



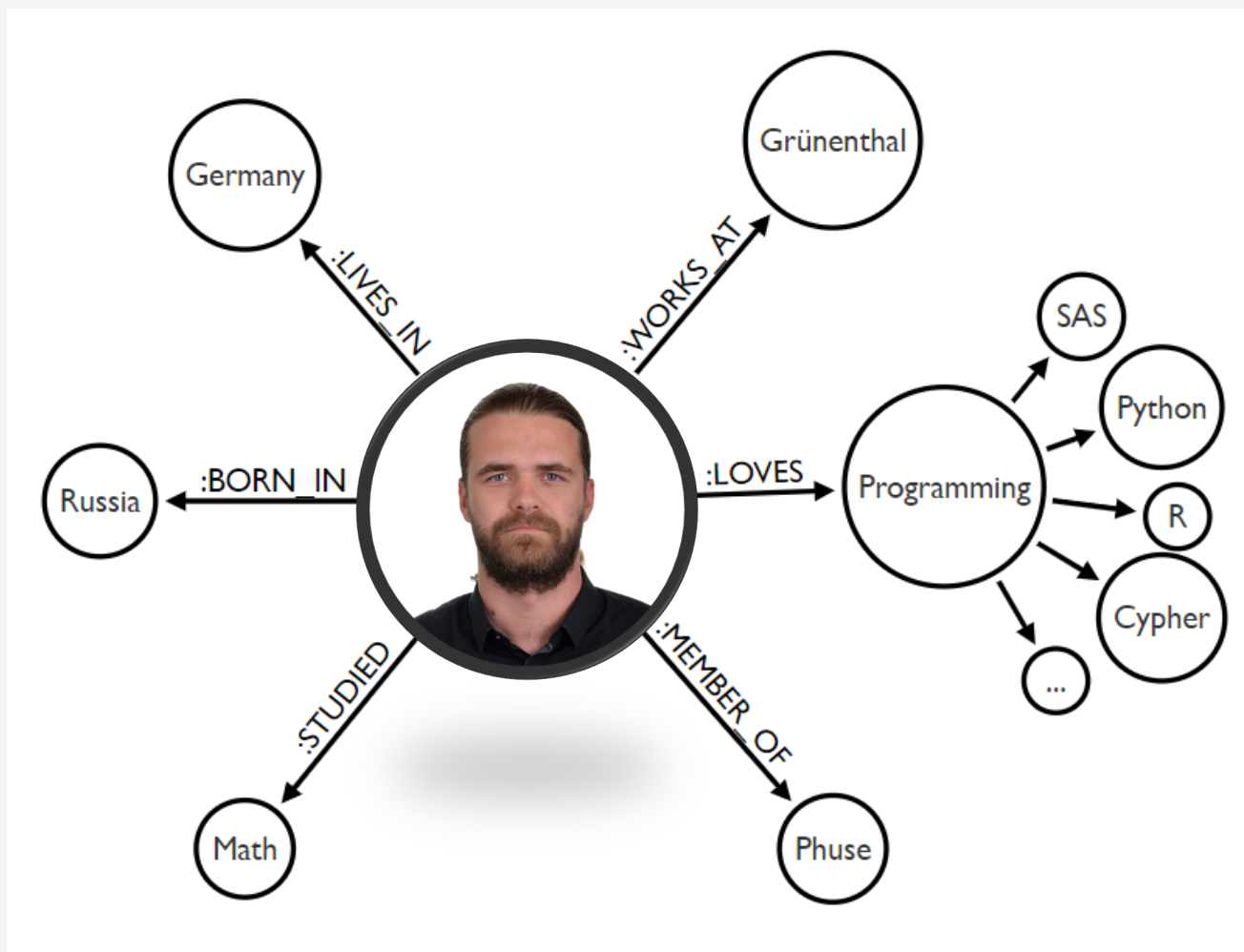
DH06: No more merging! (property graph database demystified)

Name: Alexey Kuznetsov

Title: Data Scientist

Company: Grünenthal GmbH

Alexey is working as a Data Scientist at Grünenthal, a midsize pharmaceutical company with headquarters in Germany Aachen. He joined pharmaceutical industry as a Statistical Programmer in 2011 and since then held various data science positions in CROs and pharma companies. Provided CDISC expert services, SAS programming, python development, graph database modeling and implementation in regards to handling clinical trials data as well as other data (claims data, registries data, etc.)



Data handling toolbox



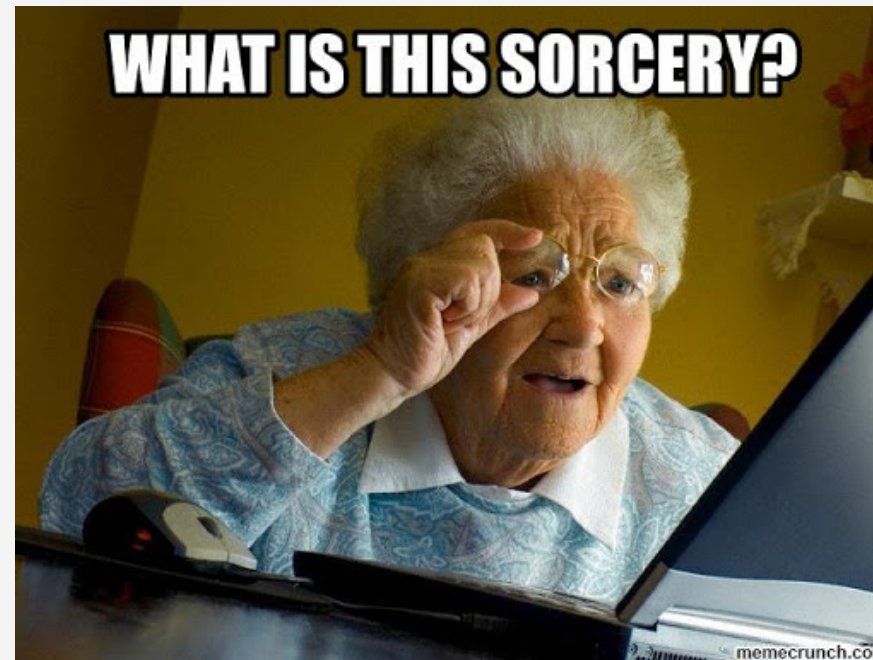
Data handling toolbox

?

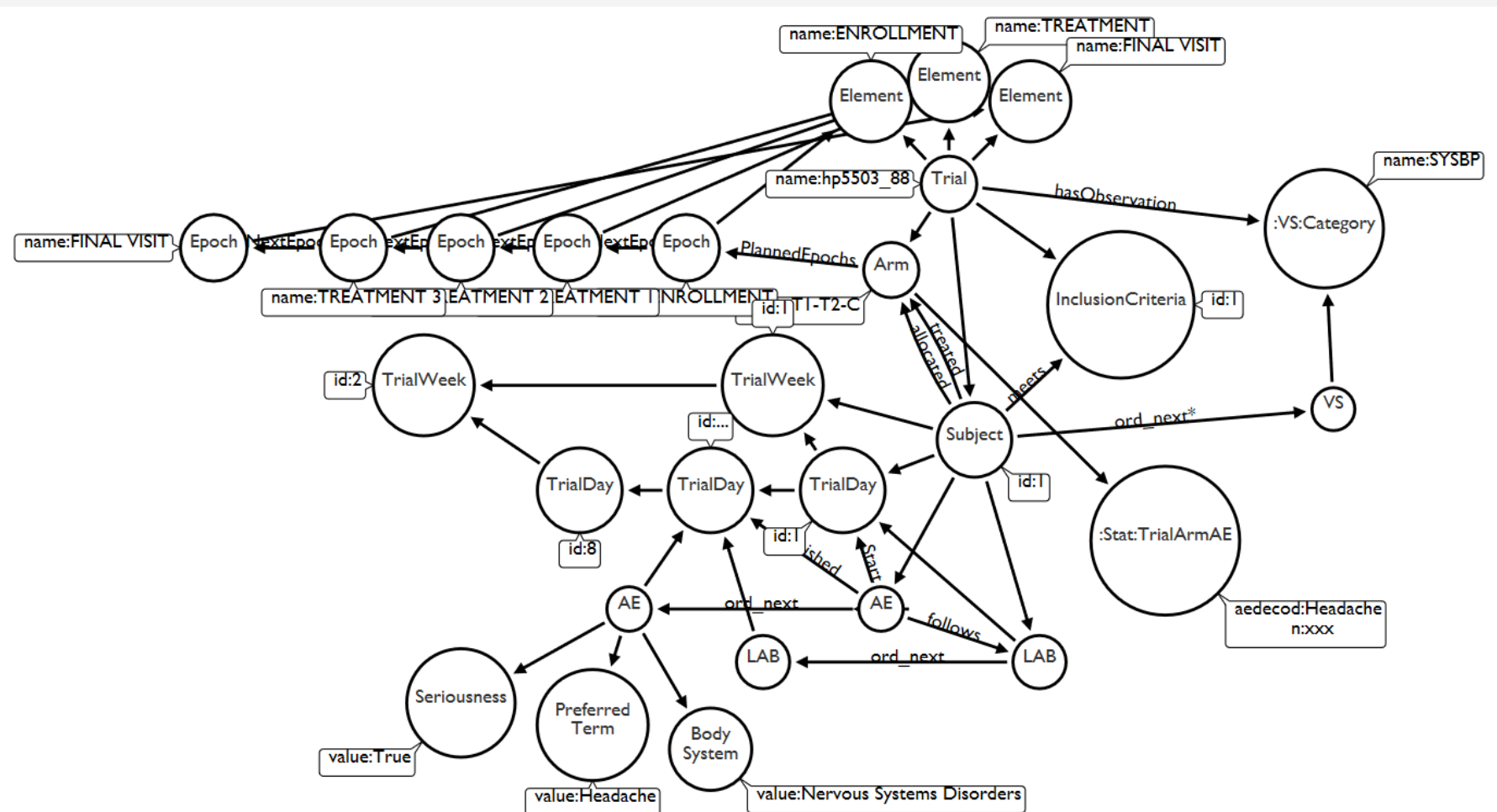
Graph Database



?



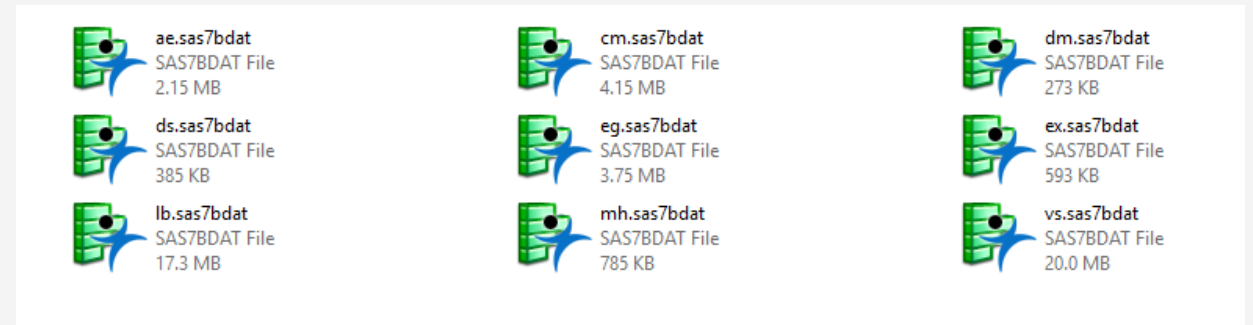
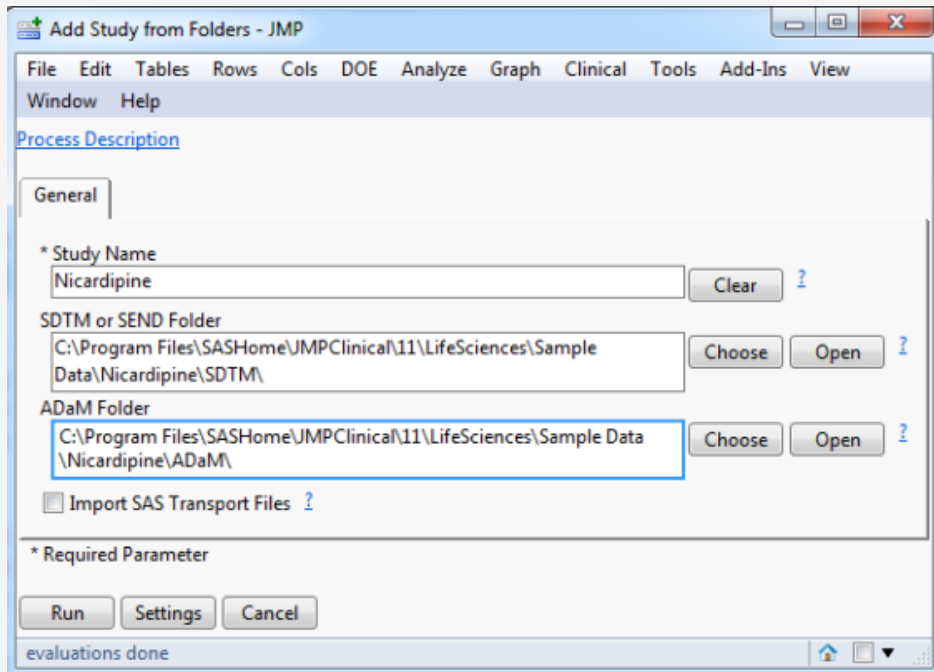
Data model



Created with <http://www.apcjones.com/arrows/#>

Source data I used

Dummy computer-generated SDTM data of a clinical trial with 906 patients. (Data shipped with JMP Clinical® software).



In the example I will walk you through DM (Demographics), AE (Adverse Events), CM (Concomitant Medications) as well as additional trial information/data from ClinicalTrials.gov

Technologies I used



- Neo4j® (graph database)
- Cypher(graph query language)



- Python®
- Python visualization packages: (plotly® dash)



Step1

Table-like Graph Explorer

Search:

pain

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

relationship dir

filter data...

☒ Take into report

Source:

filter data...

Step2

Table-like Graph Explorer

Search:

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	AE_BODY_SYSTEM	>	["BodySystem"]	1	1
☐	AE_PREFERRED_TERM	>	["AdverseEventPreferredTerm"]	1	1
☐	AE_PREFERRED_TERM	<	["AdverseEvent"]	1	11
☐	EXPERIENCED	<	["Subject"]	1	1
☐	LINK_TO_AE	<	["ConcomitantMedication"]	1	1

☒ Take into report

Source:

	id	class	AEDECOD	AEREL	DOMAIN	AESTDTC	AEENDTC	AESTDY	AESER	
	filter									
☐	1382	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1988-08-11T00:02:00	1988-08-17	1	N	GEN
☐	2824	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1988-05-01T21:00:00	1988-05-14	6	N	GEN
☐	4150	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1988-05-30T12:50:00	1988-06-01	10	N	GEN
☐	4618	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1988-01-03T05:10:00	1988-01-24	6	N	GEN
☐	6503	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1989-07-09T00:01:00	1989-07-18	4	N	GEN
☐	6655	["AdverseEvent"]	Back pain	NOT RELATED	AE	1989-01-07T00:02:00	1989-01-16	1	N	
☐	6772	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1989-06-05T22:00:00	1989-06-26	6	N	GEN
☐	7099	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1988-01-30T14:30:00	1988-02-01	4	N	GEN
☐	1832	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1987-11-25T17:40:00	1987-11-28	2	N	GEN
☐	2607	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1988-09-06T19:40:00	1988-09-08	6	N	GEN

Step2 – Relationships

`(:AdverseEvent) - [:AE_BODY_SYSTEM] -> (:BodySystem)`

`(:AdverseEvent) - [:AE_PREFERRED_TERM] -> (:AdverseEventPreferredTerm)`

`(AdverseEventPreferredTerm) <- [:AE_PREFERRED_TERM] - (:AdverseEvent)`

`(:AdverseEvent) <- [:EXPERIENCED] - (:Subject)`

`(:AdverseEvent) <- [:LINK_TO_AE] - (:ConcomitantMedication)`

Step3

Table-like Graph Explorer

Search:

pain

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	AE_BODY_SYSTEM	>	["BodySystem"]	1	1
☐	AE_PREFERRED_TERM	>	["AdverseEventPreferredTerm"]	1	1
☐	AE_PREFERRED_TERM	<	["AdverseEvent"]	1	11
☐	EXPERIENCED	<	["Subject"]	1	1
☐	LINK_TO_AE	<	["ConcomitantMedication"]	1	1

Click

☒ Take into report

Source:

	id	class	AEDECOD	AEREL	DOMAIN	AESTDTC	AEENDTC	AESTDY	AESER	
	filter									
☐	1382	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1988-08-11T00:02:00	1988-08-17	1	N	GEN
☐	2824	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1988-05-01T21:00:00	1988-05-14	6	N	GEN
☐	4150	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1988-05-30T12:50:00	1988-06-01	10	N	GEN
☐	4618	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1988-01-03T05:10:00	1988-01-24	6	N	GEN
☐	6503	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1989-07-09T00:01:00	1989-07-18	4	N	GEN
☐	6655	["AdverseEvent"]	Back pain	NOT RELATED	AE	1989-01-07T00:02:00	1989-01-16	1	N	
☐	6772	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1989-06-05T22:00:00	1989-06-26	6	N	GEN
☐	7099	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1988-01-30T14:30:00	1988-02-01	4	N	GEN
☐	1832	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1987-11-25T17:40:00	1987-11-28	2	N	GEN
☐	2607	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1988-09-06T19:40:00	1988-09-08	6	N	GEN

Step3

Table-like Graph Explorer

Search:

pain

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	CM_CATEGORY	>	["ConcomitantMedicationCategory"]	1	1
☐	CM_INDICATION	>	["ConcomitantMedicationIndication"]	1	1
☐	CM_PREFERRED_TERM	>	["ConcomitantMedicationPreferredTerm"]	1	1
☐	LINK_TO_AE	>	["AdverseEvent"]	1	1
☐	WAS_ADMINISTRED_WITH	<	["Subject"]	1	1

☒ Take into report

Source:

	id	class	CMSEQ	DOMAIN	CMDOSTOT	CMDECOD	STUDYID	CMDOSE	
	filter								
☐	12694	["ConcomitantMedication"]	4	CM	240	MAGNESIUM HYDROXIDE	NICSAH1	240	1988-05-
☐	15822	["ConcomitantMedication"]	2	CM	5	ACETAMINOPHEN	NICSAH1	5	1988-05-
☐	16931	["ConcomitantMedication"]	2	CM	1300	ACETAMINOPHEN	NICSAH1	1300	1988-01-
☐	21531	["ConcomitantMedication"]	19	CM	300	PHENOBARBITAL	NICSAH1	300	1989-06-
☐	11701	["ConcomitantMedication"]	1	CM	10	ACETAMINOPHEN	NICSAH1	10	1988-09-
☐	21610	["ConcomitantMedication"]	5	CM	8	ACETAMINOPHEN	NICSAH1	8	1989-08-

Step4

Table-like Graph Explorer

Search:

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	CM_CATEGORY	>	["ConcomitantMedicationCategory"]	1	1
☐	CM_INDICATION	>	["ConcomitantMedicationIndication"]	1	1
☐	CM_PREFERRED_TERM	>	["ConcomitantMedicationPreferredTerm"]	1	1
☐	LINK_TO_AE	>	["AdverseEvent"]	1	1
☐	WAS_ADMINISTRED_WITH	<	["Subject"]	1	1

☒ Take into report

Source:

	id	class	CMSEQ	DOMAIN	CMDOSTOT	CMDECOD	STUDYID	CMDOSE	
	filter								
☐	12694	["ConcomitantMedication"]	4	CM	240	MAGNESIUM HYDROXIDE	NICSAH1	240	1988-05-
☐	15822	["ConcomitantMedication"]	2	CM	5	ACETAMINOPHEN	NICSAH1	5	1988-05-
☐	16931	["ConcomitantMedication"]	2	CM	1300	ACETAMINOPHEN	NICSAH1	1300	1988-01-
☐	21531	["ConcomitantMedication"]	19	CM	300	PHENOBARBITAL	NICSAH1	300	1989-06-
☐	11701	["ConcomitantMedication"]	1	CM	10	ACETAMINOPHEN	NICSAH1	10	1988-09-
☐	21610	["ConcomitantMedication"]	5	CM	8	ACETAMINOPHEN	NICSAH1	8	1989-08-

Click

Step4

Table-like Graph Explorer

Search:

pain

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	DISPOSITION_EVENT	>	["DispositionEvent"]	4	4
☐	EXPERIENCED	>	["AdverseEvent"]	4	8
☐	IN_SITE	>	["Site"]	1	1
☐	TREATED_WITH	>	["Treatment"]	1	1
☐	WAS_ADMINISTRED_WITH	>	["ConcomitantMedication"]	9	38
☐	HAS_SUBJECT	<	["Study"]	1	1

☒ Take into report

Source:

	id	class	DOMAIN	INVID	BRTHDTC	SEX	ARMCD	DMDTC	AGEU	SUBJID	SITEID	INVNAM
	filter data...											
☐	304	["Subject"]	DM	1A	1939-09-23	M	2	1988-04-25	YEARS	21008	02	021A
☐	491	["Subject"]	DM	2B	1945-01-11	F	2	1988-05-20	YEARS	282023	28	282B
☐	552	["Subject"]	DM	1A	1943-05-09	M	0	1987-12-25	YEARS	31003	03	031A
☐	857	["Subject"]	DM	1A	1945-11-04	F	0	1989-05-28	YEARS	461026	46	461A
☐	252	["Subject"]	DM	1A	1951-02-22	F	2	1988-08-30	YEARS	181014	18	181A
☐	860	["Subject"]	DM	1A	1949-05-11	F	2	1989-07-27	YEARS	461029	46	461A

Step5

Table-like Graph Explorer

Search:

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	DISPOSITION_EVENT	>	["DispositionEvent"]	4	4
☐	EXPERIENCED	>	["AdverseEvent"]	4	8
☐	IN_SITE	>	["Site"]	1	1
☐	TREATED_WITH	>	["Treatment"]	1	1
☐	WAS_ADMINISTRED_WITH	>	["ConcomitantMedication"]	9	38
☐	HAS_SUBJECT	<	["Study"]	1	1

☒ Take into report

Source:

	id	class	DOMAIN	INVID	BRTHDTC	SEX	ARMCD	DMDTC	AGEU	SUBJID	SITEID	INVNAM
	filter data...											
☐	304	["Subject"]	DM	1A	1939-09-23	M	2	1988-04-25	YEARS	21008	02	021A
☐	491	["Subject"]	DM	2B	1945-01-11	F	2	1988-05-20	YEARS	282023	28	282B
☐	552	["Subject"]	DM	1A	1943-05-09	M	0	1987-12-25	YEARS	31003	03	031A
☐	857	["Subject"]	DM	1A	1945-11-04	F	0	1989-05-28	YEARS	461026	46	461A
☐	252	["Subject"]	DM	1A	1951-02-22	F	2	1988-08-30	YEARS	181014	18	181A
☐	860	["Subject"]	DM	1A	1949-05-11	F	2	1989-07-27	YEARS	461029	46	461A

Click

Step6

Table-like Graph Explorer

Search:

pain

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	CLINTRIALS_GOV	>	["ClinTrialsGovStudy"]	1	1
☐	HAS_SITE	>	["Site"]	40	40
☐	HAS_SUBJECT	>	["Subject"]	906	906
☐	HAS_TREATMENT	>	["Treatment"]	3	3

☒ Take into report

Source:

	id	class	Class	nctId	STUDYID
	filter data...				
☐	1033	["Study"]	Study	NCT00XXXXXX	NICSAH1

Click

Step7

Table-like Graph Explorer

Search:

pain

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	IN_CONDITION	>	["Condition"]	1	1
☐	CLINTRIALS_GOV	<	["Study"]	1	1

☒ Take into report

Source:

	id	class	Condition	Phase	Allocation	InterventionModel	nctId	HealthyVolunteers	
	filter								
☐	29242	["ClinTrialsGovStudy"]	Subarachnoid Hemorrhage	Phase 3	Randomized	Parallel Assignment	NCT00XXXXXX	No Unive	

Click

Step8

Table-like Graph Explorer

Search:

pain

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	IN_CONDITION	<	["ClinTrialsGovStudy"]	16	16

☒ Take into report

Source:

	id	class	Condition	Class
	filter data...			
☐	29243	["Condition"]	Subarachnoid Hemorrhage	Condition

Filtering was possible at each step

Table-like Graph Explorer

Search:

pain

SEARCH

UNDO

RESET

☐ Disregard previous

Relationships:

	relationship	dir	target_class	min_cnt	max_cnt
	filter data...				
☐	AE_BODY_SYSTEM	>	["BodySystem"]	1	1
☐	AE_PREFERRED_TERM	>	["AdverseEventPreferredTerm"]	1	1
☐	AE_PREFERRED_TERM	<	["AdverseEvent"]	1	11
☐	EXPERIENCED	<	["Subject"]	1	1
☐	LINK_TO_AE	<	["ConcomitantMedication"]	1	1

☐ Take into report

Source:

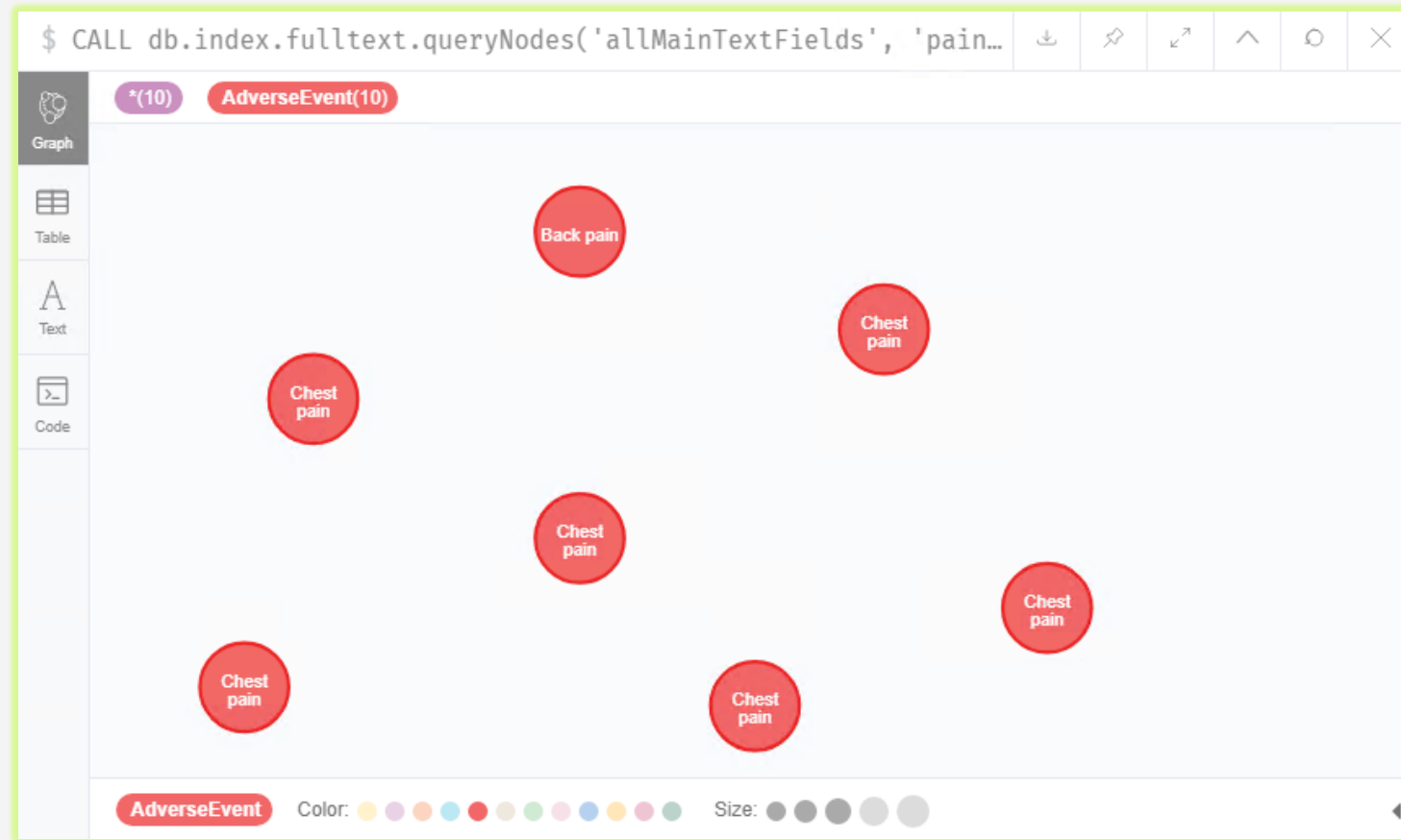
	id	class	AEDECOD	AEREL	DOMAIN	AESTDTC	AEENDTC	AESTDY	AESER
	filter							<4	
☐	1382	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1988-08-11T00:02:00	1988-08-17	1	N G
☐	6655	["AdverseEvent"]	Back pain	NOT RELATED	AE	1989-01-07T00:02:00	1989-01-16	1	N
☐	1832	["AdverseEvent"]	Chest pain	NOT RELATED	AE	1987-11-25T17:40:00	1987-11-28	2	N G
☐	4699	["AdverseEvent"]	Chest pain	UNLIKELY RELATED	AE	1988-10-06T00:02:00	1988-10-13	3	N G
☐	5391	["AdverseEvent"]	Chest pain	POSSIBLY RELATED	AE	1988-12-21T16:41:00	1988-12-29	2	N G

How does it work in the backend?



Step1

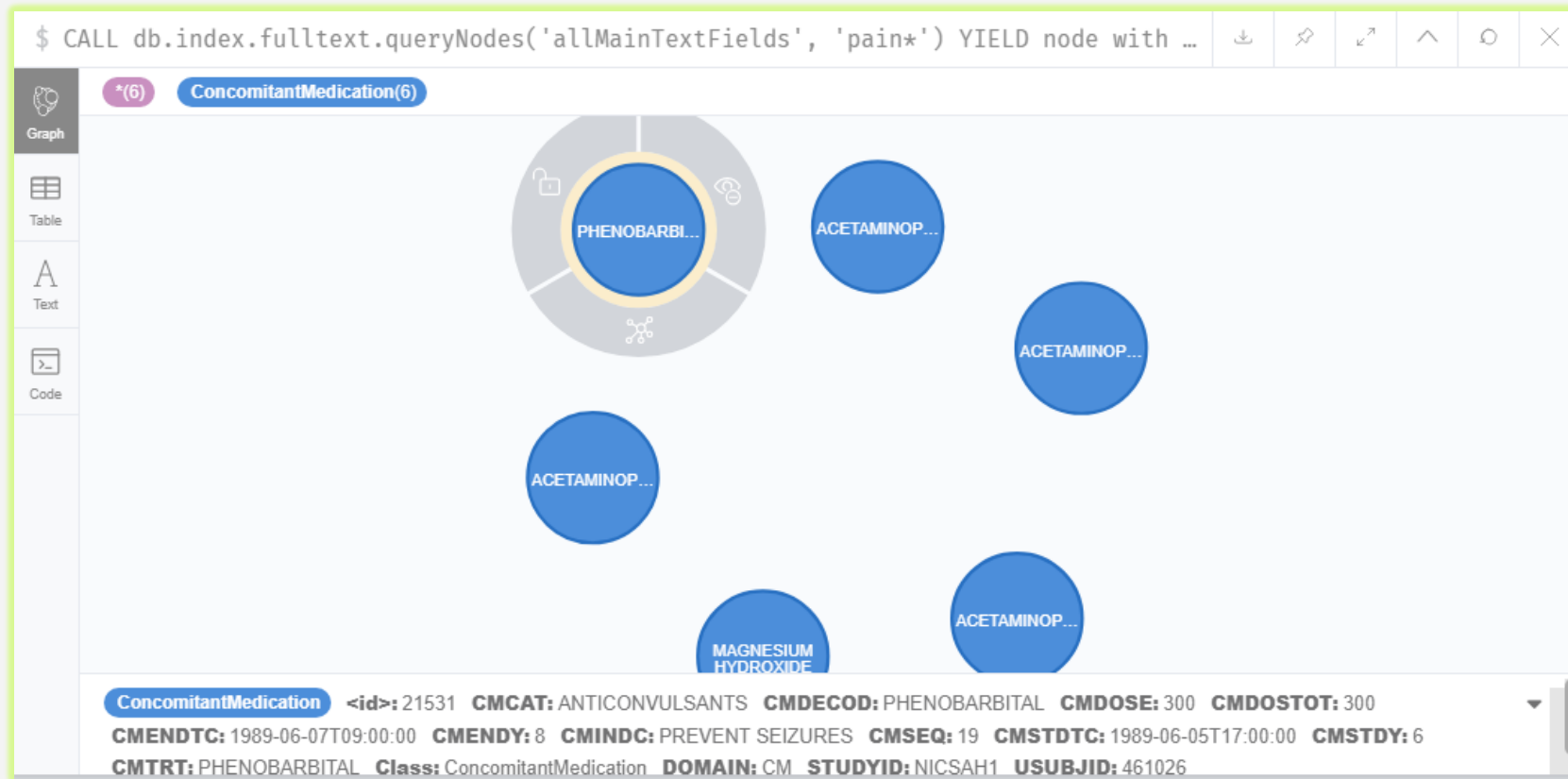
CALL db.index.fulltext.queryNodes('allMainTextFields', 'pain*') YIELD node
return node



<https://neo4j.com/developer/kb/fulltext-search-in-neo4j/>

Step2

```
CALL db.index.fulltext.queryNodes('allMainTextFields', 'pain*') YIELD node
with node as ae
MATCH (ae)<-[:LINK_TO_AE]-(cm:ConcomitantMedication)
return cm
```



Step2 - description of syntax

`() - [] -> ()`

`() - [:LINK_TO_AE] -> ()`

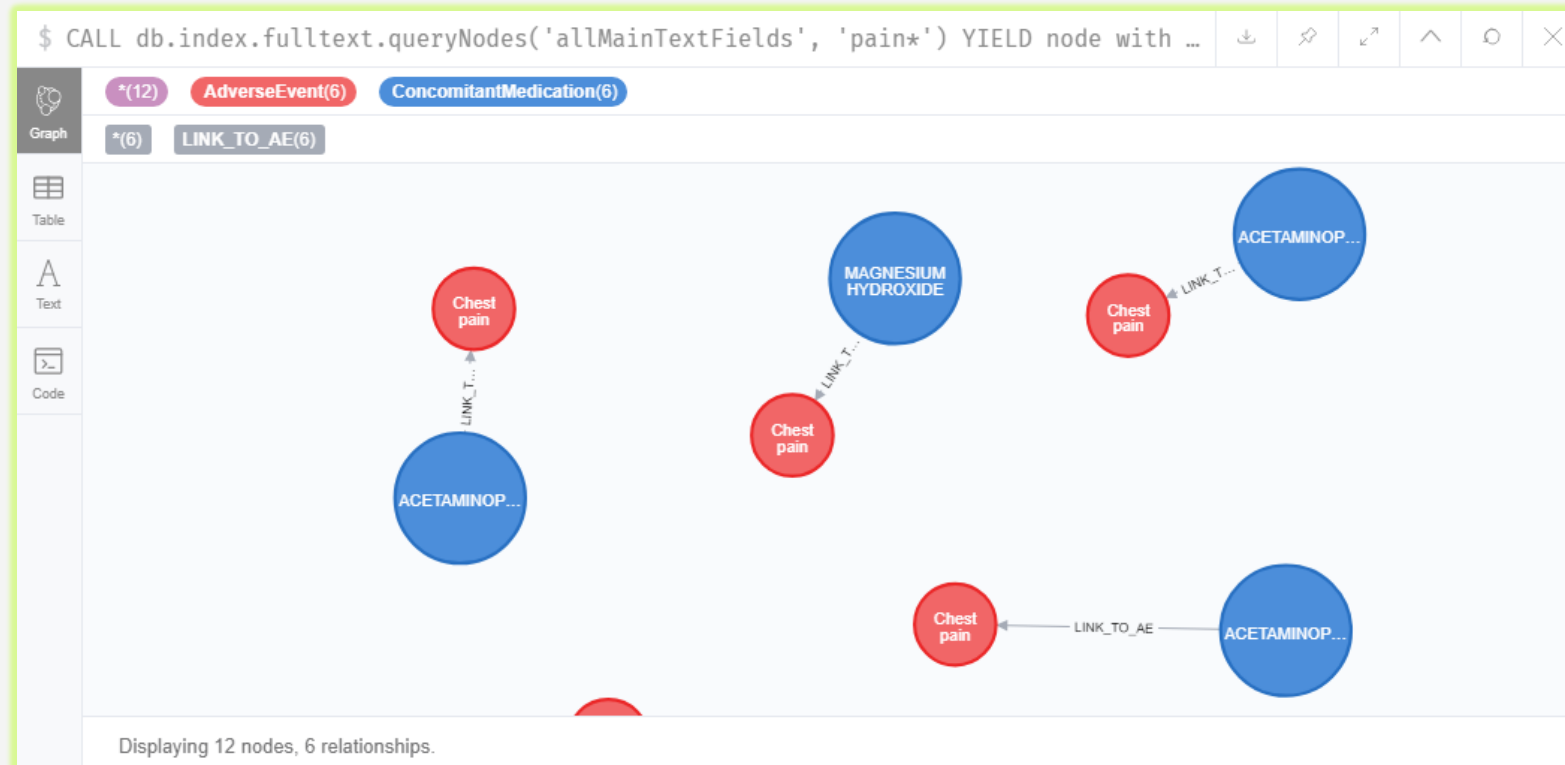
`(:ConcomitantMedication) - [:LINK_TO_AE] -> (:AdverseEvent)`

`(:ConcomitantMedication) - [:LINK_TO_AE] -> (ae)`

`(cm:ConcomitantMedication) - [:LINK_TO_AE] -> (ae)`

Step2 – displaying the path

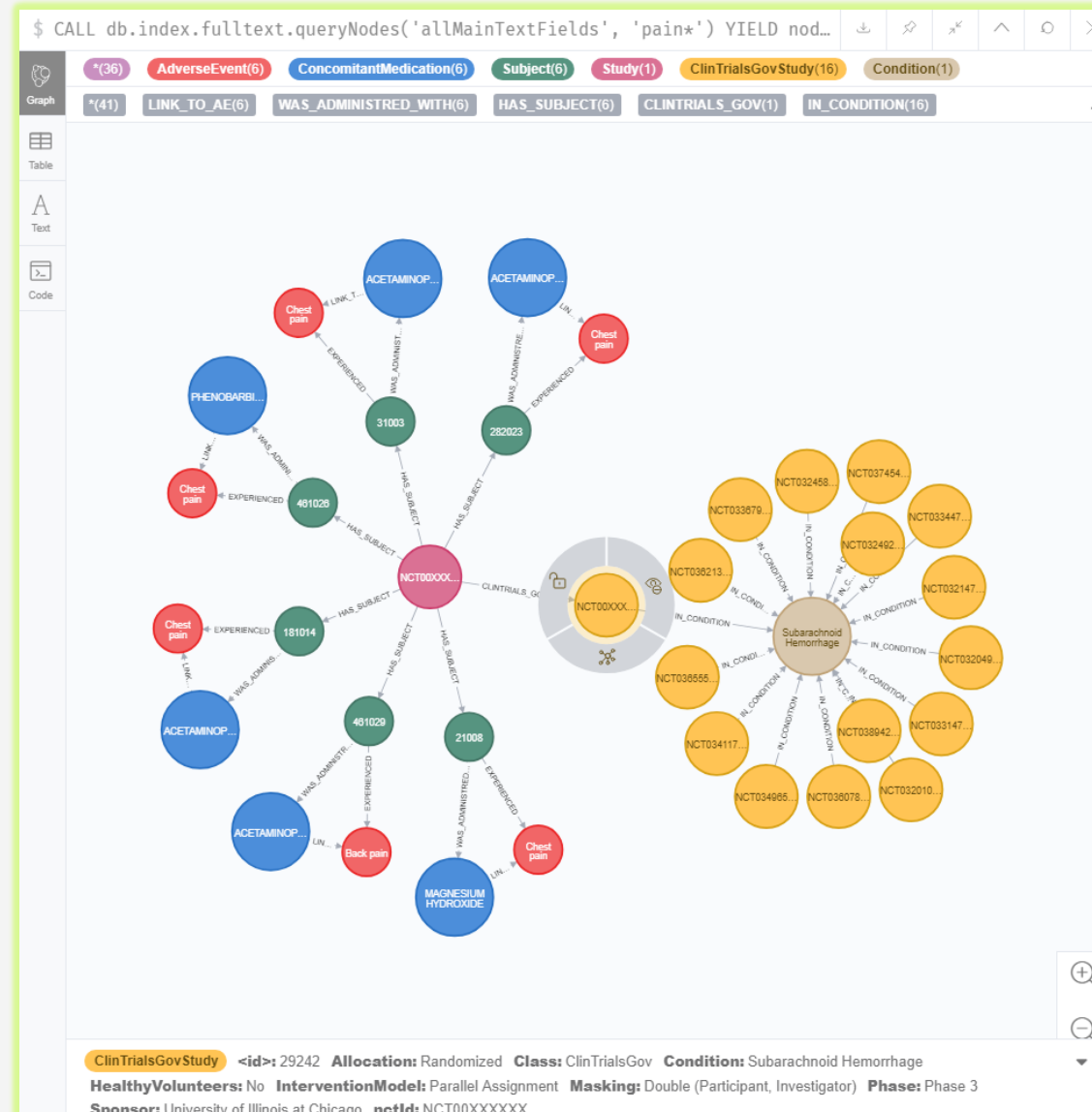
```
CALL db.index.fulltext.queryNodes('allMainTextFields', 'pain*')
YIELD node
with node as ae
MATCH p=(ae)<-[:LINK_TO_AE]-(cm:ConcomitantMedication)
return p
```



... Step8

```
CALL db.index.fulltext.queryNodes('allMainTextFields', 'pain*')
YIELD node
with node as ae
MATCH path=(ae)<-[LINK_TO_AE]-(cm:ConcomitantMedication)<-
[:WAS_ADMINISTERED_WITH]-(subj:Subject)<-[:HAS_SUBJECT]-
(study:Study)-[:CLINTRIALS_GOV]->(ctg_study:ClinTrialsGovStudy)-
[:IN_CONDITION]->(:Condition)<-[:IN_CONDITION]-
(other_studies_same_cond:ClinTrialsGovStudy)
return path
```

... Step8



... Schema

```
call db.schema()
```



Lessons learnt

- Graph databases allow faster processing of queries on highly connected data compared to relational databases
- Graph databases allow storage of connectedness as data and eliminate the necessity to perform joins (merges) during queries
- Neo4j graph database provides useful libraries and indexes (including full-text search indexing) out of the box
- Graph databases can serve as a convenient backend for web applications
- Cypher query language is intuitive and easy to learn
- Neo4j browser can be used to visualize a data model

Lessons learnt

- Web application development is possible with pure python without javascript knowledge
- The developed web application is pretty generic and would allow to walk through data of any domain after creating a graph model for the corresponding data
- As graph database technology is pretty new, it is hard to find experienced developers to support and/or further develop the code
- Plotly dash has limitations compared to pure javascript development and can be complex to learn and debug

In today's fast-moving and competitive world, the best thing you can do for your products, your company, and your people is institute a culture of experimentation and learning, and invest in the technical and management capabilities that enable it.

Nicole Forsgren, Accelerate

?