

# Exploring Ancient Texts with a User Driven Concept Search

Muhammad Faisal Cheema, Stefan Jänicke, Christoph Weilbach,  
Judith Blumenstein, Gerik Scheuermann  
Leipzig University, Germany



Bundesministerium  
für Bildung  
und Forschung

eXChange: Exploring Concept Change and Transfer in Antiquity

exchange

# Outline

- ▼ Motivation
- ▼ Concept Search Pipeline
  - ▼ Concept Editor
  - ▼ Concept Search
  - ▼ Concept Search Results Viewer
- ▼ Demo
- ▼ Results & Findings
- ▼ Conclusions

# Motivation

*“Even with a perfect search engine, a poorly constructed search question may not lead to right answer”*  
(Teevan et al. 2004)

- ▼ Keyword search
- ▼ Most popular method of everyday search
- ▼ Reformulate query to get better results
- ▼ Even in research paradigms

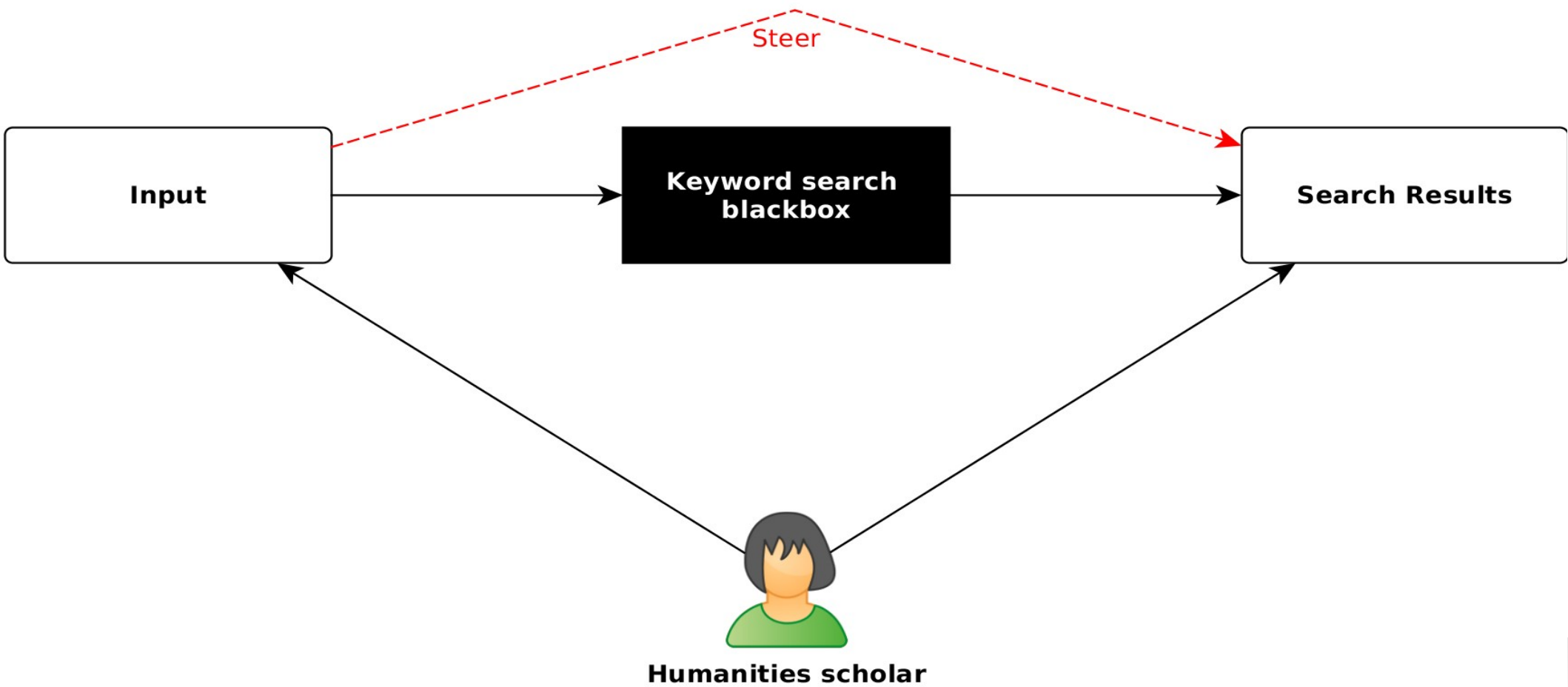
# Motivation

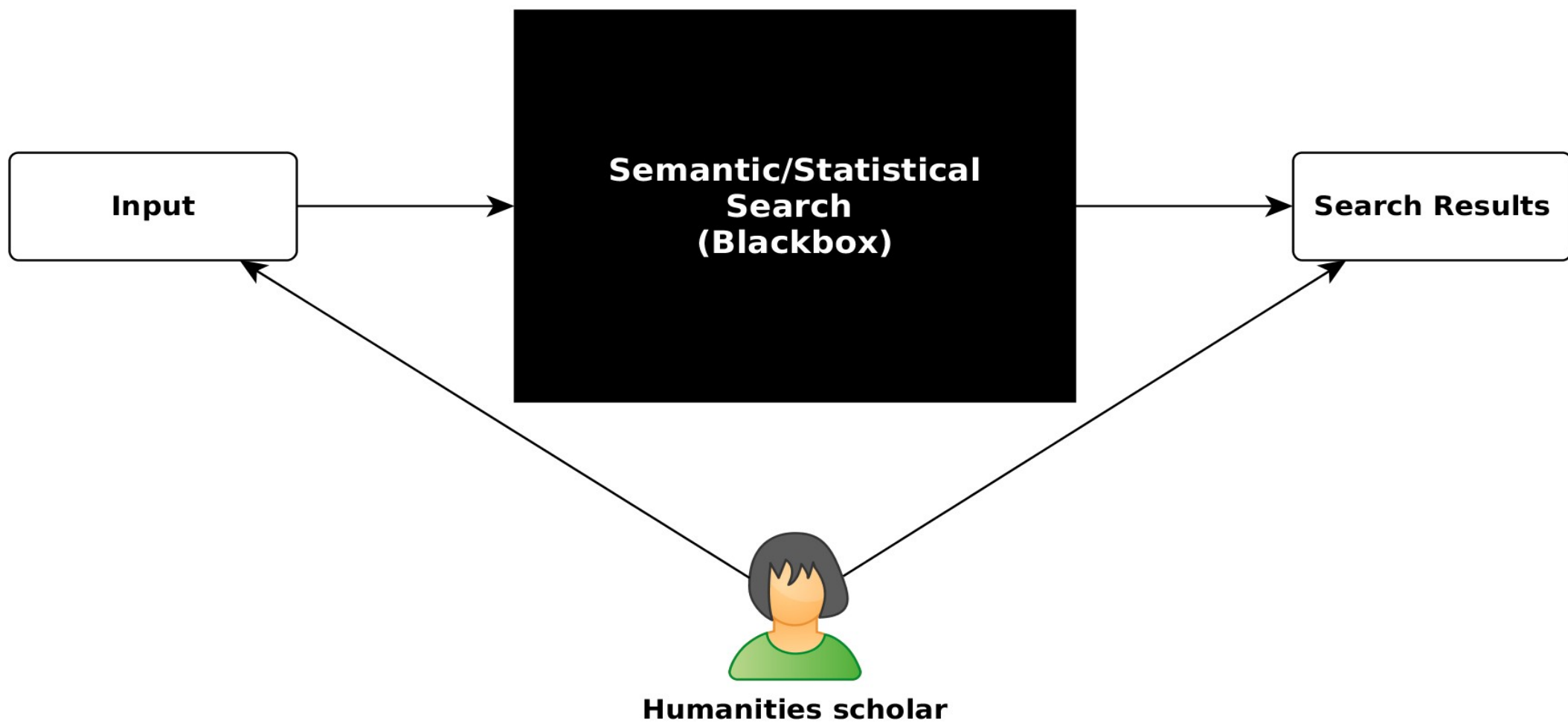
- ▼ Data in consideration: **Ancient texts**
- ▼ Keyword search yields incomplete results “e.g. concept **Epilepsy**”
  - ▼ Full word search e.g. '*morbus comitiales*'
  - ▼ Truncated search e.g. *morb\* comiti\**
  - ▼ Alternate terms e.g. '*morbus sacer*' or '*epilepsia*'
  - ▼ Descriptive terms e.g. *spuma or concidere*
- ▼ Ancient texts are quite different from modern texts in terms of:
  - ▼ ancient vocabulary
  - ▼ ideas, cultures, practices that we are no more aware of

# Existing techniques

- ▼ Semantic search approaches
  - ▼ Word sense disambiguation
  - ▼ Ontology based
- ▼ Statistical approaches
  - ▼ Topic Models
  - ▼ Latent Dirichlet allocation

## Blackbox problem





# Goal

- ▼ Ability to model a concept query that covers all dimensions of keyword search i.e. fullword, truncated, alternate and descriptive terms
- ▼ Ability to steer the search process as desired by the user
- ▼ Ability to iteratively improve results, as in keyword search

## SOLUTION

- ▼ A visual search environment that provides the user ability to:
  - ▼ Model their ideas of concepts
  - ▼ Search the concept models and iteratively improve results

Steer  
&  
Use

**Concept Model**

**Concept Search  
(Blackbox)**

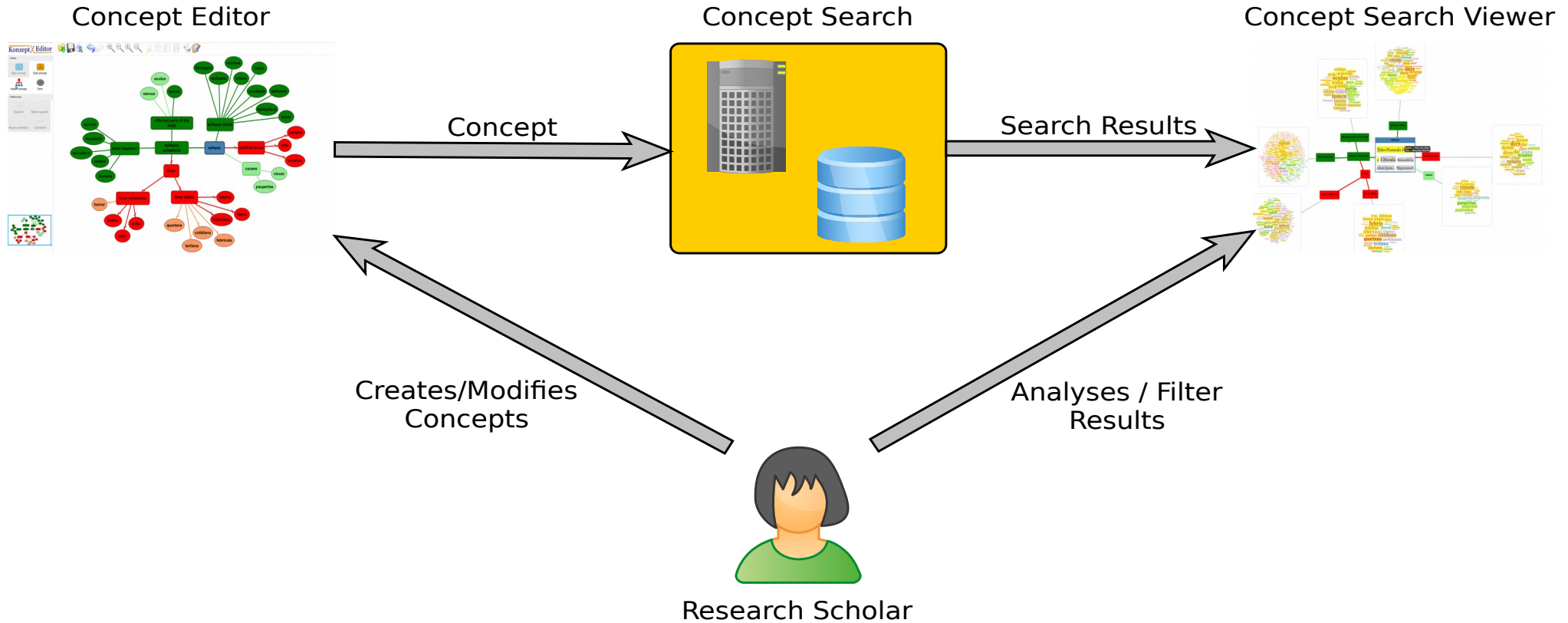
**Concept Search  
Results**



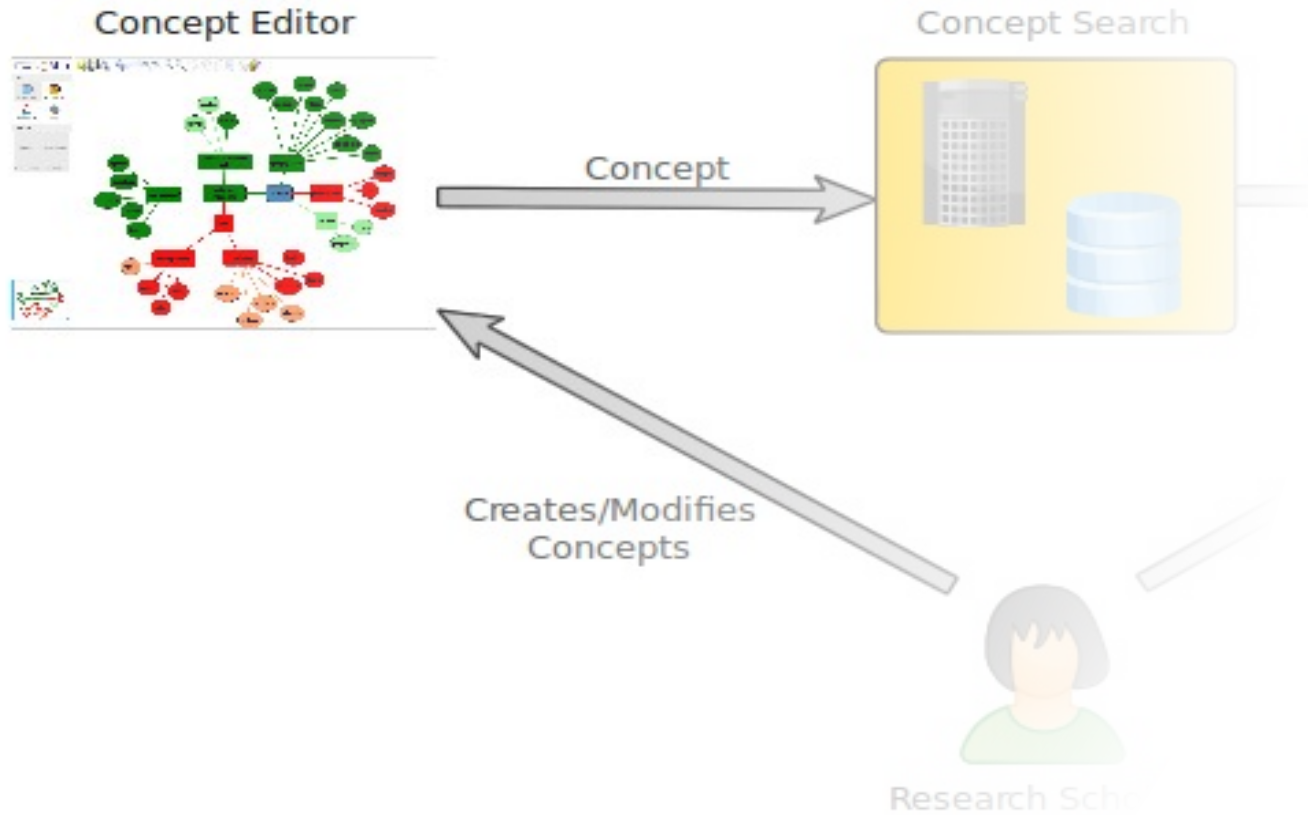
**Humanities scholar**

# Concept Search Pipeline

# Concept Search Pipeline

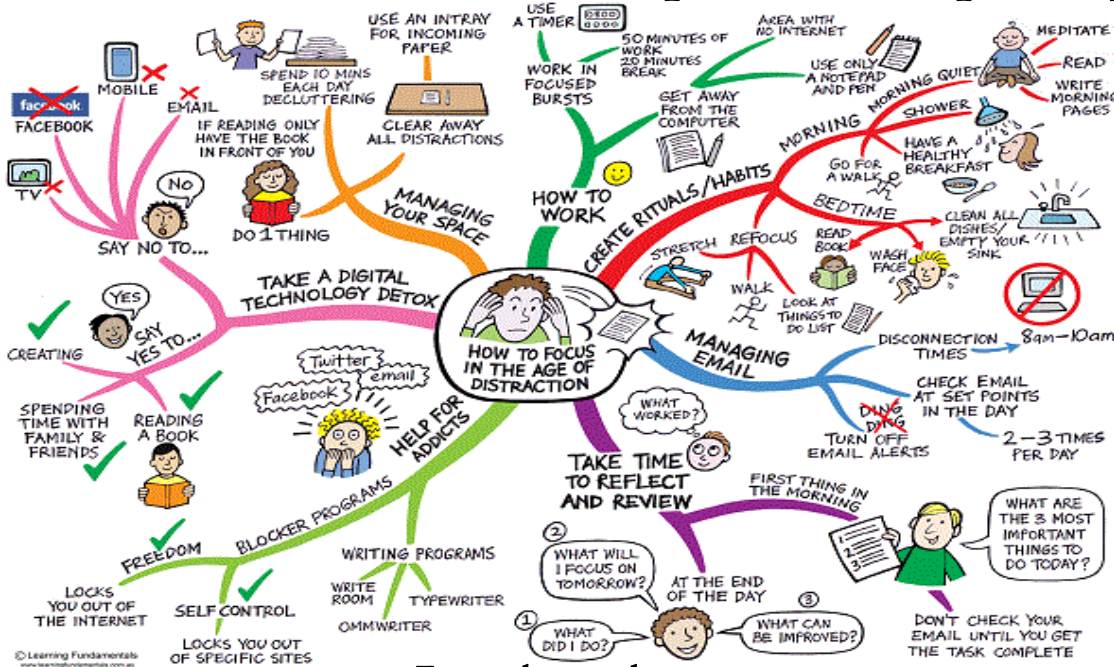


# Concept Search Pipeline

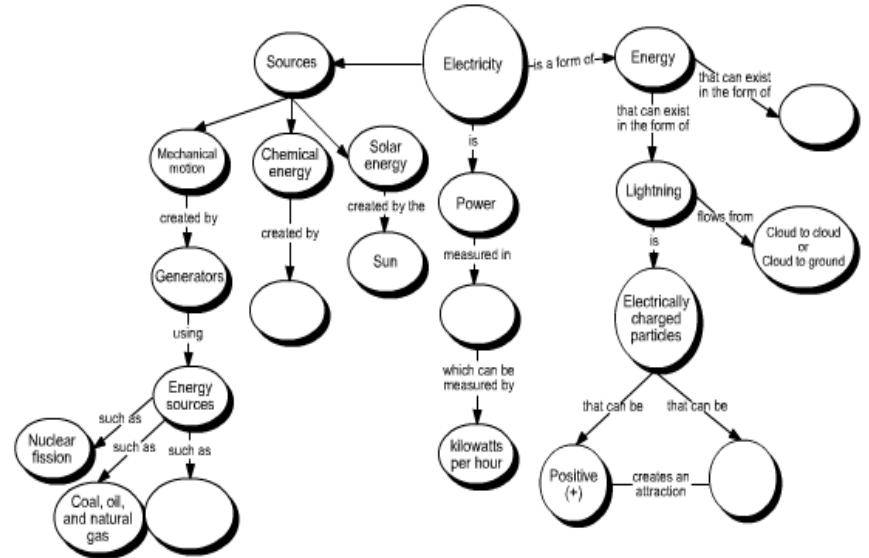


# What is our Concept model

- ▼ We define our Concept Model inspired by ideas from *mind maps* & *concept maps*



Example mind map



Example concept map

Electricity Concept Map : ([https://commons.wikimedia.org/wiki/File:Electricity\\_Concept\\_Map.gif](https://commons.wikimedia.org/wiki/File:Electricity_Concept_Map.gif))

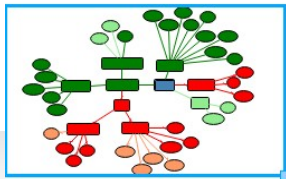
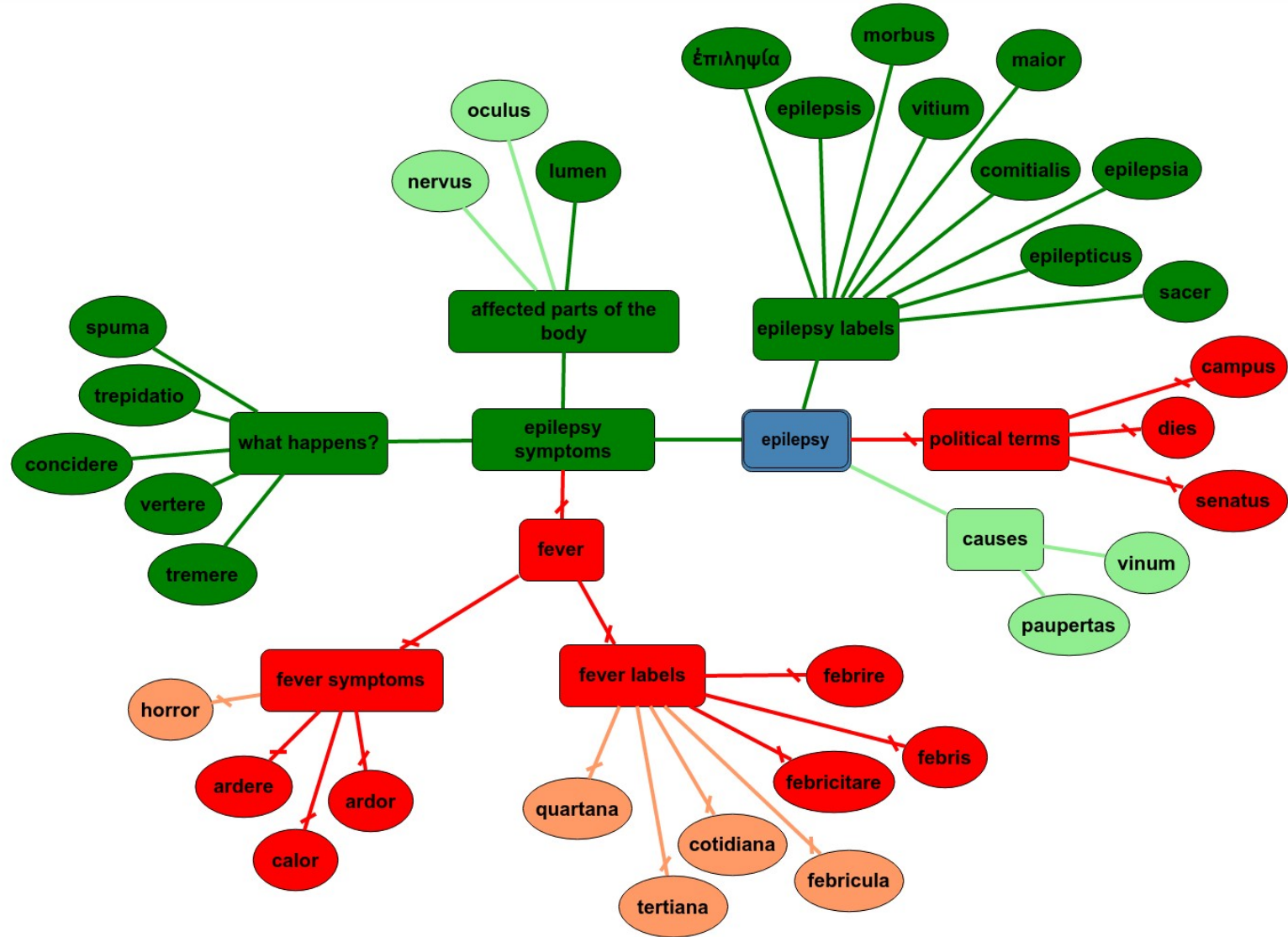


**Entities**

Main concept (blue rounded rectangle)  
Sub concept (orange rounded rectangle)  
Term (grey circle)  
Import concept (red tree icon)

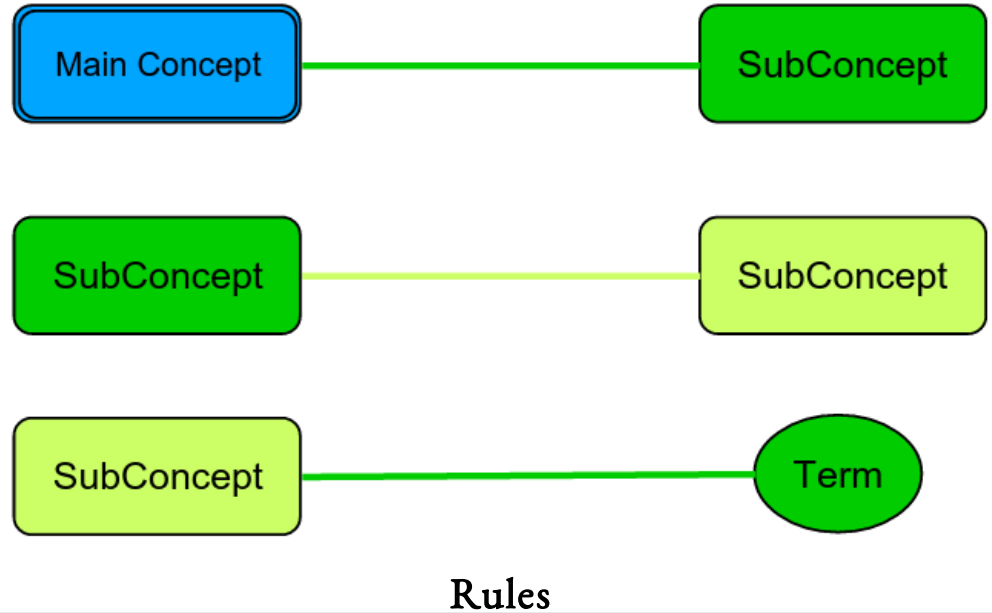
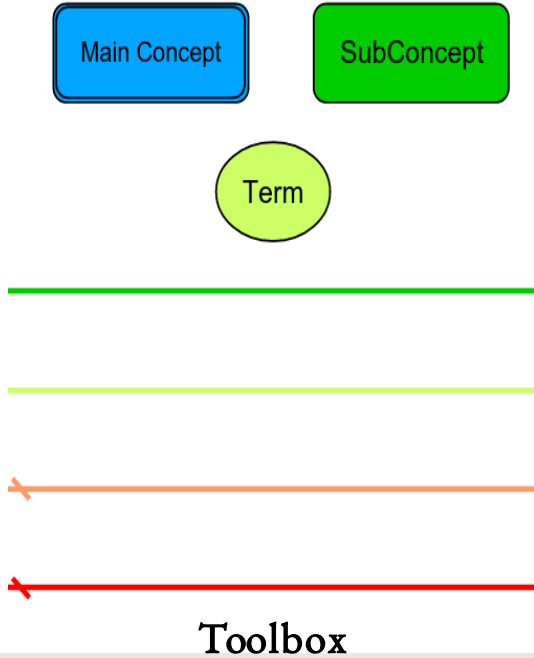
**Relationships**

Supports (green line)  
Maybe supports (yellow line)  
Maybe contradicts (orange line)  
Contradicts (red line)



# The concept model

- ▼ Concept is modeled as a **tree** data structure
  - ▼ Simpler to model for humanities scholars
  - ▼ Better (re)layout options



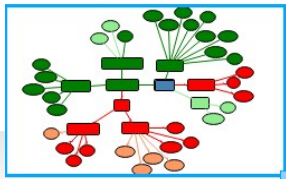
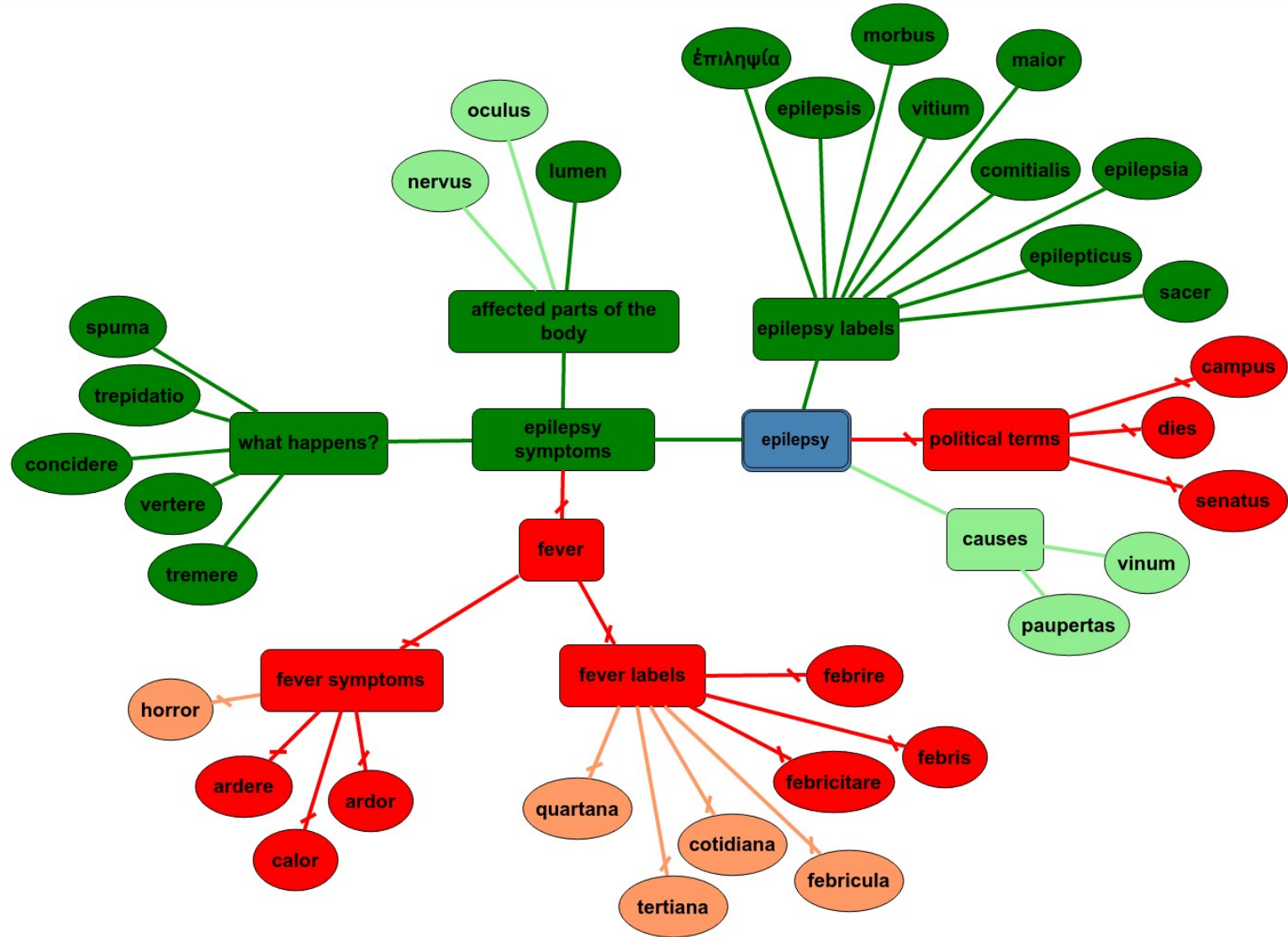


**Entities**

Main concept (blue rounded rectangle)  
Sub concept (orange rounded rectangle)  
Term (grey circle)  
Import concept (red tree icon)

**Relationships**

Supports (green line)  
Maybe supports (yellow line)  
Maybe contradicts (orange line)  
Contradicts (red line)



# Select Word Forms for :: quartana

Show 100 entries

Search:

<input type="checkbox"/>	Word Forms	Occurrences
<input checked="" type="checkbox"/>	Quartana	3
<input checked="" type="checkbox"/>	Quartanae	2
<input checked="" type="checkbox"/>	Quartanarum	1
<input checked="" type="checkbox"/>	quartana	87
<input checked="" type="checkbox"/>	quartanae	44
<input checked="" type="checkbox"/>	quartanam	24
<input checked="" type="checkbox"/>	quartanarum	6
<input checked="" type="checkbox"/>	quartanas	30
<input checked="" type="checkbox"/>	quartanasque	2
<input checked="" type="checkbox"/>	quartanis	79
<input checked="" type="checkbox"/>	quartanisque	3
<input type="checkbox"/>	Quartanus	1
<input type="checkbox"/>	quartanique	2
<input type="checkbox"/>	quartanorum	3
<input type="checkbox"/>	quartanus	3

Showing 1 to 15 of 15 entries

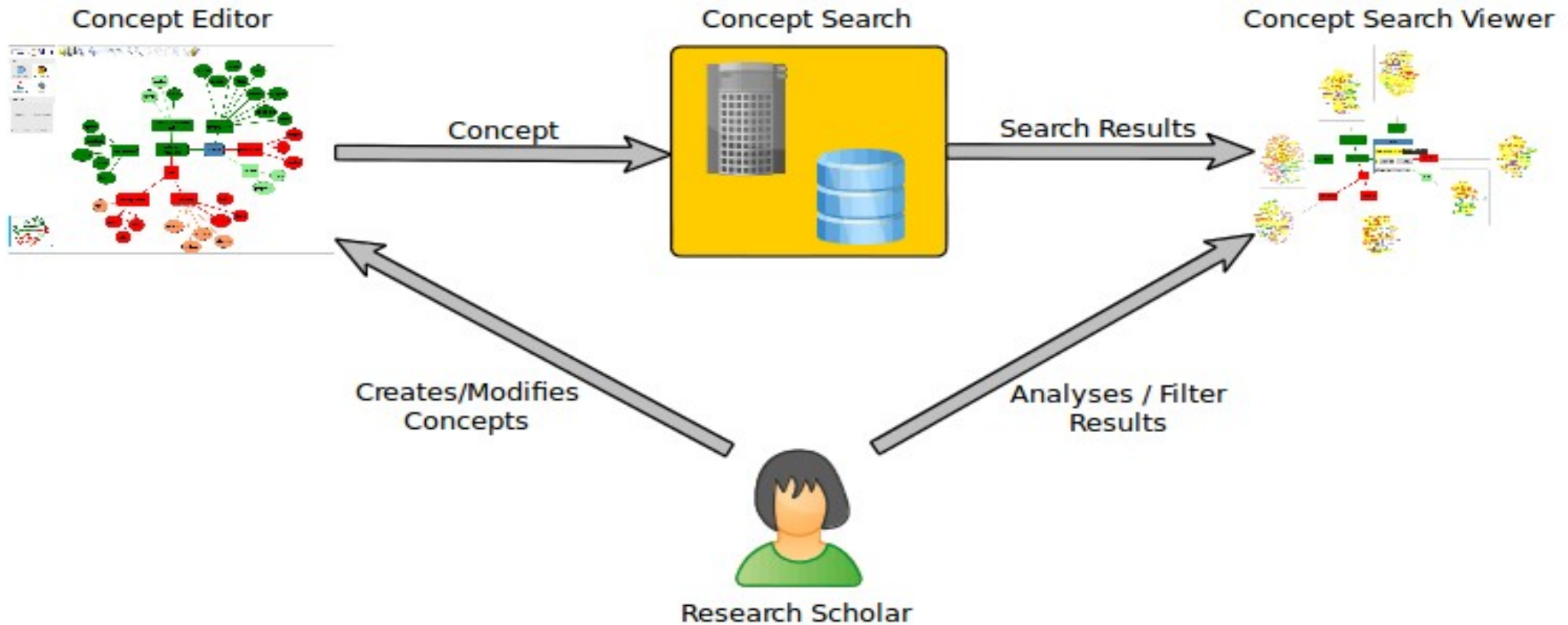
Previous Next

quartana

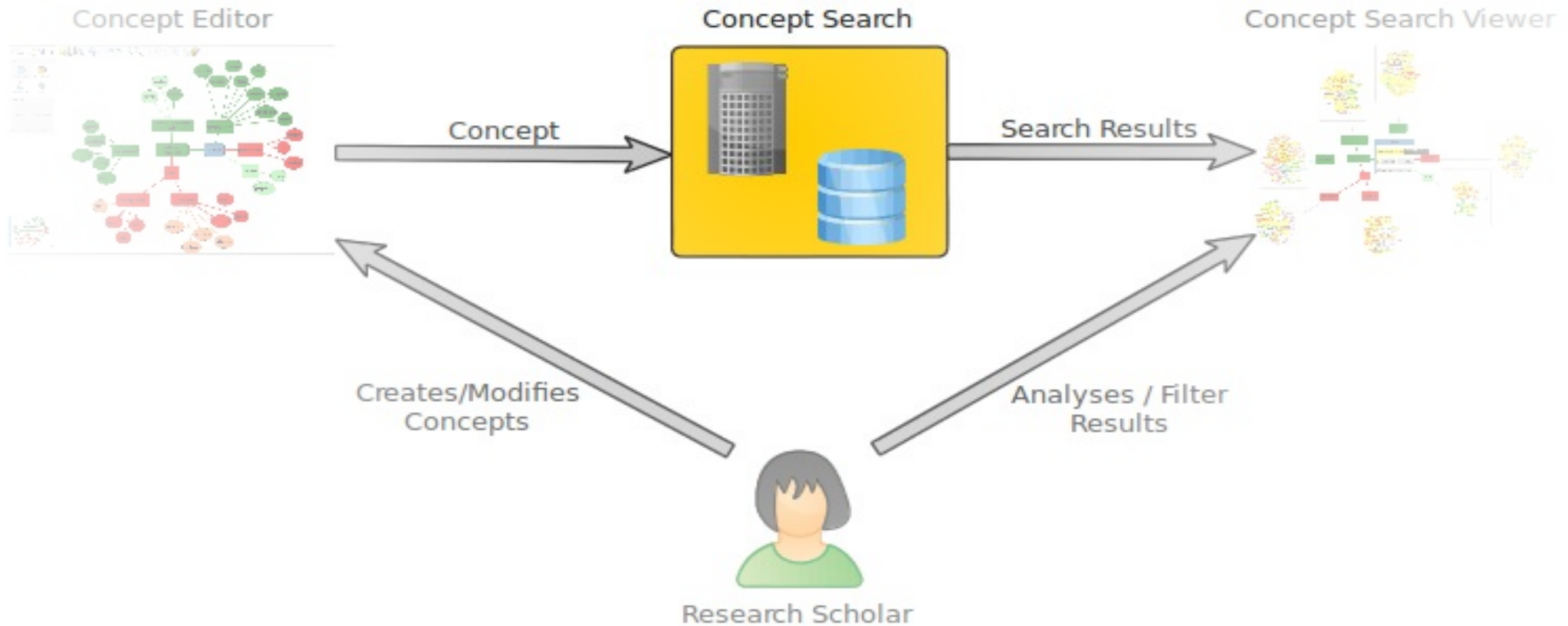
Quartana Quartanae Quartanarum **quartana**  
**quartanae** quartanam quartanarum  
quartanas quartanasque **quartanis**  
quartanisque Quartanus quartanique quartanorum quartanus

For further on concept editor details, refer to [CJBWS<sub>I5</sub>]

# Concept Search Pipeline



# Concept Search Pipeline

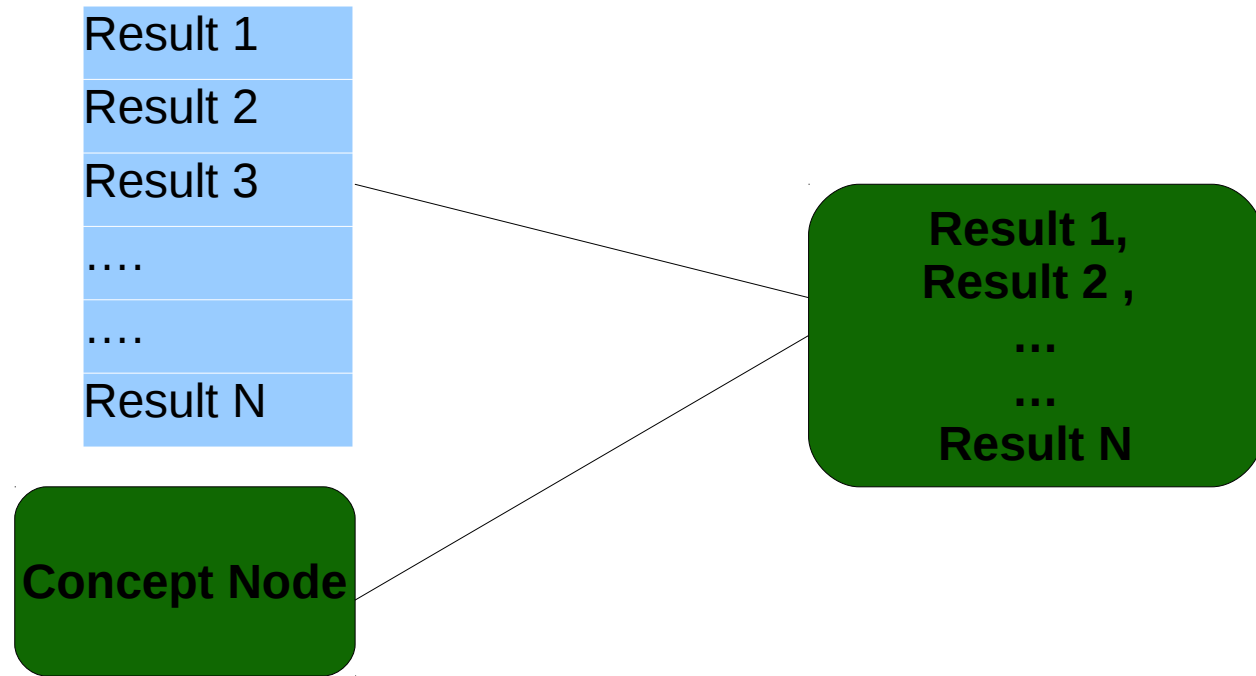


# Hierarchical Keyword Search

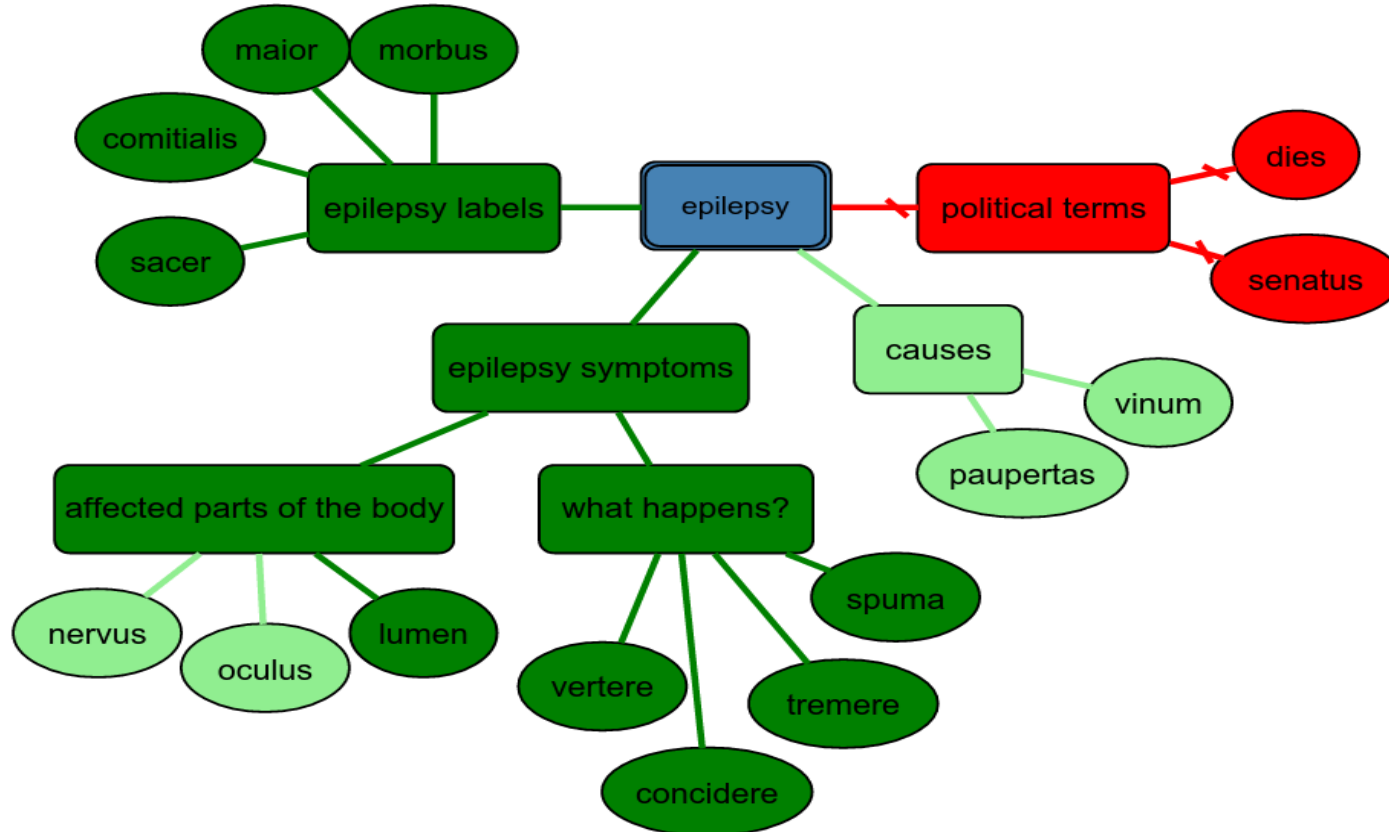
- ▼ Word forms of all terms are searched
- ▼ Each entry is searched for with classical string lookup
- ▼ Per word, all occurrences are located, grouped and counted within groups
- ▼ **Result:** Non-negative integer matrix (usually sparse) representation. Can be used to rank results according to concept hierarchy.

# Assignment of Results to Concept Nodes

- ▼ Goal: Show the search results in user created concept hierarchy

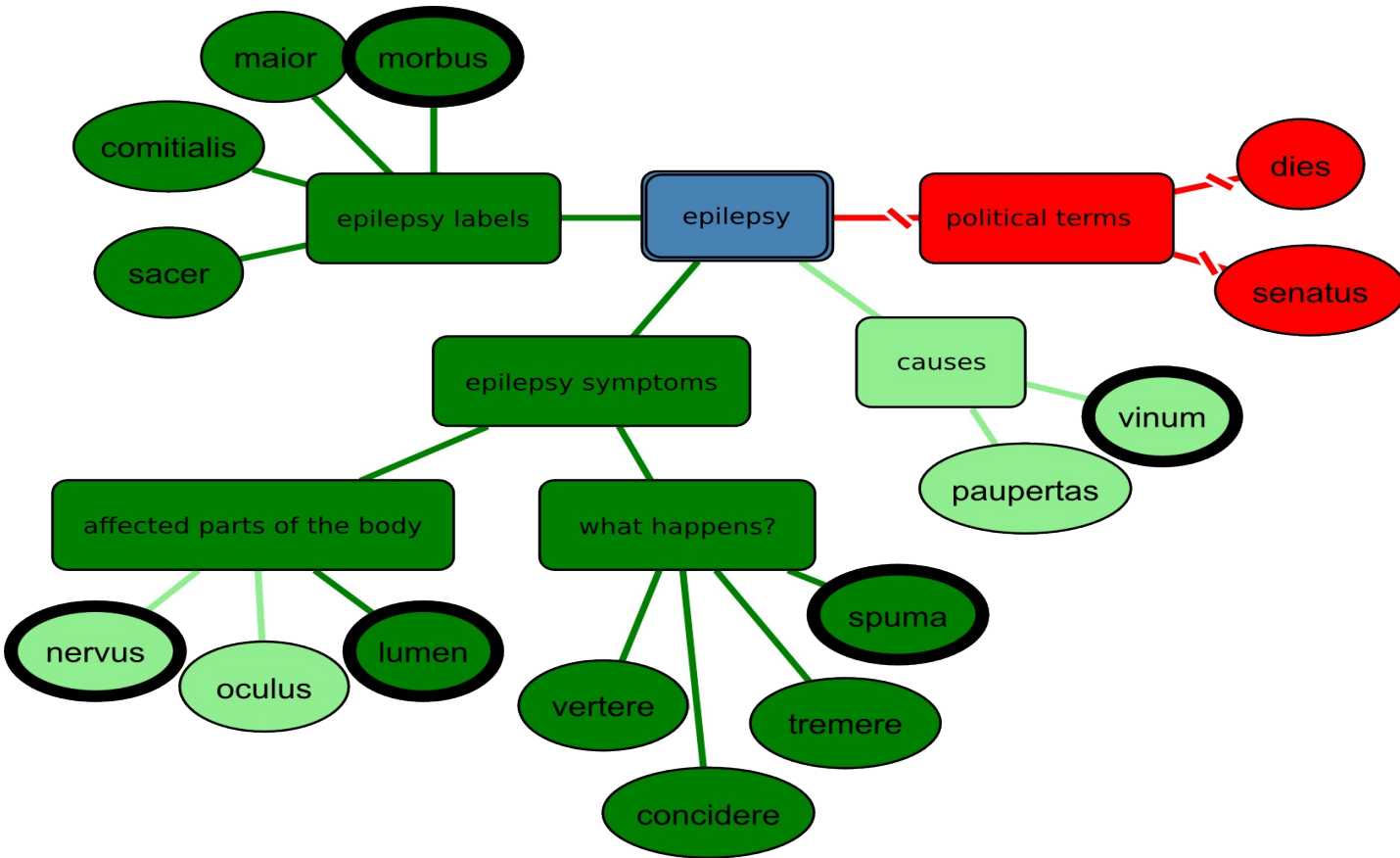


# Assignment of Results to Concept Nodes



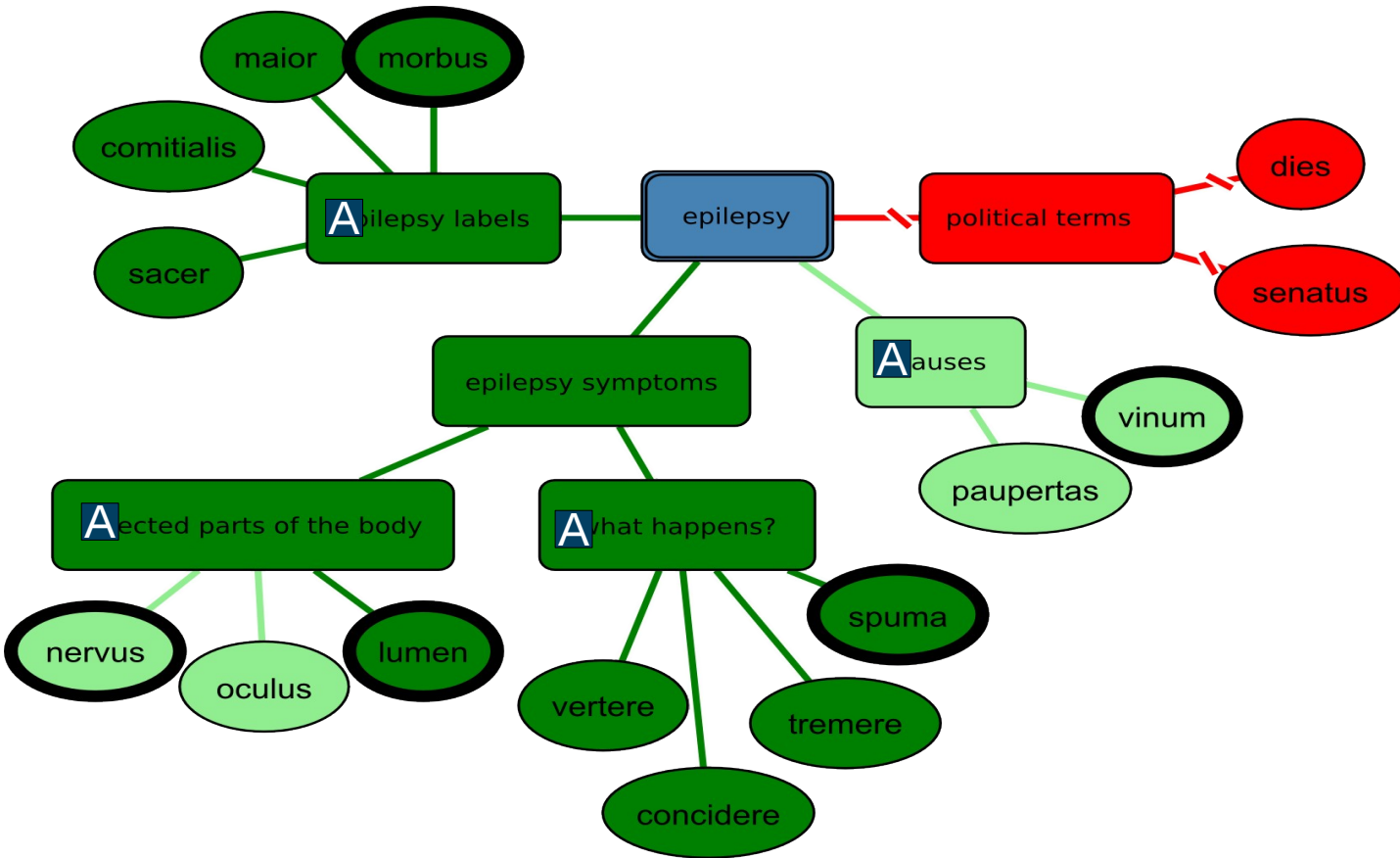
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbus, vinum
<b>B</b>	morbus, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbus, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



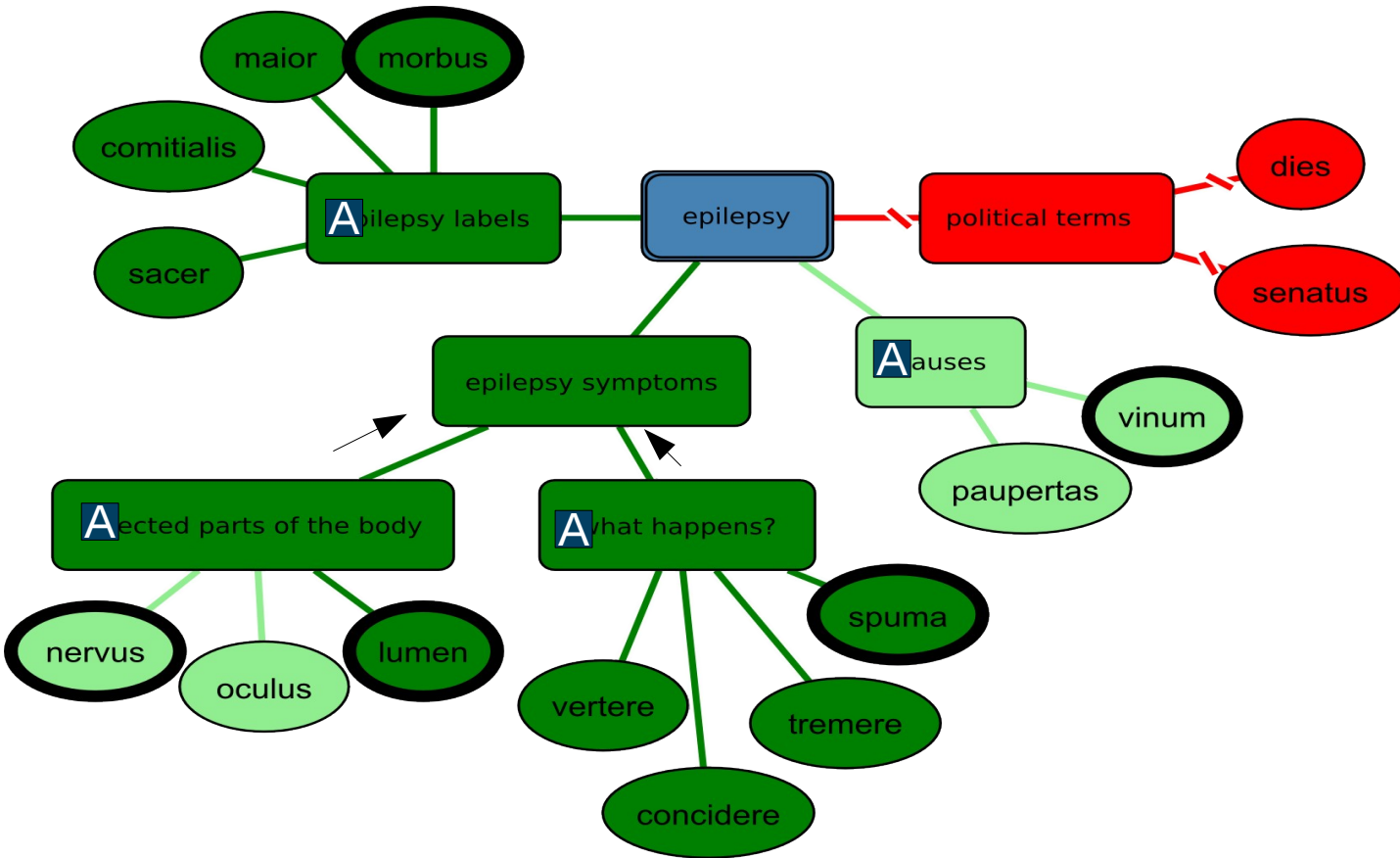
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbis, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



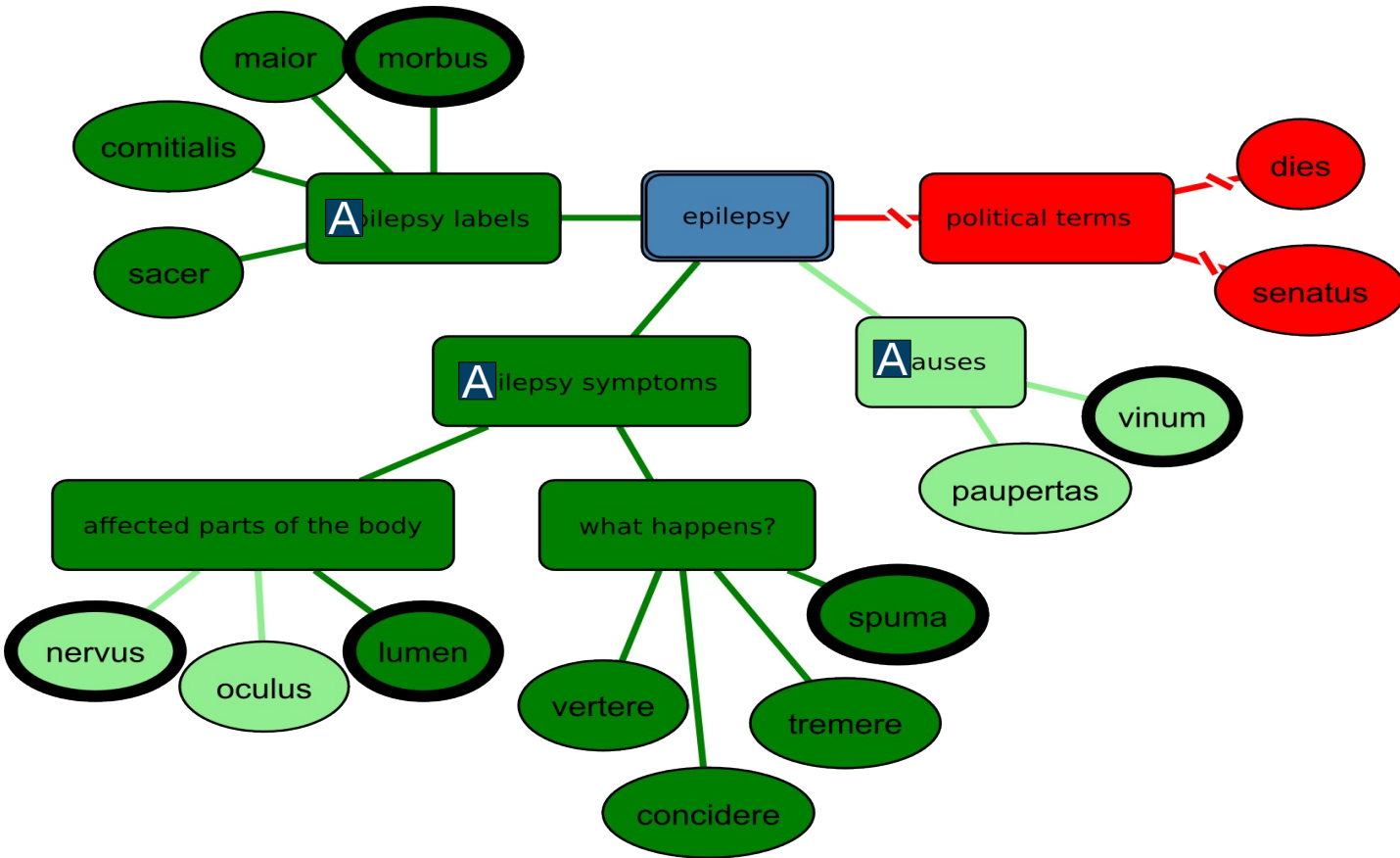
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbus, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



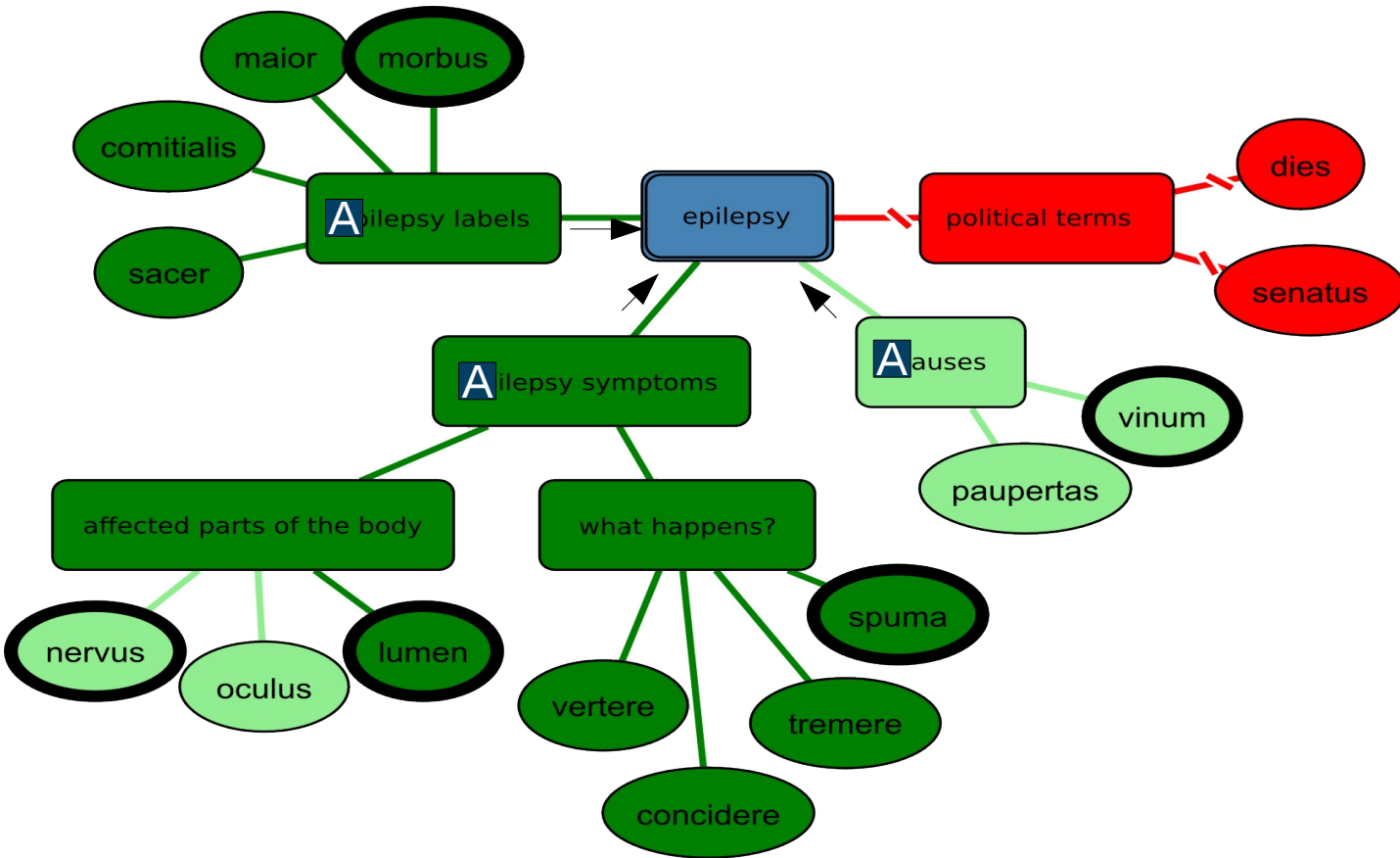
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbis, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



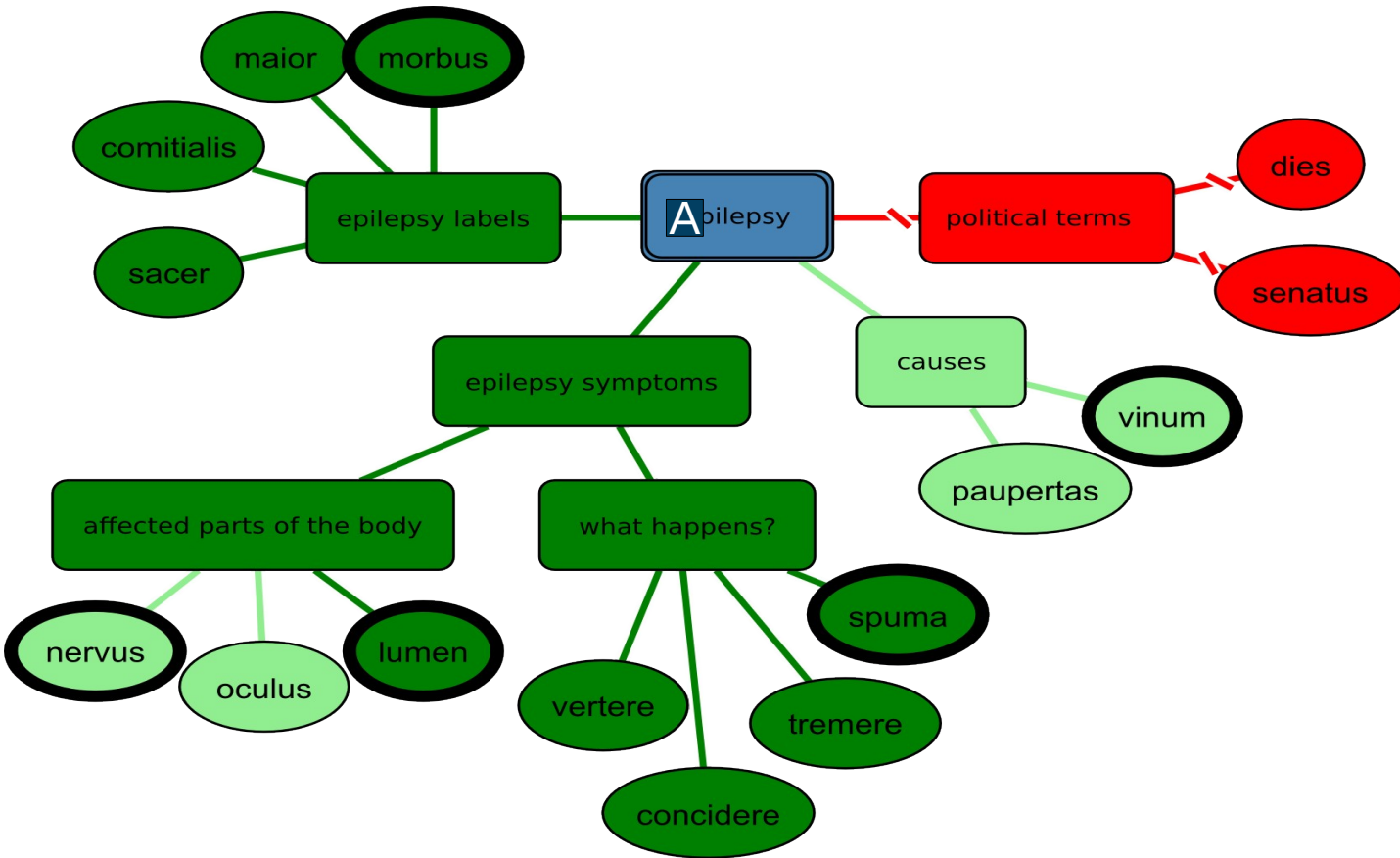
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbis, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



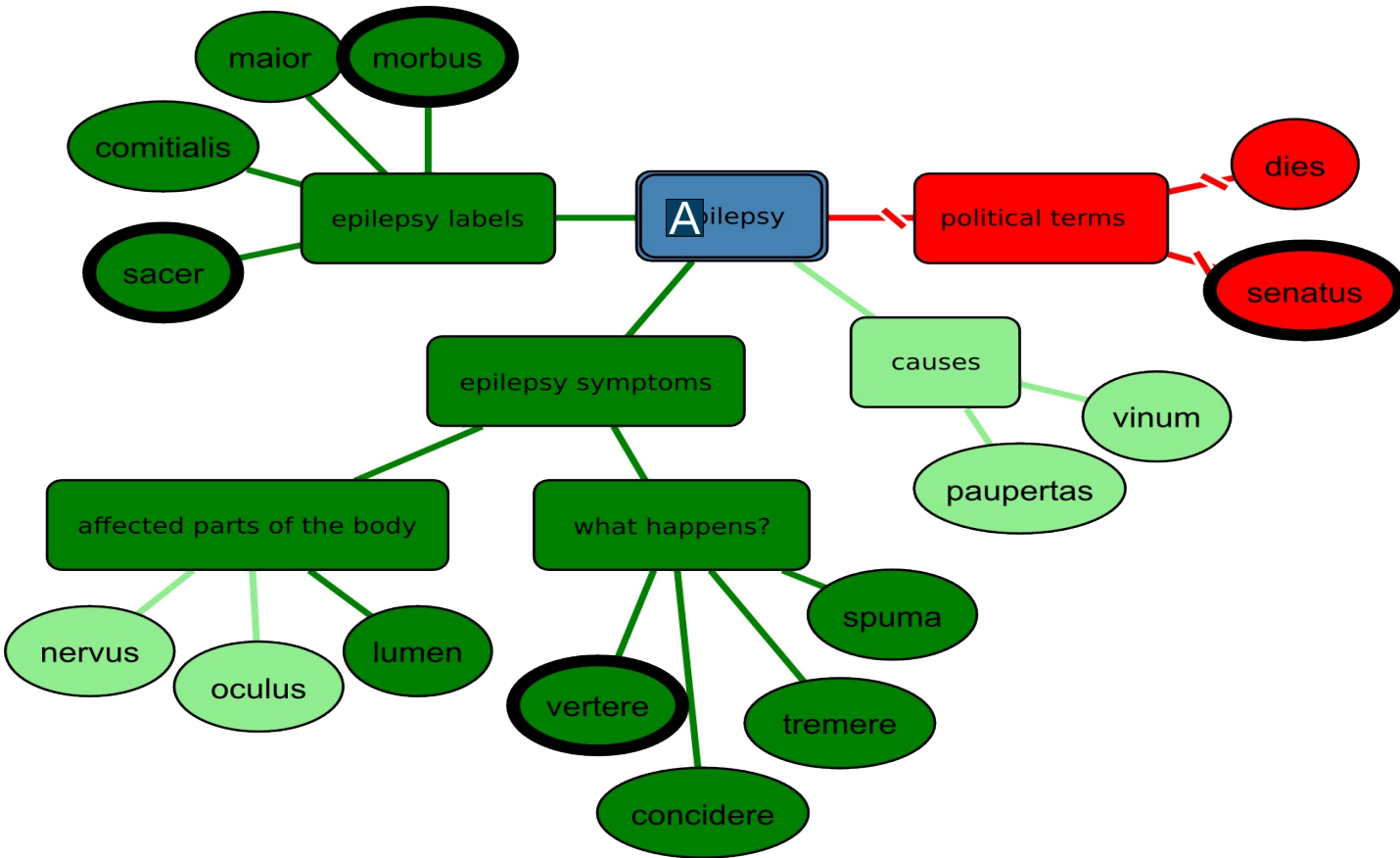
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbis, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



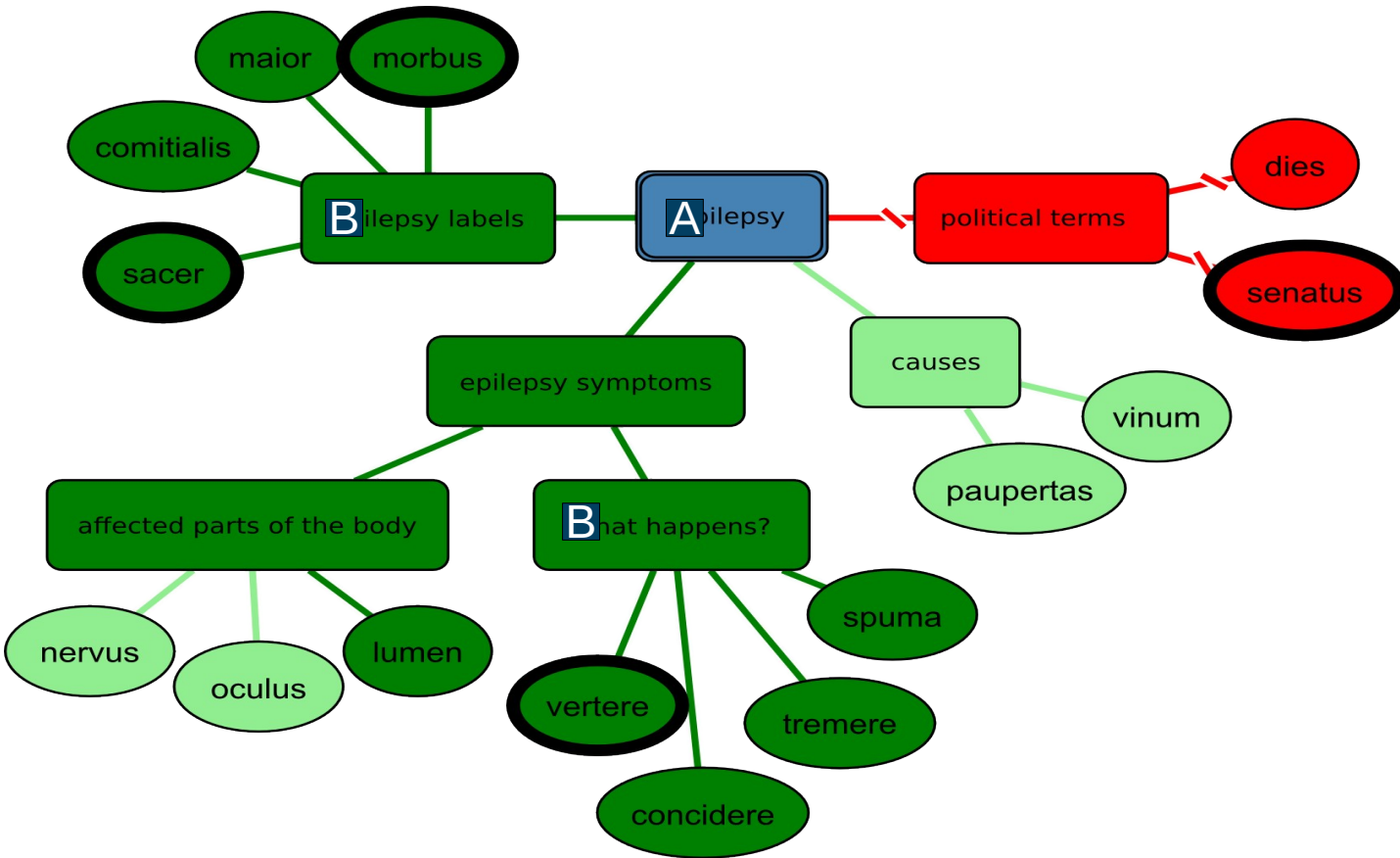
Result	Word forms
<b>A</b>	spuma, nervus, lumen, morbus, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



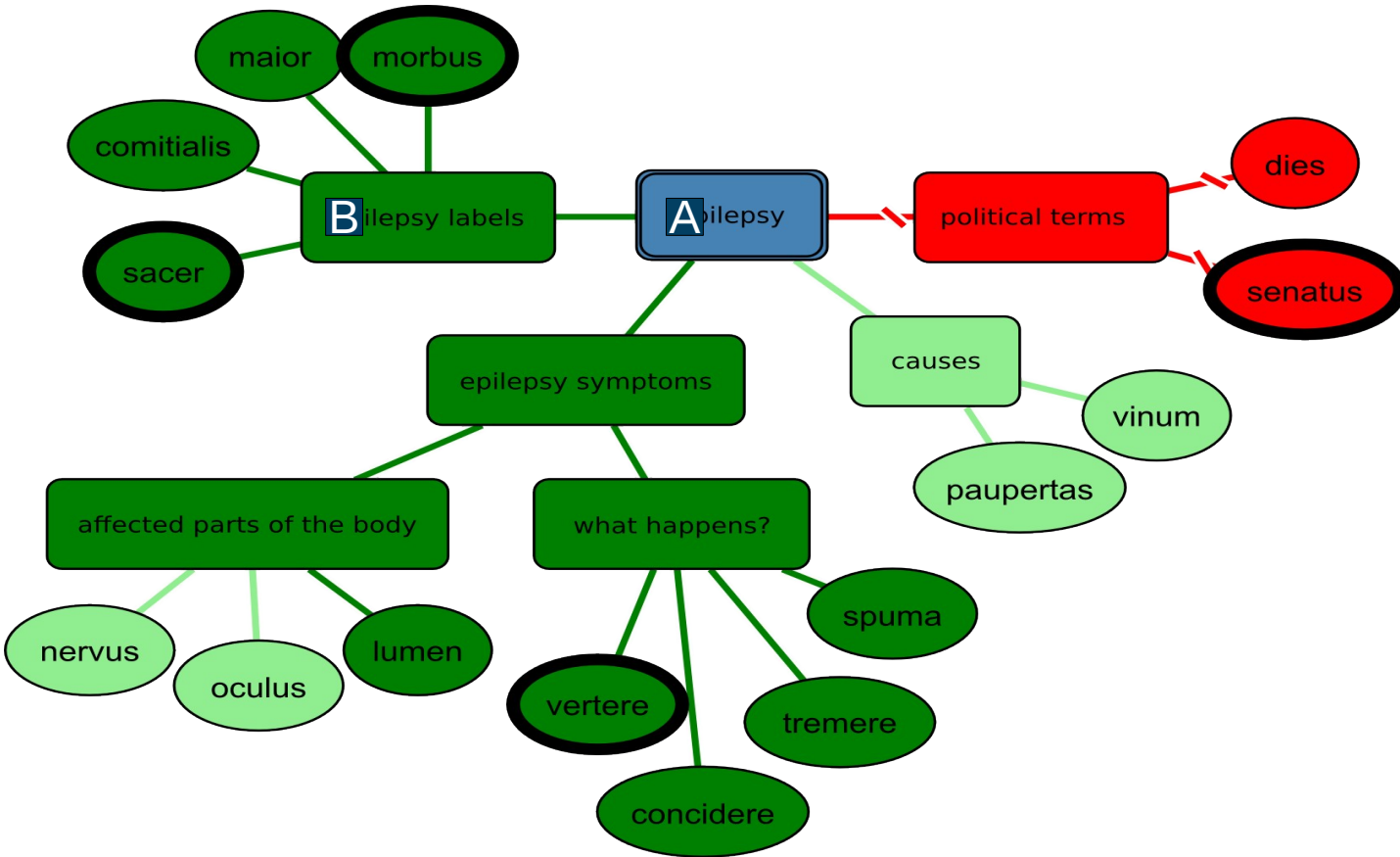
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbis, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



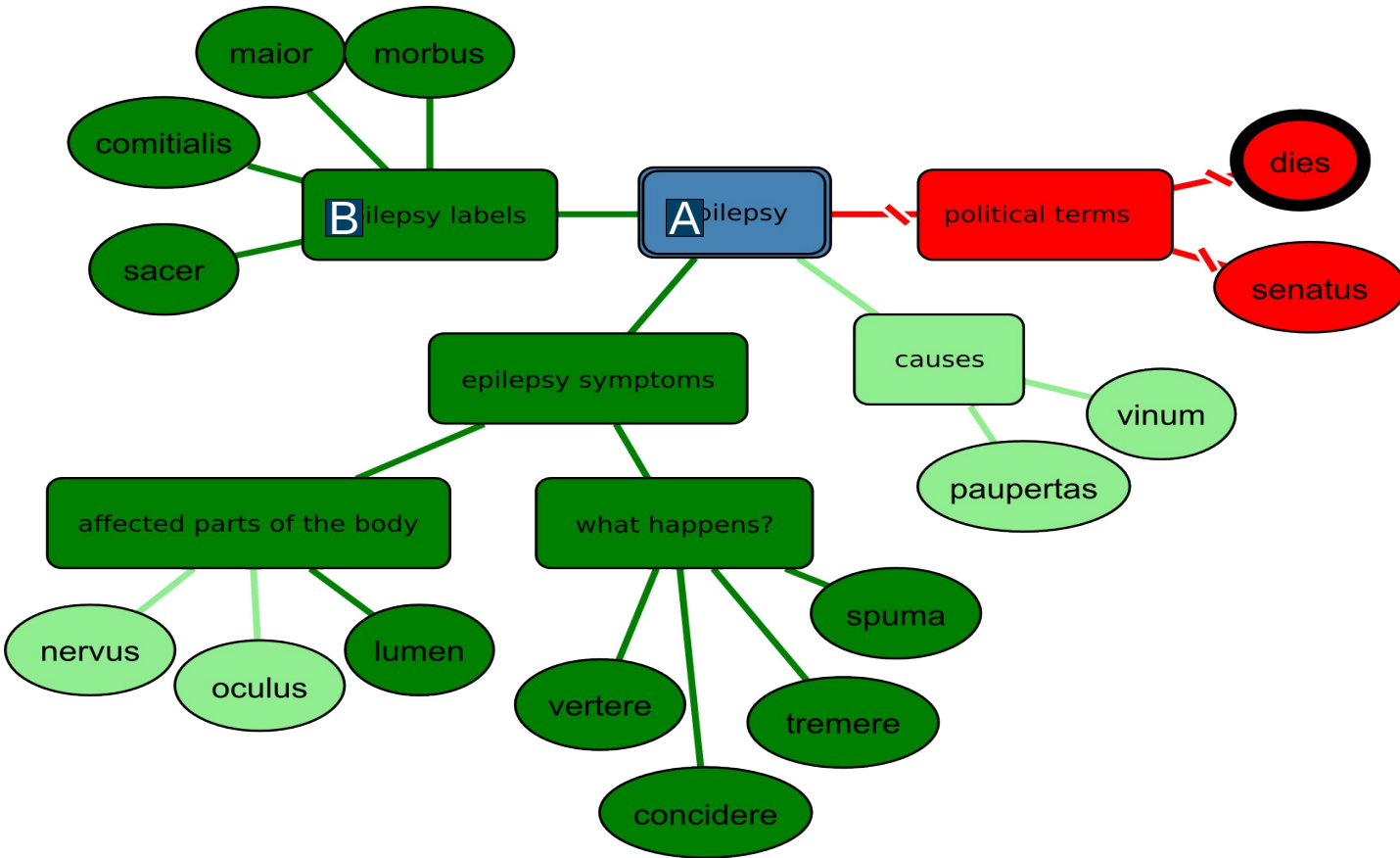
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbus, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



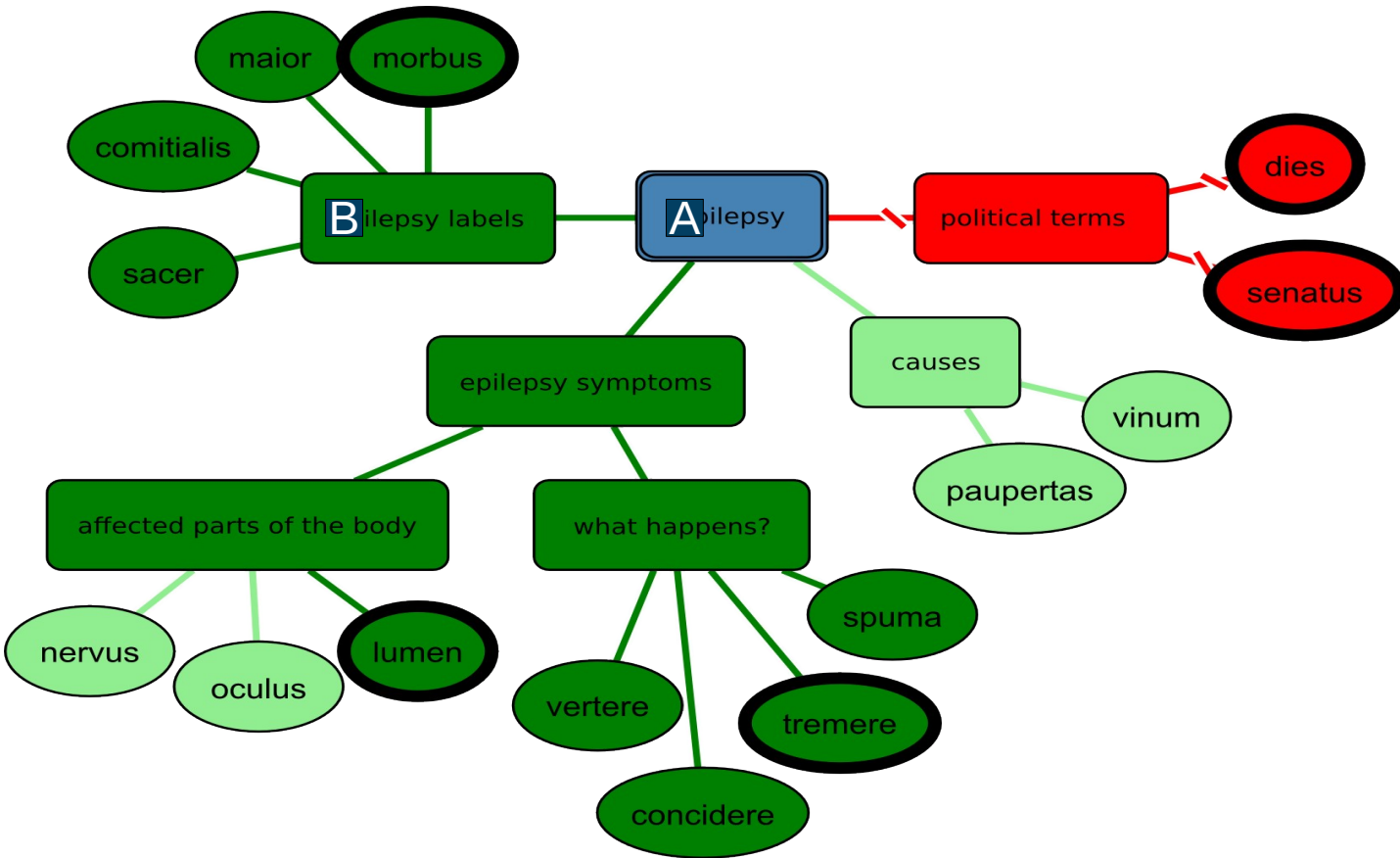
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbis, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



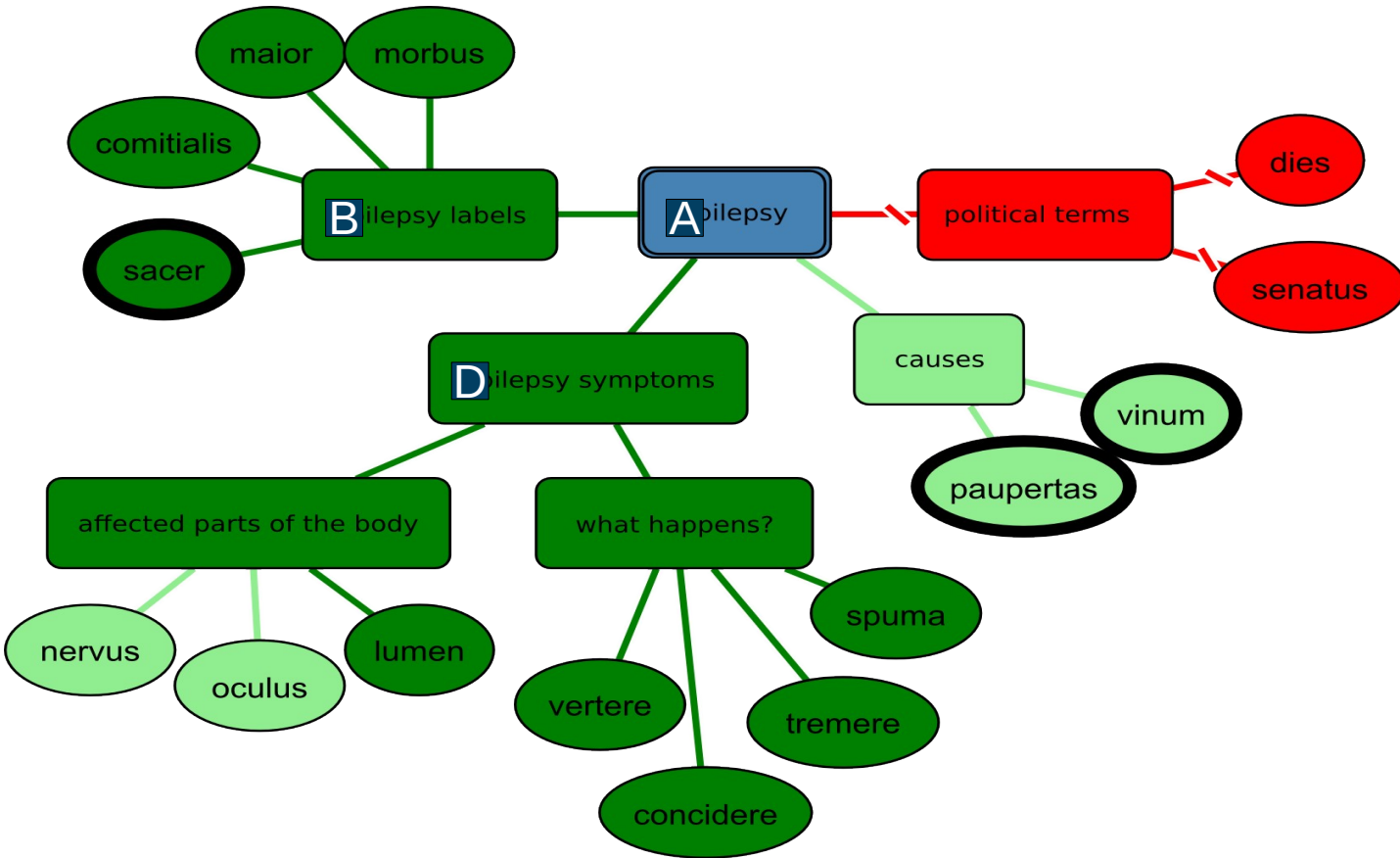
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbis, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



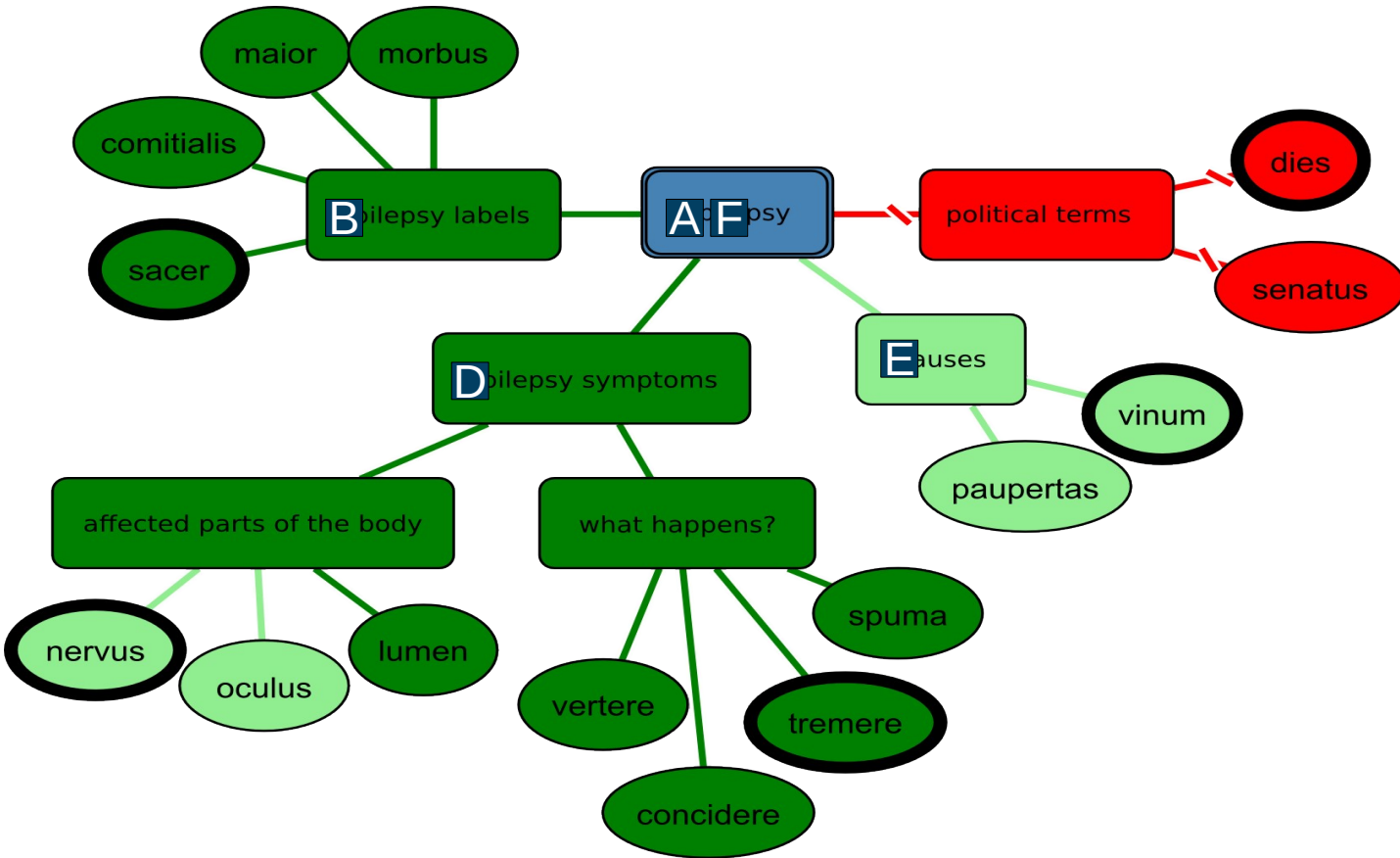
Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbis, vinum
<b>B</b>	morbis, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbis, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



Result	Word forms
<b>A</b>	spuma,nervus, lumen, morbus, vinum
<b>B</b>	morbus, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbus, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes



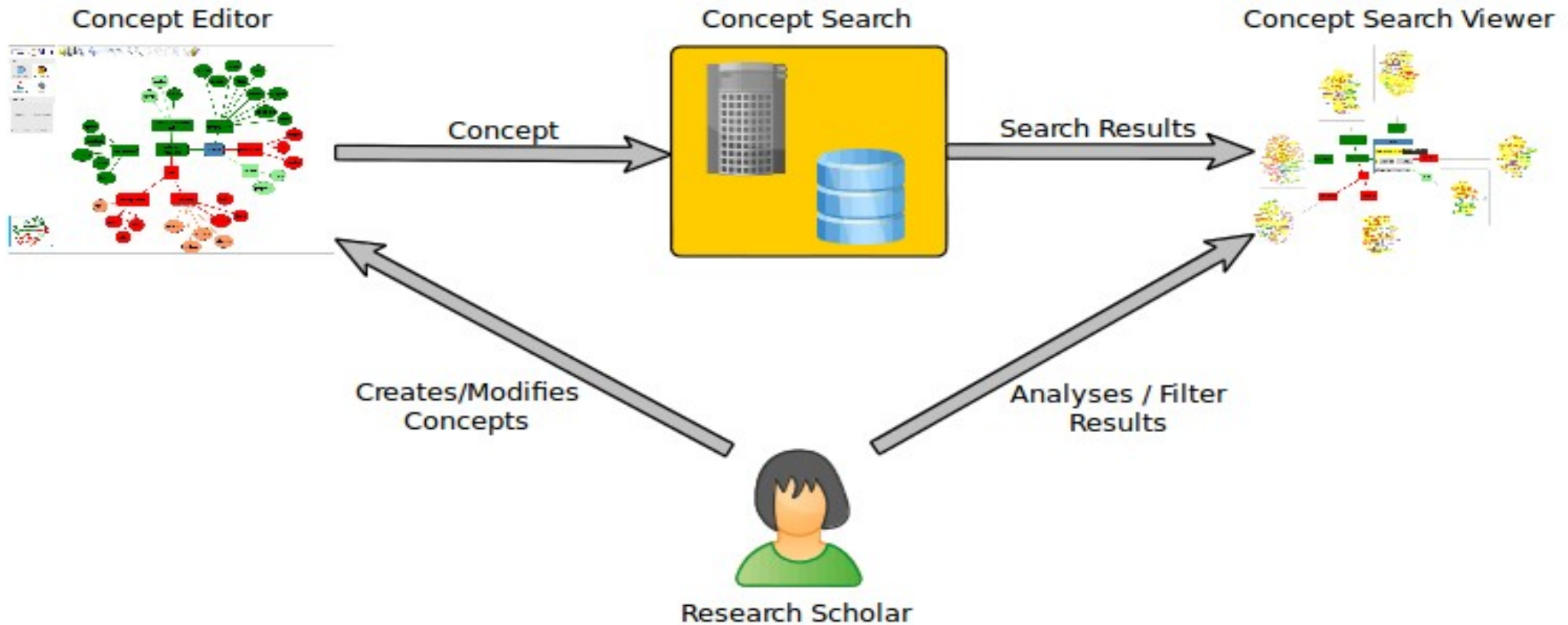
Result	Word forms
<b>A</b>	spuma, nervus, lumen, morbus, vinum
<b>B</b>	morbus, sacer, vertere, senatus
<b>C</b>	dies,
<b>D</b>	morbus, dies, senatus, lumen, tremere
<b>E</b>	Sacer, paupertas, vinum
<b>F</b>	Sacer, nervus, tremere, vium, dies

# Assignment of Results to Concept Nodes

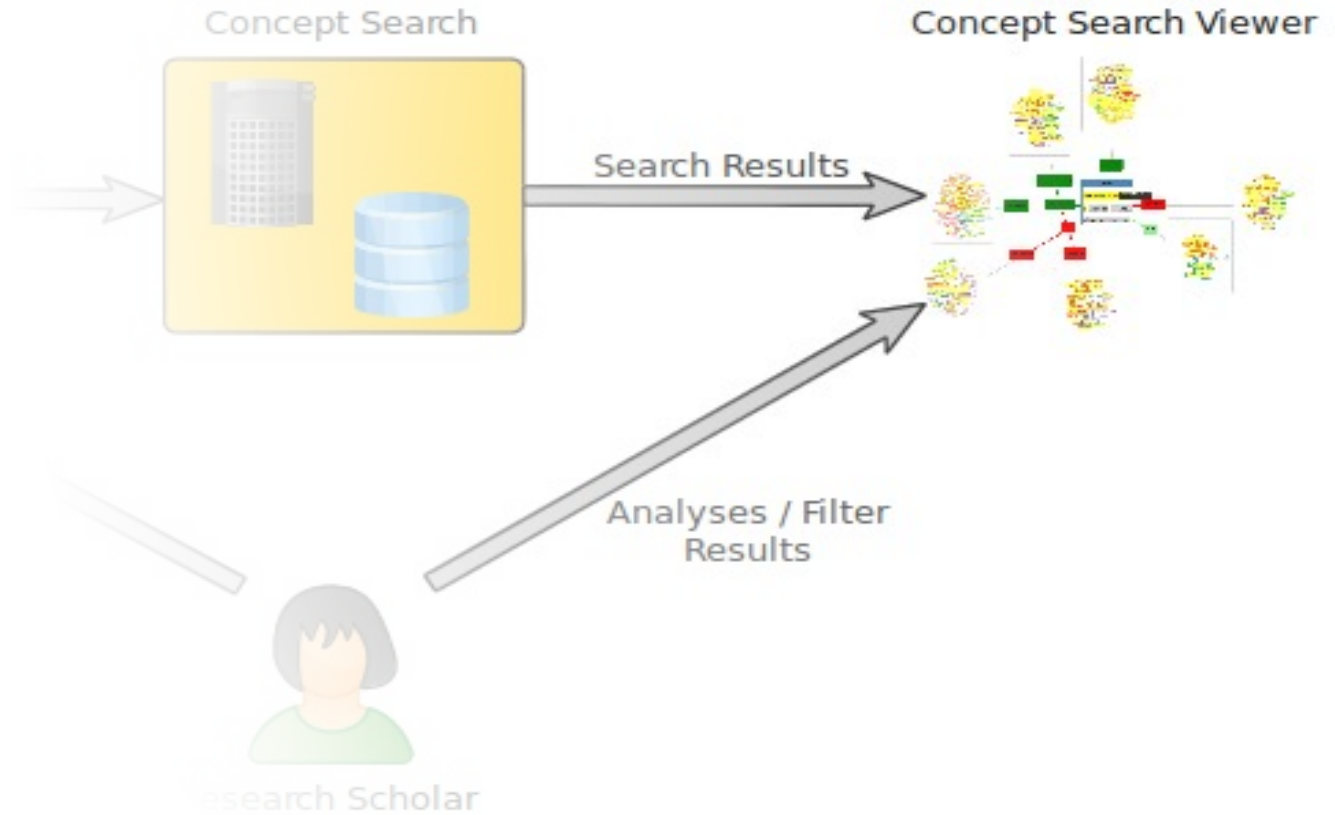
## Optional (tweakable) parameters

- ▼ Further criteria to decide when a result moves up to parent node:
  - ▼ Threshold value
  - ▼ Dark/lighter shades of concepts
  - ▼ Aggregation and intersection of results
  - ▼ Etc.
- ▼ Within a Concept node, rank texts according to
  - ▼ Term variance
  - ▼ Word counts
  - ▼ the ratio of positive to negative words
  - ▼ Titles
  - ▼ Combinations of above
  - ▼ etc.

# Concept Search Pipeline

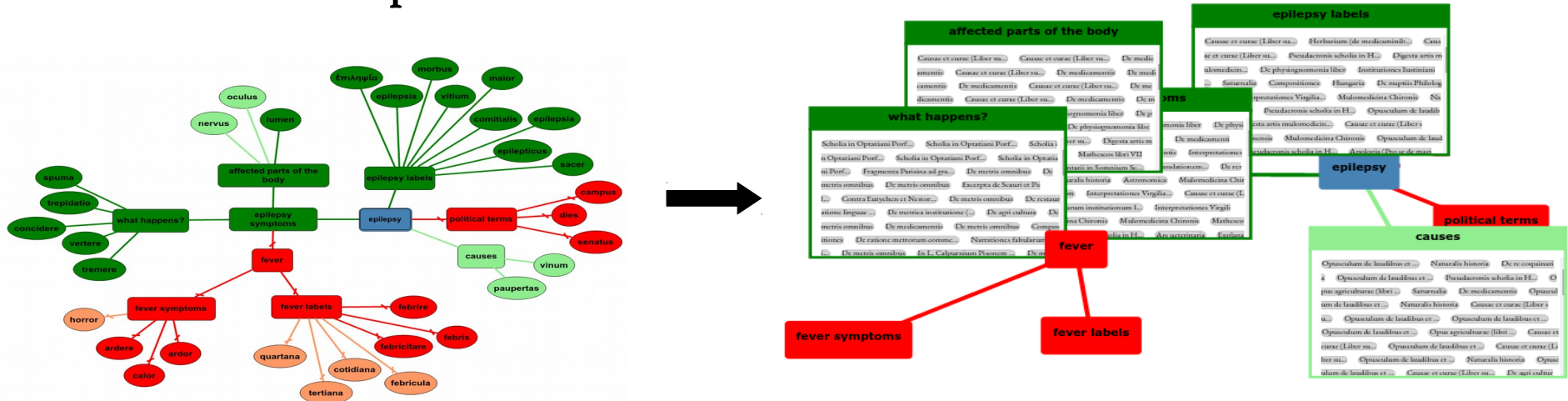


# Concept Search Pipeline

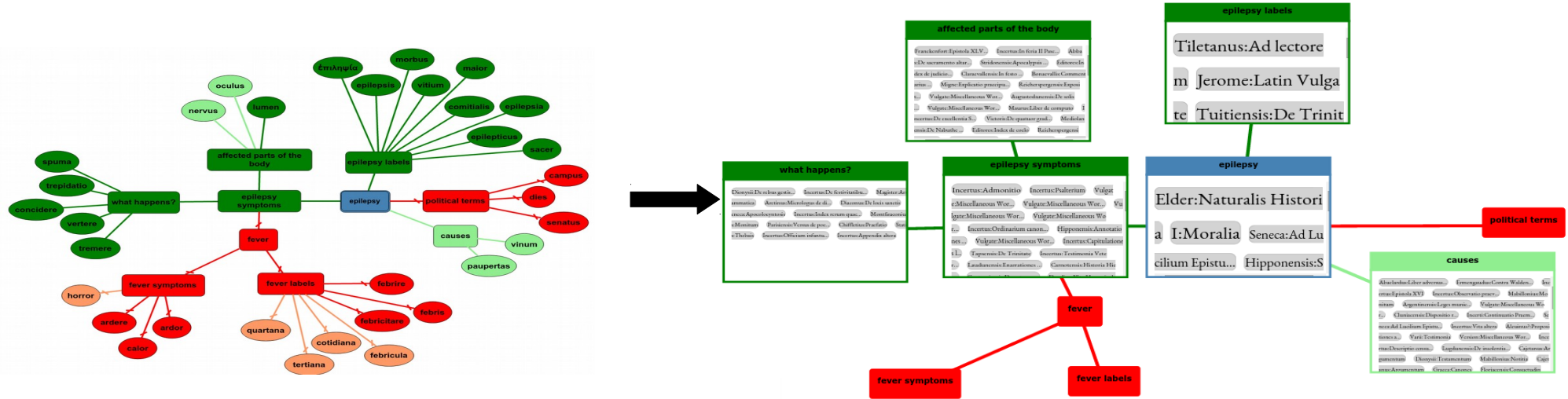


# Layout calculation

- Nodes sizes expand due to search results placements
- All expanded node sizes equal
- Overlap

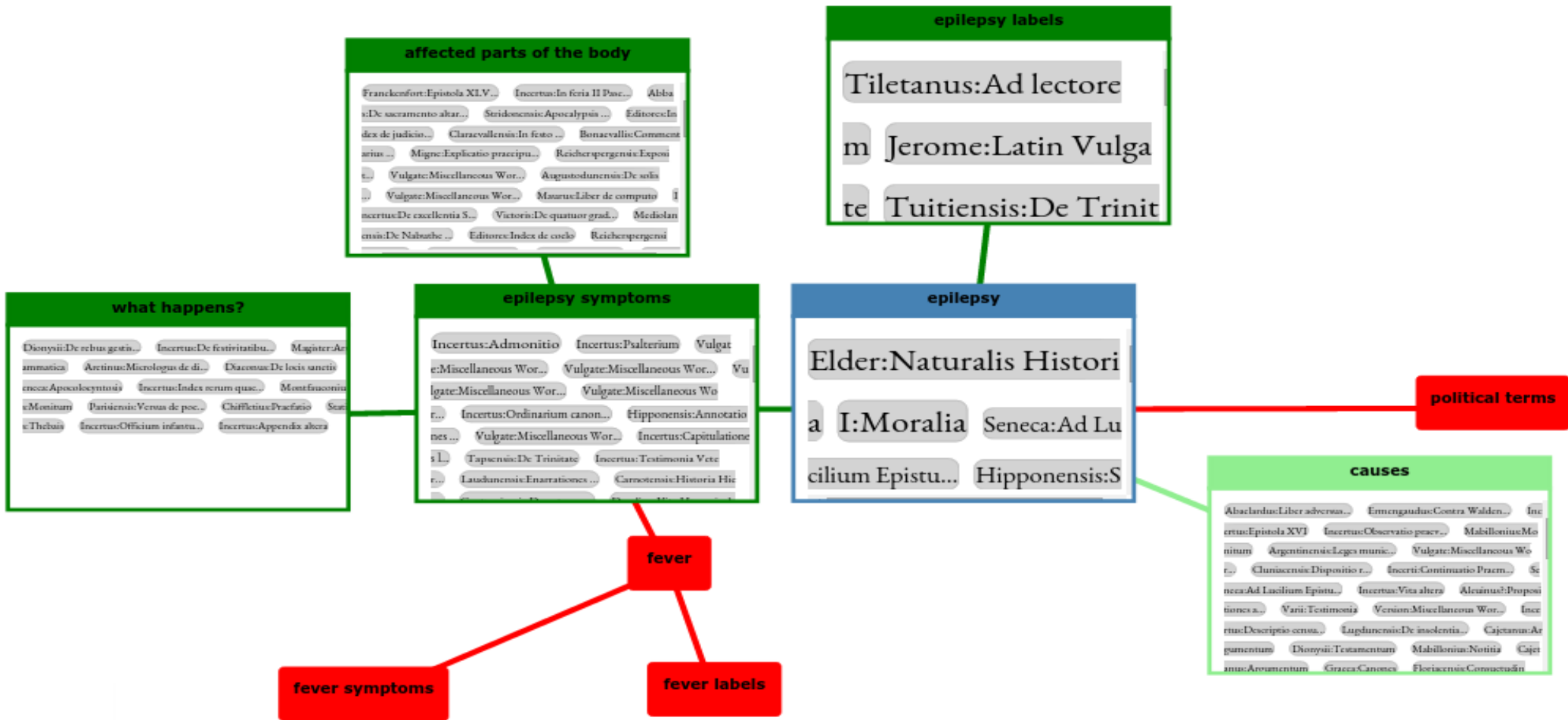


# Layout calculation (Bottom up )

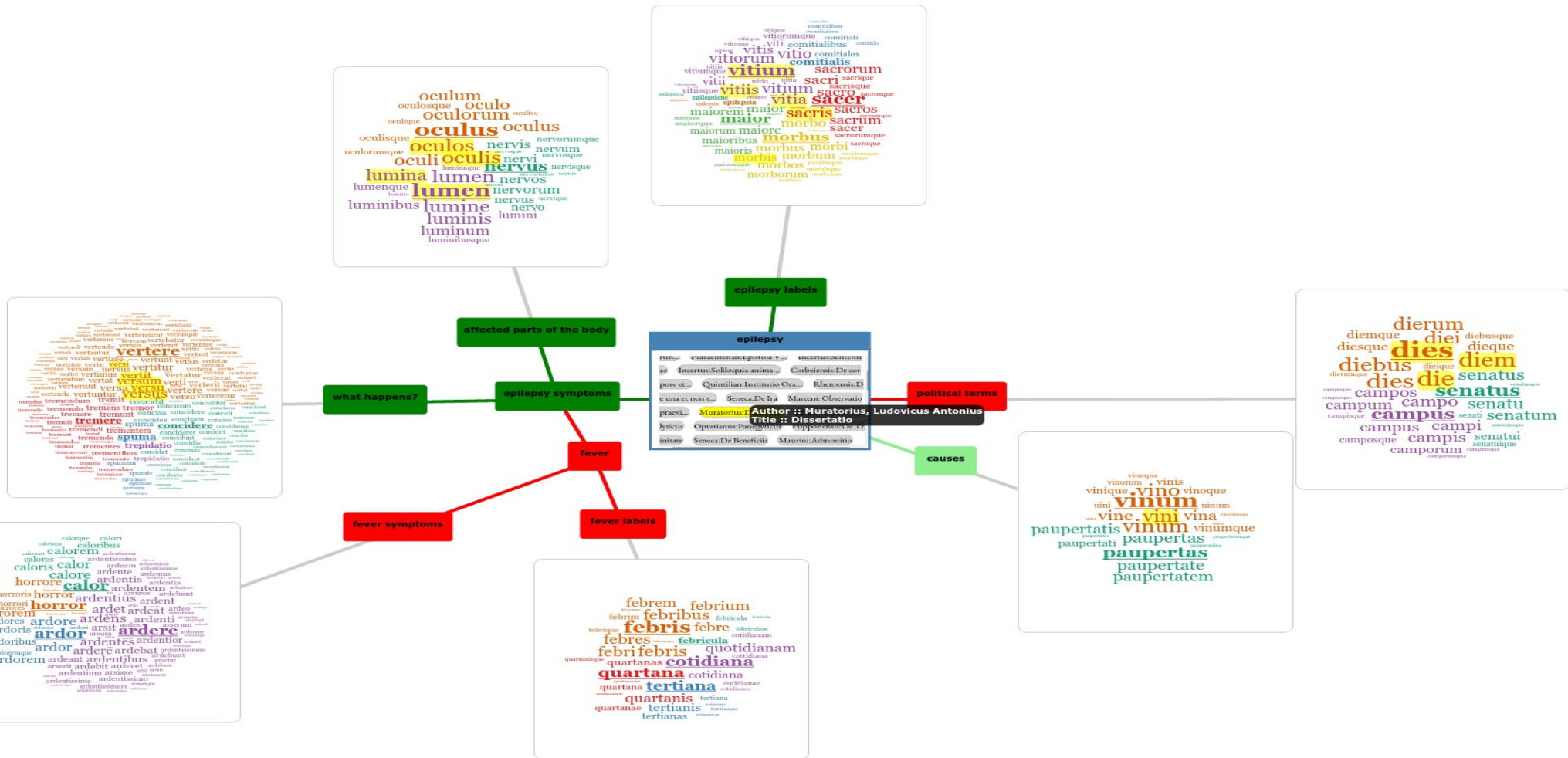


For further details refer to [CJBSI6]

# Concept Search Results (Aggregate View)



# Concept Search Results (Word Summaries View)



# Concept Search Results (Word Summaries View)



For further details refer to [JBRZS<sub>15</sub>]

# Concept Search Results (Close Reading View)

Title :: Mulomedicina Chironis

Start verse :: lib. 1, § 3, p. 4, lin. 23

No. of verses :: 2

de morbo renali iumentorum, si quid enim per ignorantiam male factum vel contrarium circa animalium corpora fuerit, emendari non potest, vincitur enim corpus anima cedente per periculum. prima ergo pars est medicinae causarum rationem reddere, deinde ipsa adiutoria medicinae docere oportet, quomodo possint auxilium salubre corpori praestare propter diuturnam vitam et integritatem membrorum. quibus itaque rebus medicinae scientia veterinae constet omnis, ex quibus adiutoris atque corpora iumentorum malam valitudinem et causa liberari possint, et omnem curam rationabiliter docere oportet. primum dico de sanguinis tractione. Firmissimum adiutorium est medicinae sanguinis detractio, verum tamen si rationabiliter et localiter fiat. licet paene in omnibus valitudinibus fieri sanguinis detractio praecipiat, saepe tamen ignorantibus rationem per sanguinis detractio nem periculum vitae fecisse contingit. non enim similiter omni aetate iumentorum vel omni tempore, sed nec omnibus sanguinis detractio fieri praecipitur. cuius ergo tamen rationem protinus reddo. primum de sanguine reddenda est ratio. quocunque in corpore per ipsum vita constat, rursus detractus a corpore sanitatem praestet. ex quacunque enim causa valitudo est, in aliquam partem corporis quae pertinet, semper indigestione fatigatus sanguis corrumpitur, discurrit per omnia membra et totius corporis languorem infert. per quem incipiet strictura et tensio nasci, per inflationem venarum. quae strictura et tensio corporis aliter laxari non potest, nisi per sanguinis detractio nem. aliud enim compendium nullum est nisi sanguinis detractio, per quam universa corporis vitia et morbi abstrahi possunt.

Notes

Is Relevant?

☒ Yes

☐ No

☐ Interesting

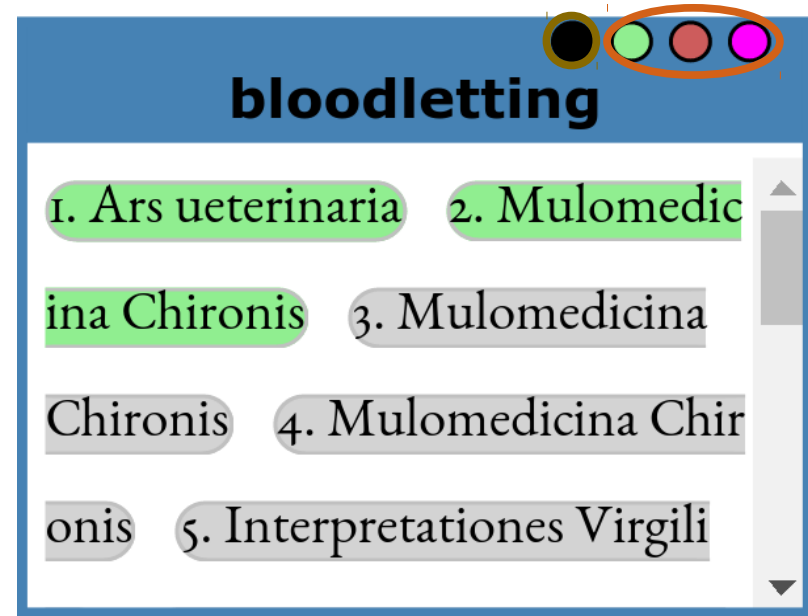
Previous Sentence

Next Sentence

# Concept Search Results

Exporting & reusing results  
(An important feature)

- ▼ Save Tables
  - ▼ Relevant
  - ▼ Irrelevant
  - ▼ Interesting
- ▼ Save All session



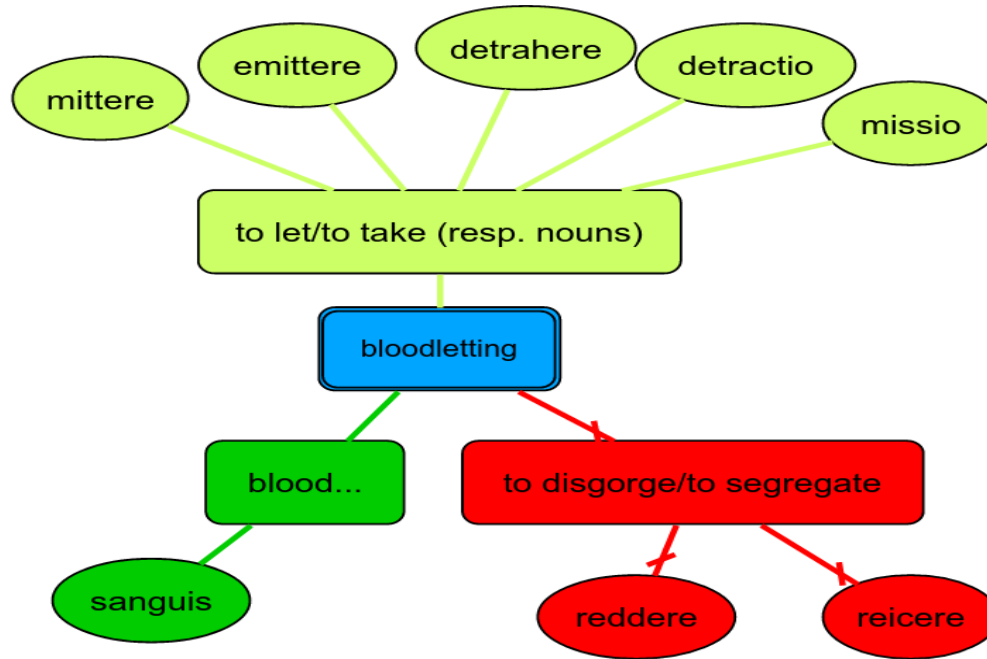
# Results, Demo and Findings

# Humanities scholars in the project

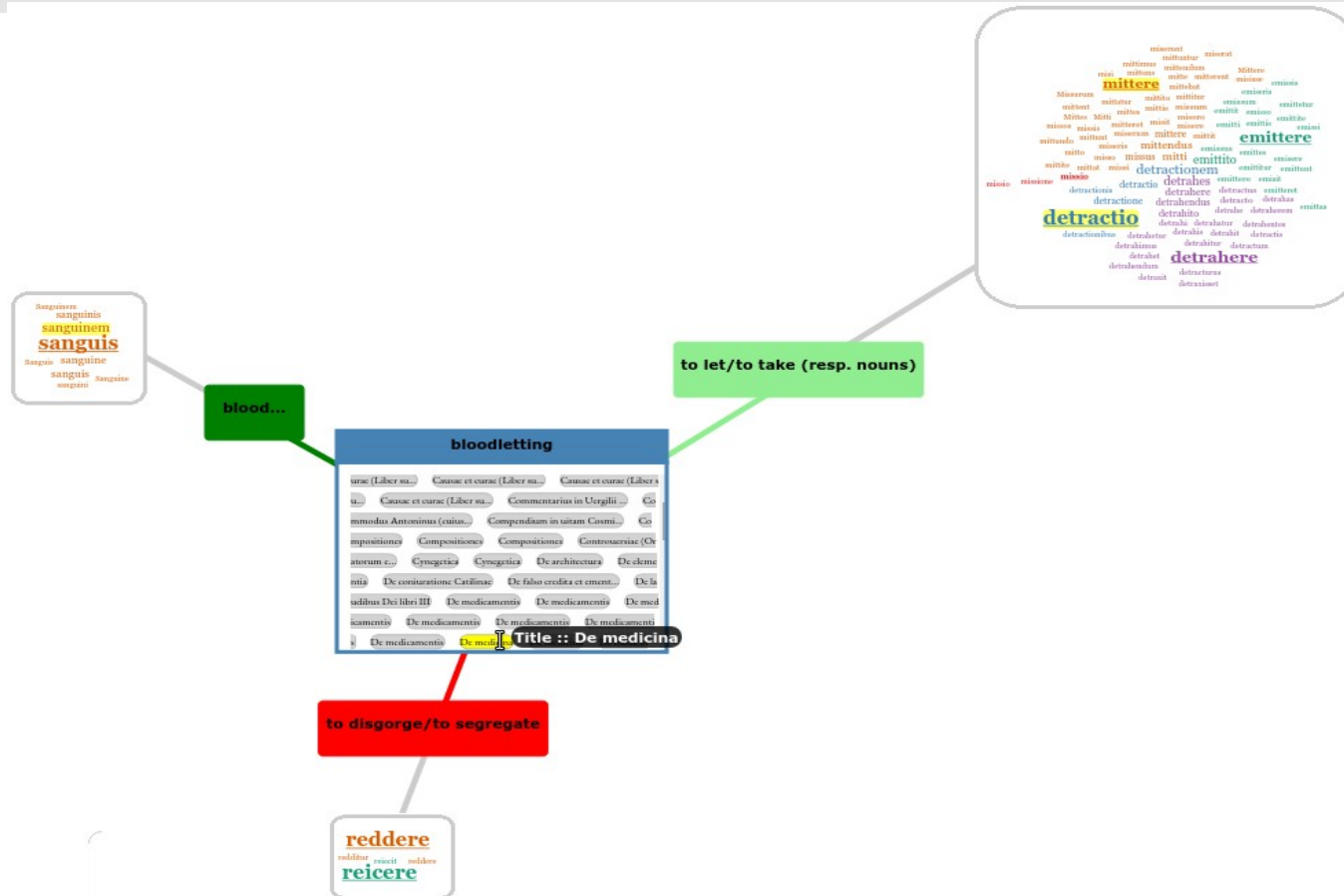
Two humanities scholars have extensively used the concept search

- ▼ *Scholar A → More Interested to explore via the distant reading aspects*
- ▼ *Scholar B → Interested to explore by reading the concerned passages directly, and not using distant reading features extensively*

# Concept - Bloodletting



# Concept - Bloodletting

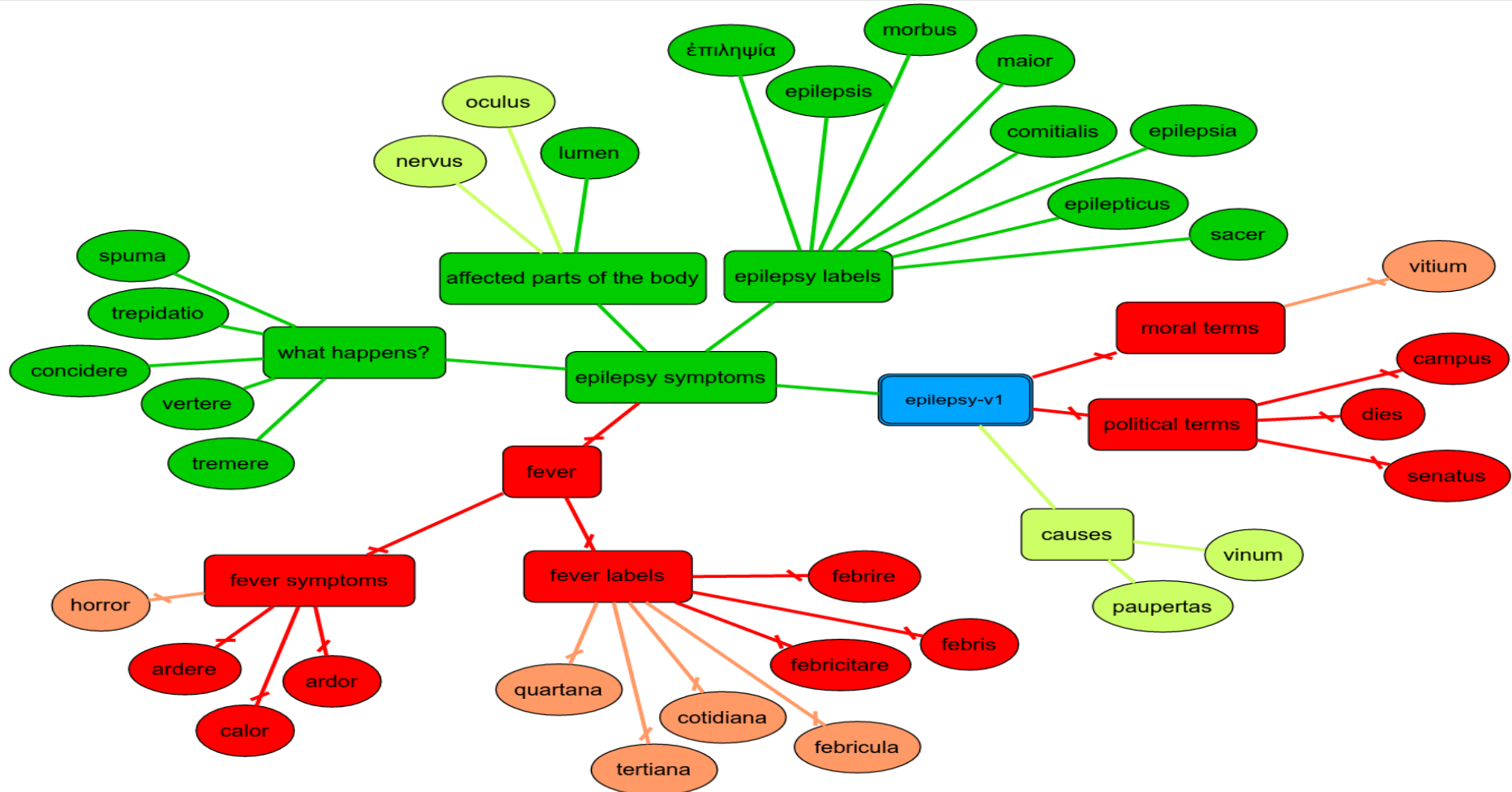


Demo

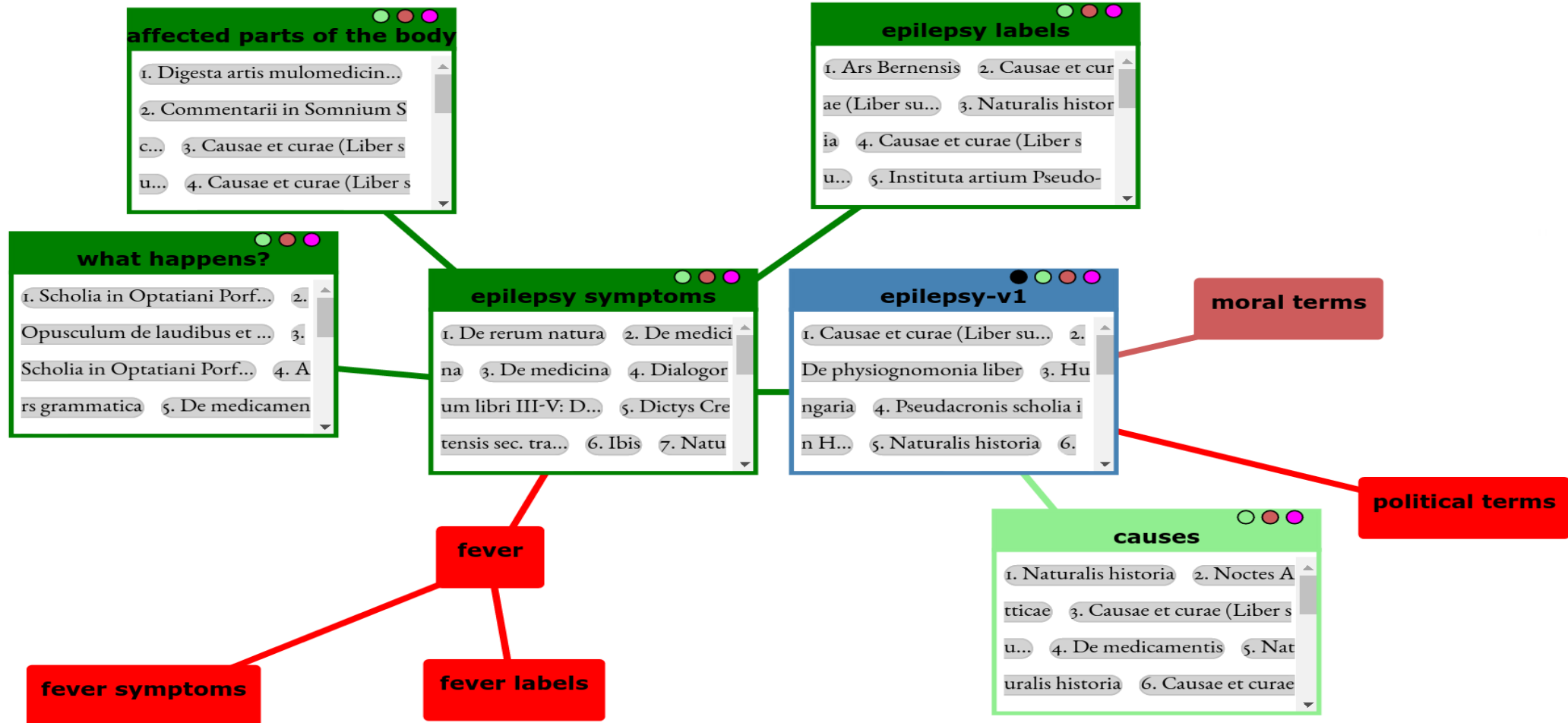
# Usage Scenario

## (Remodelling a concept)

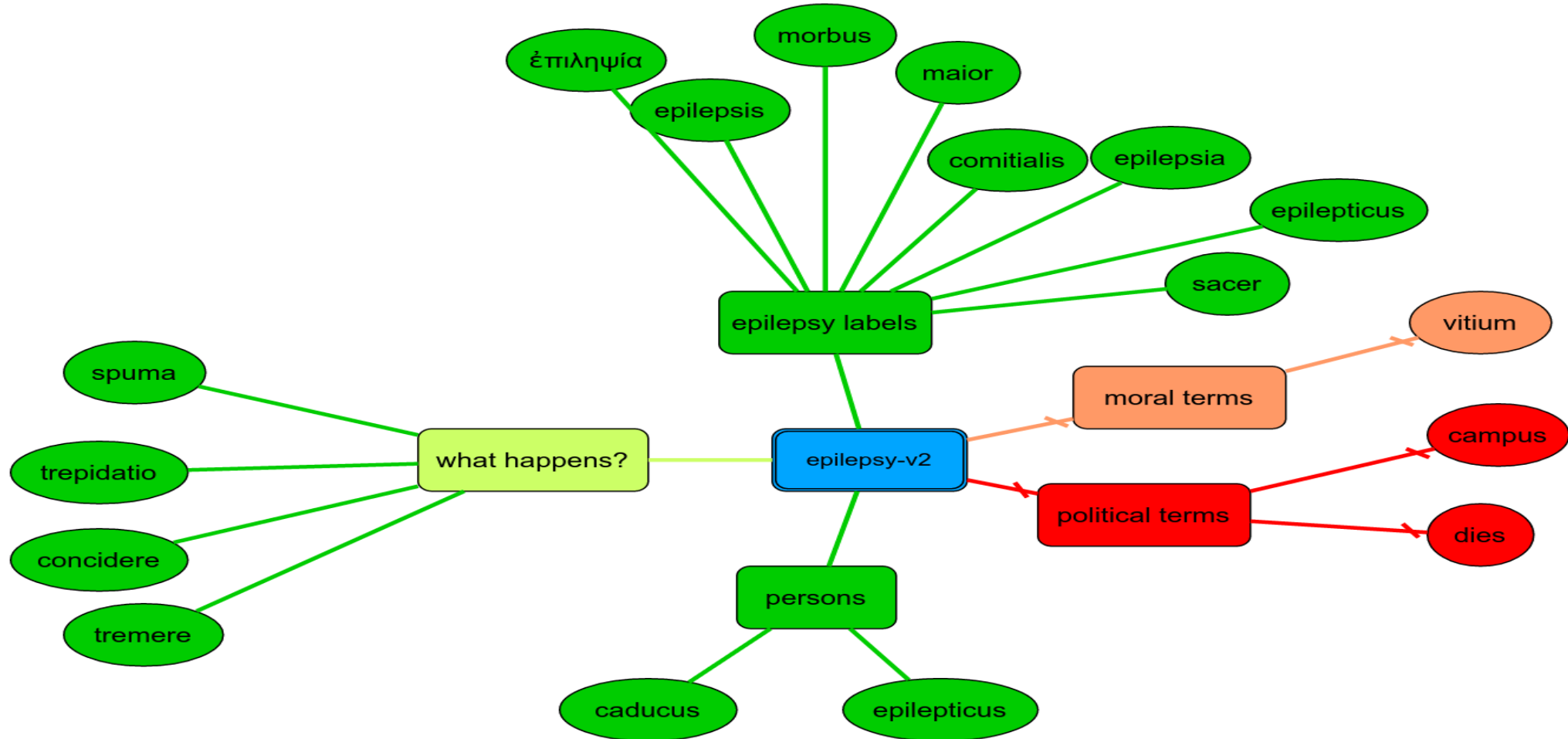
# Concept – Epilepsy – Version 1



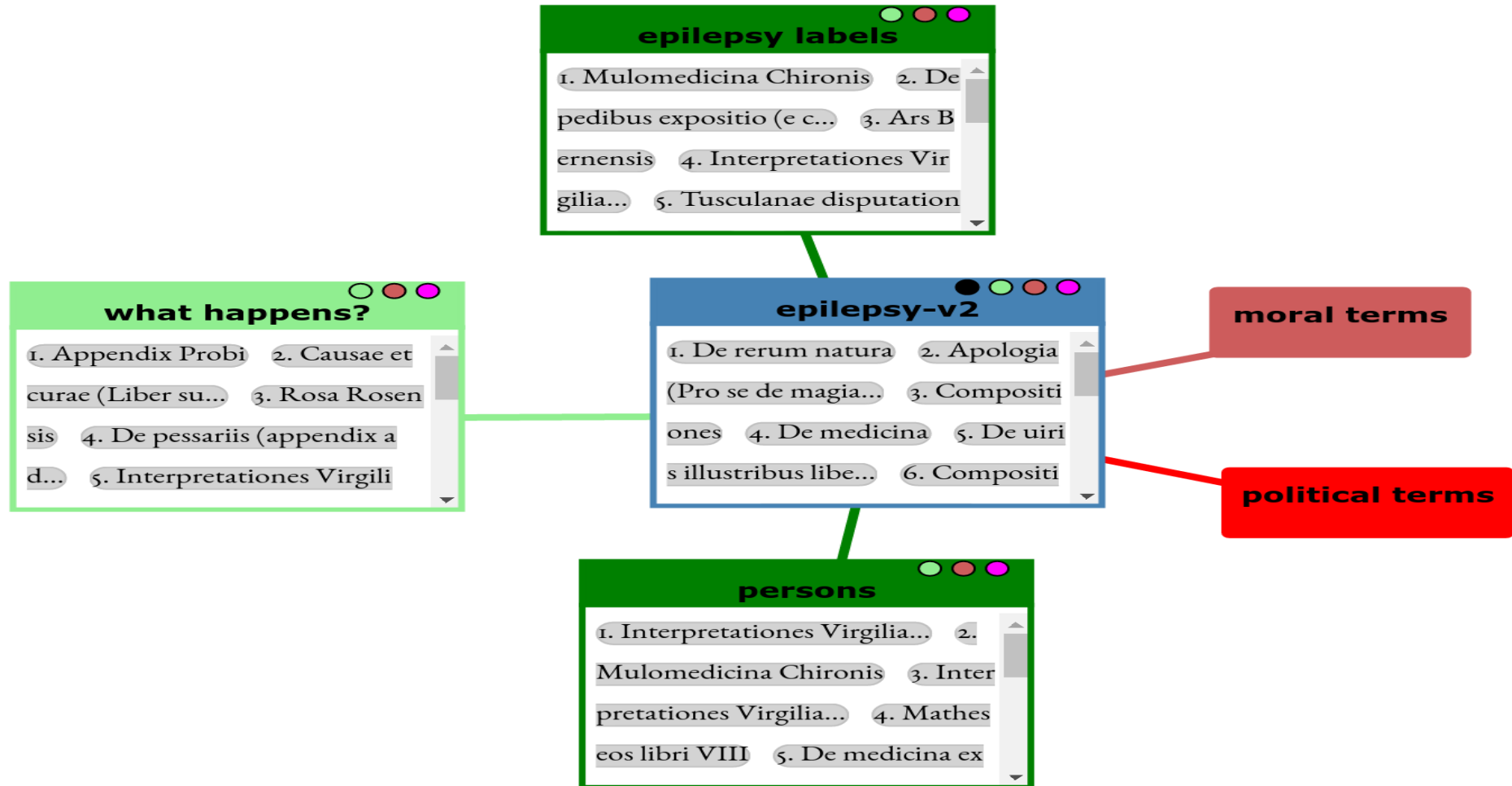
# Concept – Epilepsy – Version 1 (Search results)



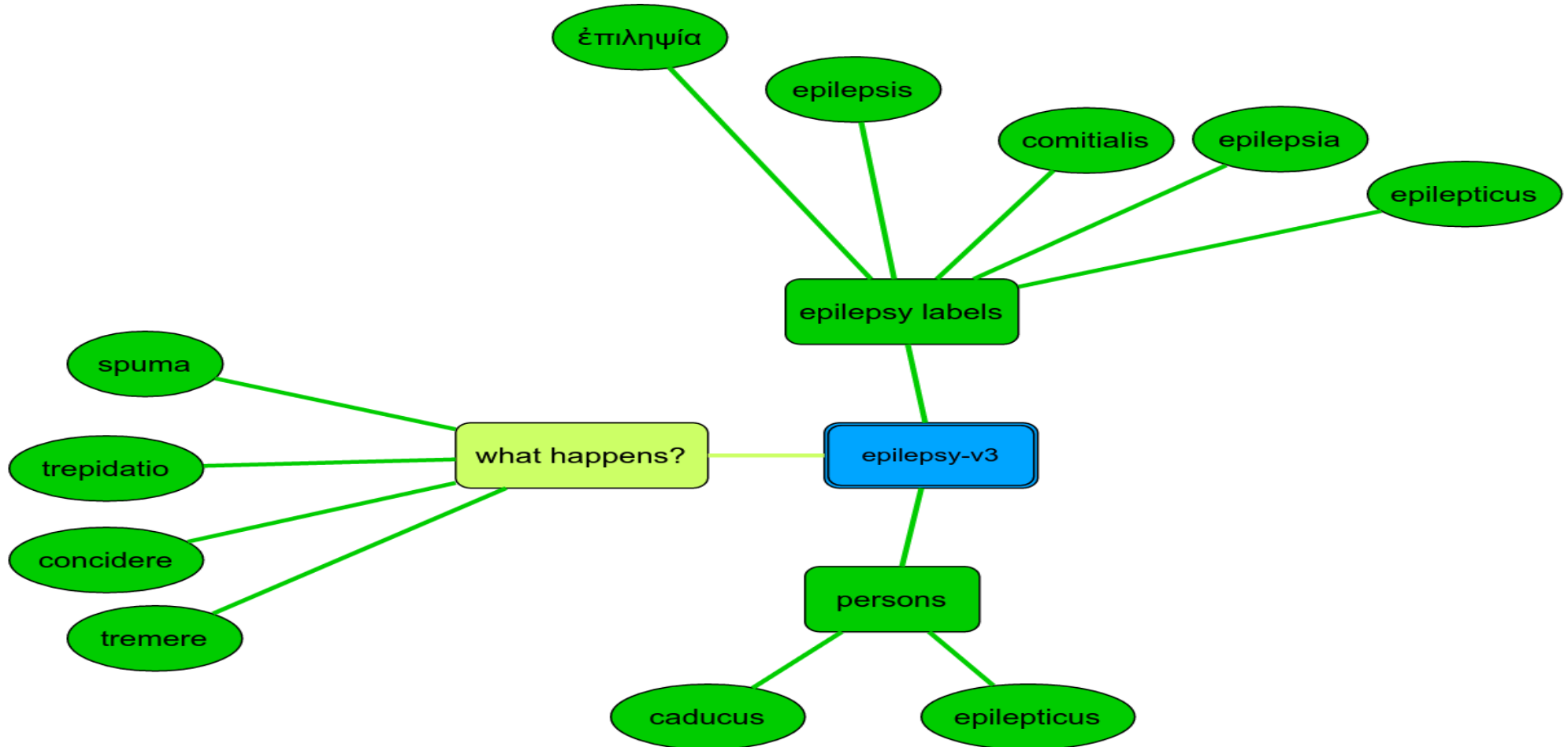
# Concept – Epilepsy – Version 2



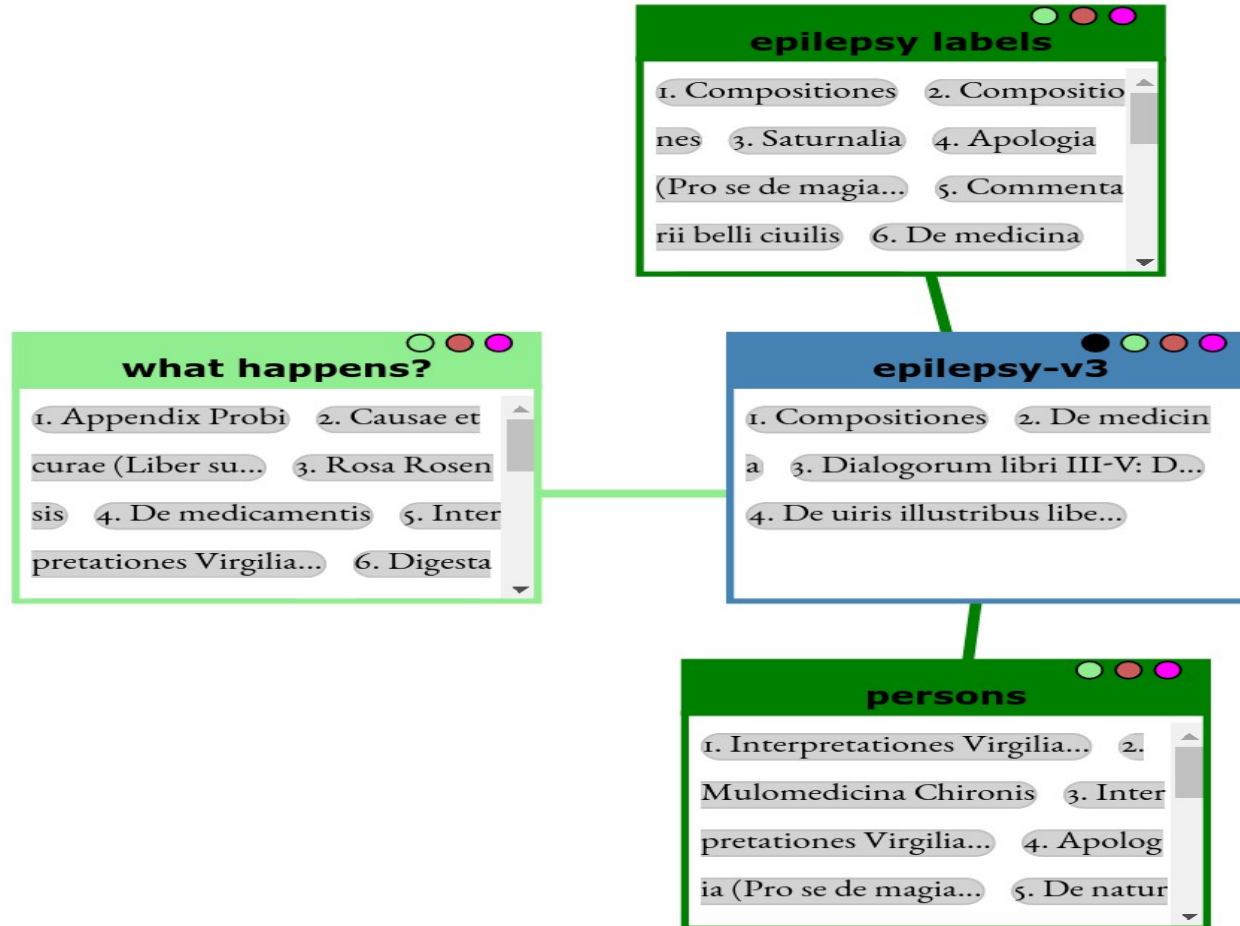
# Concept – Epilepsy – Version 2 (Search results)



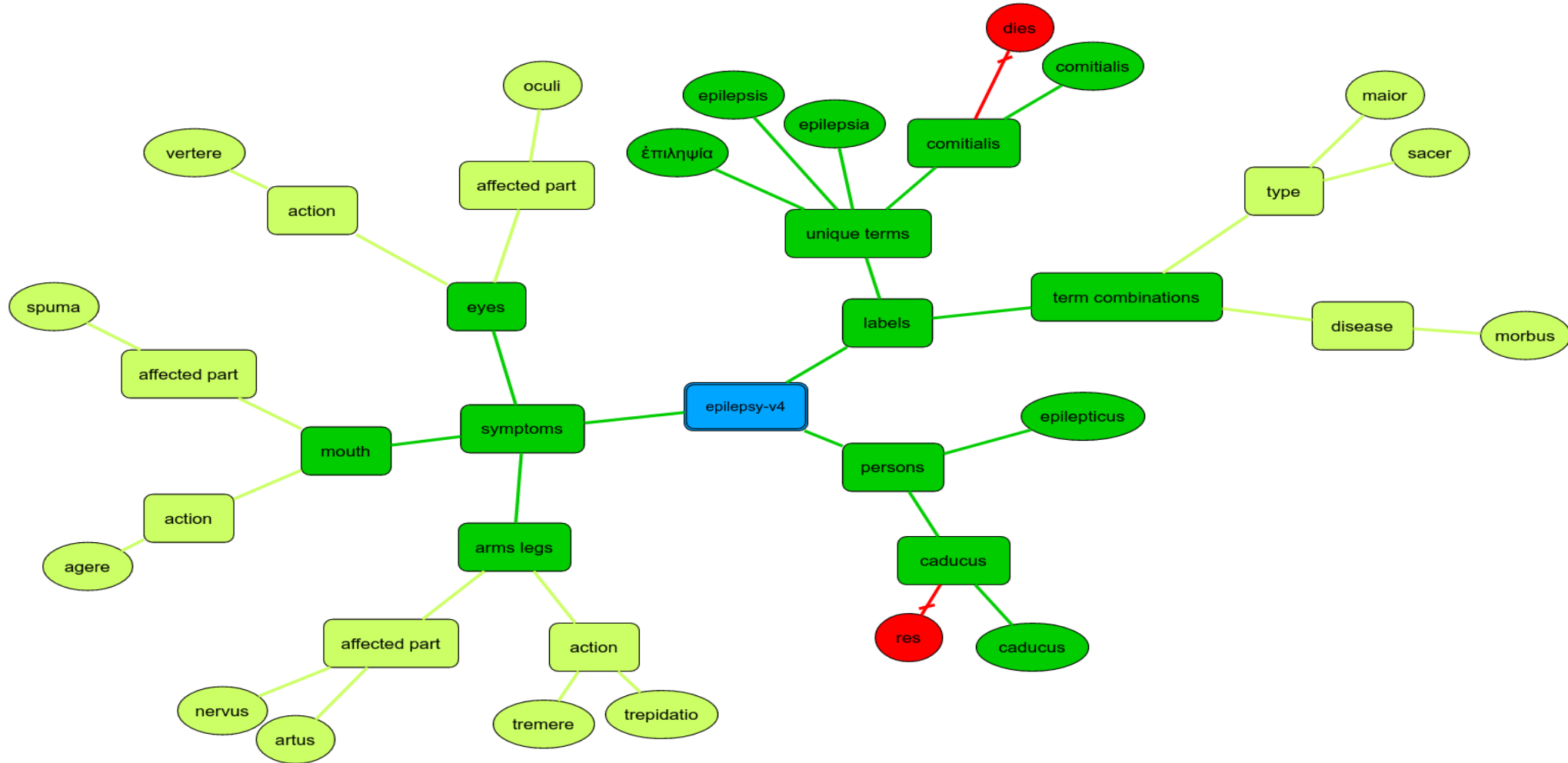
# Concept – Epilepsy – Version 3



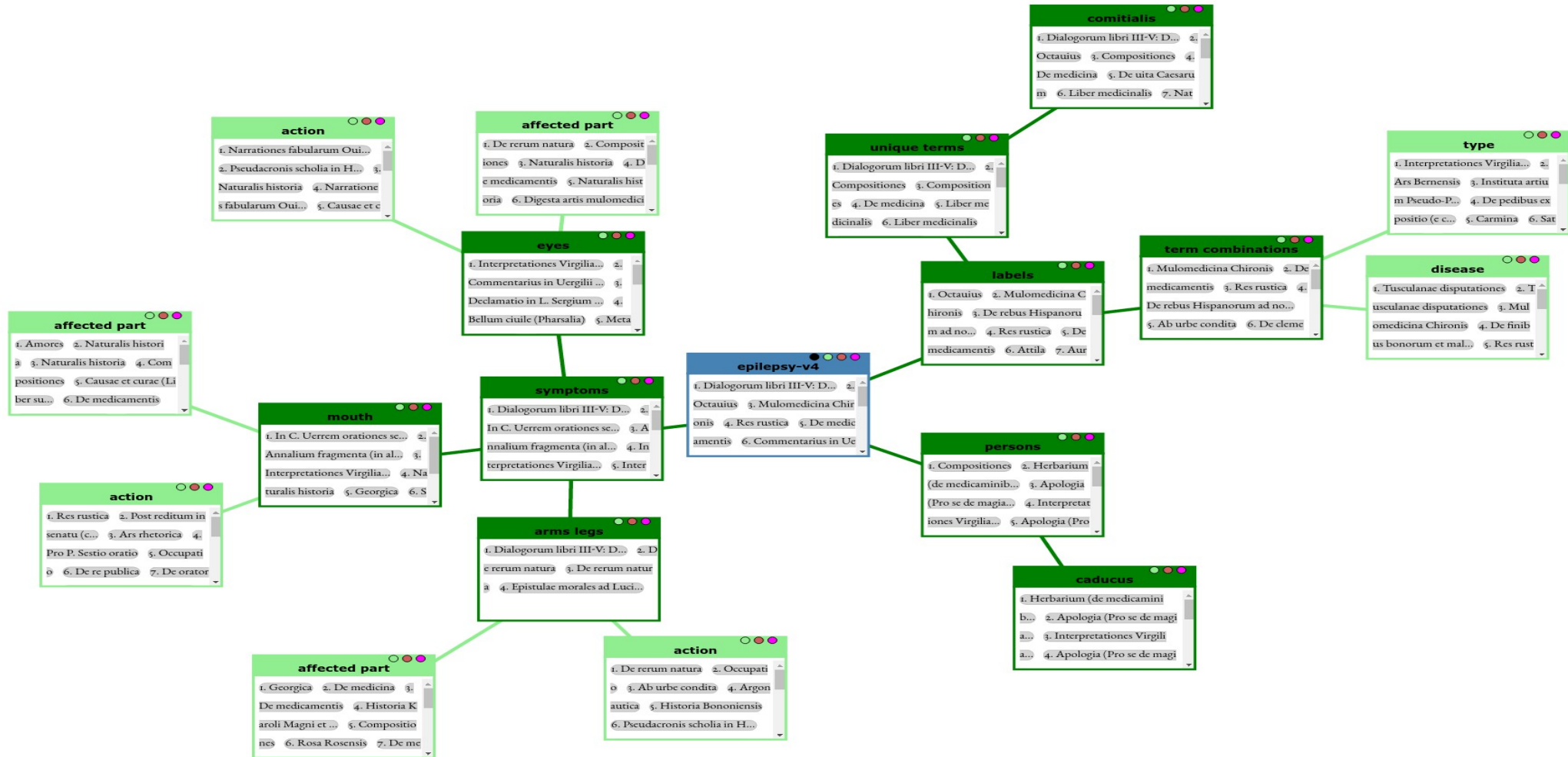
# Concept – Epilepsy – Version 3 (Search results)



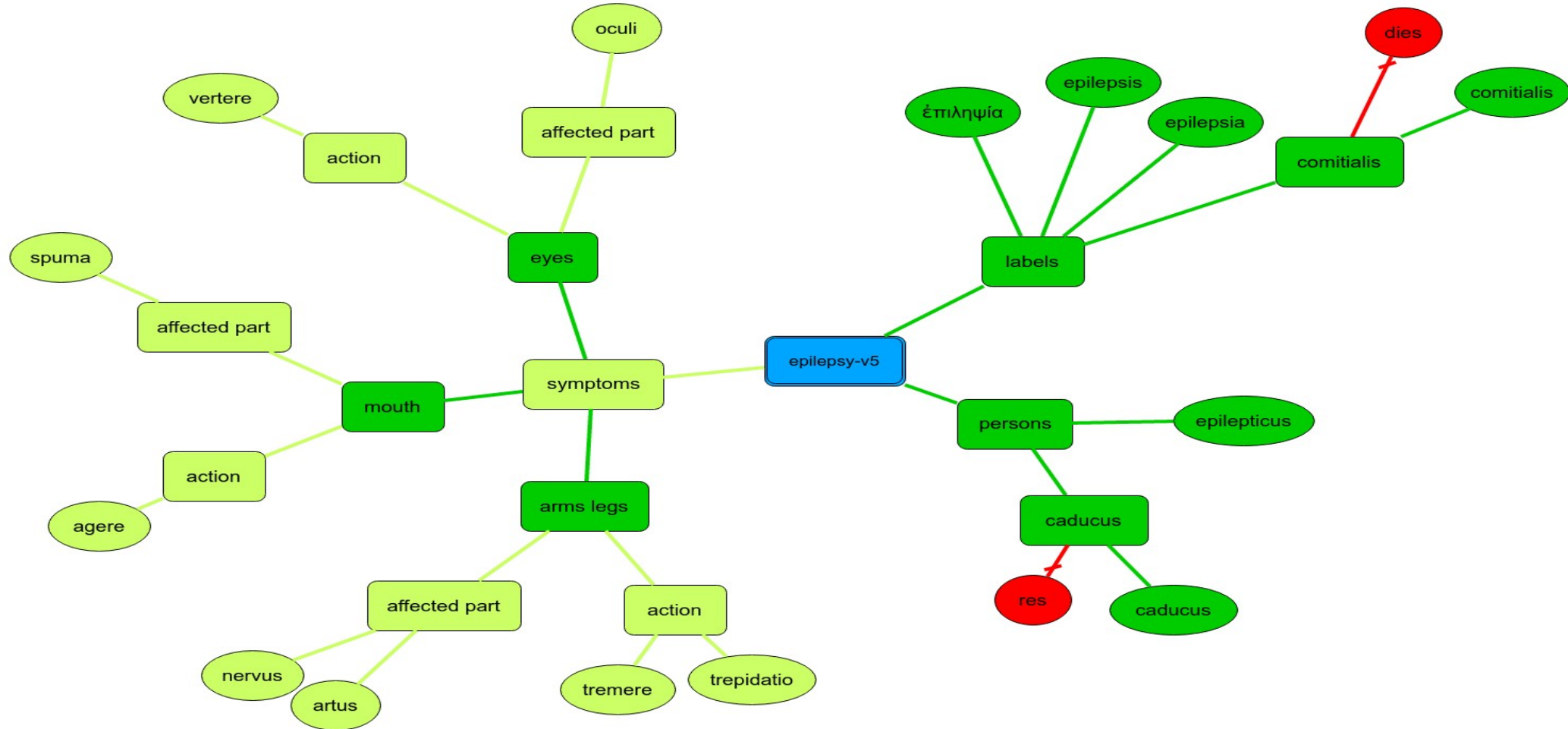
# Concept – Epilepsy – Version 4



# Concept – Epilepsy – Version 4 (Search results)



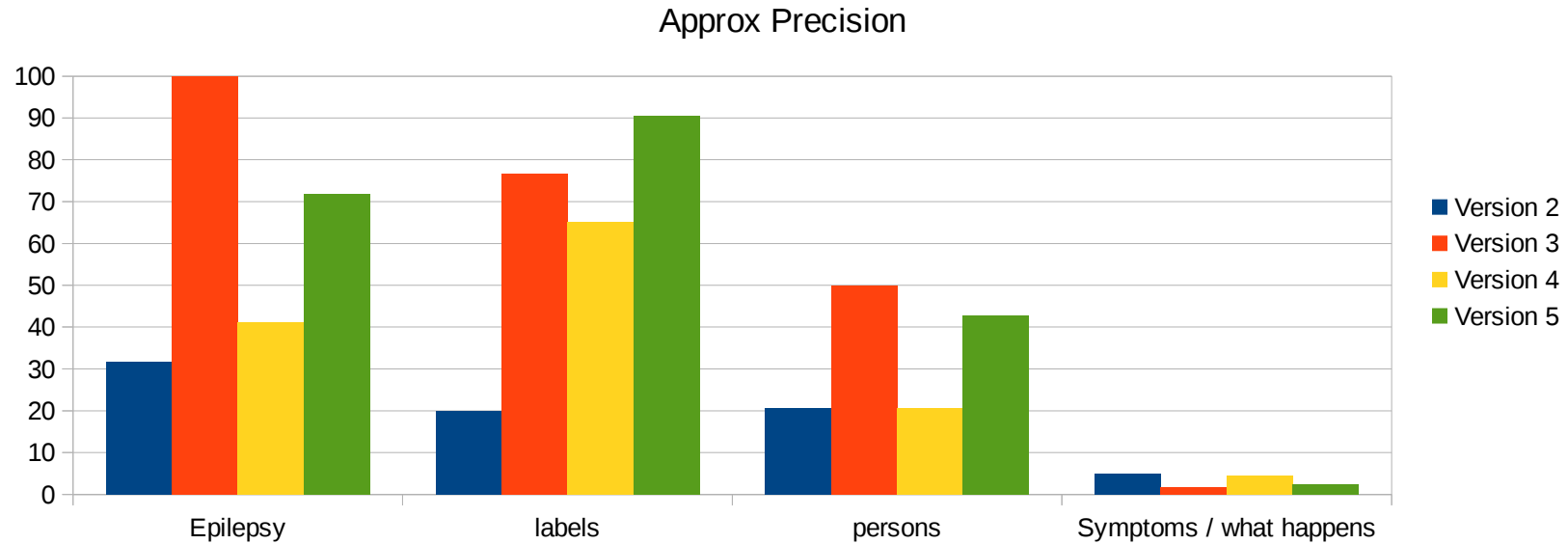
# Concept – Epilepsy – Version 5



# Concept – Epilepsy – Version 5 (Search results)



# Epilepsy - Summary of approx. precision



	Epilepsy	labels	persons	Symptoms / what happens
Version 2	38	7053	85	800
Version 3	4	128	95	824
Version 4	302	138	164	44
Version 5	174	106	68	43

Total results retrieved

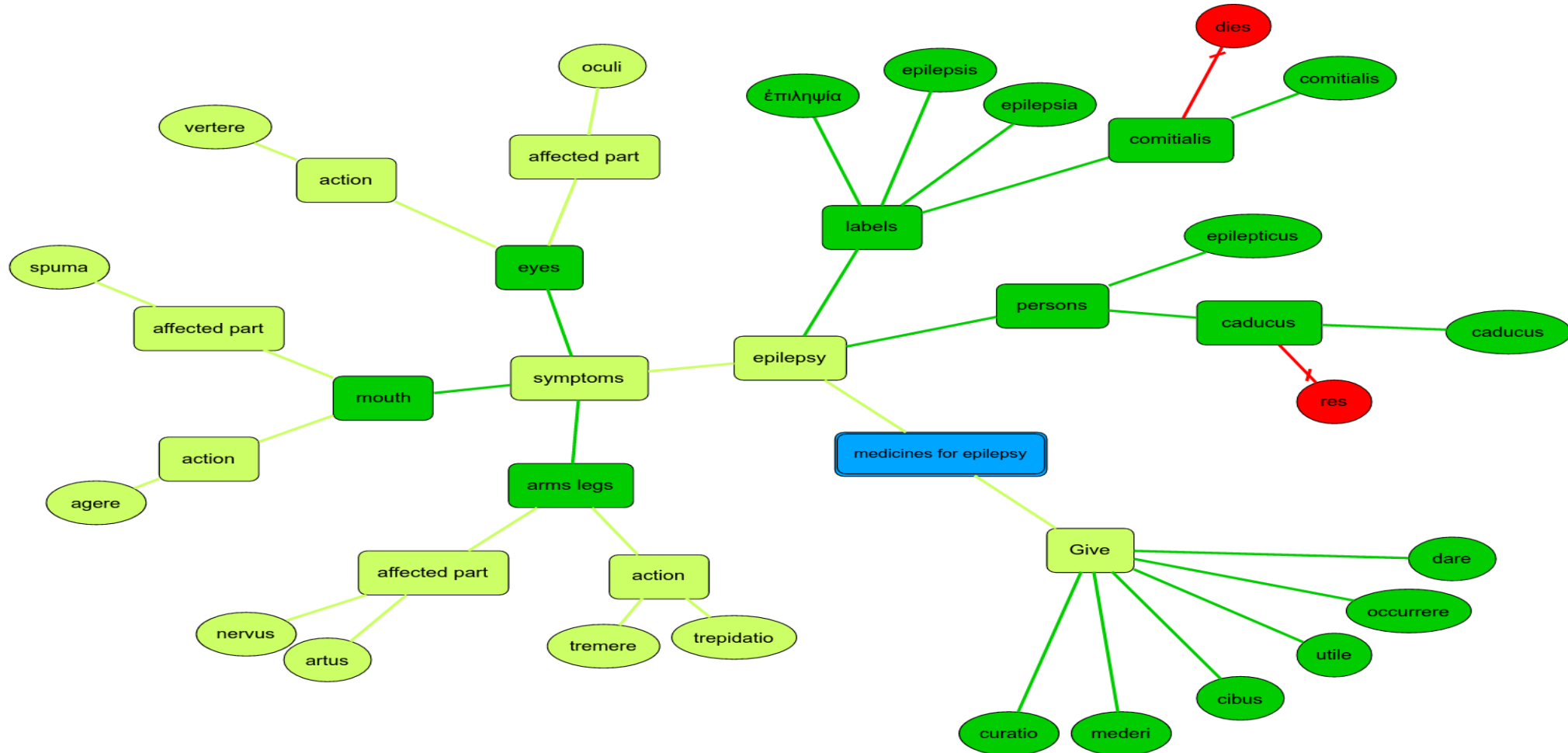
# Findings

- ▼ Ability to **steer** the search as desired – **a prominent feature of keyword search**
- ▼ Remodelling the concept iteratively can help the scholar to increase *recall* while also keeping *precision* factor high.
- ▼ Support observations concerning of a whole phrase, not only term. Is an option that dictionaries do not provide, but only commentaries.
- ▼ Support scholars to observe different uses of phrases (**phrase reuse**) depending on time, genre and author i.e. semantic development of a phrase.
- ▼ Support scholars to verify or falsify hypothesis of their defined concepts → remodel concepts → iterative process i.e. verify scholar's former knowledge, if the terms defined belong to the concept or not.

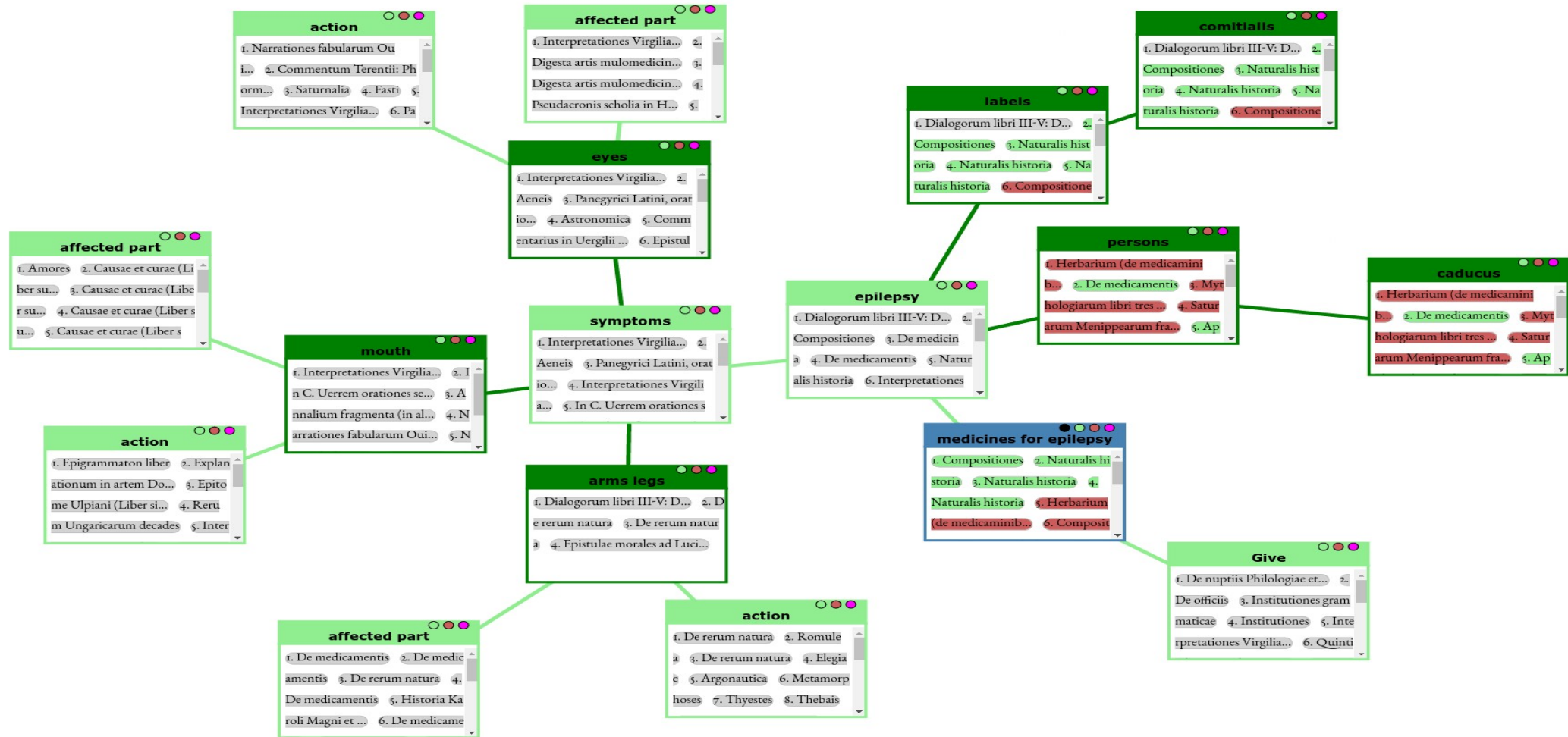
# Usage Scenario

## (Inter concept search)

# Concept – Medicines for Epilepsy



# Concept – Medicines for Epilepsy(Search results)



# Conclusion & Future Work

# Limitations of current system

- ▼ Unable to filter time, genre, author etc due to *bad metadata*.
- ▼ Ability to define maximum size between two words at sentence level.  
Defining word distance not possible due to backend.

# Conclusion & Future Work

- ▼ Visualization system turns away from being a “working black box”
- ▼ Search system able to search thousands of word forms
- ▼ Various word form combinations considered and ranked
- ▼ Step towards extracting more and accurate results (completeness)
- ▼ Iterative improvement of results (**steering the search process**)
- ▼ Phrase reuse detection
- ▼ Visual Query environment (concept level)
- ▼ Applicable to other domains in future

# References

[CJBWS<sub>15</sub>] M. F. Cheema, S. Jänicke, J. Blumenstein, C. Weilbach and G. Scheuermann. *Modeling Concepts to Improve the Search Capabilities on Ancient Corpora* in Proceedings of the Digital Humanities 2015

[JBRZS<sub>15</sub>] S. Jänicke, J. Blumenstein, M. Rücker, D. Zeckzer and G. Scheuermann (2015). Visualizing the Results of Search Queries on Ancient Text Corpora with Tag Pies. To appear in: Digital Humanities Quarterly, 2015

[CJBS<sub>16</sub>] M. F. Cheema, S. Jänicke, J. Blumenstein and G. Scheuermann. (2016). *A Directed Concept Search Environment to Visually Explore Texts Related to User-defined Concept Models*. In Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications ISBN 978-989-758-175-5, pages 72-83. DOI: 10.5220/0005727400720083

Thank You for  
listening!

Questions / Suggestions / Feedback