Generalized Semantic-to-Document Derivation

Lloyd Rutledge, Martin Alberink, Lynda Hardman and Mettina Veenstra

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

Phases

Semantics as Links

Topic Selection (Search)

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 maintain proximity (segue)
- Cross-reference
- Also conveys proximity

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

Concept Size
- Algorithm: K-means
- Generalized Proximity
- Proximity table
- Best-of search for corresponding doc struc

Concept Lattices

Example: Rijksmuseum Collection, query "water"

Proximity Table

From Rijksmuseum Example

Proximity Table

Generated Presentation

numeric Clustering

With Uniformity