
A Continuous Process From Texts to Ontology

My Thao TANG and Yannick TOUSSAINT



Outline

- Motivation and Objectives
- General Process
- From Texts to Semantic Wiki
- Building the Core of the Ontology using FCA and RCA
- Enrichment and Validation: Change in The Ontology
- Conclusion and Future Works



Motivation

- Information is continuously evolving, ontologies should also be continuously updated. Changes on ontologies can lead to updates or adaptations in the web content.

Contents and knowledge co-evolve.

- Split ontology building from texts in several subtasks leading to loss traceability between sources of knowledge and the ontology.



Objectives

- To build more complex knowledge from “simple” units.
- To assist and improve the continuous transformations of content into knowledge and its annotations in texts.
- ✓ Keeping the link between knowledge and resources.
- ✓ Human-machine collaboration should be the key to ensure the co-evolution of content and knowledge.

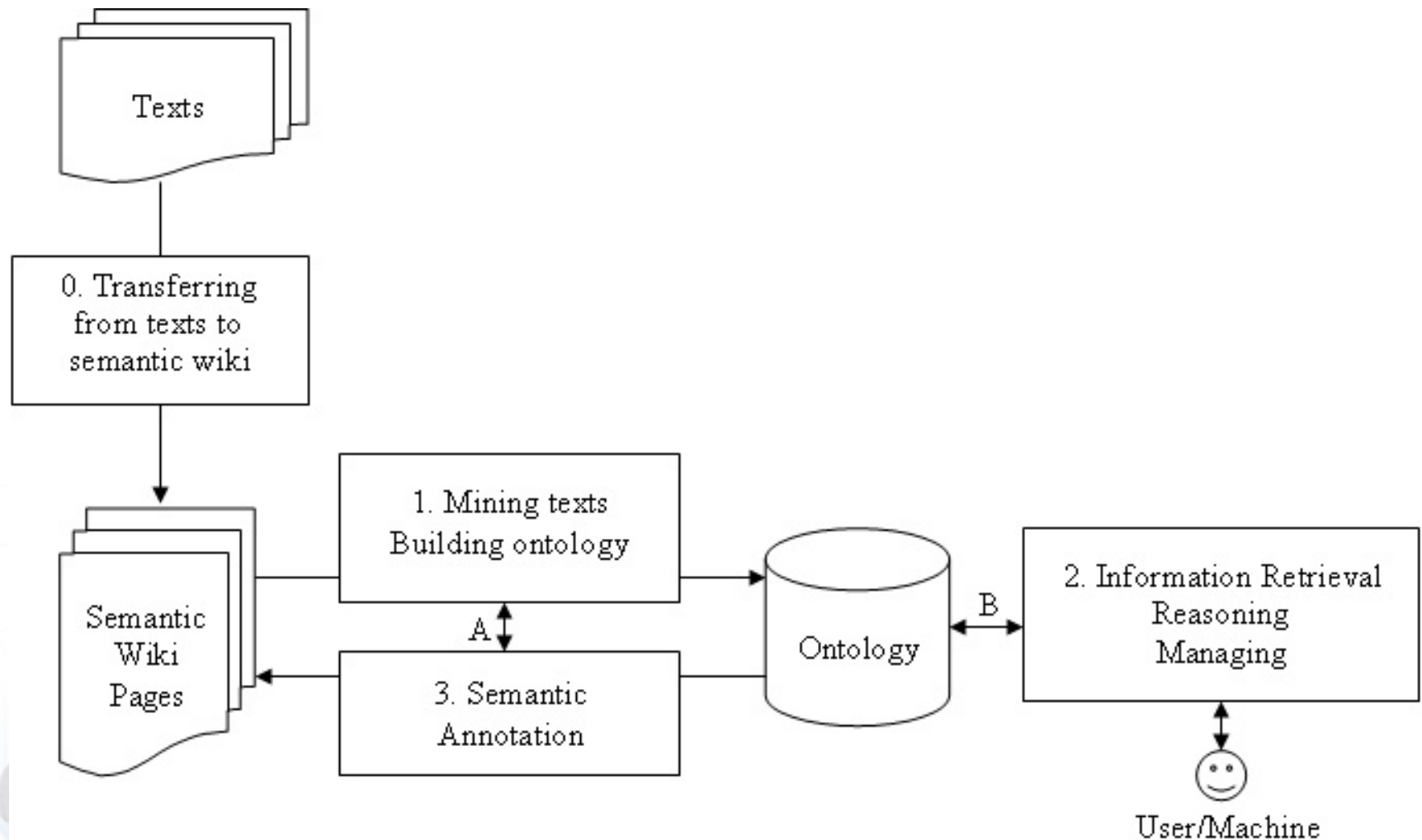


Outline

- Motivation and Objectives
- General Process
- From Texts to Semantic Wiki
- Building the Core of the Ontology using FCA and RCA
- Enrichment and Validation: Change in The Ontology
- Conclusion and Future Works



General Process



Outline

- Motivation and Objectives
- General Process
- From Texts to Semantic Wiki
- Building the Core of the Ontology using FCA and RCA
- Enrichment and Validation: Change in The Ontology
- Conclusion and Future Works



Semantic Wiki (1)

Artifacts:

- Entities
- Properties
- Relations
- Categories/subcategories



Semantic Wiki (2)

Harry Potter and the Philosopher's Stone

This movie was directed by **Chris Columbus** and produced by **David Heyman** and **J. K. Rowling**. It was starred by **Daniel Radcliffe** and **Emma Watson**. As a **Fantasy** movie, it's loved by most of children.

Running time: **152min**

Category: **Film**

Editing Harry Potter and the Philosopher's Stone



This movie was directed by `[[directedBy::Chris Columbus]]` and produced by `[[producedBy:: David Heyman]]` and `[[producedBy::J. K. Rowling]]`. It was starred by `[[Starring::Daniel Radcliffe]]` and `[[Starring::Emma Watson]]`. As a `[[hasGenre::Fantasy]]` movie, it's loved by most of children.

Running time: `[[Duration::152min]]`
`[[Category:Film]]`

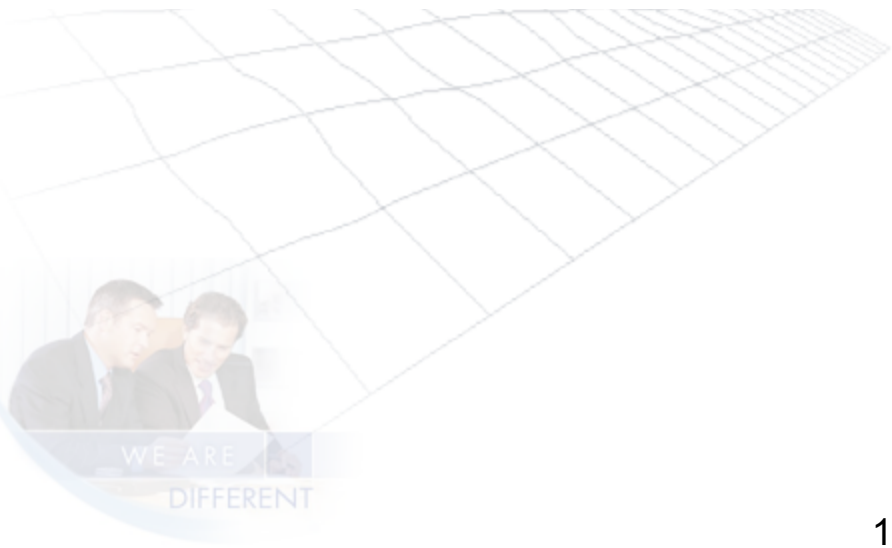
WE ARE
DIFFERENT

Example: Orphanet (1)

9888733

Fibromuscular dysplasia of the renal arteries: comparison of helical CT angiography and arteriography. Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably reveal renal artery fibromuscular dysplasia. However, because some lesions may not be shown, arteriography with pressure measurements remains the only technique that can assess the physiologic significance of the dysplasia.

Category: Text



Example: Orphanet (2)

9888733

Fibromuscular dysplasia of the renal arteries: comparison of helical CT angiography and arteriography. Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably reveal renal artery fibromuscular dysplasia. However, because some lesions may not be shown, arteriography with pressure measurements remains the only technique that can assess the physiologic significance of the dysplasia.

Category: Text

Editing 9888733



```
[[Fibromuscular dysplasia]] of the renal arteries: comparison of [[helical
CT angiography]] and [[arteriography]].
[[Helical CT angiography]], especially the combination of transverse
sections and maximum-intensity-projection reconstructions, can reliably
reveal renal artery [[fibromuscular dysplasia]]. However, because some
lesions may not be shown, [[arteriography]] with pressure measurements
remains the only technique that can assess the physiologic significance of
the [[fibromuscular dysplasia|dysplasia]].
```

```
[[Category:Text]]
__SHOWFACTBOX__
```

The text is indexed by entities

Example: Orphanet (3)

9888733

Fibromuscular dysplasia of the renal arteries: comparison of helical CT angiography and arteriography. Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably reveal renal artery fibromuscular dysplasia. However, because some lesions may not be shown, arteriography with pressure measurements remains the only technique that can assess the physiologic significance of the dysplasia.

Category: Text

Editing 9888733



```
[[Fibromuscular dysplasia]] of the renal arteries: comparison of [[helical
CT angiography]] and [[arteriography]].
[[Helical CT angiography]], especially the combination of transverse
sections and maximum-intensity-projection reconstructions, can reliably
reveal renal artery [[fibromuscular dysplasia]]. However, because some
lesions may not be shown, [[arteriography]] with pressure measurements
remains the only technique that can assess the physiologic significance of
the [[fibromuscular dysplasia|dysplasia]].
```

```
[[Category:Text]]
__SHOWFACTBOX__
```

The text is indexed by entities

Example: Orphanet (4)

9888733

Fibromuscular dysplasia of the renal arteries: comparison of helical CT angiography and arteriography. Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably **reveal** renal artery **fibromuscular dysplasia**. However, because some lesions may not be shown, arteriography with pressure measurements remains the only technique that can **assess** the physiologic significance of the **dysplasia**.

Category: Text

Extracting the text

→ To build the page for entity “Fibromuscular dysplasia”

Example: Orphanet (5)

9888733

Fibromuscular dysplasia of the renal arteries: comparison of helical CT angiography and arteriography.

Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably reveal renal artery fibromuscular dysplasia. However, because some lesions may not be shown, arteriography with pressure measurements remains the only technique that can assess the physiologic significance of the dysplasia.

Category: Text

Fibromuscular dysplasia

The part of the text
should be extracted

Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably reveal renal artery **fibromuscular dysplasia**. However, because some lesions may not be shown, arteriography with pressure measurements remains the only technique that can assess the physiologic significance of the **dysplasia**. (9888733)

Example: Orphanet (6)

Fibromuscular dysplasia

Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably **reveal** renal artery **fibromuscular dysplasia**. However, because some lesions may not be shown, **arteriography** with pressure measurements remains the only technique that can **assess** the physiologic significance of the **dysplasia**. (9888733)

Editing Fibromuscular dysplasia



```
[[Revealed by:: Helical CT angiography]], especially the combination of
transverse sections and maximum-intensity-projection reconstructions, can
reliably reveal renal artery [[fibromuscular dysplasia]]. However, because
some lesions may not be shown, [[assessed_by::arteriography]] with pressure
measurements remains the only technique that can assess the physiologic
significance of the [[Fibromuscular dysplasia|dysplasia]]
([[Coming from::9888733]])
```

The page of the entity is annotated
with its relations

Example: Orphanet (7)

The part of the text should be extracted

9588747

Segmental mediolytic arteriopathy of the splenic and hepatic arteries mimicking systemic necrotizing vasculitis. Segmental mediolytic arteriopathy a rare, noninflammatory arterial disease is fundamentally a variant of fibromuscular dysplasia. The characteristic angiographic findings of segmental mediolytic arteriopathy include the "string of beads" and microaneurysms which are indistinguishable from those of vasculitis, and the correct diagnosis can be made only after histopathologic evaluation of the arterial lesions. Thrombosis, arterial wall hemorrhage, and dissection are among the complications of segmental mediolytic arteriopathy. We describe herein a patient with segmental mediolytic arteriopathy who presented with hemoperitoneum. The patient underwent urgent surgical repair of a ruptured hepatic artery aneurysm. The postoperative visceral arteriography findings led to a clinical diagnosis of polyarteritis nodosa, and immunosuppressive therapy was initiated. This treatment was stopped as soon as the correct biopsy diagnosis of segmental mediolytic arteriopathy was obtained through outside consultation. The patient recovered without drug treatment and was spared the potentially life-threatening complications of immunosuppression.

Category: Text

Example: Orphanet (8)

Fibromuscular dysplasia

Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably reveal renal artery **fibromuscular dysplasia**. However, because some lesions may not be shown, **arteriography** with pressure measurements remains the only technique that can assess the physiologic significance of the **dysplasia**. (9888733)

Segmental mediolytic arteriopathy is fundamentally a variant of **fibromuscular dysplasia**. (9588747)

Editing Fibromuscular dysplasia



```
[[Revealed by:: Helical CT angiography]], especially the combination of
transverse sections and maximum-intensity-projection reconstructions, can
reliably reveal renal artery [[fibromuscular dysplasia]]. However, because
some lesions may not be shown, [[assessed_by::arteriography]] with pressure
measurements remains the only technique that can assess the physiologic
significance of the [[Fibromuscular dysplasia|dysplasia]].
([[Coming from::9888733]])
```

```
[[Varied by::Segmental mediolytic arteriopathy]] is fundamentally a variant
of [[fibromuscular dysplasia]]. ([[Coming from::9588747]])
```

DIFFERENT

Example: Orphanet (9)

9588747

Segmental mediolytic arteriopathy of the splenic and hepatic arteries mimicking systemic necrotizing vasculitis. Segmental mediolytic arteriopathy, a rare, noninflammatory arterial disease, is fundamentally a variant of fibromuscular dysplasia. The characteristic angiographic findings of segmental mediolytic arteriopathy include the "string of beads" and microaneurysms which are indistinguishable from those of vasculitis, and the correct diagnosis can be made only after histopathologic evaluation of the arterial lesions. Thrombosis, arterial wall hemorrhage, and dissection are among the complications of segmental mediolytic arteriopathy. We describe herein a patient with segmental mediolytic arteriopathy who presented with hemoperitoneum. The patient underwent urgent surgical repair of a ruptured hepatic artery aneurysm. The postoperative visceral arteriography findings led to a clinical diagnosis of polyarteritis nodosa, and immunosuppressive therapy was initiated. This treatment was stopped as soon as the correct biopsy diagnosis of segmental mediolytic arteriopathy was obtained through outside consultation. The patient recovered without drug treatment and was spared the potentially life-threatening complications of immunosuppression.

Category: Text

Example: Orphanet (10)

Segmental mediolytic arteriopathy

Segmental mediolytic arteriopathy of the splenic and hepatic arteries mimicking systemic necrotizing vasculitis.

Segmental mediolytic arteriopathy, a rare, noninflammatory arterial disease, is fundamentally a variant of fibromuscular dysplasia. (9588747)

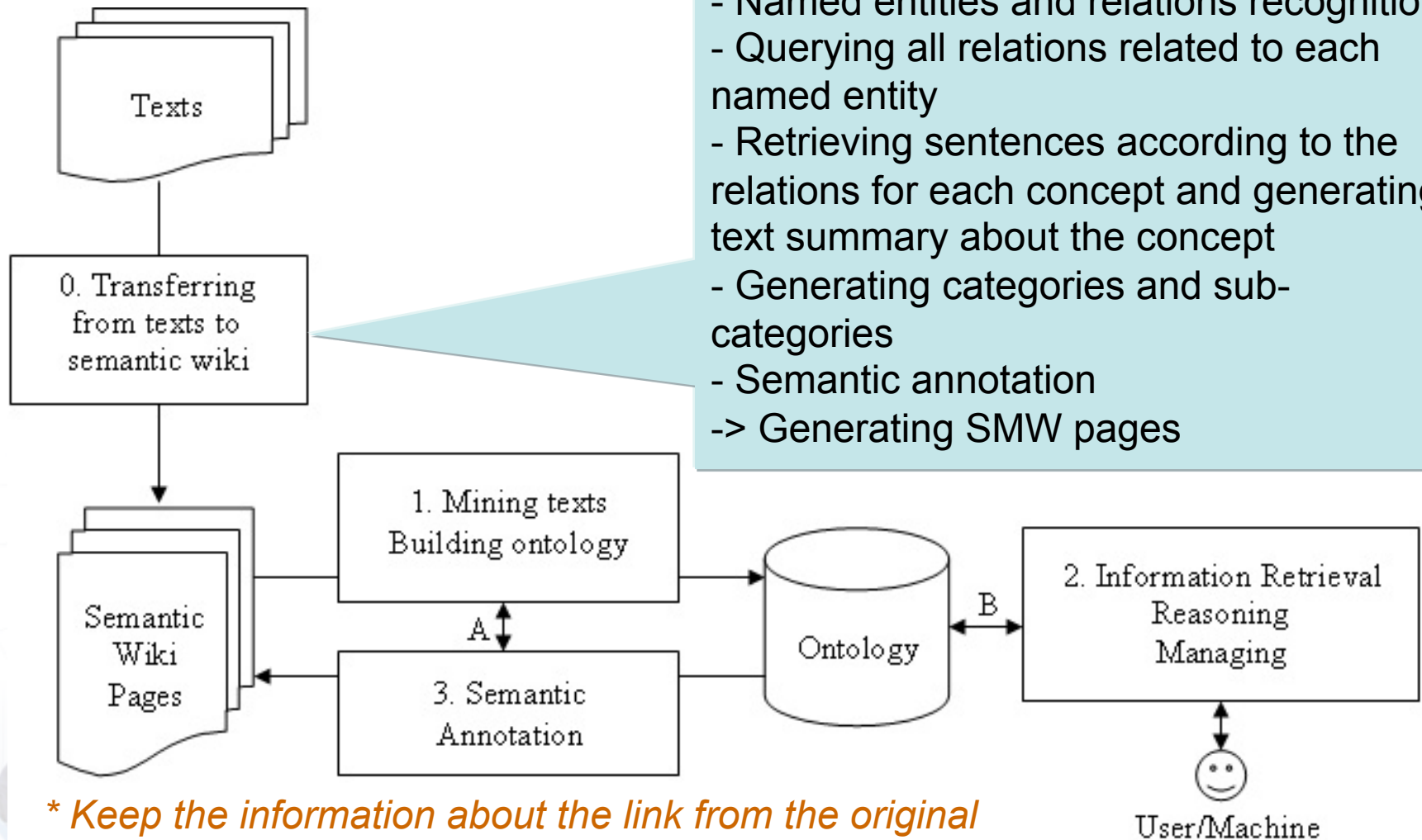
Editing Segmental mediolytic arteriopathy



```
[[Segmental mediolytic arteriopathy]] of the splenic and hepatic arteries  
mimicking [[Mimic::systemic necrotizing vasculitis]].
```

```
[[Segmental mediolytic arteriopathy]], a rare, noninflammatory [[Varied  
by::arterial disease]], is fundamentally a variant of [[varied  
by::fibromuscular dysplasia]]. ([[Coming from::9588747]])
```

Texts to SMW



** Keep the information about the link from the original texts to wiki pages*

DIFFERENT

Texts to SMW: Related works (1)

Plantec, A.; Ribaud, V. & Varma, V.

Building a semantic virtual museum: from Wiki to semantic Wiki using named entity recognition
ACM, 2009, 769-77

Method:

- Automatic Named entities recognition (NER) + Human verification
- Class recognition + Human verification
- Assisted triples generation



Texts to SMW: Related works (2)

Shang, Y.; Li, Y.; Lin, H. & Yang, Z.

Enhancing Biomedical Text Summarization Using Semantic Relation Extraction

PLoS ONE, Public Library of Science, 2011

Generating text summary for a given biomedical concept from multiple documents

Method:

- Extract semantic relations
- Select the relations most relevant to each query concept and visualize them in a graphic representation
- For relations in the relevant set, extract informative sentences that can interpret them from the document collection to generate text summary

Outline

- Motivation and Objectives
- General Process
- From Texts to Semantic Wiki
- Building the Core of the Ontology using FCA and RCA
- Enrichment and Validation: Change in The Ontology
- Conclusion and Future Works

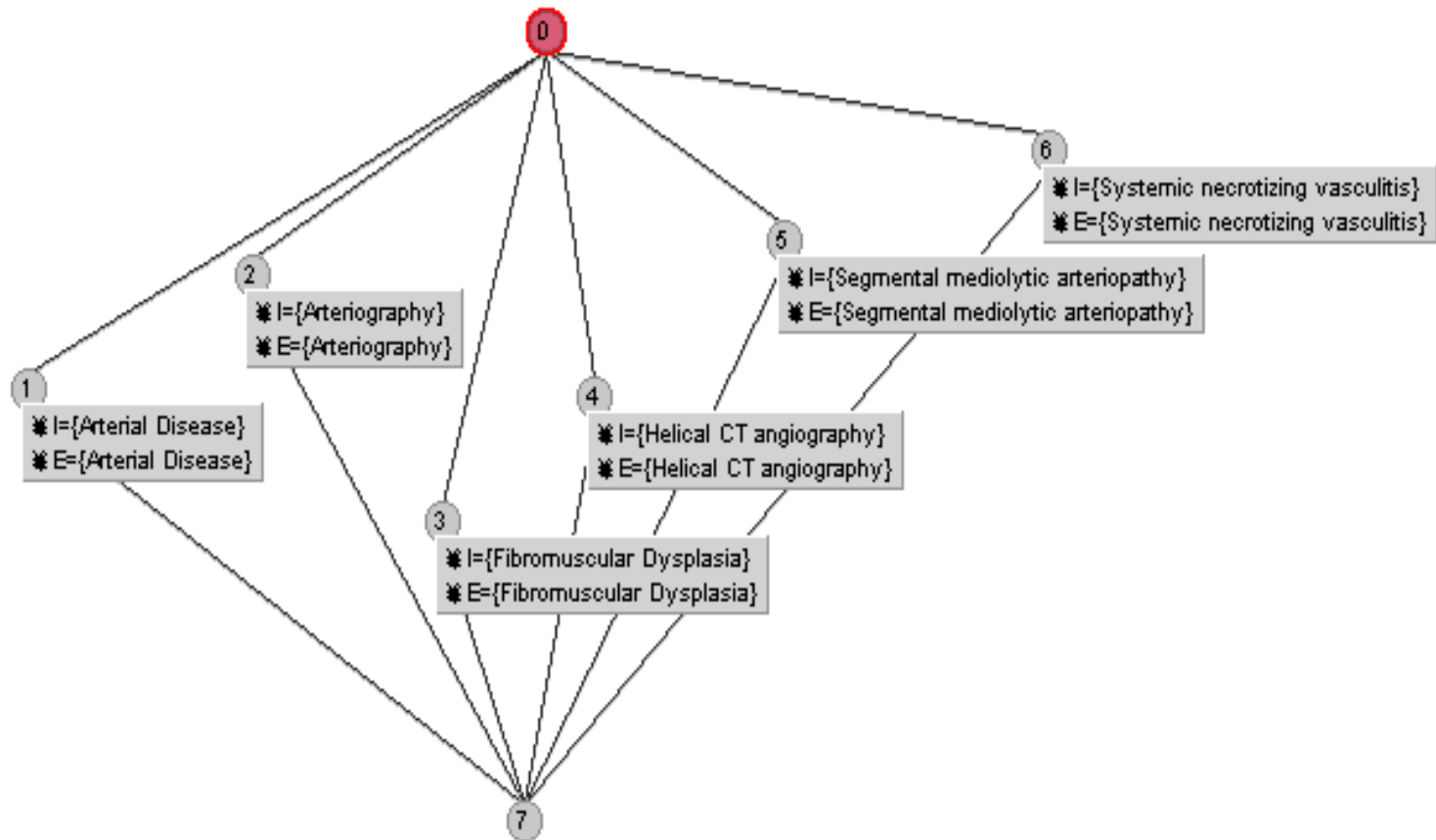


Building the Core of the Ontology using FCA and RCA: Example (1)

Object	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia	x					
Helical CT angiography		x				
Arteriography			x			
Segmental mediolytic arteriopathy				x		
Systemic necrotizing vasculitis					x	
Arterial Disease						x

The binary context of objects

Building the Core of the Ontology using FCA and RCA: Example (2)



The initial lattice



Building the Core of the Ontology using FCA and RCA: Example (3)

	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
access						
Fibromuscular Dysplasia			x			
Helical CT angiography						
Arteriography	x					
Segmental mediolytic arteriopathy						
Systemic necrotizing vasculitis						
Arterial Disease						

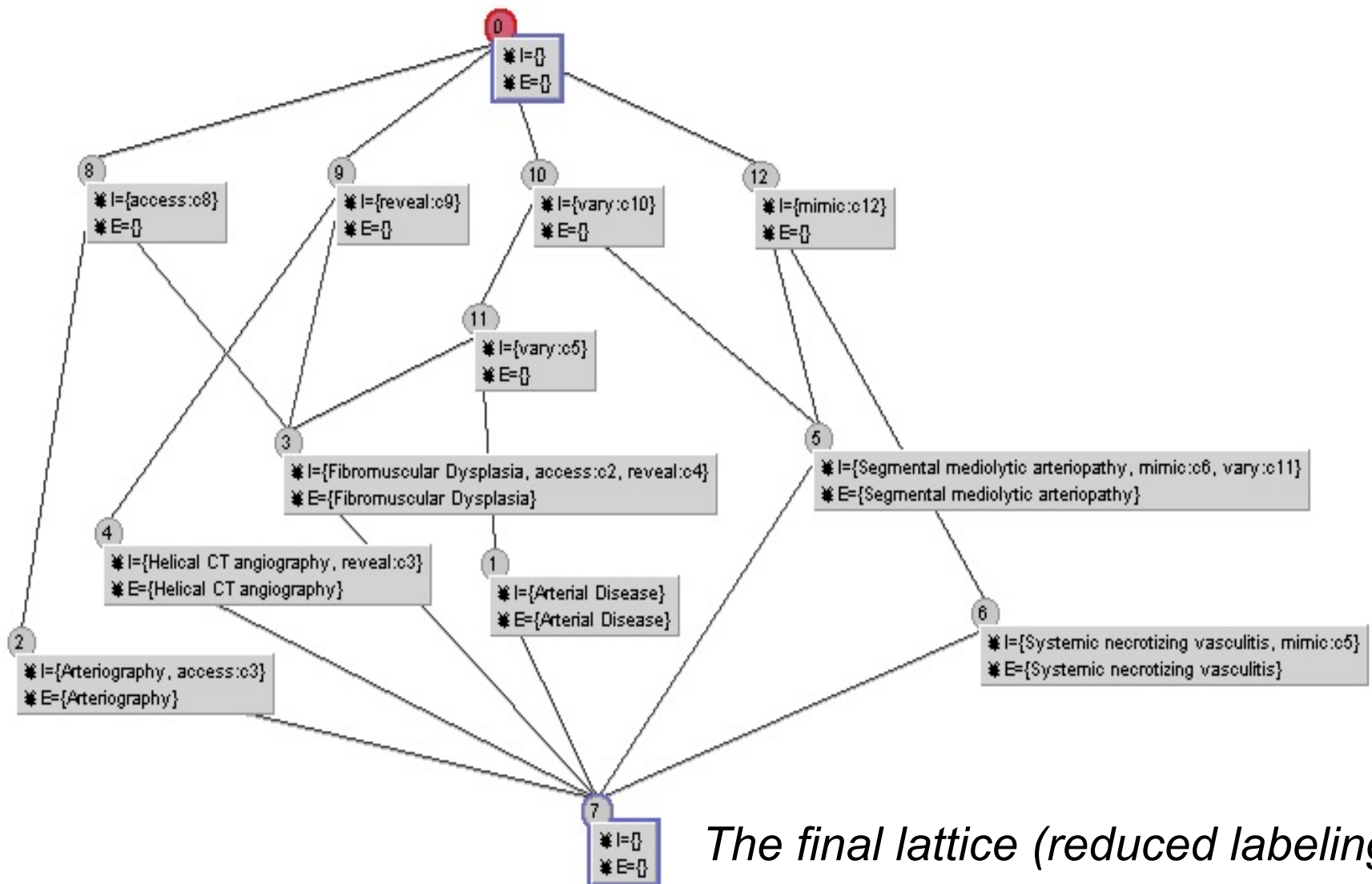
	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
reveal						
Fibromuscular Dysplasia		x				
Helical CT angiography	x					
Arteriography						
Segmental mediolytic arteriopathy						
Systemic necrotizing vasculitis						
Arterial Disease						

	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
vary						
Fibromuscular Dysplasia				x		
Helical CT angiography						
Arteriography						
Segmental mediolytic arteriopathy	x					x
Systemic necrotizing vasculitis						
Arterial Disease			x			

	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
mimic						
Fibromuscular Dysplasia						
Helical CT angiography						
Arteriography						
Segmental mediolytic arteriopathy					x	
Systemic necrotizing vasculitis			x			
Arterial Disease						

The relational contexts

Building the Core of the Ontology using FCA and RCA: Example (4)

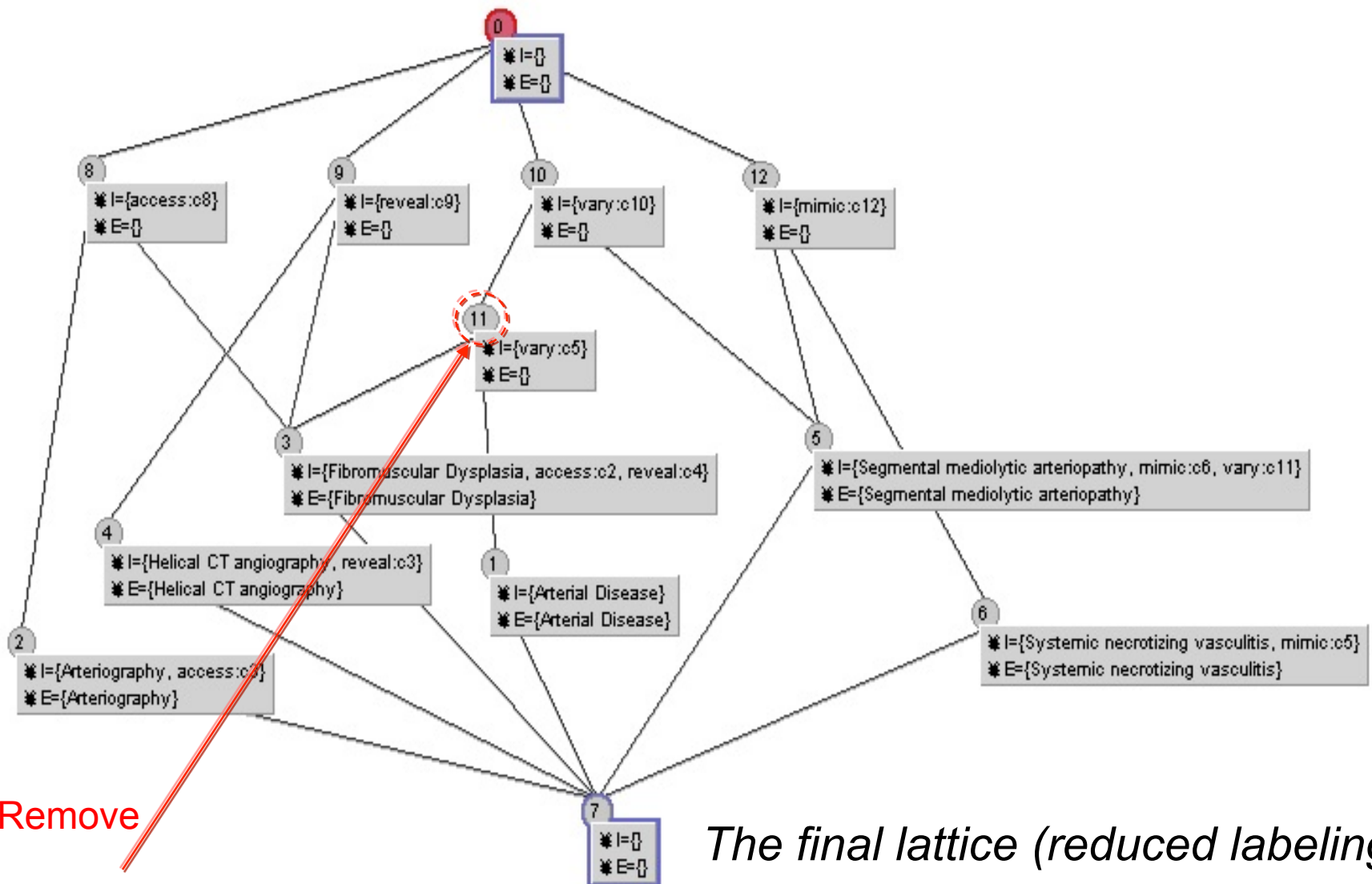


Outline

- Motivation and Objectives
- General Process
- From Texts to Semantic Wiki
- Building the Core of the Ontology using FCA and RCA
- Enrichment and Validation: Change in The Ontology
- Conclusion and Future Works



Enrichment and Validation: Example (1)



A user wishes to make a change

Enrichment and Validation: Ontology Evolution

“Ontology Evolution is the timely adaptation of an ontology to the arisen changes and the consistent propagation of these changes to dependent artefacts.” ([Stojanovic 2004])

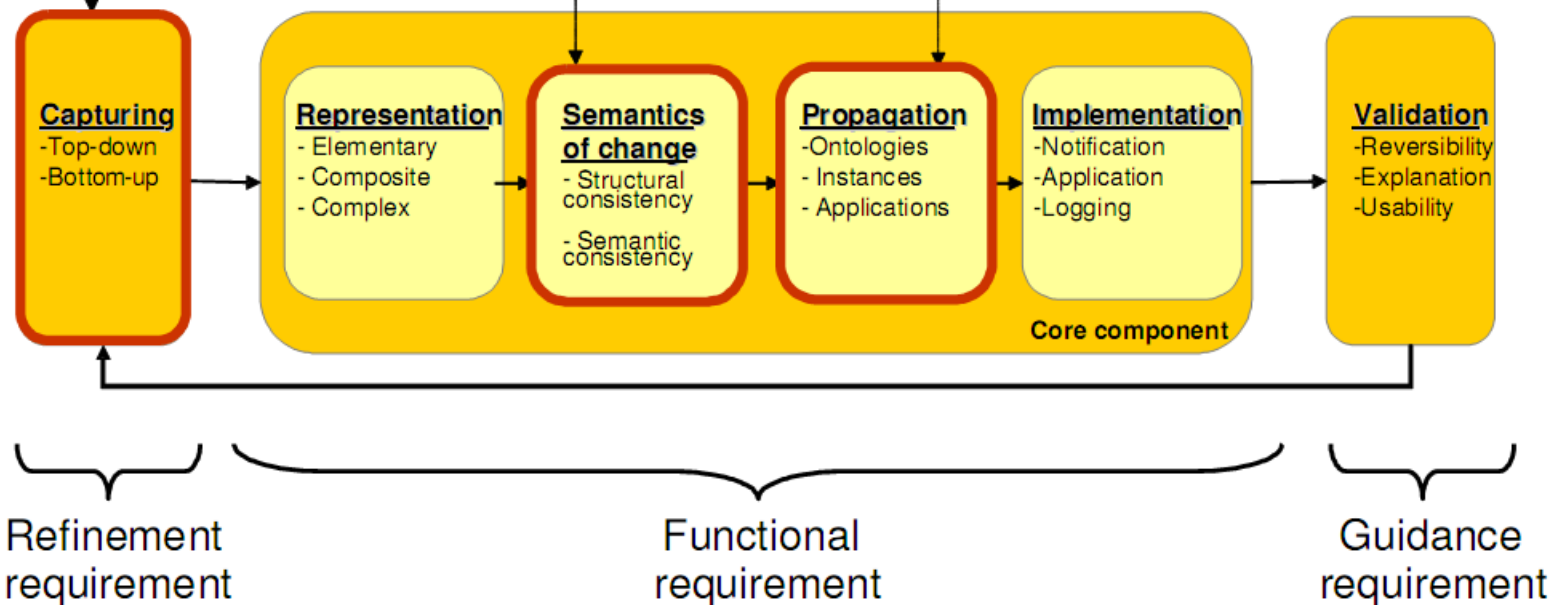


Enrichment and Validation: Ontology Evolution Process

How to discover a change?

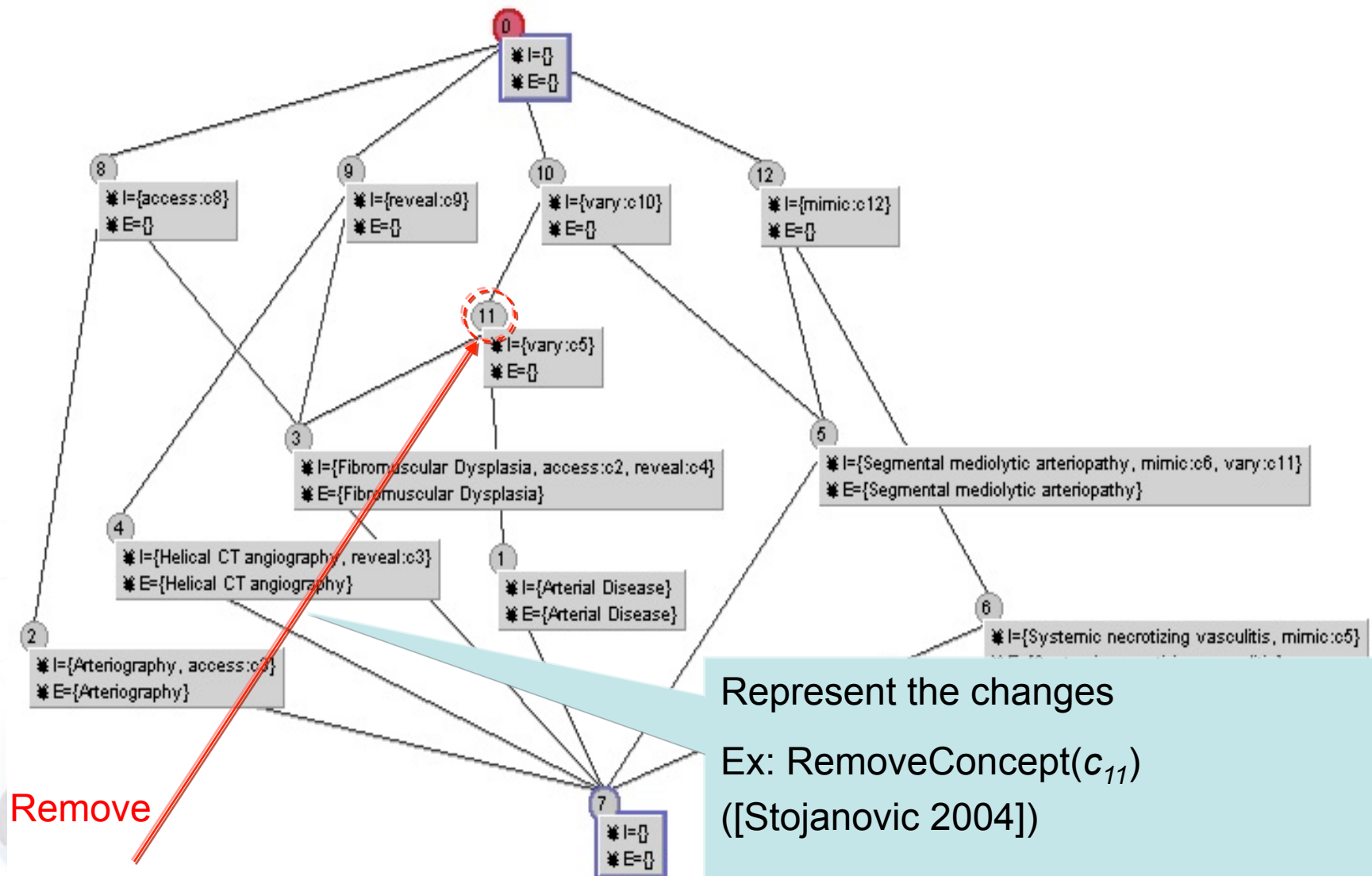
How to resolve a change?

How to ensure the consistency?



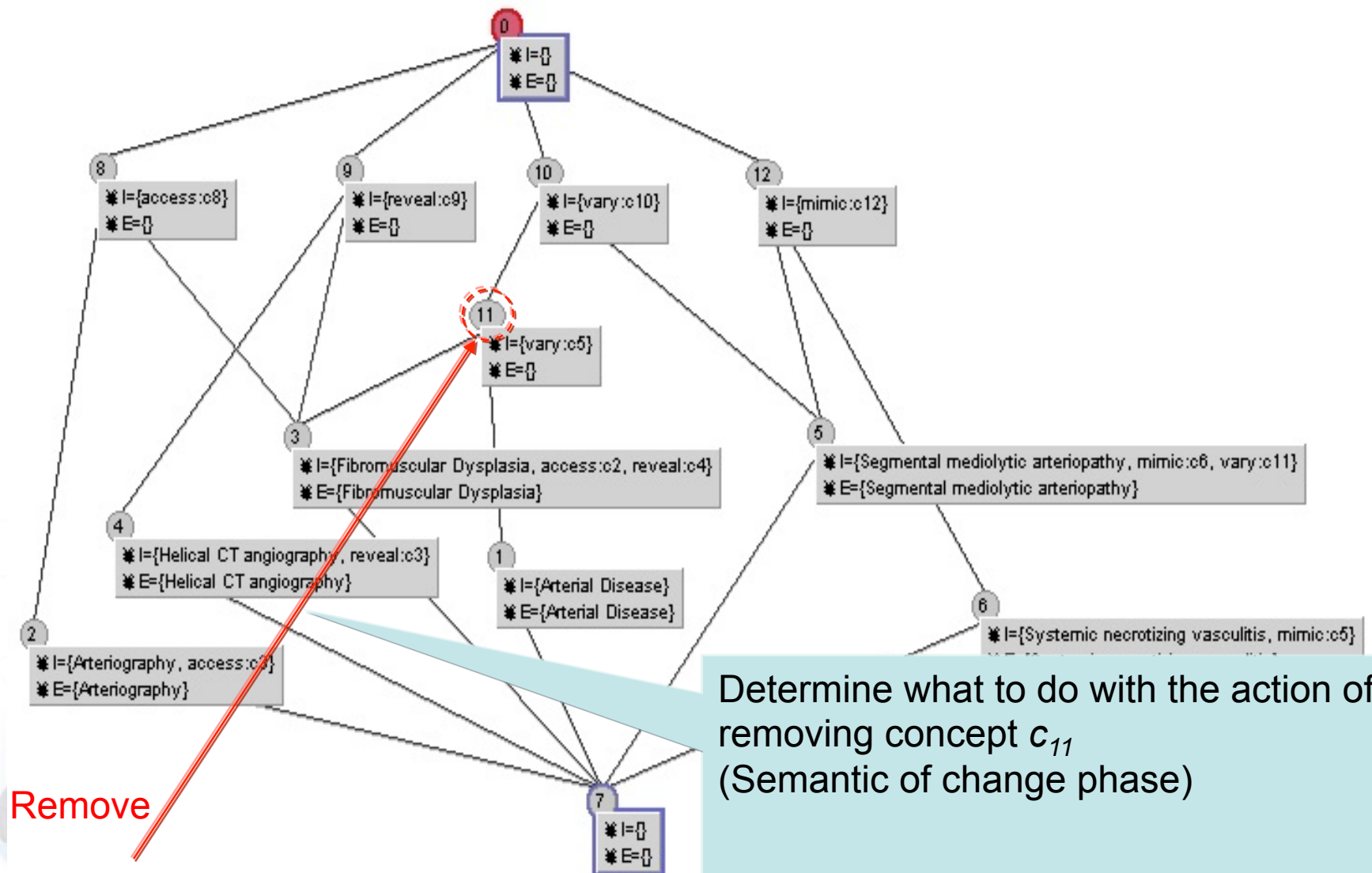
[Stojanovic et al. 2004]

Enrichment and Validation: Example (2)

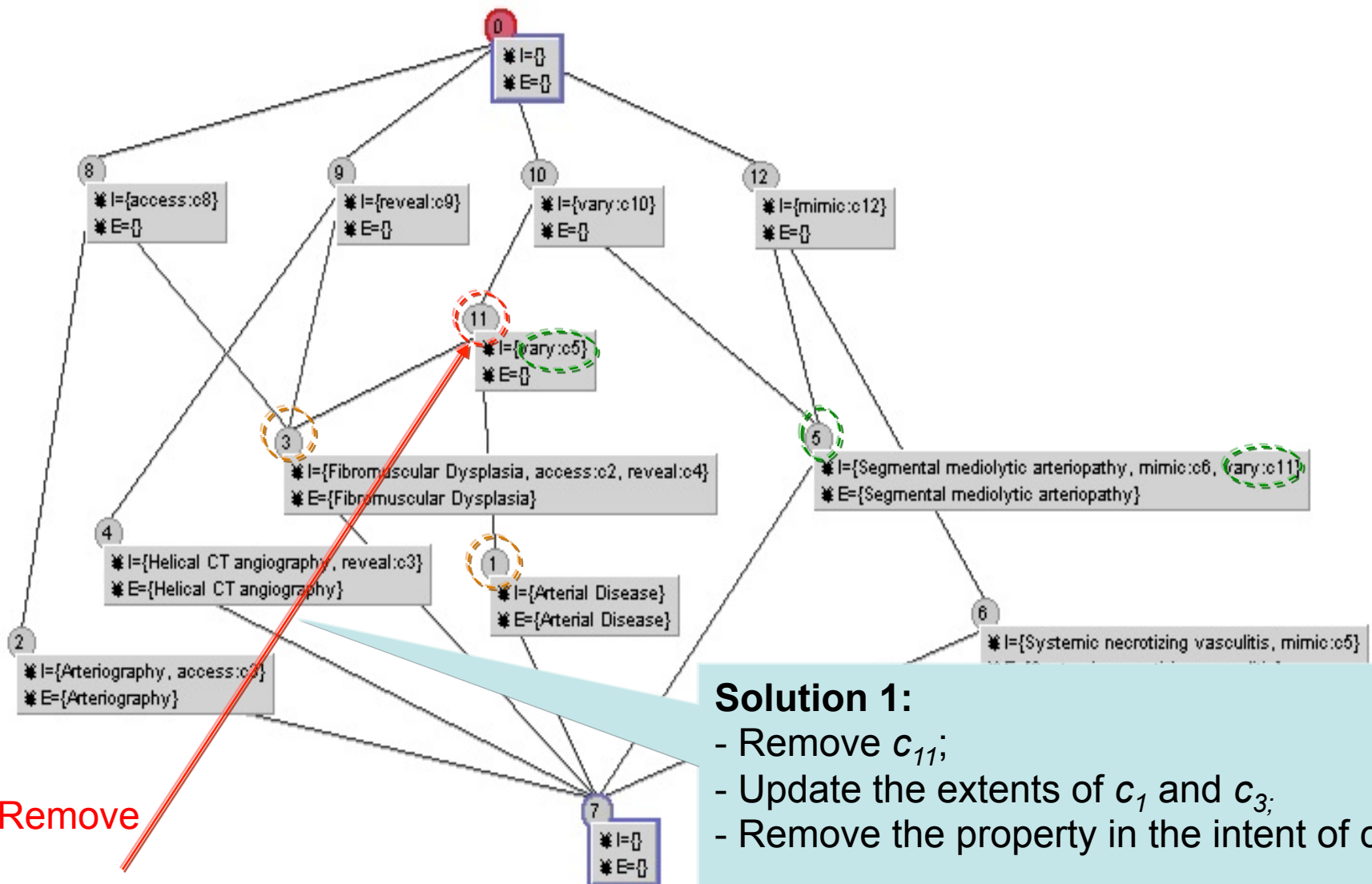


A user wishes to make a change

Enrichment and Validation: Example (3)



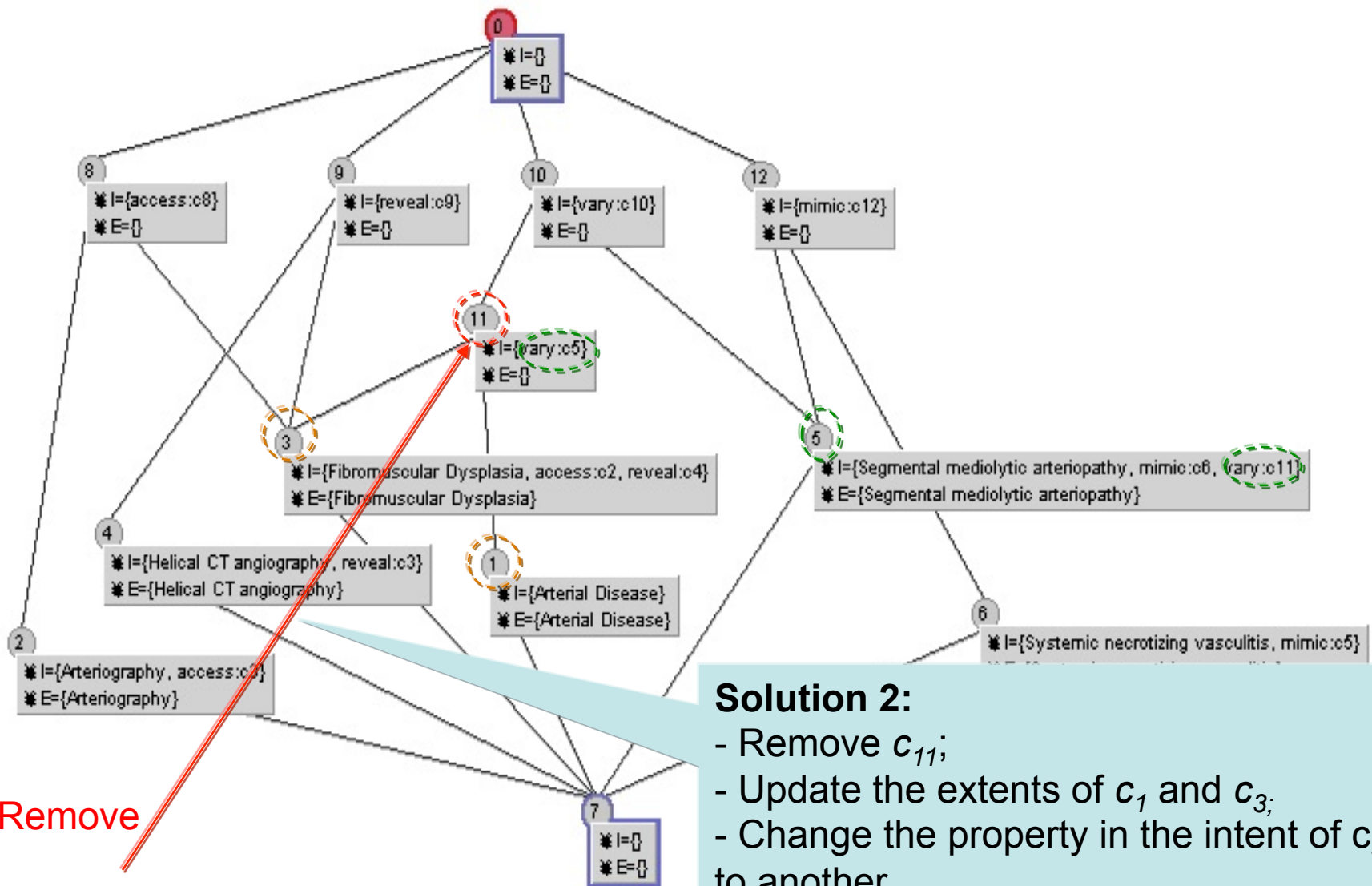
Enrichment and Validation: Example (4)



Remove

A user wishes to make a change

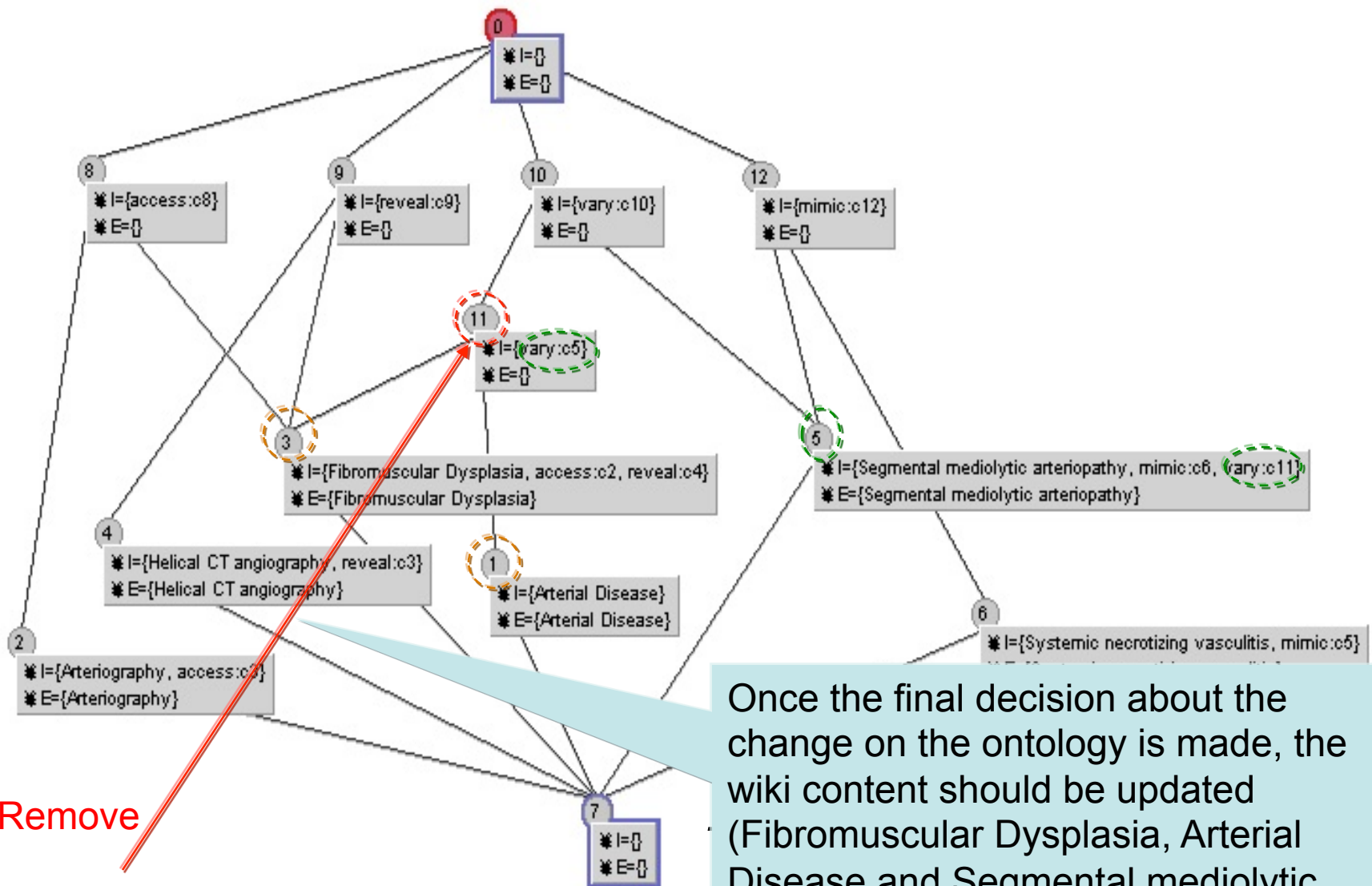
Enrichment and Validation: Example (5)



Remove

A user wishes to make a change

Enrichment and Validation: Example (6)



A user wishes to make a change

Enrichment and Validation: Example (7)

	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease	vary:c0	vary:c1	vary:c2	vary:c3	vary:c4	vary:c5	vary:c6	vary:c7	vary:c8	vary:c9	vary:c10	vary:c11	...	mimic:c6	...
var																					
Fibromuscular Dysplasia	x											x									
Helical CT angiography		x																			
Arteriography			x																		
Segmental mediolytic arteriopathy				x														x		x	
Systemic necrotizing vasculitis					x																
Arterial Disease						x						x									

C_{11}

C_5

Enrichment and Validation: Example (8)

access	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia			x			
Helical CT angiography						
Arteriography	x					
Segmental mediolytic arteriopathy						
Systemic necrotizing vasculitis						
Arterial Disease						

reveal	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia		x				
Helical CT angiography	x					
Arteriography						
Segmental mediolytic arteriopathy						
Systemic necrotizing vasculitis						
Arterial Disease						

vary	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia			x			
Helical CT angiography						
Arteriography						
Segmental mediolytic arteriopathy	x					x
Systemic necrotizing vasculitis						
Arterial Disease			x			

mimic	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia						
Helical CT angiography						
Arteriography						
Segmental mediolytic arteriopathy					x	
Systemic necrotizing vasculitis			x			
Arterial Disease						

The relational contexts

Enrichment and Validation: Example (9)

access	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia			x			
Helical CT angiography						
Arteriography	x					
Segmental mediolytic arteriopathy						
Systemic necrotizing vasculitis						
Arterial Disease						

reveal	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia		x				
Helical CT angiography	x					
Arteriography						
Segmental mediolytic arteriopathy						
Systemic necrotizing vasculitis						
Arterial Disease						

vary	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia						
Helical CT angiography						
Arteriography						
Segmental mediolytic arteriopathy						
Systemic necrotizing vasculitis						
Arterial Disease						

mimic	Fibromuscular Dysplasia	Helical CT angiography	Arteriography	Segmental mediolytic arteriopathy	Systemic necrotizing vasculitis	Arterial Disease
Fibromuscular Dysplasia						
Helical CT angiography						
Arteriography						
Segmental mediolytic arteriopathy				x		
Systemic necrotizing vasculitis			x			
Arterial Disease						

*The updated
relational contexts*

Enrichment and Validation: Example (10)

Fibromuscular dysplasia

Helical CT angiography, especially the combination of transverse sections and maximum-intensity-projection reconstructions, can reliably reveal renal artery **fibromuscular dysplasia**. However, because some lesions may not be shown, arteriography with pressure measurements remains the only technique that can assess the physiologic significance of the **dysplasia**. (9888733)

Segmental mediolytic arteriopathy is fundamentally a variant of **fibromuscular dysplasia**. (9588747)

Editing Fibromuscular dysplasia



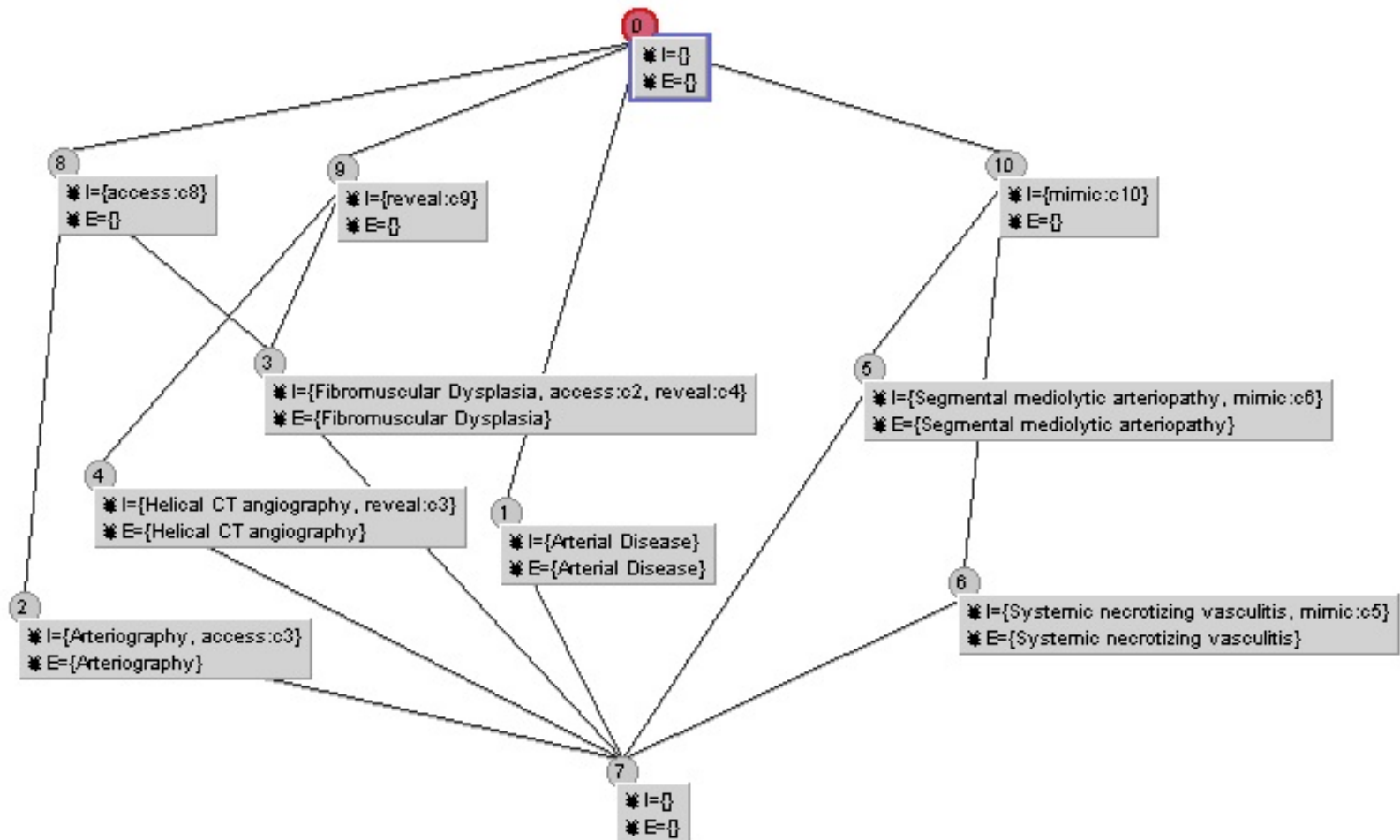
```
[[Revealed by:: Helical CT angiography]], especially the combination of  
transverse sections and maximum-intensity-projection reconstructions, can  
reliably reveal renal artery [[fibromuscular dysplasia]]. However, because  
some lesions may not be shown, [[assessed_by::arteriography]] with pressure  
measurements remains the only technique that can assess the physiologic  
significance of the [[Fibromuscular dysplasia|dysplasia]].  
([[Coming from::9888733]])
```

```
[[Varied by::Segmental mediolytic arteriopathy]] is fundamentally a variant  
of [[fibromuscular dysplasia]]. ([[Coming from::9588747]])
```

WE ARE
DIFFERENT

The corresponding part should be updated

Enrichment and Validation: Example (11)



The rebuilt lattice after the data were updated (Solution 1)

DIFFERENT

Outline

- Motivation and Objectives
- General Process
- From Texts to Semantic Wiki
- Building the Core of the Ontology using FCA and RCA
- Enrichment and Validation: Change in The Ontology
- Conclusion and Future Works

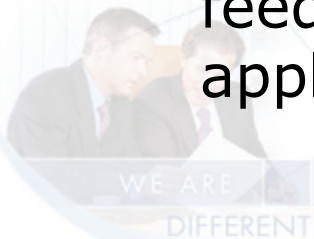


Conclusion and Future Works (1)

- **Conclusion:**

The global architecture of a continuous process from texts to ontology:

- Texts are encoded and semantically annotated in a semantic wiki.
- The core process is a conceptualization tool which uses formal concept analysis.
- The system is able to propose strategies to get feedback from an expert or from a task of an application.



Conclusion and Future Works (2)

- **Future Works:**

- Define the set of operations, define their semantics and implement them in order to preserve the link between semantic annotations and the ontology.



Thank You!

