Generalized Semantic-to-Document Derivation
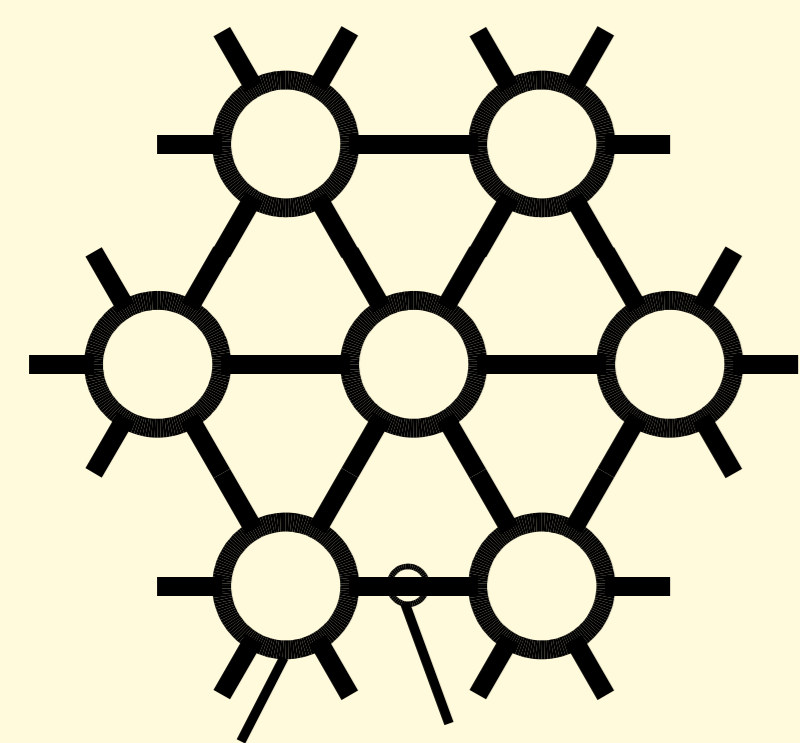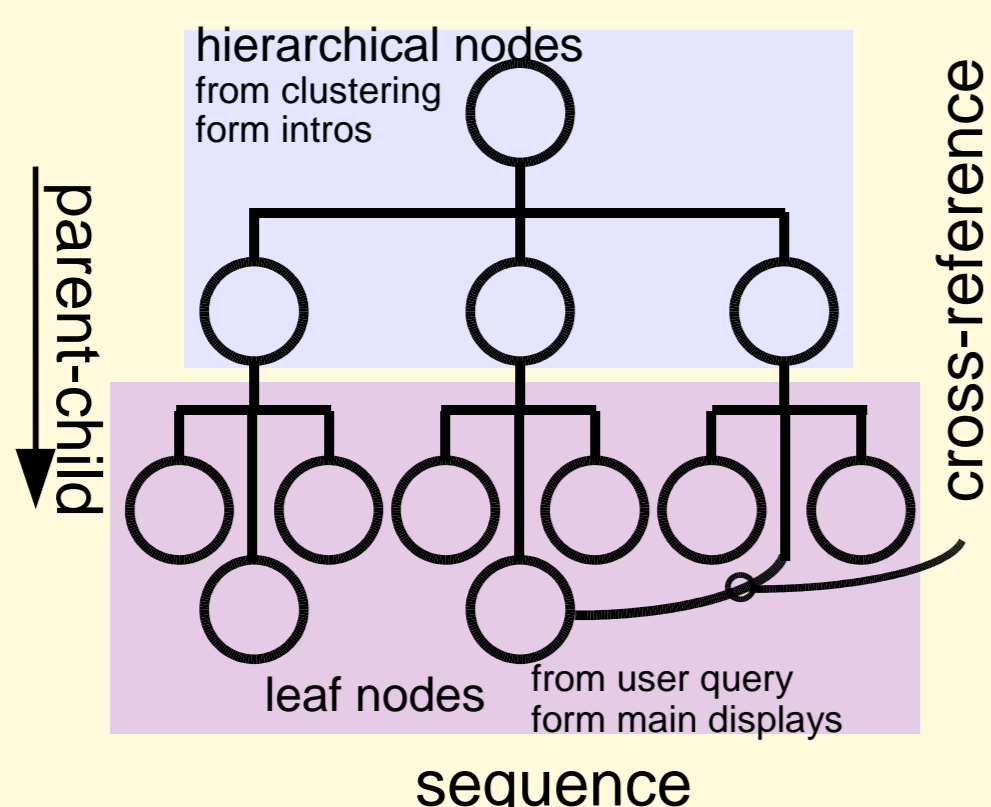
Lloyd Rutledge, Martin Alberink, Lynda Hardman and Mettina Veenstra

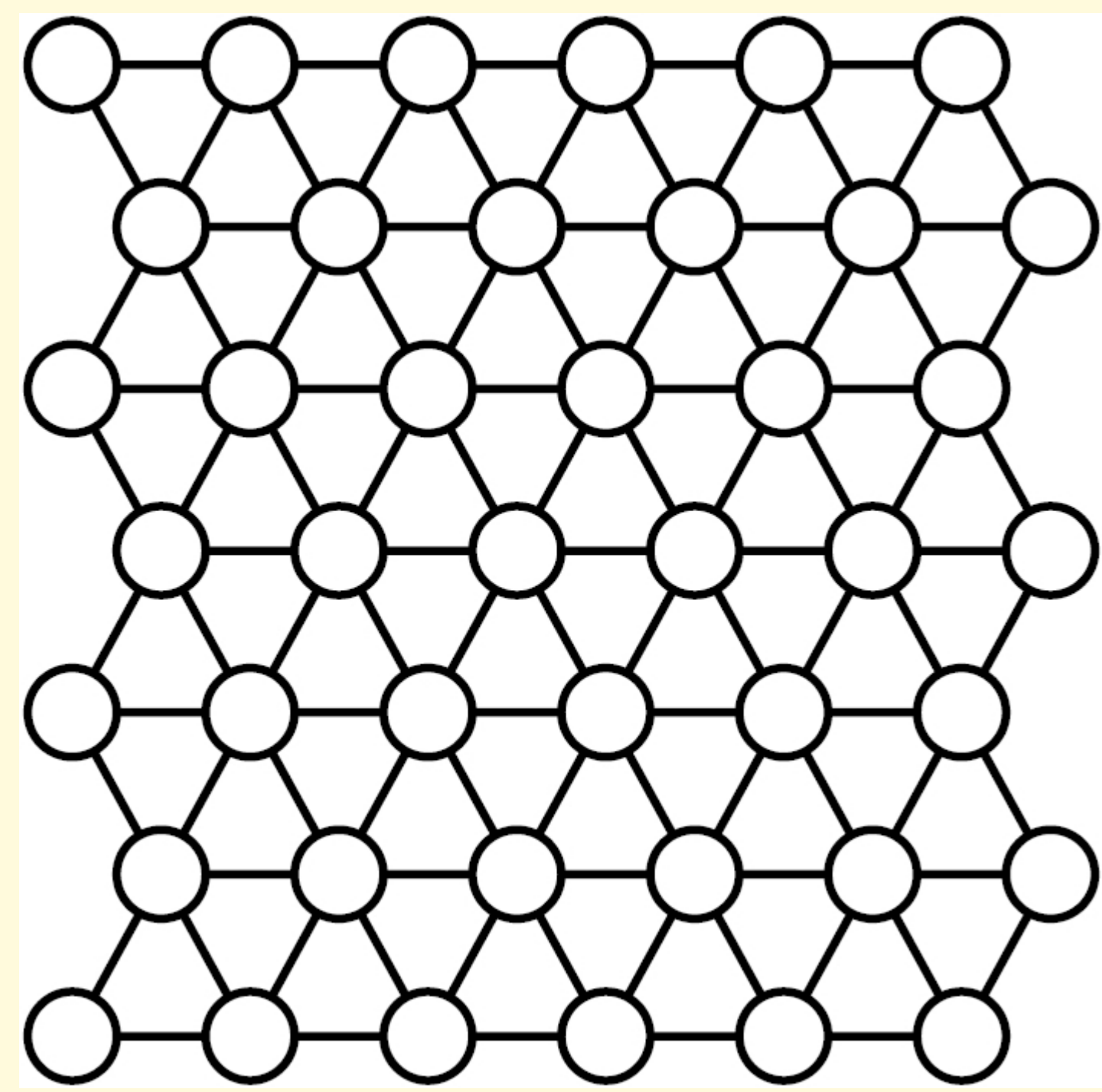## From this ...   ... to this

hierarchical nodes from clustering form intros

parent-child

cross-reference

leaf nodes   from user query form main displays

sequence

URI (RDF Resource)   RDF predicate between resources in triple

Proximity

Semantic Proximity
- Concepts in Knowledge base interrelate
- Interconcept relation has proximity (distance)

Document Proximity
- Document structure conveys conceptual proximity
- Goal is to have generate documents reflect underlying semantic proximity
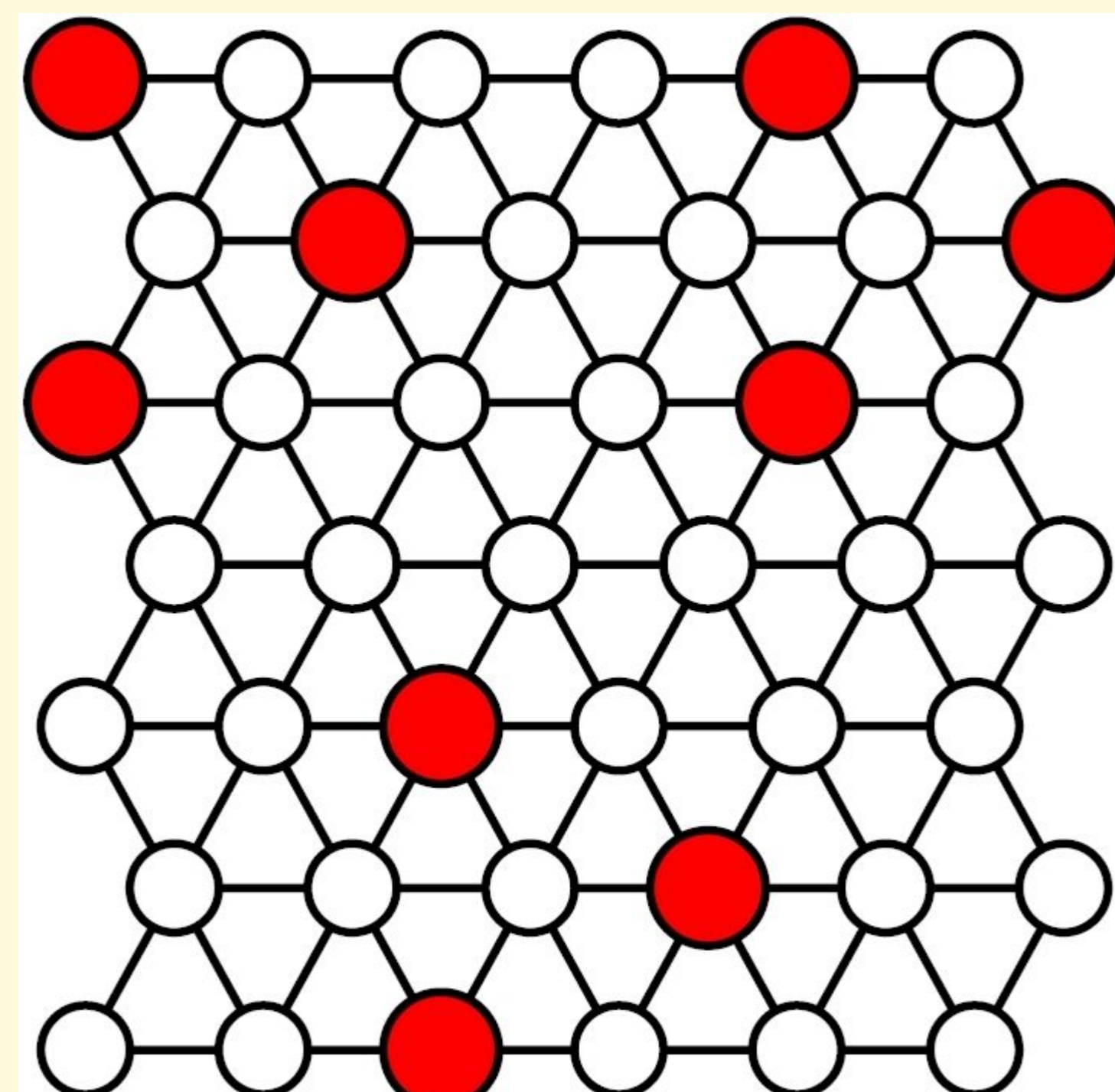
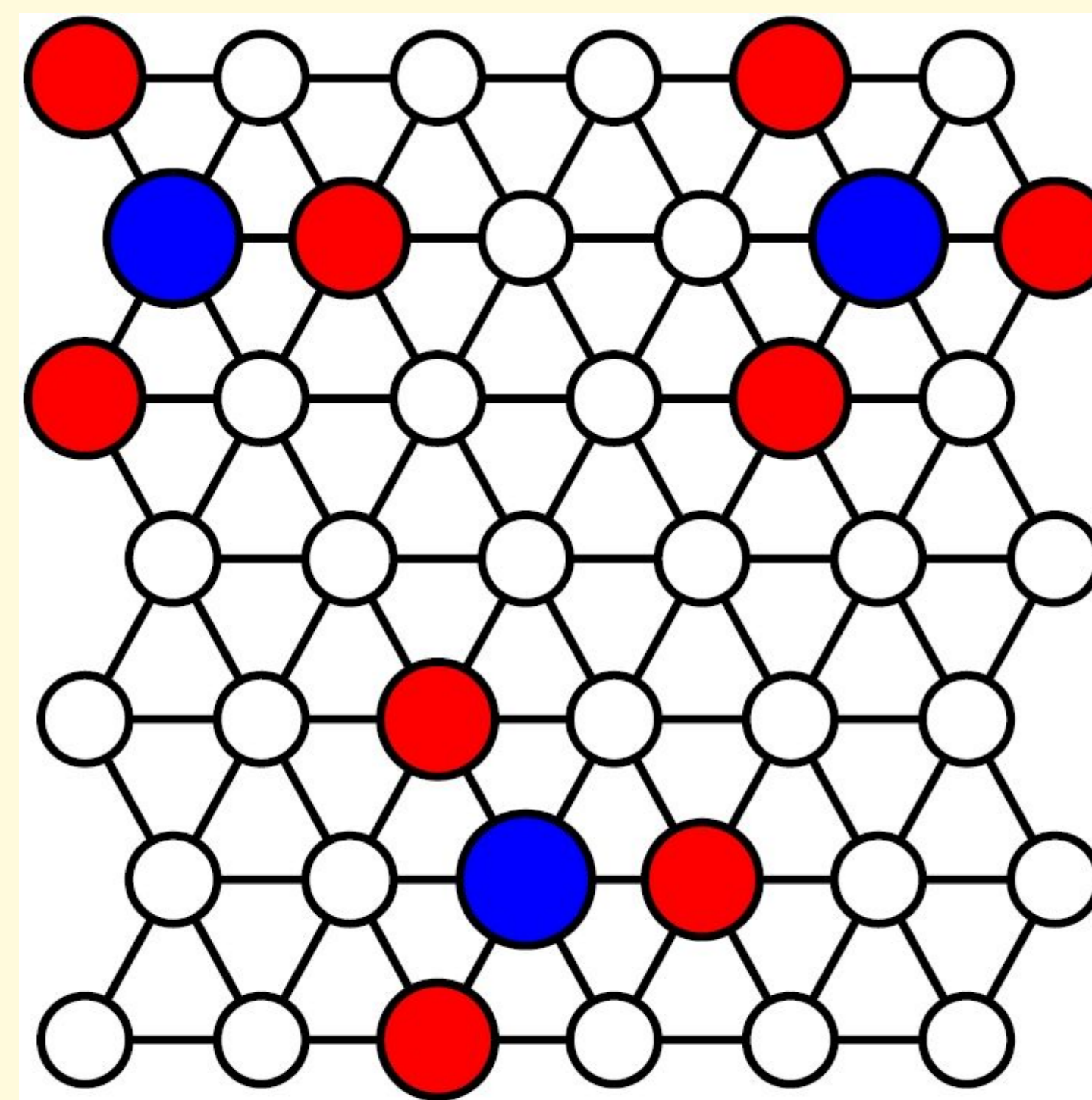## Generated Presentation

Phases

Semantics as Links

Concepts related to each other in semantic network

Topic Selection (Search)

Particular concepts selected as relevant to topic of desired document

Document Structure

Hierarchy
- Clustering into groups
- Siblings are proximate
- Each cluster node is section intro
- Uniformity

Sequence
- From uniform numeric property
- Adjacency is proximity
- To maintiain proximity (segue)

Cross-reference
- Also conveys proximity

Numeric Clustering

Clustering

Concepts found for groupings of selected concepts

Document Structure
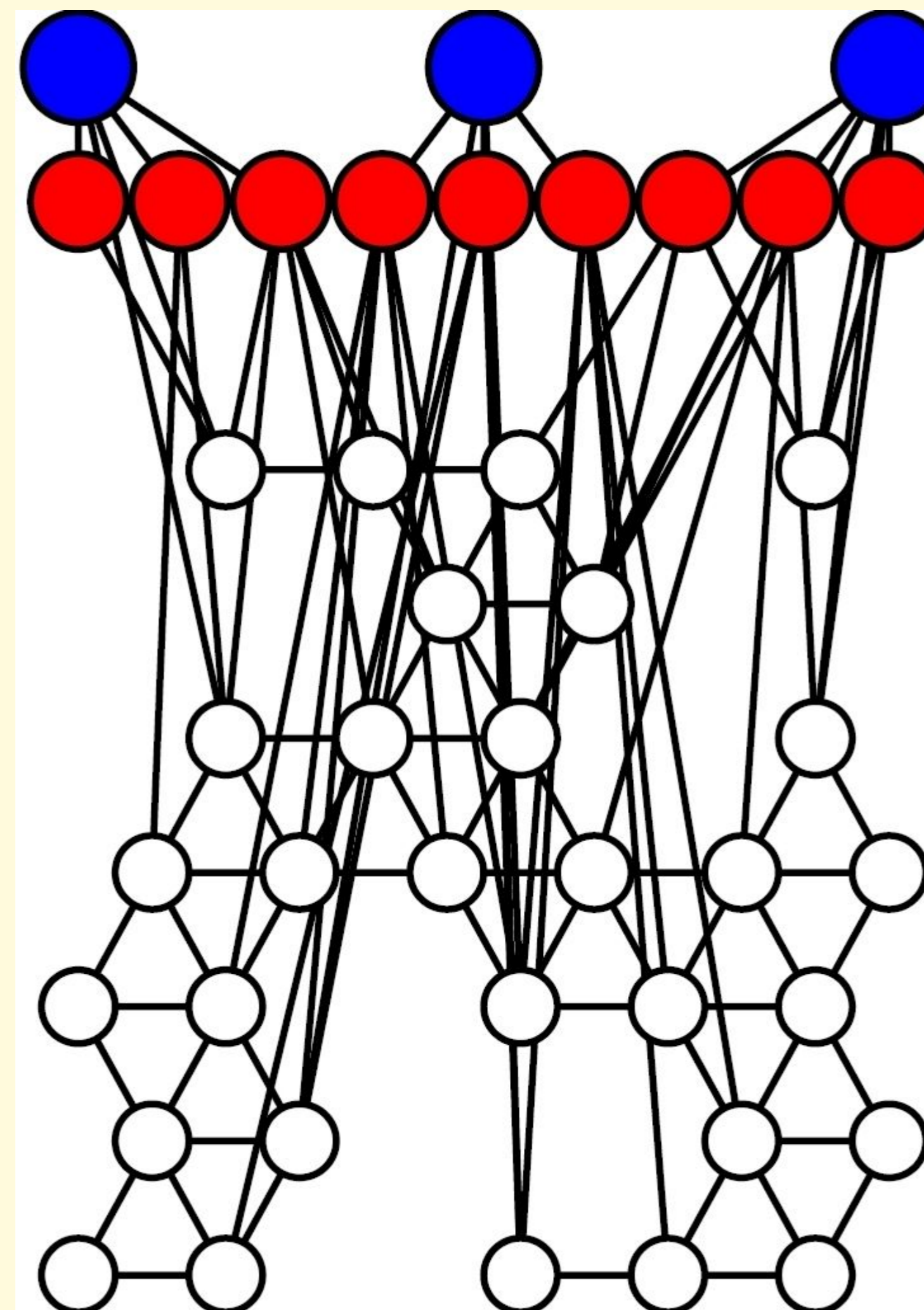
Cluster concepts form basis for hierarchy of generated document

Means of Clustering

Common Properties
- Algorithm: Concept lattice
- Previous work: Topia

Link Hubs
- Algorithm: recursive concept lattice
- Superset of property-based

Numeric (Axis)
- Algorithm: K-means
- As with astronomy

Generalized Proximity
- Proximity table
- Best-of search for corresponding doc struc

With Uniformity

Concept Lattices

Example: Rijksmuseum Collection, query "water"

Resulting Document Structure

Properties

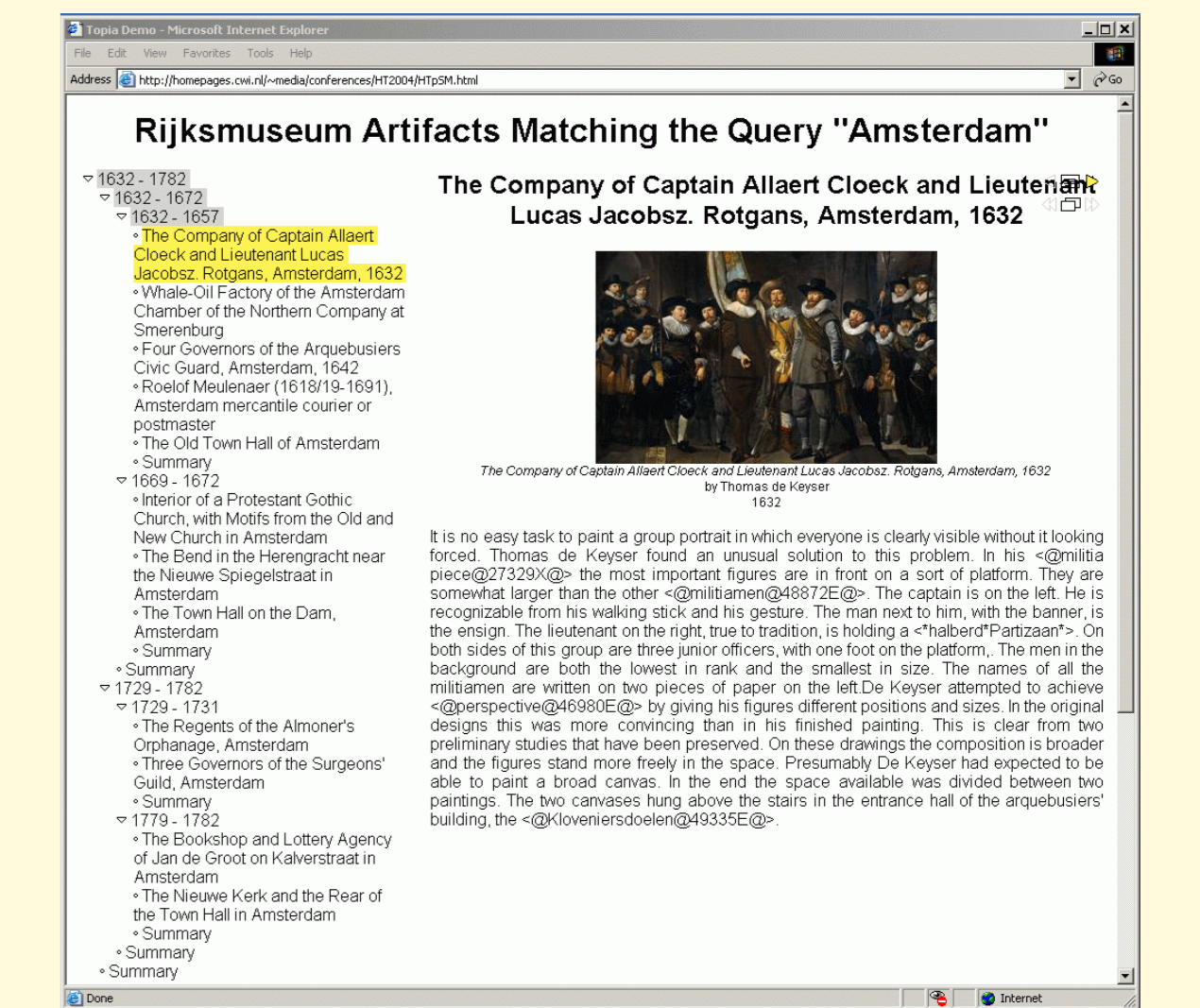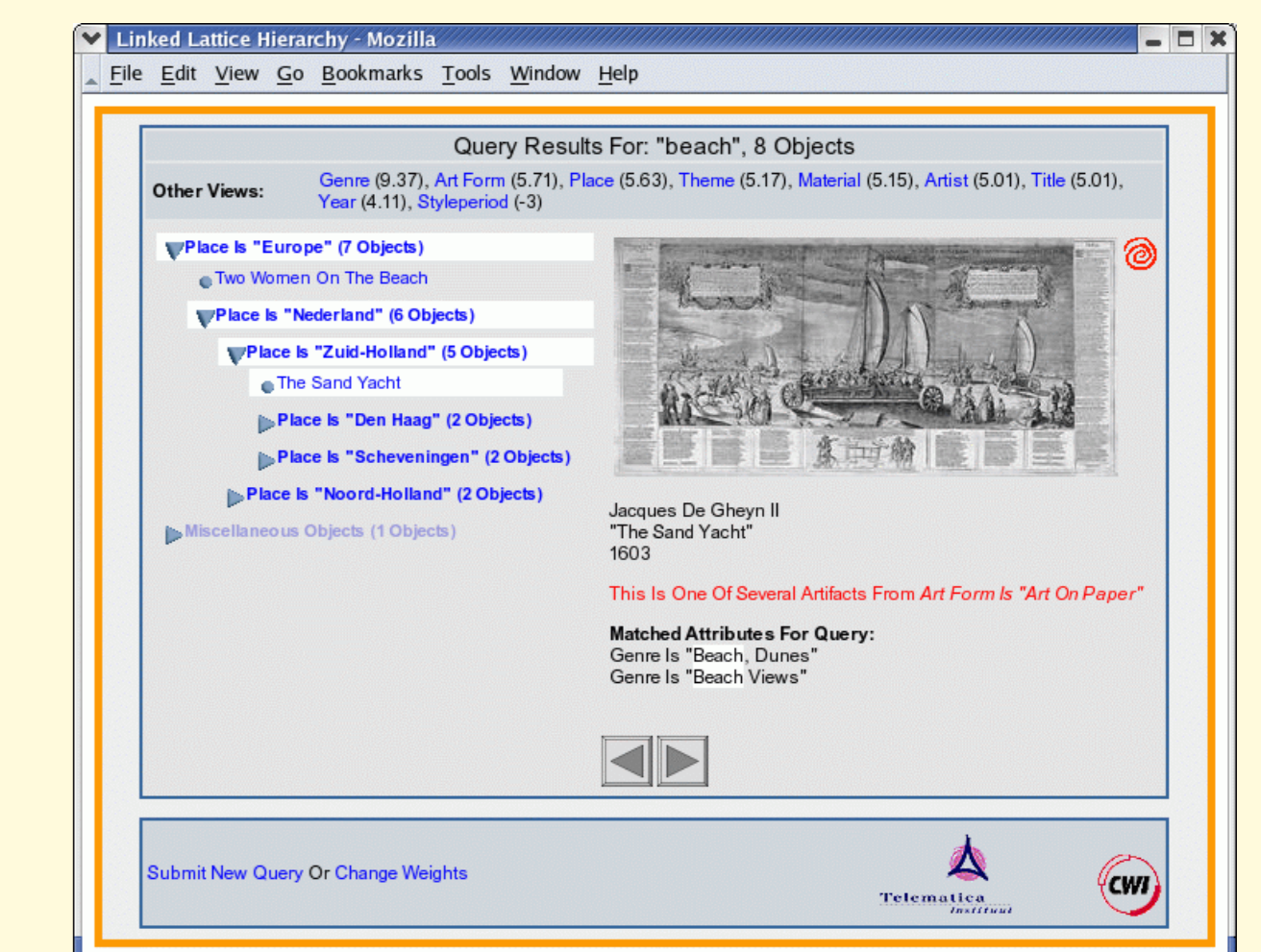| Objects	(C2) Genre: Water ice and snow	(C3) Genre: Buildings in landscapes	(C4) Genre: Field meadows	(C5) Genre: Dutch Landscapes	(C6) Artist: Jacob van Ruisdael	(C7) Genre: Buildings in landscapes	(C8) Genre: Riverscapes	(C9) Artist: Paul Joseph Constantin Gabriel
(C1) "Water"								
Watercourse at Abcoude (A1)	X		X	X				X
Watercourse near 's-Graveland (A2)	X							
Mountainous landscape with waterfall (A3)	X				X	X	X	
A water mill (A4)	X				X	X	X	
Landscape with waterfall (A5)	X				X	X	X	
Water mill (A6)		X			X			
Windmill on a polder waterway (A7)		X	X	X				X
A waterside ruin in Italy (A8)		X						
The battle of Waterloo, 18 june 1815 (A9)			X					
Concept Size	5	3	3	3	3	3	3	2

Proximity Table

From Rijksmuseum Example

```
A2 2
A3 2 2
A4 2 2 2
A5 2 2 2 2
A6 4 4 2 4 2
A7 2 2 4 4 4 2
A8 4 4 4 6 6 2 2
A9 2 4 4 4 4 4 4 4
C2 1 1 1 1 1 3 5 3
C3 3 3 3 3 3 1 1 3 4
C4 1 3 3 3 5 1 3 3 2 2
C5 1 1 3 3 5 1 3 3 2 2 4
C6 3 3 1 3 1 1 3 3 5 2 2 4 4
C7 3 3 1 1 1 3 5 3 5 2 4 4 4 2
C8 3 3 1 1 1 3 5 5 5 2 4 4 4 2 2
C9 1 3 3 3 3 3 1 3 3 2 2 2 2 4 4 4
   A1 A2 A3 A4 A5 A6 A7 A8 A9 C2 C3 C4 C5 C6 C7 C8
```

Telematica Instituut