# The power of the Cognitive Probability Graph

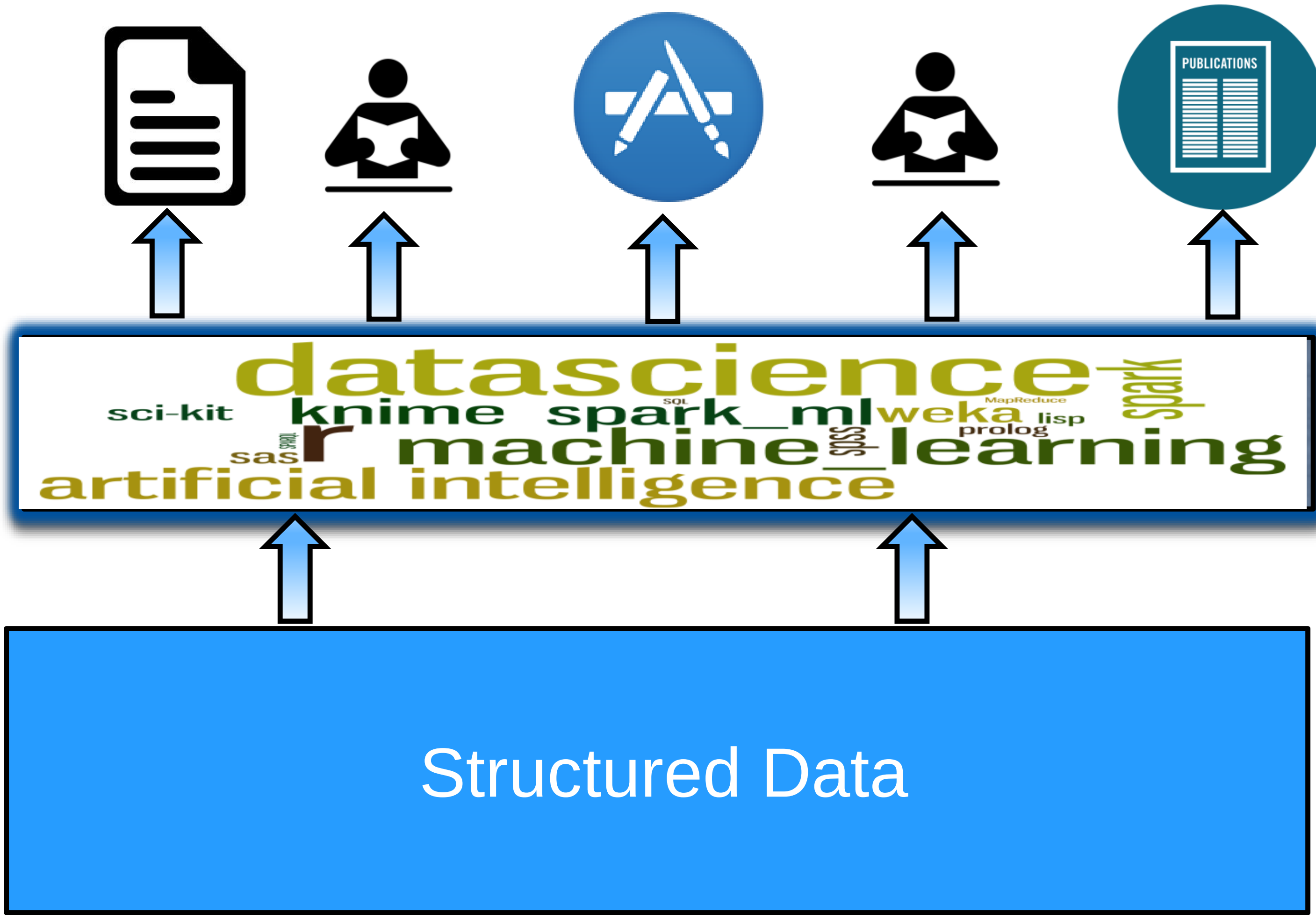
## (aka Cognitive Computing)

June 2016

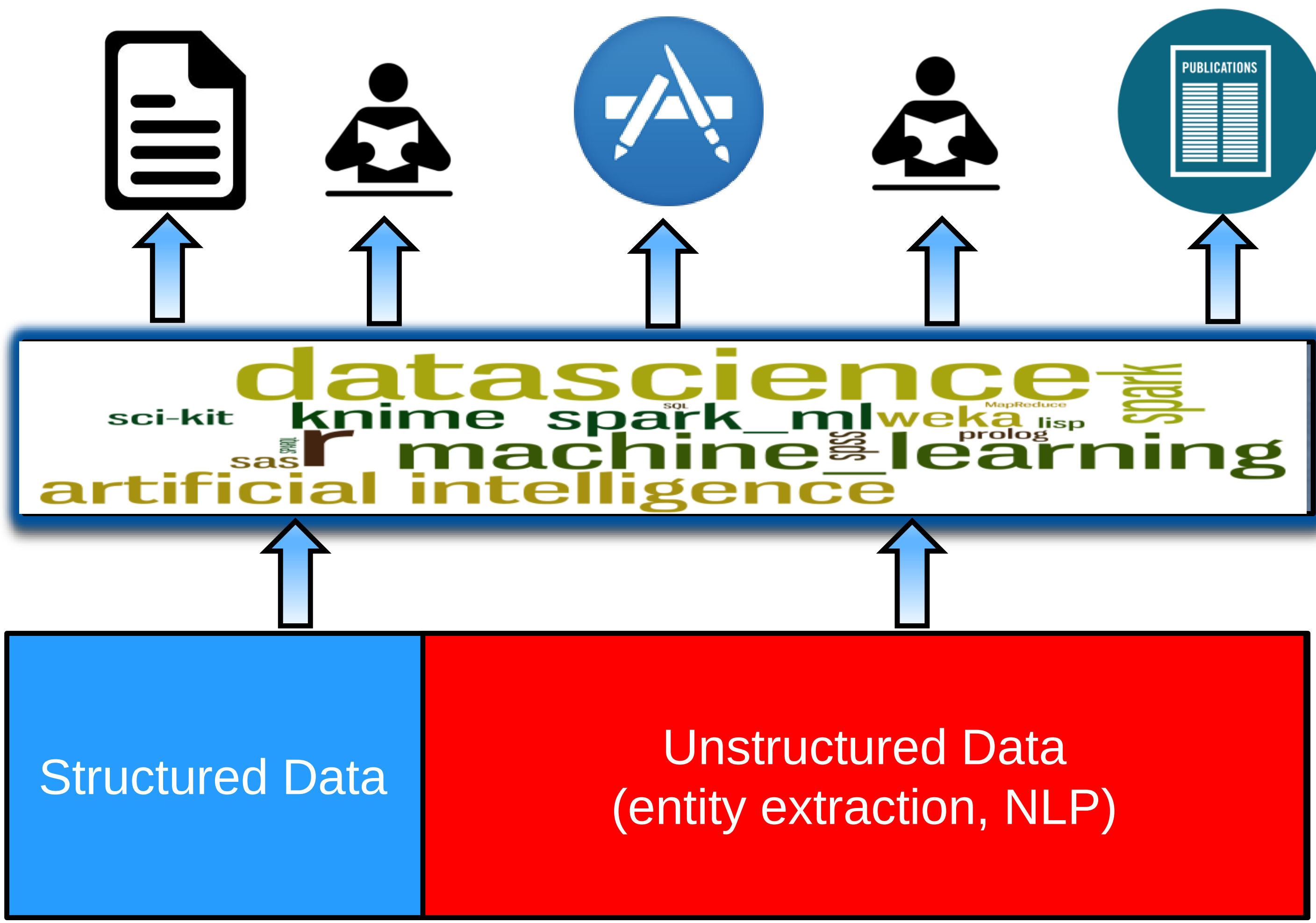
Jans Aasman
ja@franz.com

Franz Inc.

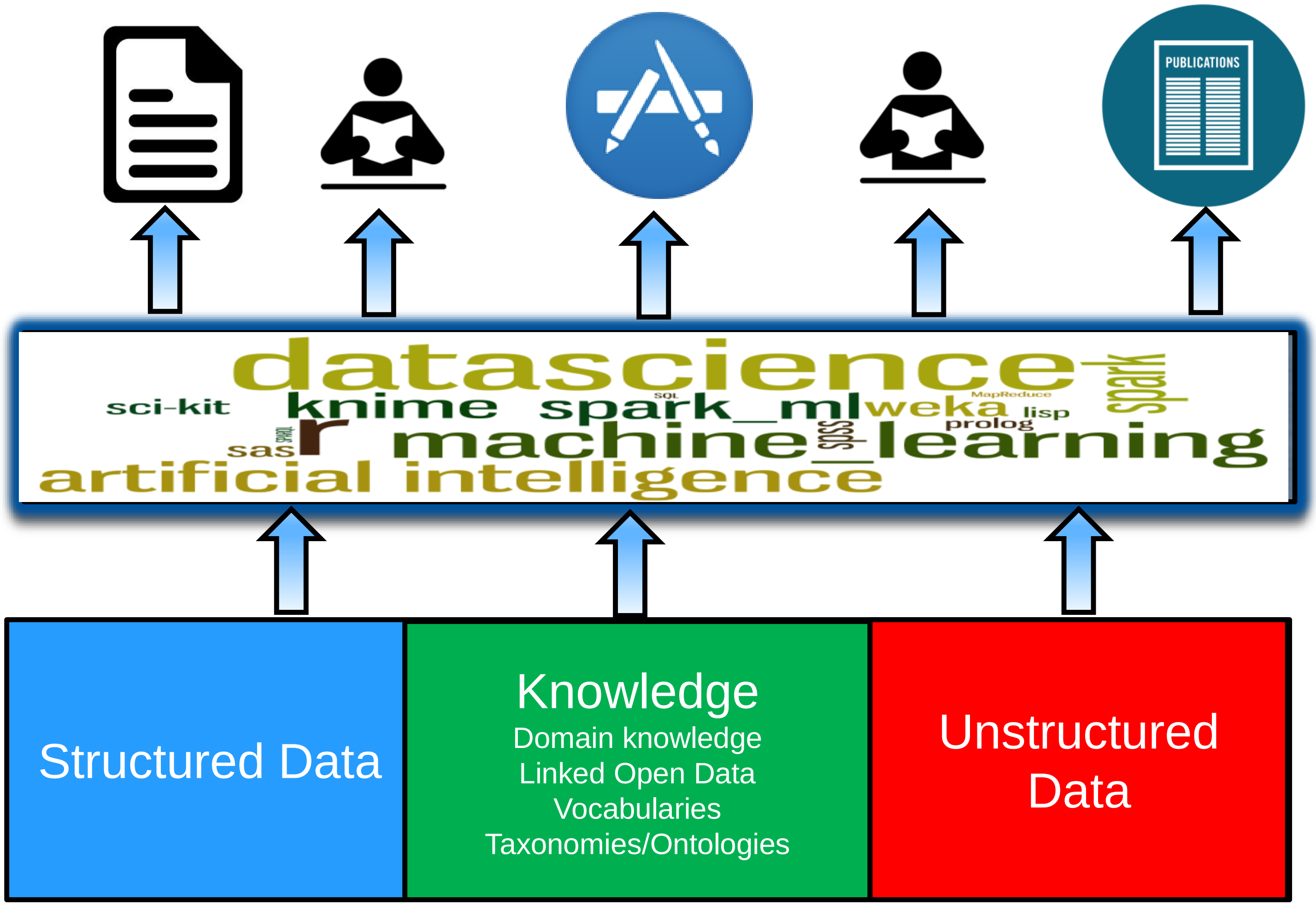
# 10 years ago


PUBLICATIONS
datascience
spark
sci-kit
knime
spark_ml
weka
lisp
prolog
r
sas
spss
machine_learning
artificial intelligence
Structured Data

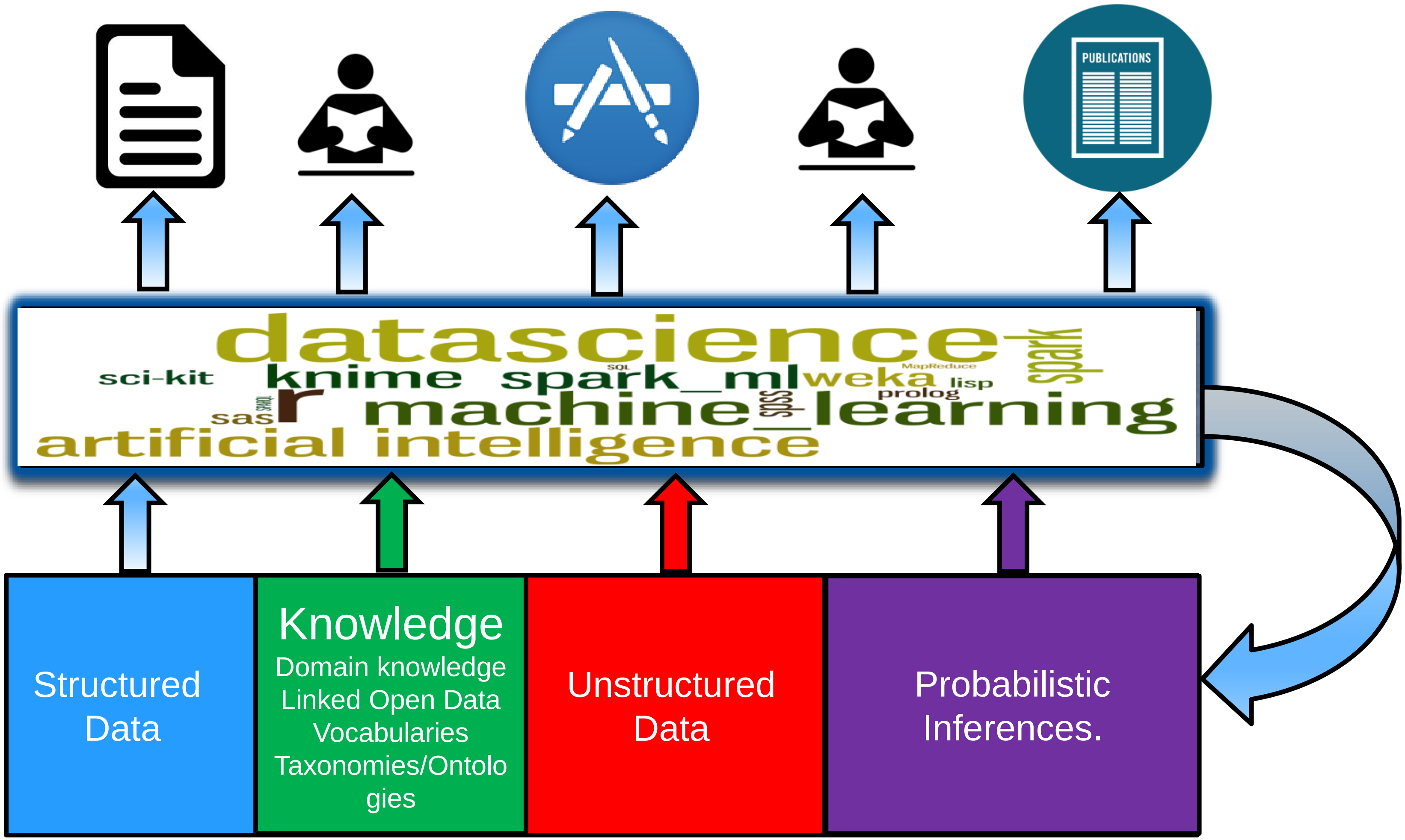
# 7 years ago


PUBLICATIONS
datascience
sci-kit
knime
spark_ml
weka
lisp
spark
prolog
r
sas
machine_learning
spss
artificial intelligence
Structured Data
Unstructured Data
(entity extraction, NLP)

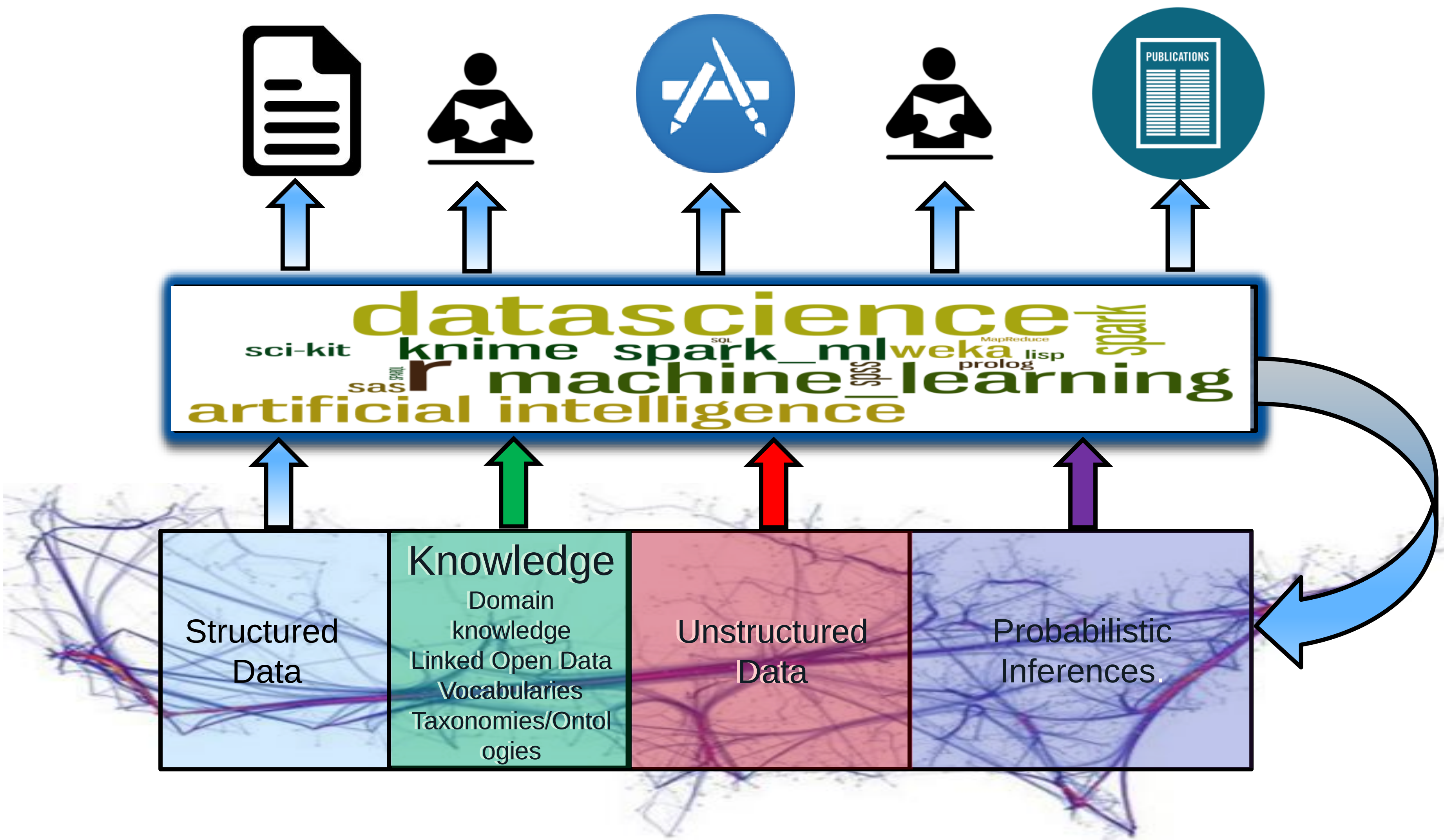
# 4 to 5 years ago



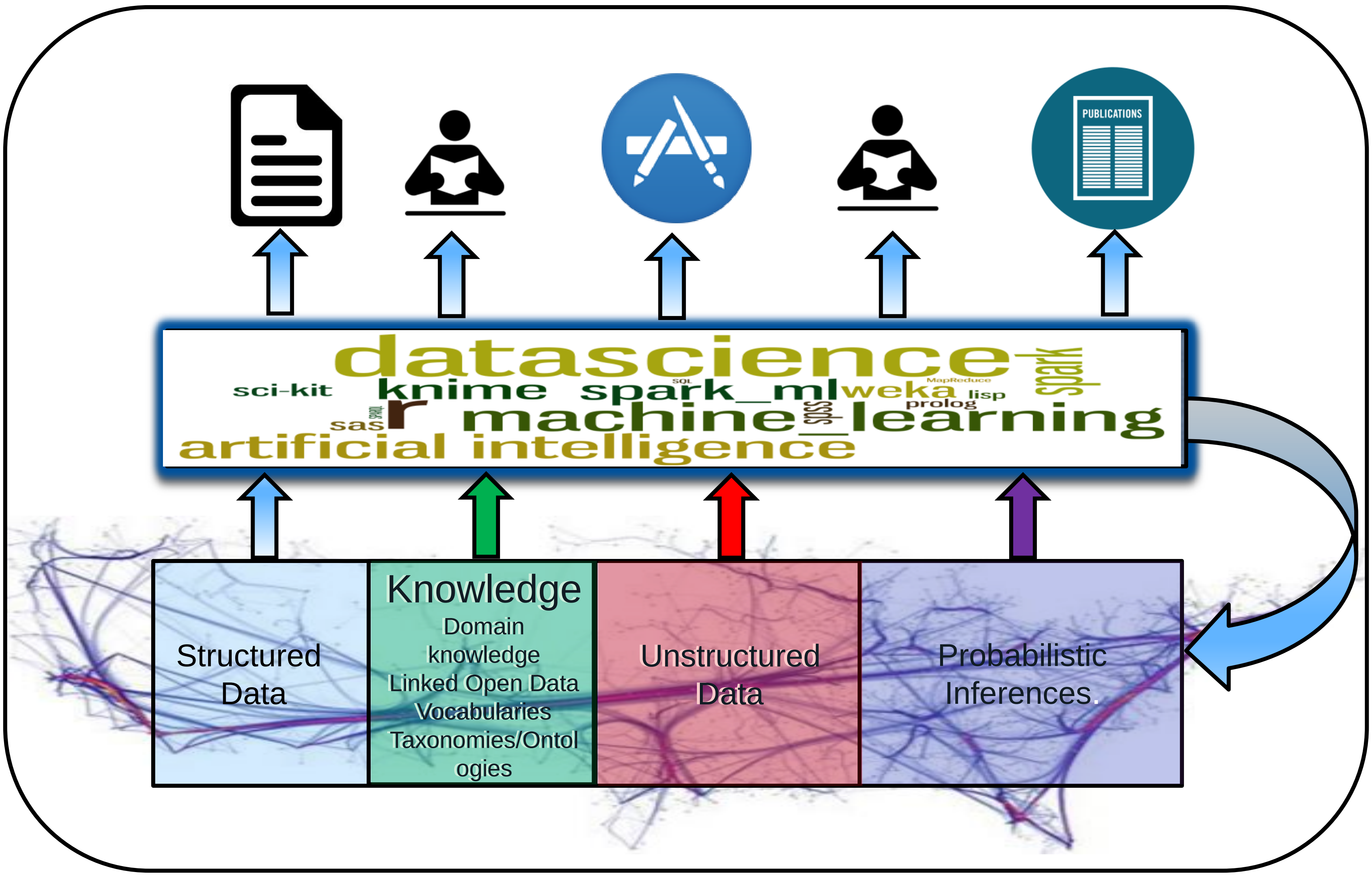
# New #1: Learning. Feed output of data science back into data infrastructure



# New # 2: everything in one (distributed) semantic graph



# AKA: Cognitive Computing

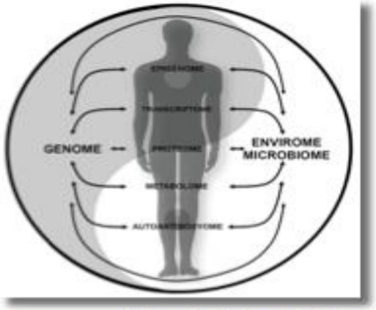


# Examples of (various degrees of) cognitive computing

Examples

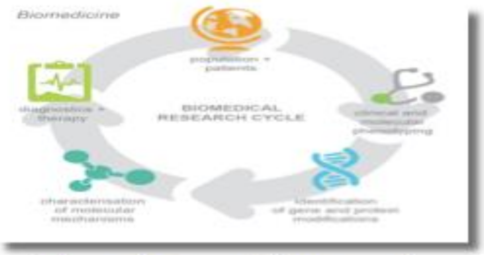
- Healthcare: If I have this class of diagnostics and I get this procedure what are some of the new symptoms I might get in the next two years.
- eCommerce and brand protection: find all my products based on product similarity
- Logistics: what can I statistically predict about part P breaking down and what other parts do I usually buy after that part breaks down.
- Police Intelligence: find the most plausible story of a temporally ordered shortest path between two criminals through observed (hard) facts and inferred (soft) facts.
- Fraud detection: find links between your local chamber of commerce and the Panama papers through similar names and addresses.

# One cognitive computing platform for all healthcare analytics




Personalized Medicine

Predictive Modeling




Translational Research

Fraud Detection

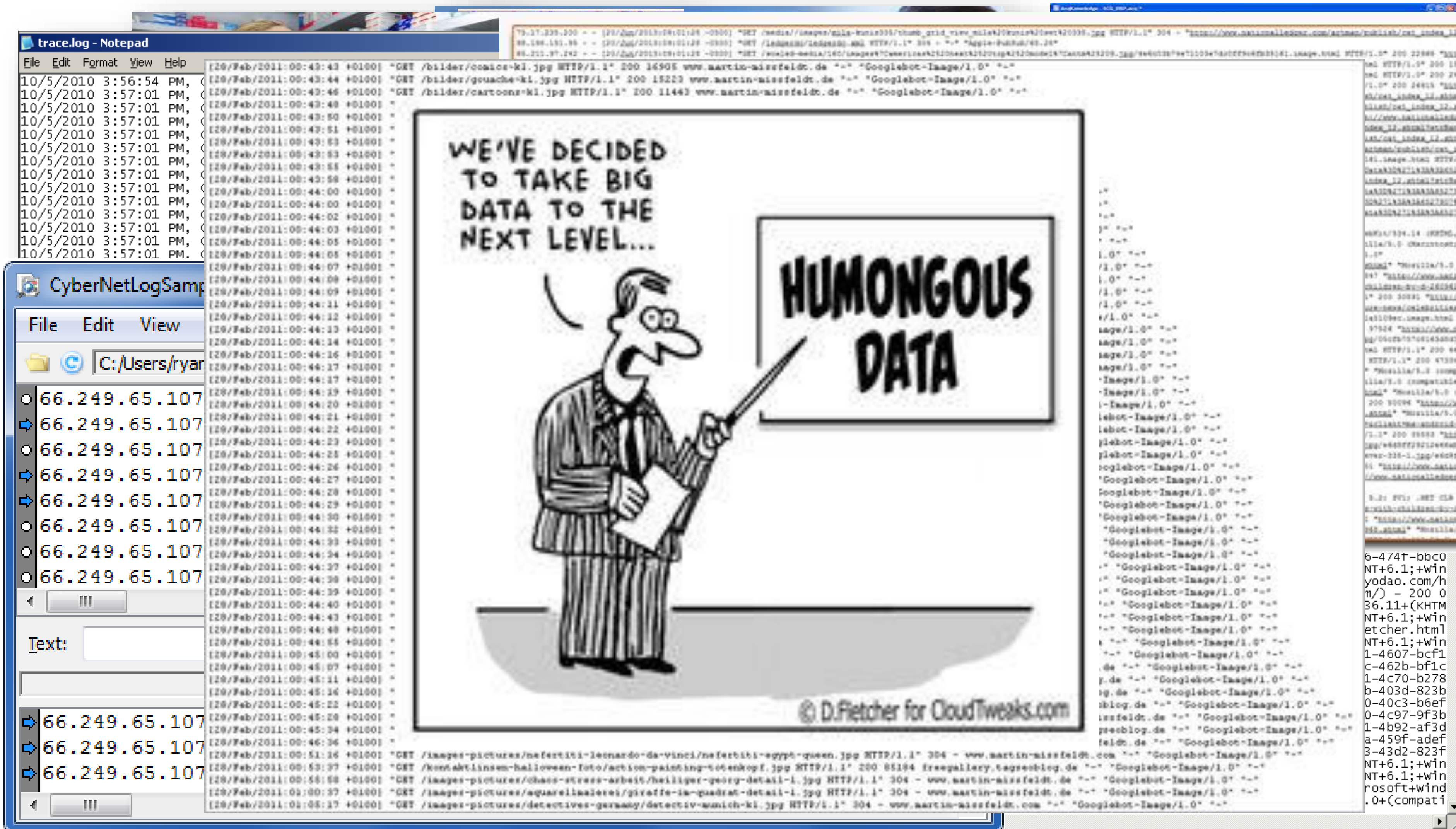
Risk Assessment

Business Intelligence

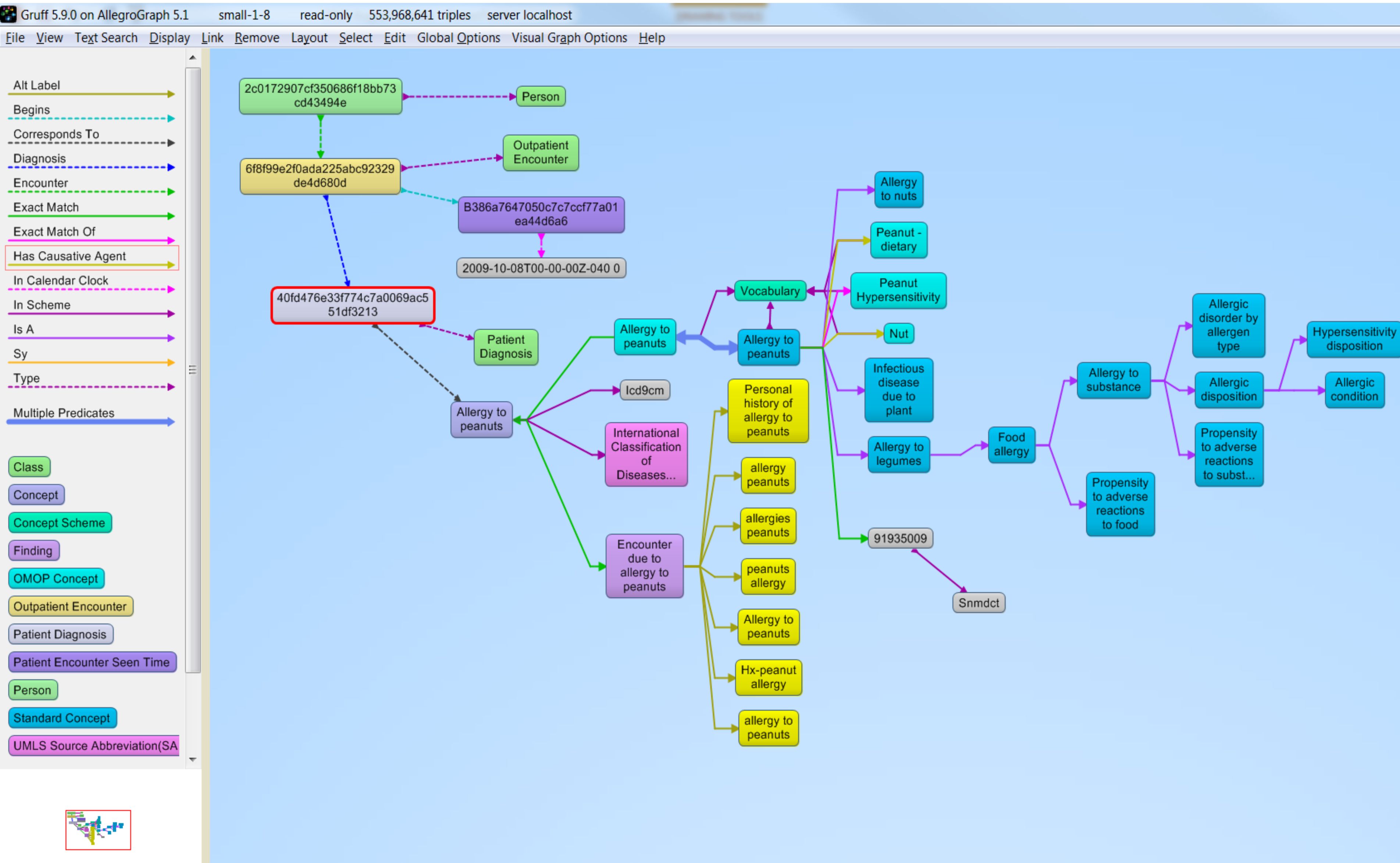

Public Health

Mobile Health

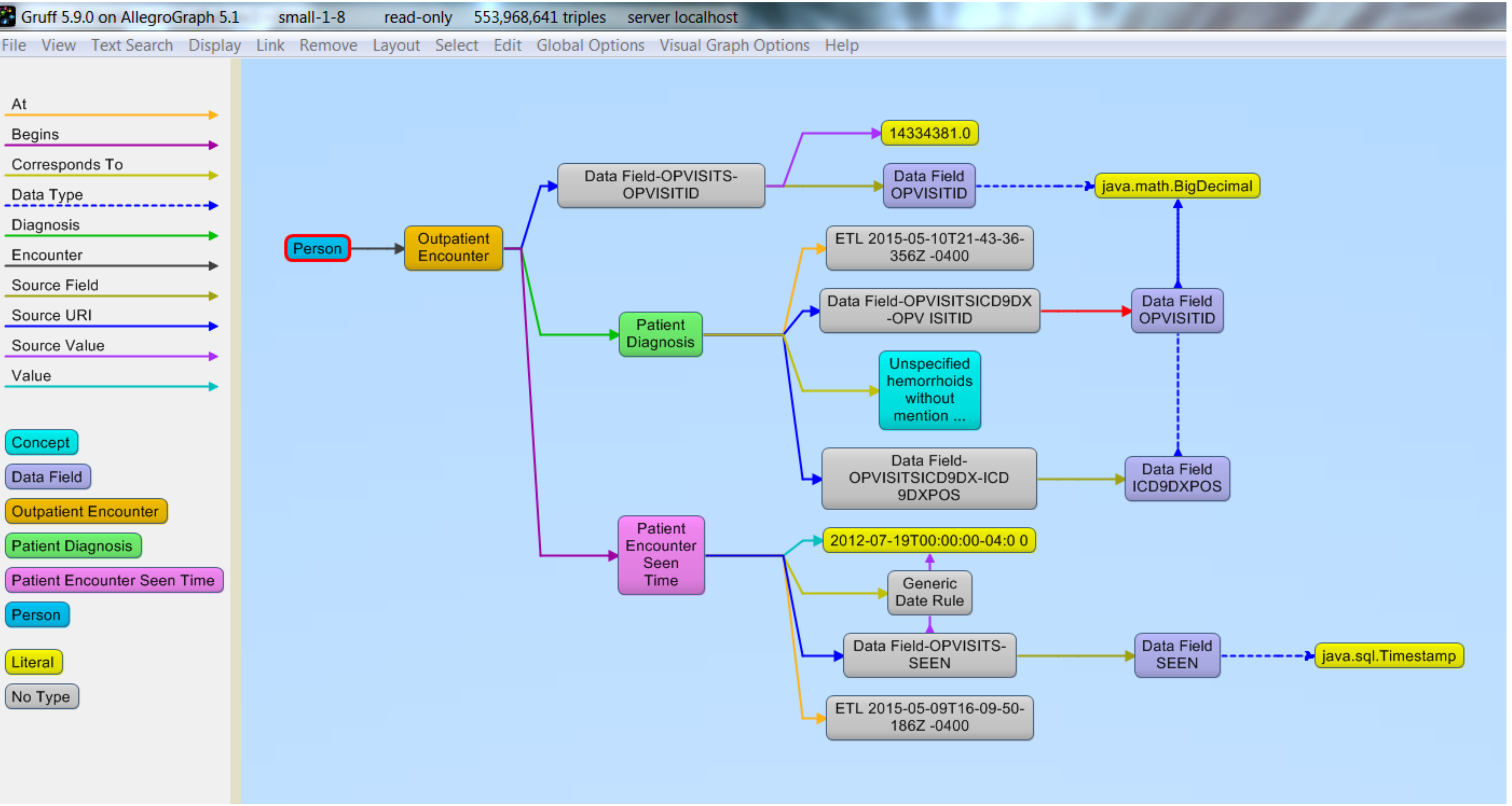
Decision Support


trace.log - Notepad
File Edit Format View Help
CyberNetLogSamp
File Edit View
Text:
WE'VE DECIDED TO TAKE BIG DATA TO THE NEXT LEVEL...
HUMONGOUS DATA
© D.Fletcher for CloudTweaks.com

# Structured patient data combined with complex integrated terminology

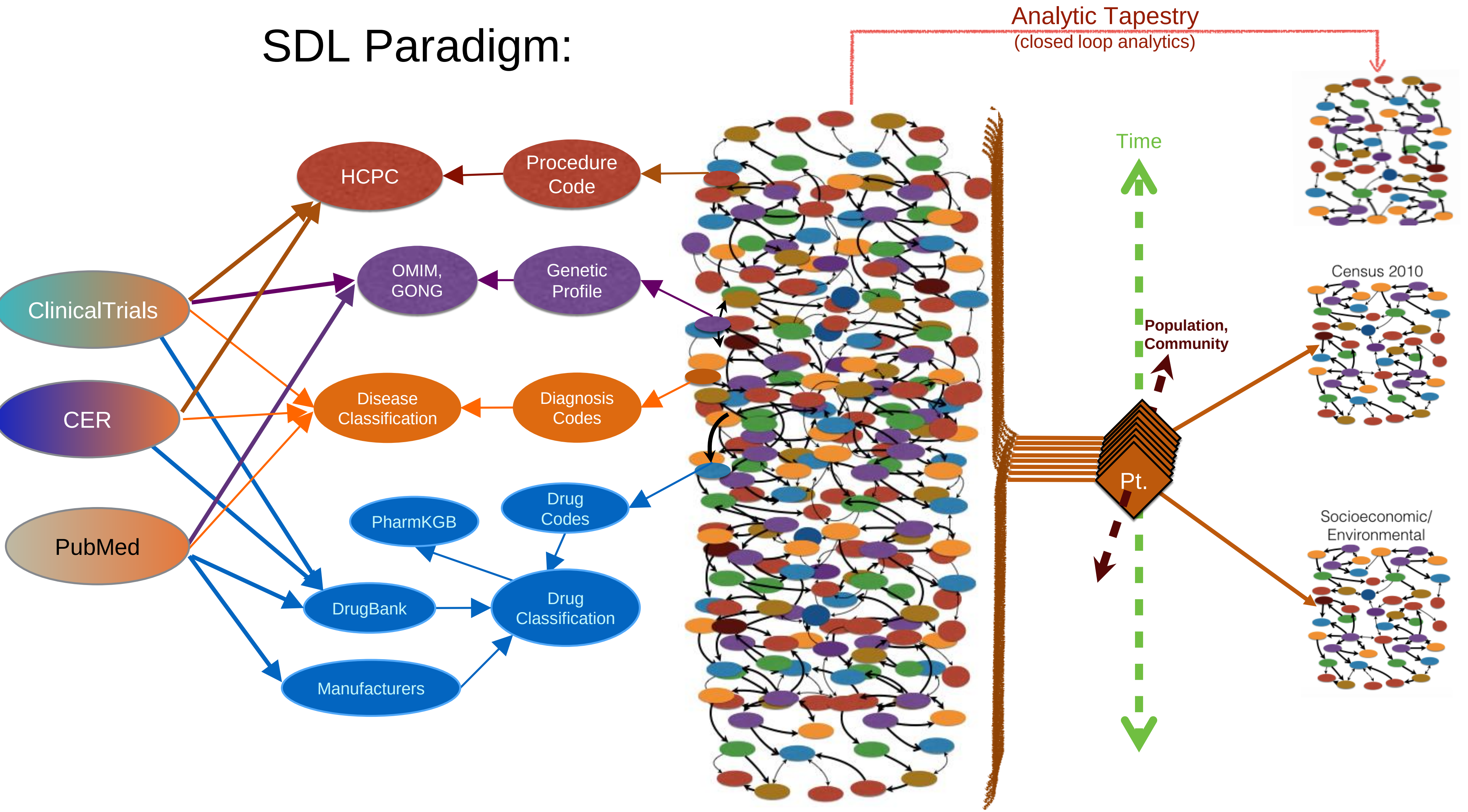


# Provenance for every value




FRANZ INC.

# SDL Paradigm:



# Healthcare: probabilistic inferences

## Why is this so important?

- Usually the output of data science results in reports and publications but
  - No formal trace where the data came from
  - No formal link to the actual methods you used, or who did it, or when you did it
  - Cannot be compared to earlier results
  - Cannot be used as building blocks for further research
  - In general : the output is not queryable
- This is not good for delivery of care, reproducibility of research findings, security and compliance, and results in loss of value-added information, and enterprise intellectual property and assets, and unnecessary duplication of efforts

# Odds ratio

In statistics, the **odds ratio**[1][2][3] (usually abbreviated "OR") is one of three main ways to quantify how strongly the presence or absence of property A is associated with the presence or absence of property B in a given population. If each individual in a population either does or does not have a property "A", (e.g. "high blood pressure"), and also either does or does not have a property "B" (e.g. "moderate alcohol consumption") where both properties are appropriately defined, then a ratio can be formed which quantitatively describes the association between the presence/absence of "A" (high blood pressure) and the presence/absence of "B" (moderate alcohol consumption) for individuals in the population. This ratio is the odds ratio (OR) and can be computed following these steps:

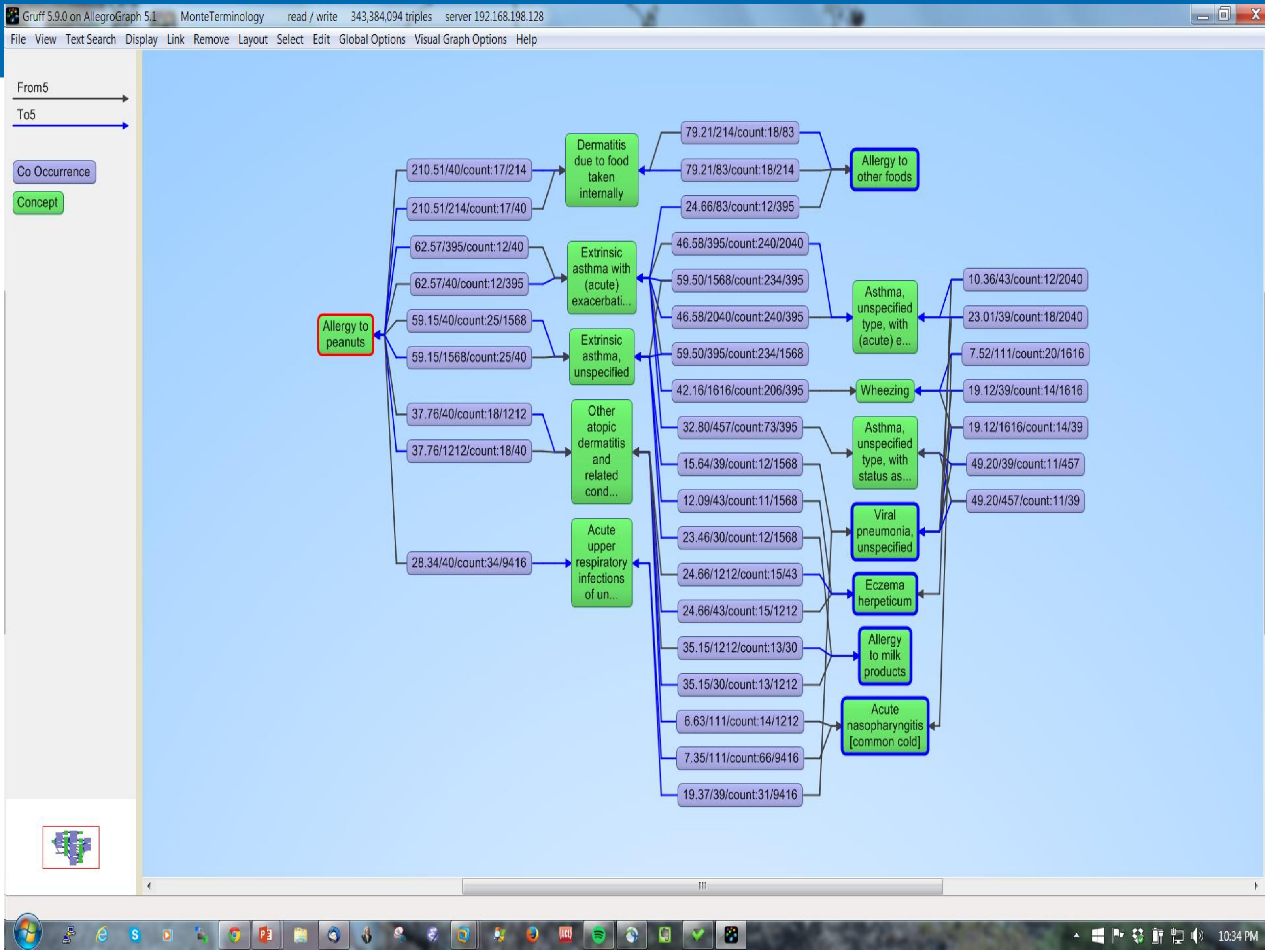
1. For a given individual that has "B" compute the odds that the same individual has "A"
2. For a given individual that does not have "B" compute the odds that the same individual has "A"
3. Divide the odds from step 1 by the odds from step 2 to obtain the odds ratio (OR).

**Patient Population** **1,802,464**

| | | Ingestion Dermatitis | | |
|---|---|---|---|---|
| | | **TO+** | **TO-** | |
| **Peanut Allergy** | **FROM+** | 544 (5) | 736 | **1,280** |
| | **FROM-** | 6304 | 1,795,424 | |
| | | **6,848** | | |

| | |
|---|---|
| **Odds Ratio** | **210.51** |
| **95% CI Lower** | **187.91** |
| **95% CI Upper** | **235.82** |

FRANZ INC.



Gruff 5.9.0 on AllegroGraph 5.1 MonteTerminology read / write 343,384,094 triples server 192.168.198.128
File View Text Search Display Link Remove Layout Select Edit Global Options Visual Graph Options Help
From5
To5
Co Occurrence
Concept
Allergy to peanuts
210.51/40/count:17/214
210.51/214/count:17/40
62.57/395/count:12/40
62.57/40/count:12/395
59.15/40/count:25/1568
59.15/1568/count:25/40
37.76/40/count:18/1212
37.76/1212/count:18/40
28.34/40/count:34/9416
Dermatitis due to food taken internally
Extrinsic asthma with (acute) exacerbati...
Extrinsic asthma, unspecified
Other atopic dermatitis and related cond...
Acute upper respiratory infections of un...
79.21/214/count:18/83
79.21/83/count:18/214
24.66/83/count:12/395
46.58/395/count:240/2040
59.50/1568/count:234/395
46.58/2040/count:240/395
59.50/395/count:234/1568
42.16/1616/count:206/395
32.80/457/count:73/395
15.64/39/count:12/1568
12.09/43/count:11/1568
23.46/30/count:12/1568
24.66/1212/count:15/43
24.66/43/count:15/1212
35.15/1212/count:13/30
35.15/30/count:13/1212
6.63/111/count:14/1212
7.35/111/count:66/9416
19.37/39/count:31/9416
Allergy to other foods
Asthma, unspecified type, with (acute) e...
Wheezing
Asthma, unspecified type, with status as...
Viral pneumonia, unspecified
Eczema herpeticum
Allergy to milk products
Acute nasopharyngitis [common cold]
10.36/43/count:12/2040
23.01/39/count:18/2040
7.52/111/count:20/1616
19.12/39/count:14/1616
19.12/1616/count:14/39
49.20/39/count:11/457
49.20/457/count:11/39
10:34 PM


FRANZ INC.

# Association rules

**Association rule learning** is a method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness.[1] Based on the concept of strong rules, Rakesh Agrawal et al.[2] introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule $\{\mathrm{onions}, \mathrm{potatoes}\} \Rightarrow \{\mathrm{burger}\}$ found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, they are likely to also buy hamburger meat. Such information can be used as the basis for decisions

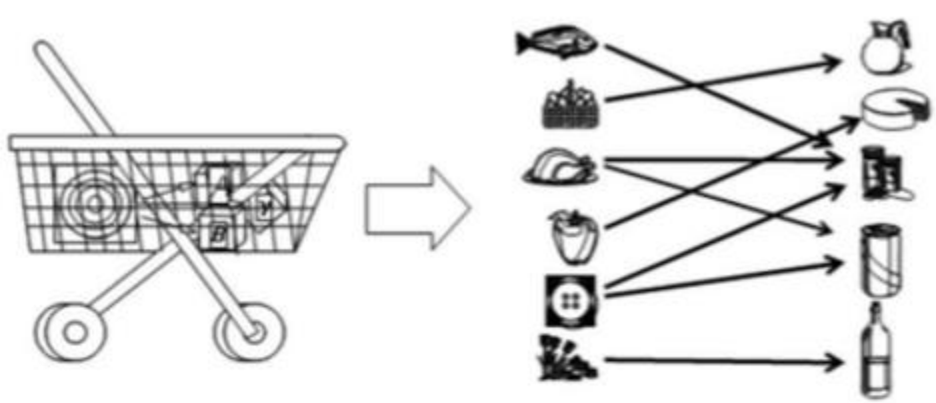
$$Rule\text{: } X \Rightarrow Y$$

$$Support = \frac{frq(X,Y)}{N}$$

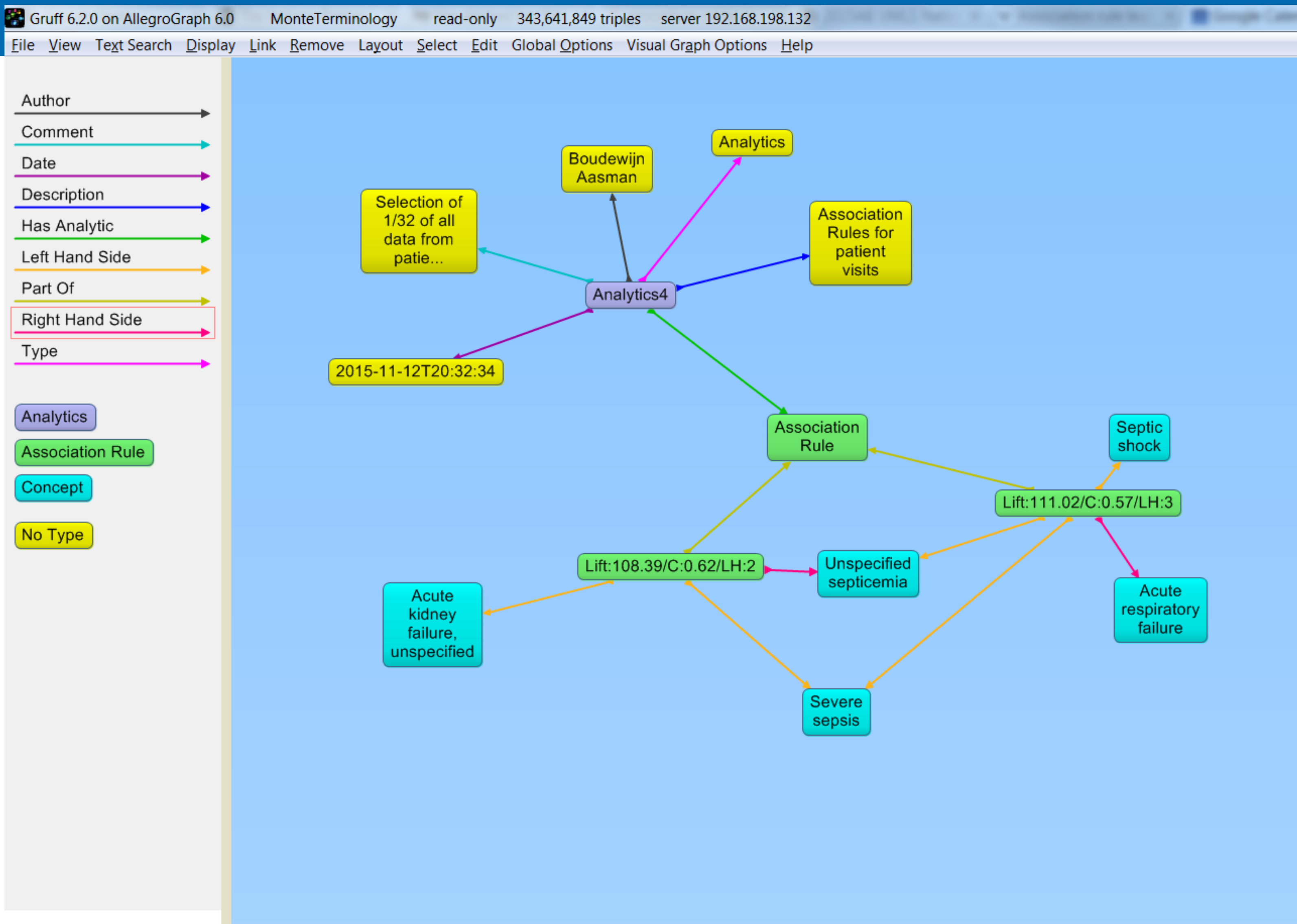
$$Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

MARKET BASKET ANALYSIS



*98% of people who purchased items A and B also purchased item C*

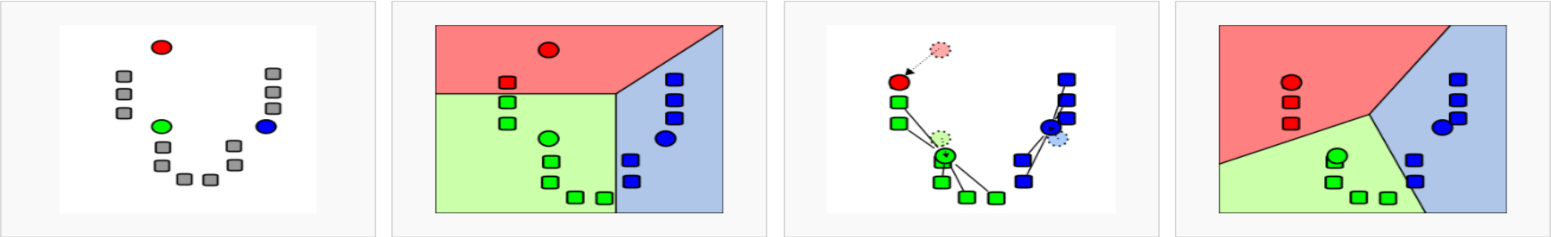

Gruff 6.2.0 on AllegroGraph 6.0
MonteTerminology
read-only
343,641,849 triples
server 192.168.198.132
File
View
Text Search
Display
Link
Remove
Layout
Select
Edit
Global Options
Visual Graph Options
Help
Author
Comment
Date
Description
Has Analytic
Left Hand Side
Part Of
Right Hand Side
Type
Analytics
Association Rule
Concept
No Type
Boudewijn Aasman
Analytics
Selection of 1/32 of all data from patie...
Association Rules for patient visits
Analytics4
2015-11-12T20:32:34
Association Rule
Septic shock
Lift:111.02/C:0.57/LH:3
Lift:108.39/C:0.62/LH:2
Unspecified septicemia
Acute kidney failure, unspecified
Acute respiratory failure
Severe sepsis

FRANZINC.

# K-means clustering

***k*-means clustering** is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. *k*-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

**Demonstration of the standard algorithm**

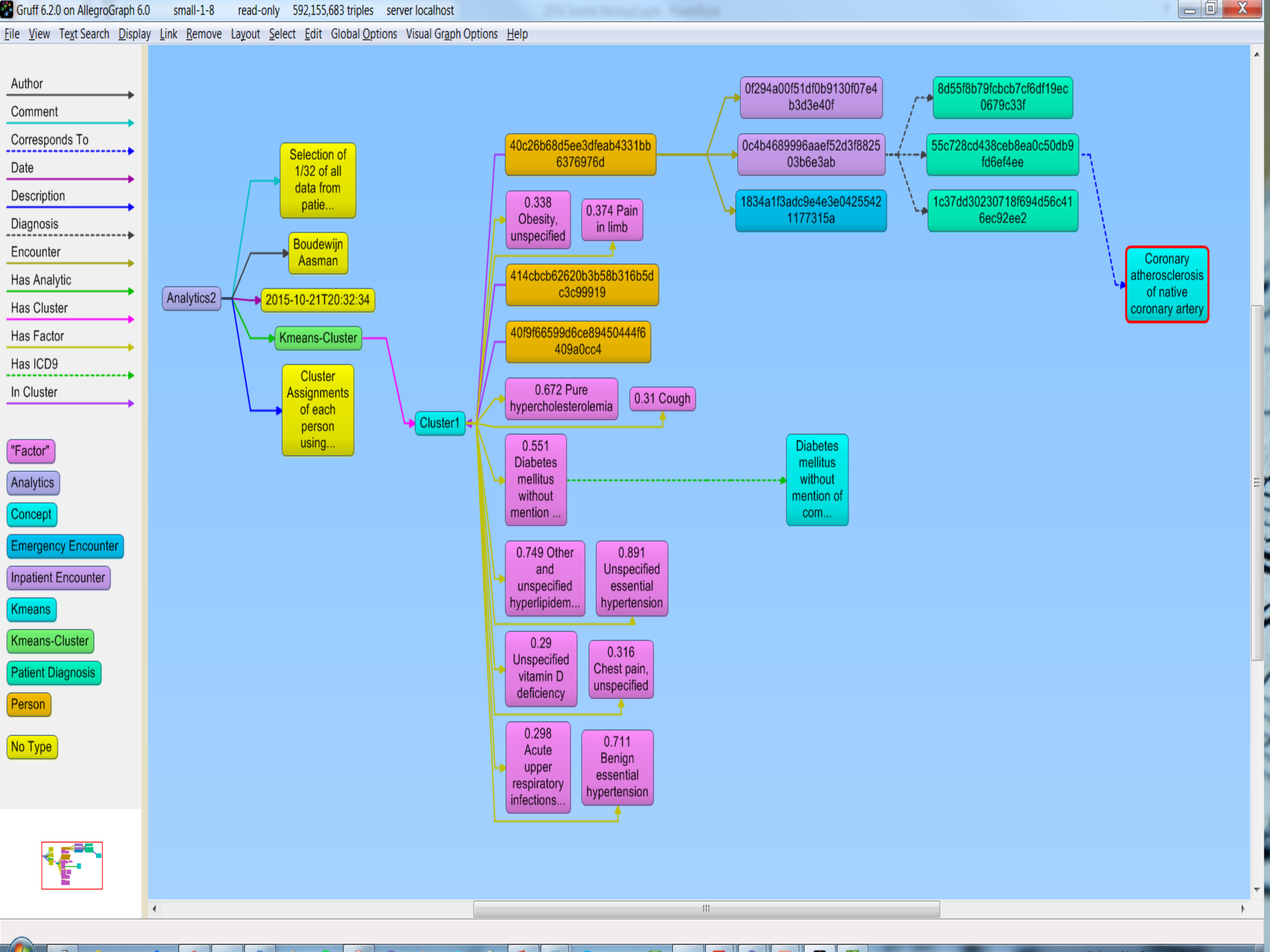


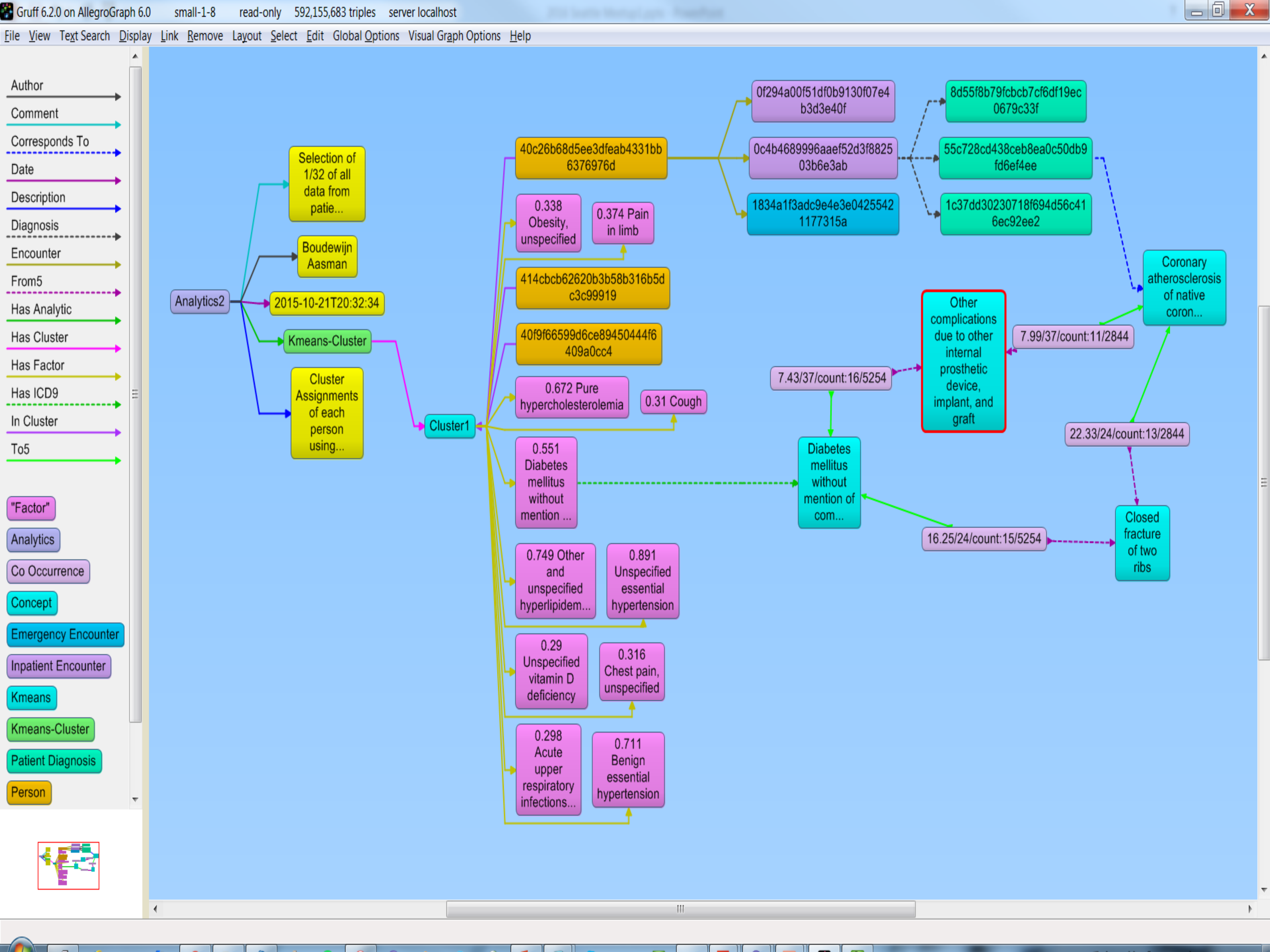
1. *k* initial "means" (in this case *k*=3) are randomly generated within the data domain (shown in color).

2. *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

3. The centroid of each of the *k* clusters becomes the new mean.

4. Steps 2 and 3 are repeated until convergence has been reached.


Gruff 6.2.0 on AllegroGraph 6.0 small-1-8 read-only 592,155,683 triples server localhost
File View Text Search Display Link Remove Layout Select Edit Global Options Visual Graph Options Help
Author
Comment
Corresponds To
Date
Description
Diagnosis
Encounter
Has Analytic
Has Cluster
Has Factor
Has ICD9
In Cluster
"Factor"
Analytics
Concept
Emergency Encounter
Inpatient Encounter
Kmeans
Kmeans-Cluster
Patient Diagnosis
Person
No Type
Analytics2
Selection of 1/32 of all data from patie...
Boudewijn Aasman
2015-10-21T20:32:34
Kmeans-Cluster
Cluster Assignments of each person using...
Cluster1
40c26b68d5ee3dfeab4331bb 6376976d
0.338 Obesity, unspecified
0.374 Pain in limb
414cbcb62620b3b58b316b5d c3c99919
40f9f66599d6ce89450444f6 409a0cc4
0.672 Pure hypercholesterolemia
0.31 Cough
0.551 Diabetes mellitus without mention ...
Diabetes mellitus without mention of com...
0.749 Other and unspecified hyperlipidem...
0.891 Unspecified essential hypertension
0.29 Unspecified vitamin D deficiency
0.316 Chest pain, unspecified
0.298 Acute upper respiratory infections...
0.711 Benign essential hypertension
0f294a00f51df0b9130f07e4 b3d3e40f
0c4b4689996aaef52d3f8825 03b6e3ab
1834a1f3adc9e4e3e0425542 1177315a
8d55f8b79fcbcb7cf6df19ec 0679c33f
55c728cd438ceb8ea0c50db9 fd6ef4ee
1c37dd30230718f694d56c41 6ec92ee2
Coronary atherosclerosis of native coronary artery


Gruff 6.2.0 on AllegroGraph 6.0 small-1-8 read-only 592,155,683 triples server localhost
File View Text Search Display Link Remove Layout Select Edit Global Options Visual Graph Options Help
Author
Comment
Corresponds To
Date
Description
Diagnosis
Encounter
From5
Has Analytic
Has Cluster
Has Factor
Has ICD9
In Cluster
To5
"Factor"
Analytics
Co Occurrence
Concept
Emergency Encounter
Inpatient Encounter
Kmeans
Kmeans-Cluster
Patient Diagnosis
Person
Analytics2
Selection of 1/32 of all data from patie...
Boudewijn Aasman
2015-10-21T20:32:34
Kmeans-Cluster
Cluster Assignments of each person using...
Cluster1
40c26b68d5ee3dfeab4331bb 6376976d
0.338 Obesity, unspecified
0.374 Pain in limb
414cbcb62620b3b58b316b5d c3c99919
40f9f66599d6ce89450444f6 409a0cc4
0.672 Pure hypercholesterolemia
0.31 Cough
0.551 Diabetes mellitus without mention ...
0.749 Other and unspecified hyperlipidem...
0.891 Unspecified essential hypertension
0.29 Unspecified vitamin D deficiency
0.316 Chest pain, unspecified
0.298 Acute upper respiratory infections...
0.711 Benign essential hypertension
0f294a00f51df0b9130f07e4 b3d3e40f
0c4b4689996aaef52d3f8825 03b6e3ab
1834a1f3adc9e4e3e0425542 1177315a
8d55f8b79fcbcb7cf6df19ec 0679c33f
55c728cd438ceb8ea0c50db9 fd6ef4ee
1c37dd30230718f694d56c41 6ec92ee2
Coronary atherosclerosis of native coron...
Other complications due to other internal prosthetic device, implant, and graft
7.99/37/count:11/2844
7.43/37/count:16/5254
22.33/24/count:13/2844
Diabetes mellitus without mention of com...
16.25/24/count:15/5254
Closed fracture of two ribs

# And then a query you could do never before

- Using the Knowledge Base, the Structured Data and the Probabilistic inferences all at the same time.

- To find the statistical links between Diabetes and Vision problems in our Semantic Data Lake
  - Find the set of ICD9s that are connected via one or more steps to concepts in the KB that mention Diabetes
  - Find the set of ICD9s that are connected via one or more steps to vision* or eye* or retinal*
  - An show how those two sets are related in the space of odds ratios

Gruff 6.2.0 on AllegroGraph 6.0 MonteTerminology read-only 343,641,849 triples server 192.168.198.132

File View Text Search Display Edit Global Options Query Options Table Options Help

SPARQL Use Planner Reindent Name Query Revisit Graph View Table View

Prolog Run Query Select All Graphical Query View

Query

```
select ?chd1 ?map1 ?icd1 ?map2 ?icd2 ?oddsratio ?chd2 where  {
   { select distinct ?map1 ?chd1 ?icd1 where {
       ?chd1 fti:match ('diabetes') .
       ?map1 mth:chd* ?chd1 .
       ?icd1 skos:exactMatch ?map1 . } }
   { select distinct ?map2 ?chd2 ?icd2 where {
       ?chd2 fti:match ('vision* | eye* | retin*') .
       ?map2 mth:chd* ?chd2 .
       ?icd2 skos:exactMatch ?map2 . }}
   ?oddsratio franz:to5 ?icd1 ; franz:from5 ?icd2 .
   }
limit 32
```

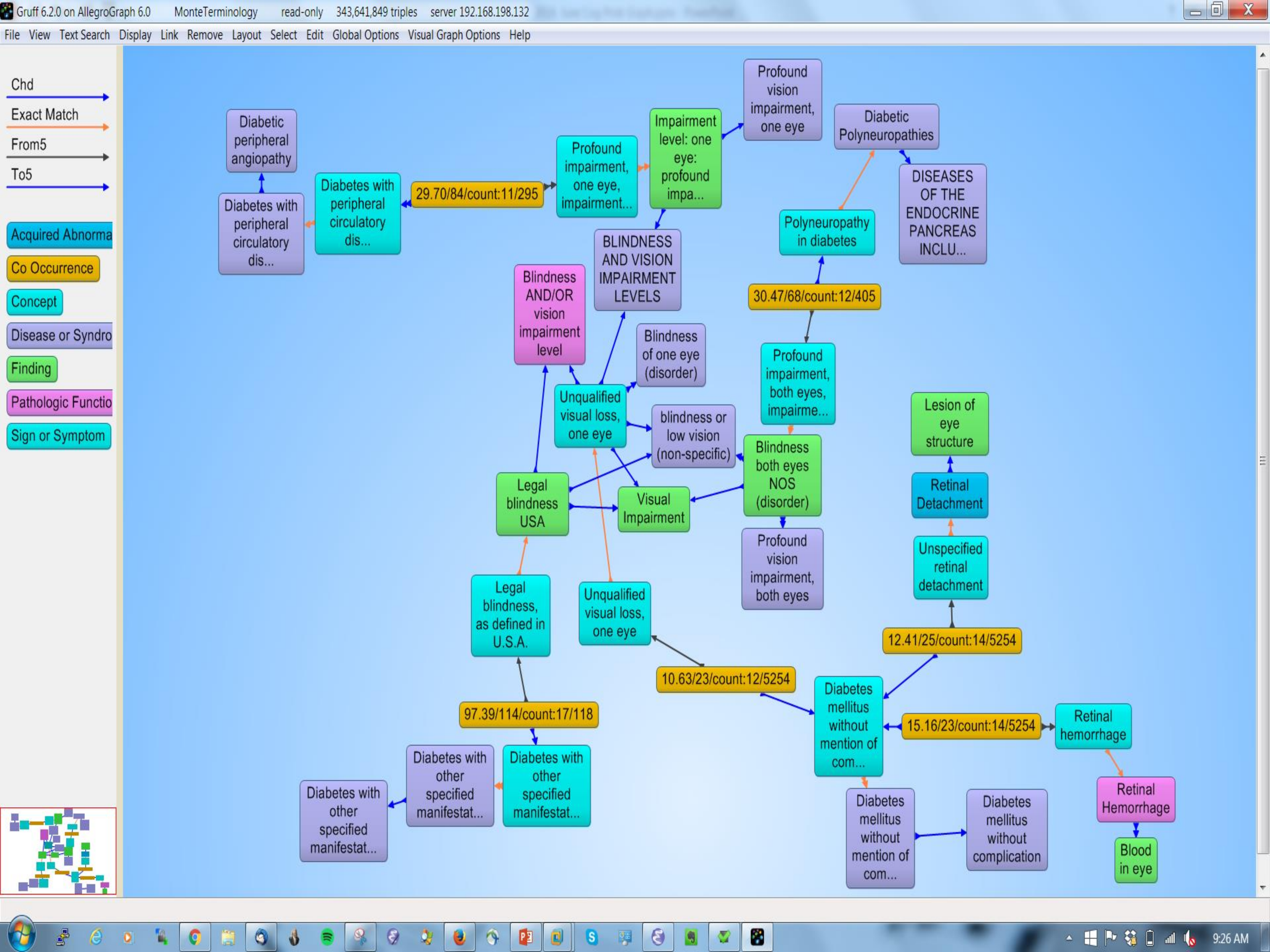
32 Results

Create Visual Graph Add to Visual Graph Write Text Report Save as CSV

| ?chd1 | ?map1 | ?icd1 | ?map2 | ?icd2 | ?oddsratio | ?chd2 |
|---|---|---|---|---|---|---|
| DISEASES OF THE ENDO | Diabetic Polyneuropathies | Polyneuropathy in diabetes | Blindness both eyes NOS (d | Profound impairment, both e | 30.47/68/count:12/405 | Visual Impairment |
| DISEASES OF THE ENDO | Diabetic Polyneuropathies | Polyneuropathy in diabetes | Blindness both eyes NOS (d | Profound impairment, both e | 30.47/68/count:12/405 | blindness or low vision (non- |
| DISEASES OF THE ENDO | Diabetic Polyneuropathies | Polyneuropathy in diabetes | Blindness both eyes NOS (d | Profound impairment, both e | 30.47/68/count:12/405 | Profound vision impairment, |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Unqualified visual loss, one | Unqualified visual loss, one | 10.63/23/count:12/5254 | Visual Impairment |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Unqualified visual loss, one | Unqualified visual loss, one | 10.63/23/count:12/5254 | Blindness AND/OR vision im |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Unqualified visual loss, one | Unqualified visual loss, one | 10.63/23/count:12/5254 | blindness or low vision (non- |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Unqualified visual loss, one | Unqualified visual loss, one | 10.63/23/count:12/5254 | BLINDNESS AND VISION I |
| Diabetic peripheral angiopat | Diabetes with peripheral circ | Diabetes with peripheral circ | Impairment level: one eye: p | Profound impairment, one e | 29.70/84/count:11/295 | Profound vision impairment, |
| Diabetic peripheral angiopat | Diabetes with peripheral circ | Diabetes with peripheral circ | Impairment level: one eye: p | Profound impairment, one e | 29.70/84/count:11/295 | BLINDNESS AND VISION I |
| Diabetes with other specifie | Diabetes with other specifie | Diabetes with other specifie | Legal blindness USA | Legal blindness, as defined i | 97.39/114/count:17/118 | Visual Impairment |
| Diabetes with other specifie | Diabetes with other specifie | Diabetes with other specifie | Legal blindness USA | Legal blindness, as defined i | 97.39/114/count:17/118 | Blindness AND/OR vision im |
| Diabetes with other specifie | Diabetes with other specifie | Diabetes with other specifie | Legal blindness USA | Legal blindness, as defined i | 97.39/114/count:17/118 | blindness or low vision (non- |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Retinal Hemorrhage | Retinal hemorrhage | 15.16/23/count:14/5254 | Eye Hemorrhage |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Retinal Hemorrhage | Retinal hemorrhage | 15.16/23/count:14/5254 | Blood in eye |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Unqualified visual loss, one | Unqualified visual loss, one | 10.63/23/count:12/5254 | Disorder of eye |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Unqualified visual loss, one | Unqualified visual loss, one | 10.63/23/count:12/5254 | Unqualified visual loss, one |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Unqualified visual loss, one | Unqualified visual loss, one | 10.63/23/count:12/5254 | Blindness of one eye (disord |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Retinal Detachment | Unspecified retinal detachm | 12.41/25/count:14/5254 | Lesion of eye structure |
| Diabetes mellitus without co | Diabetes mellitus without me | Diabetes mellitus without m | Retinal Detachment | Unspecified retinal detachm | 12.41/25/count:14/5254 | Eye injuries NEC |

Type or paste a SPARQL query here, then press Run Query.

9:20 AM



Gruff 6.2.0 on AllegroGraph 6.0
MonteTerminology
read-only
343,641,849 triples
server 192.168.198.132
File View Text Search Display Link Remove Layout Select Edit Global Options Visual Graph Options Help
Chd
Exact Match
From5
To5
Acquired Abnorma
Co Occurrence
Concept
Disease or Syndro
Finding
Pathologic Functio
Sign or Symptom
Diabetic peripheral angiopathy
Diabetes with peripheral circulatory dis...
Diabetes with peripheral circulatory dis...
29.70/84/count:11/295
Profound impairment, one eye, impairment...
Impairment level: one eye: profound impa...
Profound vision impairment, one eye
Diabetic Polyneuropathies
DISEASES OF THE ENDOCRINE PANCREAS INCLU...
Polyneuropathy in diabetes
BLINDNESS AND VISION IMPAIRMENT LEVELS
Blindness AND/OR vision impairment level
30.47/68/count:12/405
Blindness of one eye (disorder)
Profound impairment, both eyes, impairme...
Unqualified visual loss, one eye
blindness or low vision (non-specific)
Lesion of eye structure
Blindness both eyes NOS (disorder)
Legal blindness USA
Visual Impairment
Retinal Detachment
Profound vision impairment, both eyes
Unspecified retinal detachment
Legal blindness, as defined in U.S.A.
Unqualified visual loss, one eye
12.41/25/count:14/5254
10.63/23/count:12/5254
Diabetes mellitus without mention of com...
97.39/114/count:17/118
15.16/23/count:14/5254
Retinal hemorrhage
Diabetes with other specified manifestat...
Diabetes with other specified manifestat...
Diabetes with other specified manifestat...
Retinal Hemorrhage
Diabetes mellitus without mention of com...
Diabetes mellitus without complication
Blood in eye
9:26 AM