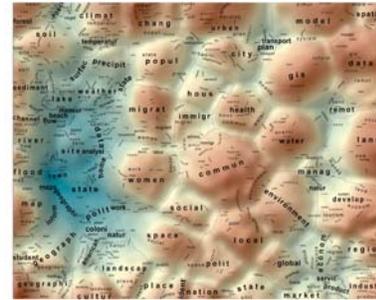
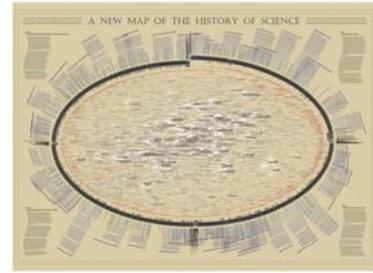


# Representations of Topical Data

- **Charts:** Word cloud
- **Tables:** GRIDL
- **Graphs:** MDS plots, circular visualization, Crossmaps →
- **Geospatial maps:** SOM maps →
- **Network graphs:** Tree visualizations, word co-occurrence networks, concept maps, science map overlays



9

## Information Visualization MOOC

### Unit 4 – “What”: Topical Data

#### Overview and Terminology

##### Relevant Research Disciplines:

Linguistics, Computer Science, Artificial Intelligence

##### Reference

Börner, Katy, Chaomei Chen, and Kevin W. Boyack. 2003. [“Visualizing Knowledge Domains.”](#) Chap. 5 in *Annual Review of Information Science & Technology*, edited by Blaise Cronin, 37:179-255. Medford, NJ: American Society for Information Science and Technology.

# Topical Analysis and Visualization Goals

**Main goals** are to understand

- Topical distribution of a dataset, e.g., what topics are covered and how much.
- How topics emerge, merge, split, or die.
- Bursts of topics, see Unit 2 on 'Temporal Analysis'
- Topical change over time, see Unit 2

Topical analyses at different levels of aggregation are common. Analyses may range from micro to macro—e.g.

- single documents (micro), journal/book volumes, scientific disciplines (macro), or
- single individuals (micro), institutions, or countries (macro)

11

## Terminology

**Text:** A sequence of written or spoken words.

**Text corpus:** A large and structured set of texts (e.g., tweets, emails, books).

**Topic:** A noun phrase that expresses what a sentence is about.

**N-gram:** A subsequence of  $n$  items (e.g., phonemes, syllables, letters, words) from a given sequence.

**Stop words:** Very commonly used words (e.g., a, and, in) that are excluded from topical analysis.

**Stemming:** Process for reducing inflected (or sometimes derived) words to their stem, base, or root form.

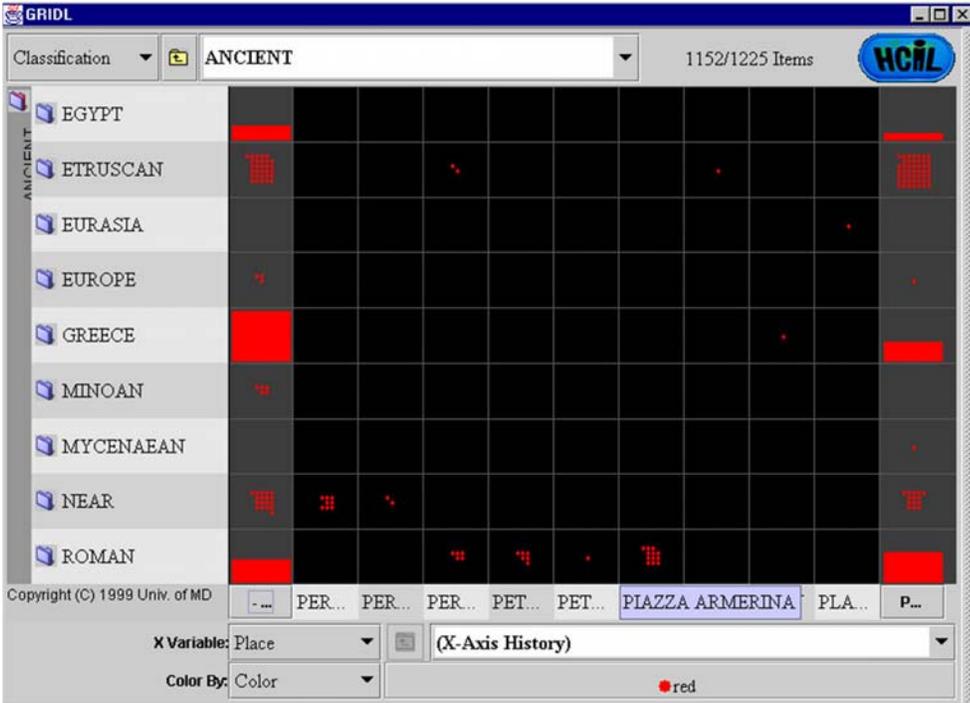
**Synonymy:** Words or phrases alike in meaning or significance (e.g., happy, joyful, elated or close, shut).

**Polysemy:** The same word having many meanings (e.g., bank, crane).

12

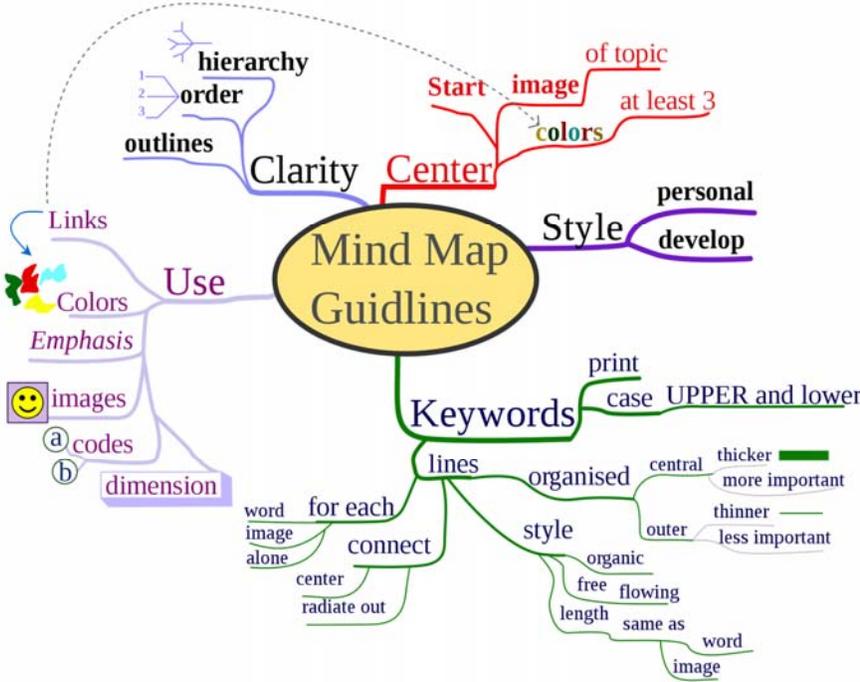


GRIDL, developed at HCIL, uses categorical and hierarchical axes to support categorical zooming.



15

Concept maps are network graphs that show the relationships among concepts.



16