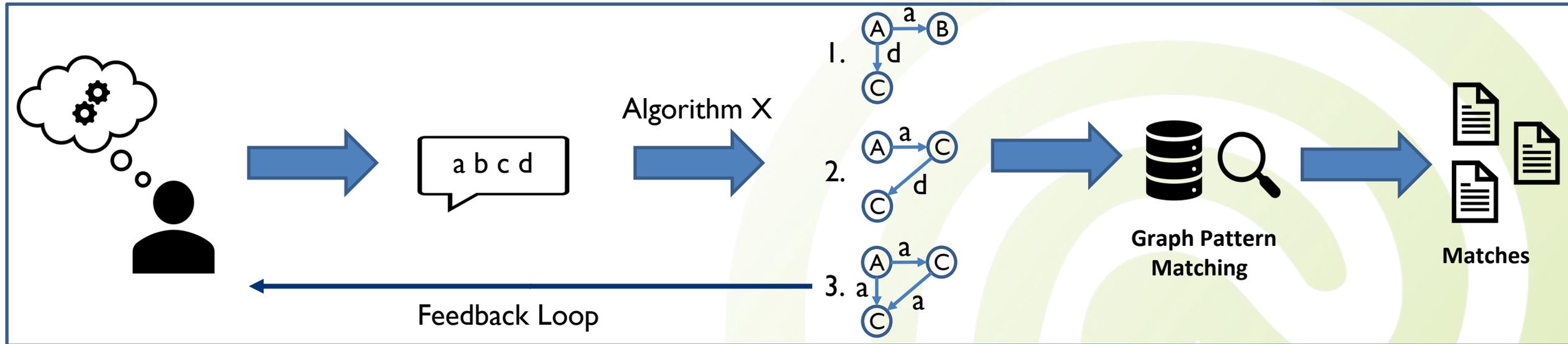


Findings (1/3)

- **Graph-based representation** most suitable for our users
 - **Pharmacists** were **familiar** with graph representations
 - Graphs were **easier** and **faster** to understand than natural language



System's Satisfaction





User Interface

Narrative Service Last updated 01.11.2022 Search Drug Overviews Long COVID Overview Help Impressum

1. Long Covid Fatigue Vaccine Search

2. **Queries:**
Click on one query to search.

Terms:

Terms: covid, long

Terms: covid, long

3. **Results:**
[15 Documents](#)

Orthostatic Intolerance in Long-Haul COVID after SARS-CoV-2: A Case-Control Comparison with Post-EBV and Insidious-Onset Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients.
in: Healthcare (Basel, Switzerland), Vol. 10 No. 10 (Oct 2022) | 10/2022
by: van Campen, C | Visser, F
PMID: 36688581



System Evaluation

- Think-Aloud + semi-structured interviews
 - Time: 30 minutes each
 - Participants: 10 pharm. researchers (PhD Students, PhDs, Profs.)





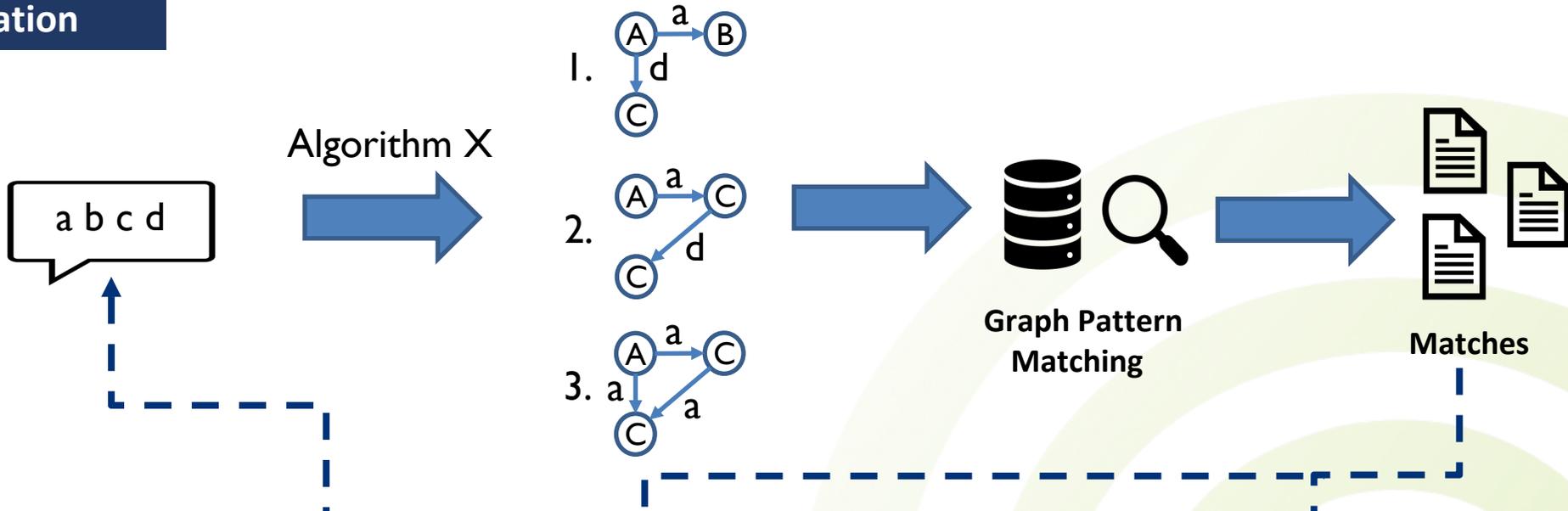
Findings (2/3)

- **System was appreciated** by our users:
 - It eases narrative retrieval and yield quickly relevant documents
 - But, the system was too slow for usage and needs UI improvements

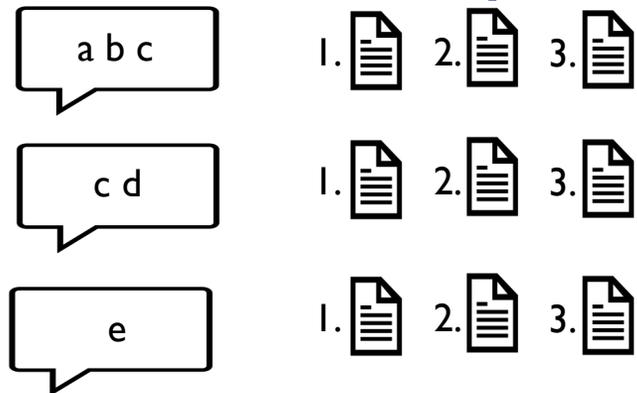
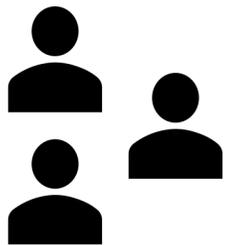


“Translation” Effectiveness

Evaluation



Benchmark





Findings (3/3)

- Query model **boosts precision**
in biomedical document retrieval
 - And indeed, our strategies selected those queries
 - Details and statistics can be found in our paper



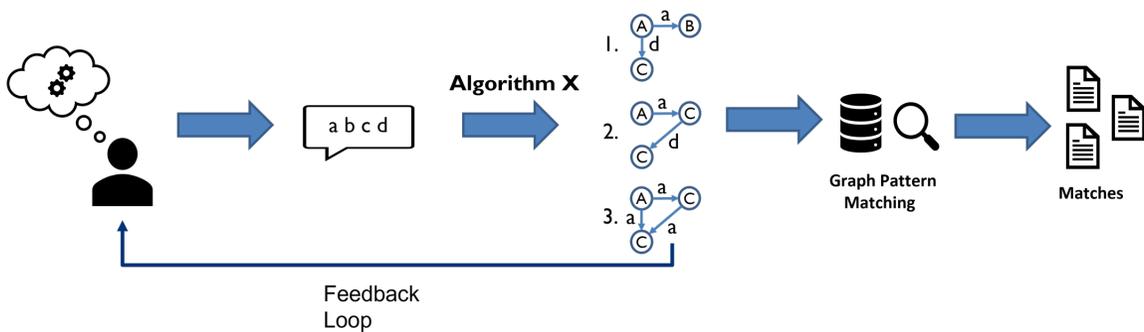
Summary

- **Graph-based representation**
most suitable for our users
 - Fast and easy to understand
- **System was appreciated** by our users:
 - It eases narrative retrieval for users
 - But, performance and UI needs to be improved
- Query model **boosts precision** in biomedical inf. retrieval
 - And indeed, our strategies selected those queries



Contributions

Unsupervised Translation Algorithm



Query Representation



- (a) Graph query of 'Through which target do Simvastatin and Amiodarone interact so that Simvastatin may induce a Muscular Disease?'
Metformin administered ?X(DosageForm)
?X(DosageForm) administered Patients
Patients associated Diabetes
- (b) Structured query of 'As which dosage forms can Metformin be administered to diabetic patients?'
 Some vaccine is associated with COVID 19.
 The same vaccine is administered to patients.
 These patients suffer from CVST.
- (c) Natural language query of 'Which COVID 19 vaccines may make patients suffer from cerebral venous sinus thrombosis (CVST)?'

User Studies



The screenshot shows the Narrative Service web application. It displays search results for 'Long Covid Fatigue Vaccine'. The interface includes a search bar, a list of queries, and a list of results. The results section shows 15 documents, with the first document being a PubPharm article titled 'Orthostatic Intolerance in Long-Haul COVID after SARS-CoV-2: A Case-Control Comparison with Post-EBV and Insidious-Onset Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients'.

Systematic Evaluation

	Metric	TermB	BestP.	BestR.	BestF1
Abstract-Only Retrieval					
PM2020	Prec.	0.48	0.84	0.51	0.54
	Rec.	0.24	0.06	0.41	0.40
	F1	0.27	0.10	0.40	0.41
COVID	Prec.	0.33	0.40	0.33	0.34
	Rec.	0.26	0.20	0.31	0.29
	F1	0.22	0.17	0.24	0.24
Tripl.	Prec.	0.44	0.51	0.44	0.47
	Rec.	0.85	0.75	0.87	0.85
	F1	0.53	0.50	0.53	0.55
Full-text Retrieval					
COVID	Prec.	0.16	0.26	0.16	0.18
	Rec.	0.45	0.32	0.49	0.44
	F1	0.18	0.14	0.19	0.21
Genom.	Prec.	0.23	0.42	0.23	0.30
	Rec.	0.23	0.11	0.26	0.20
	F1	0.14	0.12	0.14	0.18

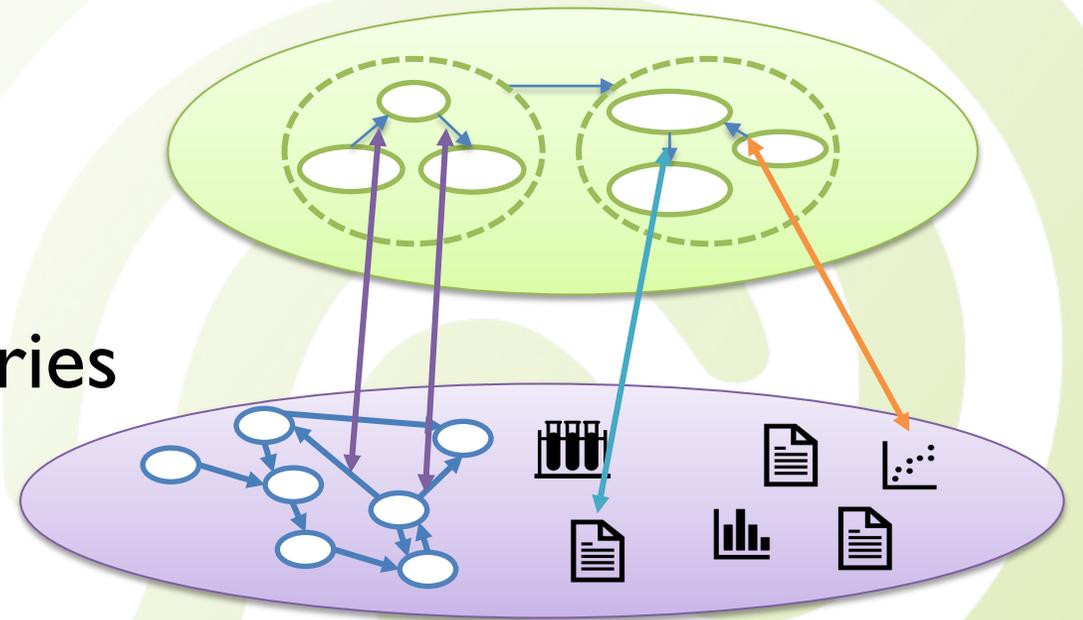
Benchmark	QJ	BestP.	BestR.	BestF1	Any
Exact Query Found					
PM2020	31	4	21	10	22
COVID	50	20	40	34	40
Tripl.	1136	548	806	579	849
COVID+F	50	22	45	29	45
Genom.	36	12	23	16	25
One Allowed Edit in Terms/Concepts					
PM2020	31	12	24	20	25
COVID	50	36	46	43	46
Tripl.	1136	804	926	839	947
COVID+F	50	38	50	45	50
Genom.	36	16	30	23	30
One Allowed Edit in Predicates					
PM2020	31	16	21	10	25
COVID	50	21	40	34	41
Tripl.	1136	569	807	594	854
COVID+F	50	25	46	32	46
Genom.	36	15	24	18	27



Conclusion

- Can a user's search intent be deduced from a keyword query?
 - **Yes**, a rather simple translation algorithm did already the job

- This research eases **narrative information access** in digital libraries by deducing narrative patterns from well-known keyword queries





Thank You!



FACHINFORMATIONSDIENST
PHARMAZIE
TU Braunschweig

If you have any questions,
contact me via:



@hkroll@fosstodon.org



kroll@ifis.cs.tu-bs.de



[@HermannKroll](https://twitter.com/HermannKroll)

