Introduction to Content Analysis

Academy of Management Professional Development Workshop

Chicago 2009

Different Styles: Which is me?

plus: Introducing ~~æ~~, an 'association engine'

Within content analysis broadly defined, what are the various styles?

Broad Definition — "a research technique for making replicable and valid inferences from texts … to the contexts of their use" (Krippendorf 2004: 18)

4 Styles … — We can distinguish 4 (overlapping) styles of content analysis

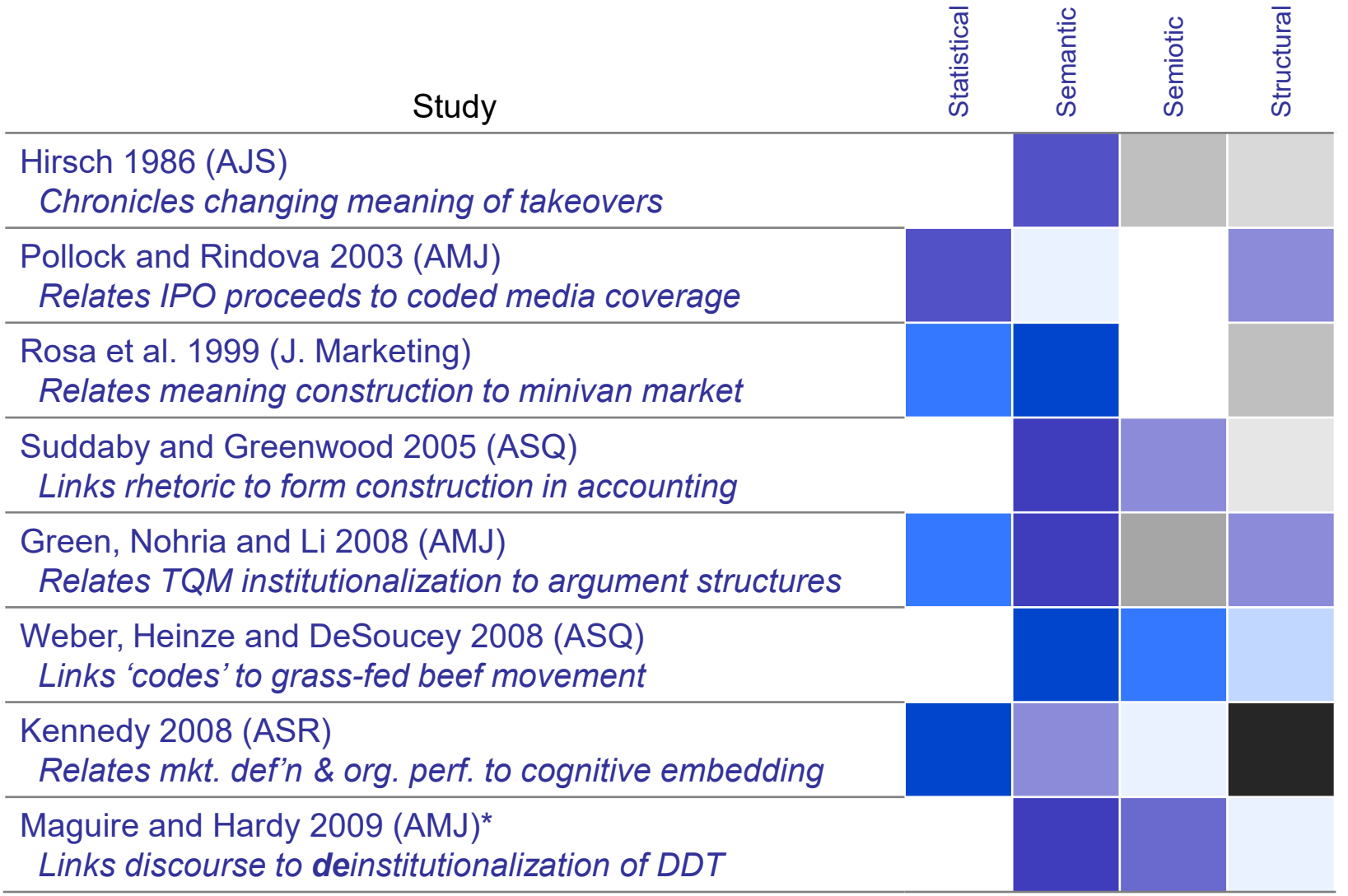
Statistical
- Show trends of growing or declining attention to a topic
- Relate text coding to changing sentiment about a topic

Semantic
- Explore subtleties or changes in meaning of an idea

Semiotic
- Find sources of meaning or action in relations among ideas

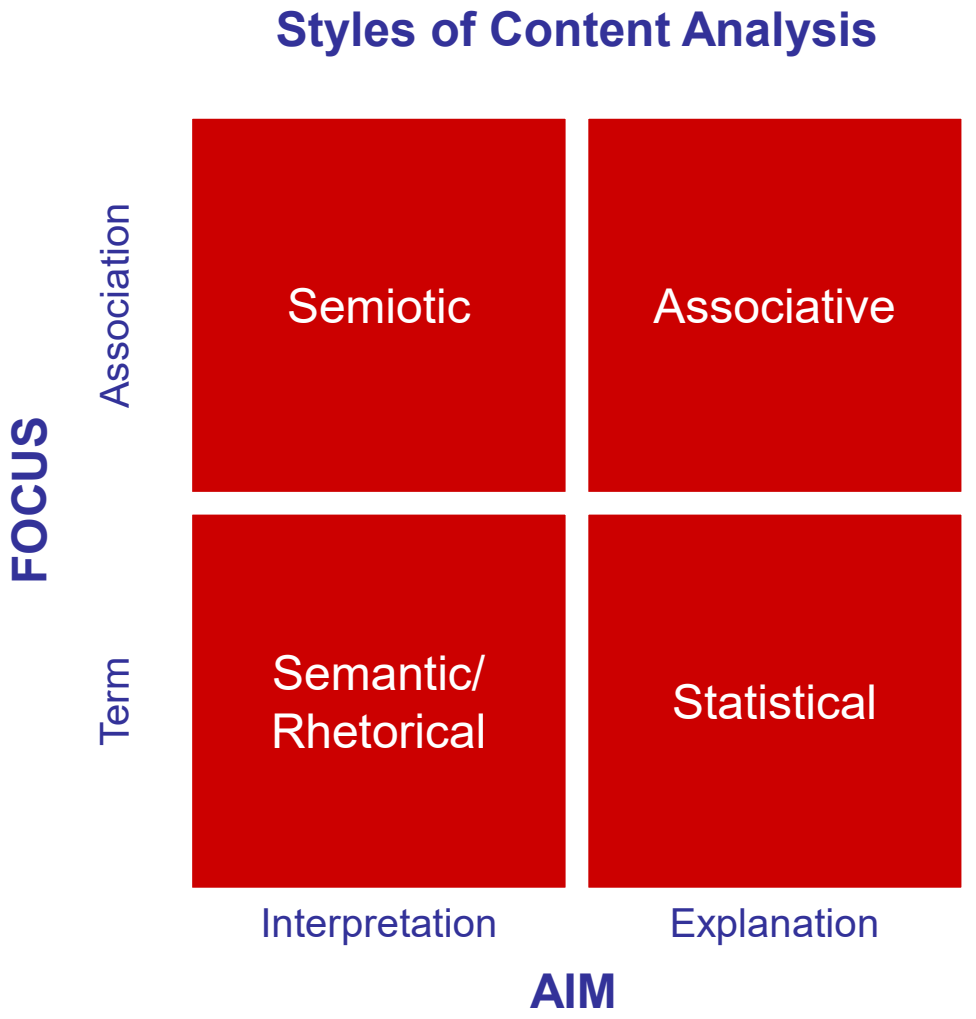
Structural
- Relate association among ideas to various macro outcomes

How about some examples?

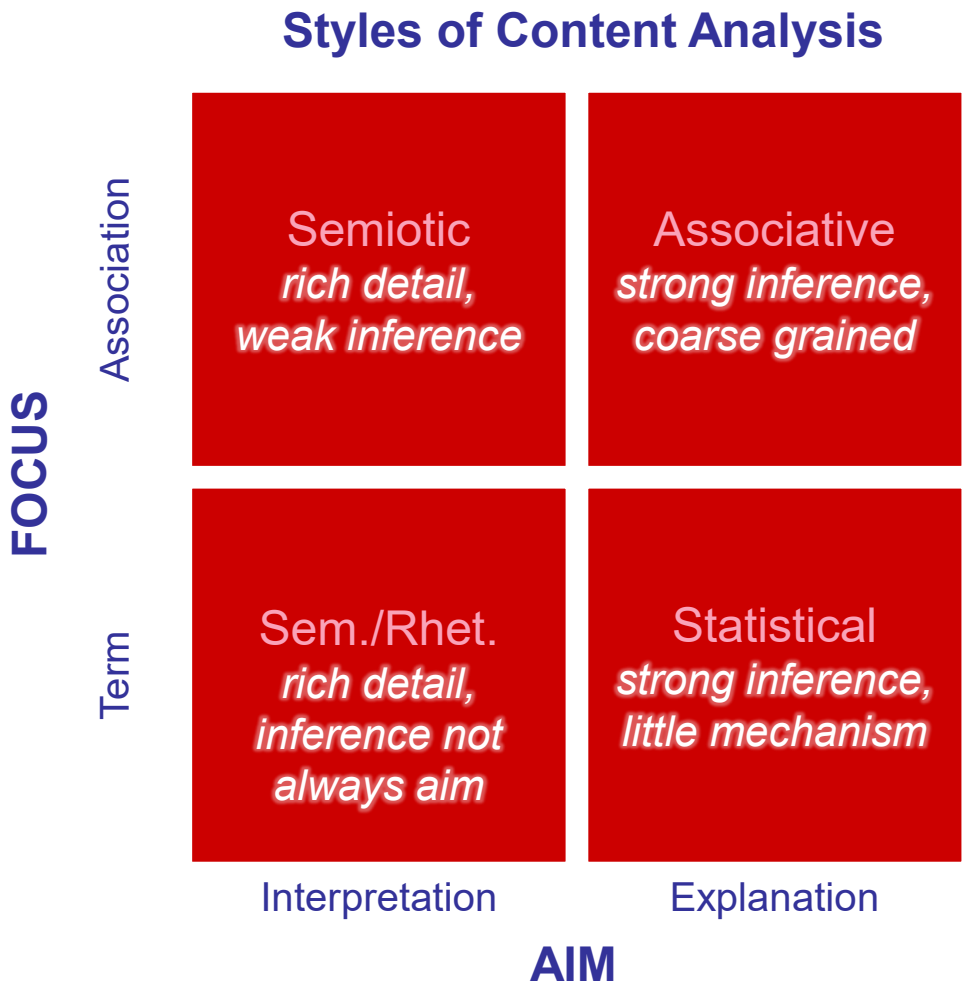


* Not saying discourse analysis = content analysis!

Which style best fits my research?
That's determined by study focus and aim



Each approach has its pros and cons



Each approach has charms, but each also attracts predictable criticisms from others

Statistical 'Shallow'	Counts of terms criticized as explaining little, but recent studies offer clear mechanisms (e.g., see Pollock et al. 2008)
Semantic 'Fuzzy'	Rich tales of meaning change are seen as revealing (Hirsch 1986), but some audiences are hungry for greater mechanism detail
Semiotic / Rhetoric 'Obtuse'	Semiotics and rhetoric offer powerful approaches to meaning and structure, but papers draw on diverse traditions not widely known to organization scholars (e.g., Green et al. 2008; Suddaby and Greenwood 2005; Weber 2005)
Associative 'Imperious'	Structural approaches remain an ambiguous middle in persistent debates over meta-theorical and methodological commitments – Despite dual constitution of structure & meaning (Mohr 1997), divides persist between qual. / quant. approaches, esp. when equated with constructivism vs. realism (Hardy et al. 2004)

Good news

Institutional theory provides a helpful common ground

So, while text analysis is a great match for today's hot topics, it's still quite risky!

Hot Topics — New theory links discursive sensemaking of categories and related identities to dynamics of important social structures …

- Markets and industries
- Organizational practices and forms
- Social and political movements
- Academic disciplines and interdisciplinary fields
- Genres and styles in cultural industries and the arts

Challenges — Text archives offer vast data resources for cool studies, but …

- Projects are risky and expensive
- The work is laborious
- Results are misunderstood

Needed … — Methods for finding changing pattern of association among items

- Representing the defining attributes of constitutive instances
- Related by refined selection of co-occurrences in a corpus

Viewpoint: simple extraction logic meets key needs while staying clear of nasty traps

Luddite Power	Human coding yields the most precise, nuanced insights, so "with a big enough team (budget)", but beware scale-up melt-down!
Android Dreams	AI-like mining of a very large corpus to capture changing usage patterns (denotation, connotation) overpromises, under-delivers CS researchers are into building ontologies, but explaining meaning construction requires taking a step back from what is
Stuck-in-the-Middle	Thin-but-huge-N and tiny-N-but-really-rich both have strengths, but reviewers may not be crazy about hybrid combinations …
Simple Logic	Inspired by advances in search but informed by tough lessons suggest a keep-it-simple-stupid approach for detail *and* scale

Getting detail & scale does entail a (reasonable) tradeoff

Conceptually, specifying a bit more up front makes for an "association engine"

Search Engine
Simplicity

From a very comprehensive document collection, find the subset containing items (terms) of interest
… return it as a ranked list

(items as search keywords) Find

Association Engine
Power

From a corpus selected for relevance or authority, associate items of interest based on co-occurrences that meet desired rules
… and return as a graph (1 period) or graphs (multiple periods)

In: (corpus of relevant or authoritative sources)

Associate: (items of interest to researcher) Associate

Based on: (rules about co-occurrence context)

Practically, the tool I'm developing is not (yet) so user-friendly

But it is FREE to academics* … Introducing æ an association engine tool

Who might use æ? Researchers hoping to use text (discourse) to study construction of categories or identities and their effects on social structure

What is it? æ is a tool for extracting models of meaning and structure based on patterns of association among select items such as ..

- Category attributes / identity characteristics
- Category instances (members) / identity examples

Output… From a longitudinally corpus and a class or classes of items, æ …

- Extracts periodic observations items occurrence
- Extracts relations among items for each period
- Produces an analysis-ready item-period dataset (Stata, etc.)
- Produces networks for easy visualization and animation

Benefits **STATA-ready data from M's of docs in ~1 day, not 6+ months!**

* Free for use in academic research by university-based researchers.

æ mines network models of categories or identities from a relevant corpus

Core Concepts

Building on the idea of a semantic network, categories, identities and related social structures can be modeled as graphs based on:

- Adjacency relations among items as *instances*
- Adjacency relations among items as *attributes*
- Affiliation relations among items of different classes, e.g., actors x events, instances x attributes (transitive products, too)

Builds on Experience

Extends methods used in Kennedy 2005, 2008

- *New:* network analysis integrated for stata-ready data sets
- *New:* dynamic graphs in Pajek .net format for easy animation

Advanced Features

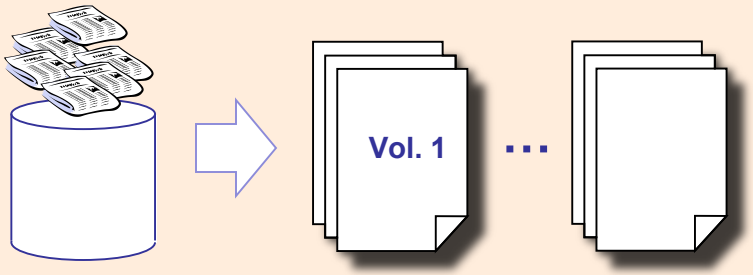
æ finds identities using methods that go way beyond co-mention

- Items can be linked by flexible measures of proximity
- Links can be based on "mediating" terms (~ auto-coding)
- Links can be developed from products of 2-mode matrices

1º

æ is an *association engine*, a tool for finding changing patterns of association among co-occurring items in a particular corpus. æ is useful for studying the meaning of categories and related identities that underlie fundamental social structure such as markets, organizational practices and forms, academic fields, social and political movements, literary and artistic genres, and so forth.

[1] Gather a **corpus** to analyze from appropriate sources*



* supported formats include Lexis-Nexis and Factiva--more soon

Organize in time-based "volume" files, e.g., mycorpus-vol1.txt, mycorpus-vol2.txt, etc.

[2] Write script with (i) **items** to associate and (ii) **rules** of association

+ items to associate by class or category.

  e.g., items in a class of competing producers are names specified as literals or regular expressions.

+ rules for associating co-occurrences.

  e.g., requests links between items that co-occur in paragraphs, sentences, within X words of each other, or near a third "mediating" term that suggests a specific relationship such as membership, cooperation, etc.

myscript.ae

[3] Run your script: **ae myscript.ae** … adapt the tutorial script to your study design to quickly produce an **æ** script that produces data for statistical analyses and graph visualization

associate

slicer
"slices" volume files into single stories for analysis

dicer
"dices" a story into body-text and meta-data for later æDB queries

mention-finder
finds item mentions and adds within-story address to graph meta-data

aelink / aemerge / aewrite

aelink
make graph using gmd, rules of association

aemerge
merge into dynamic graph

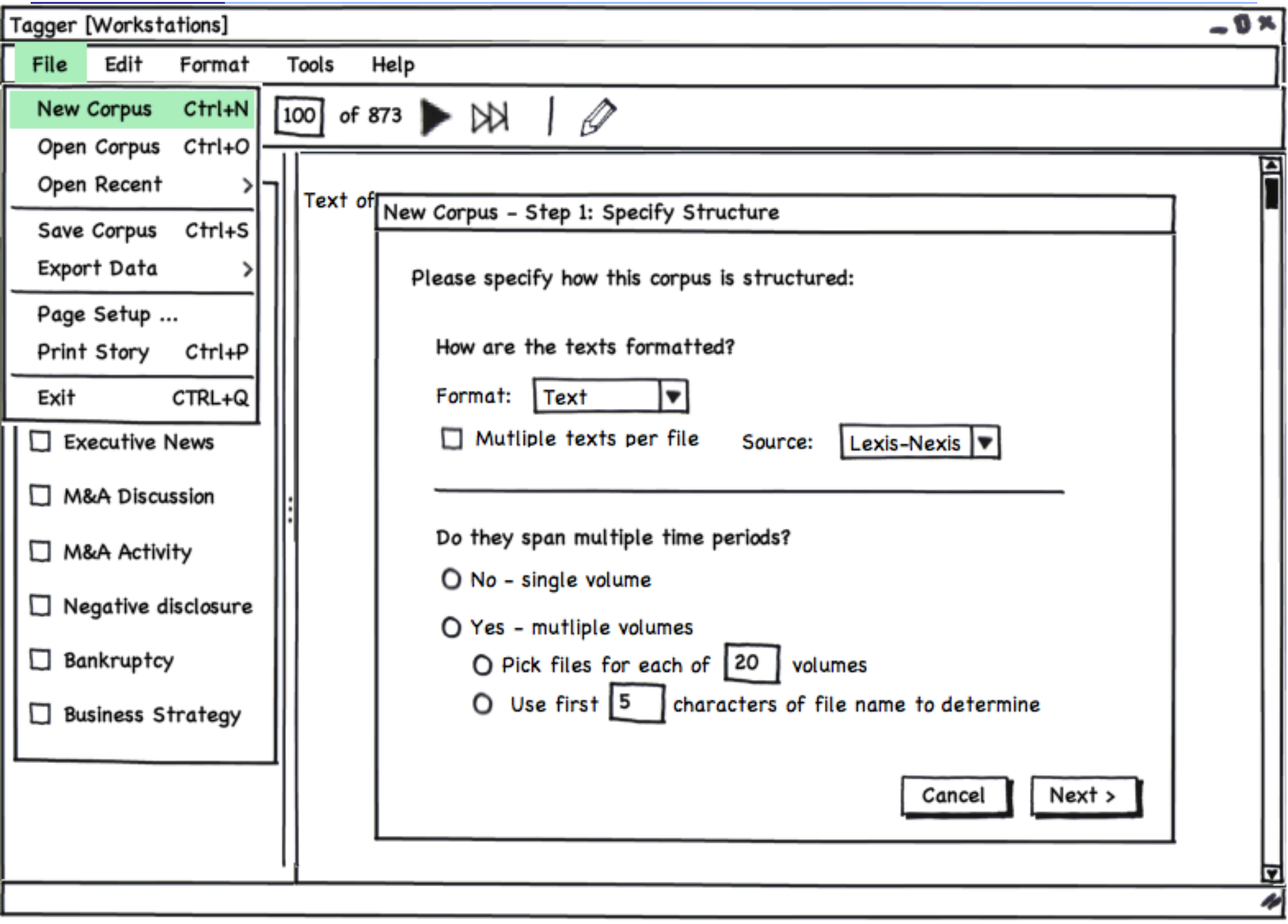
aewrite
write out dynamic graph, data set

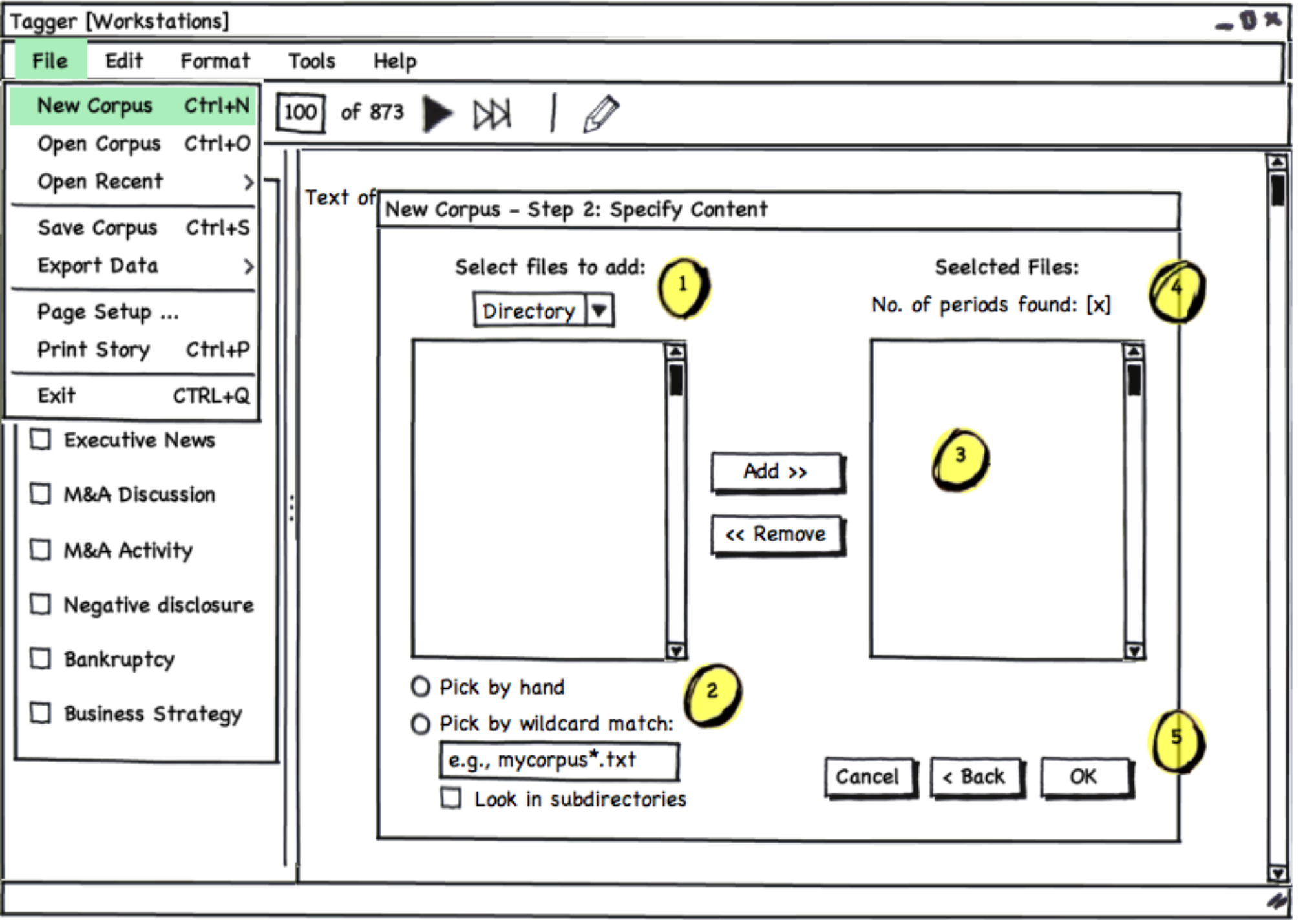
story text (.txt)

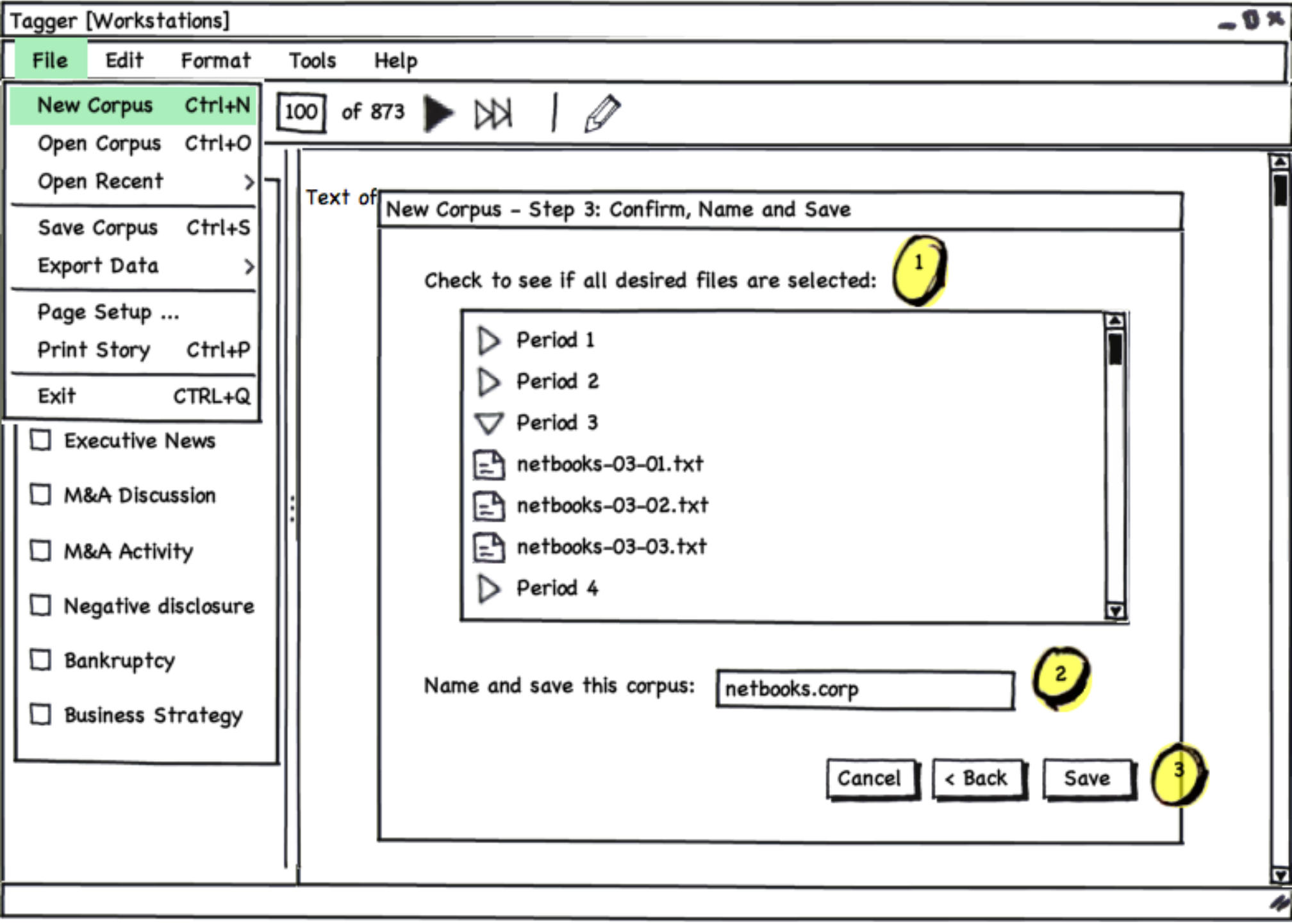
story profile (.spf)

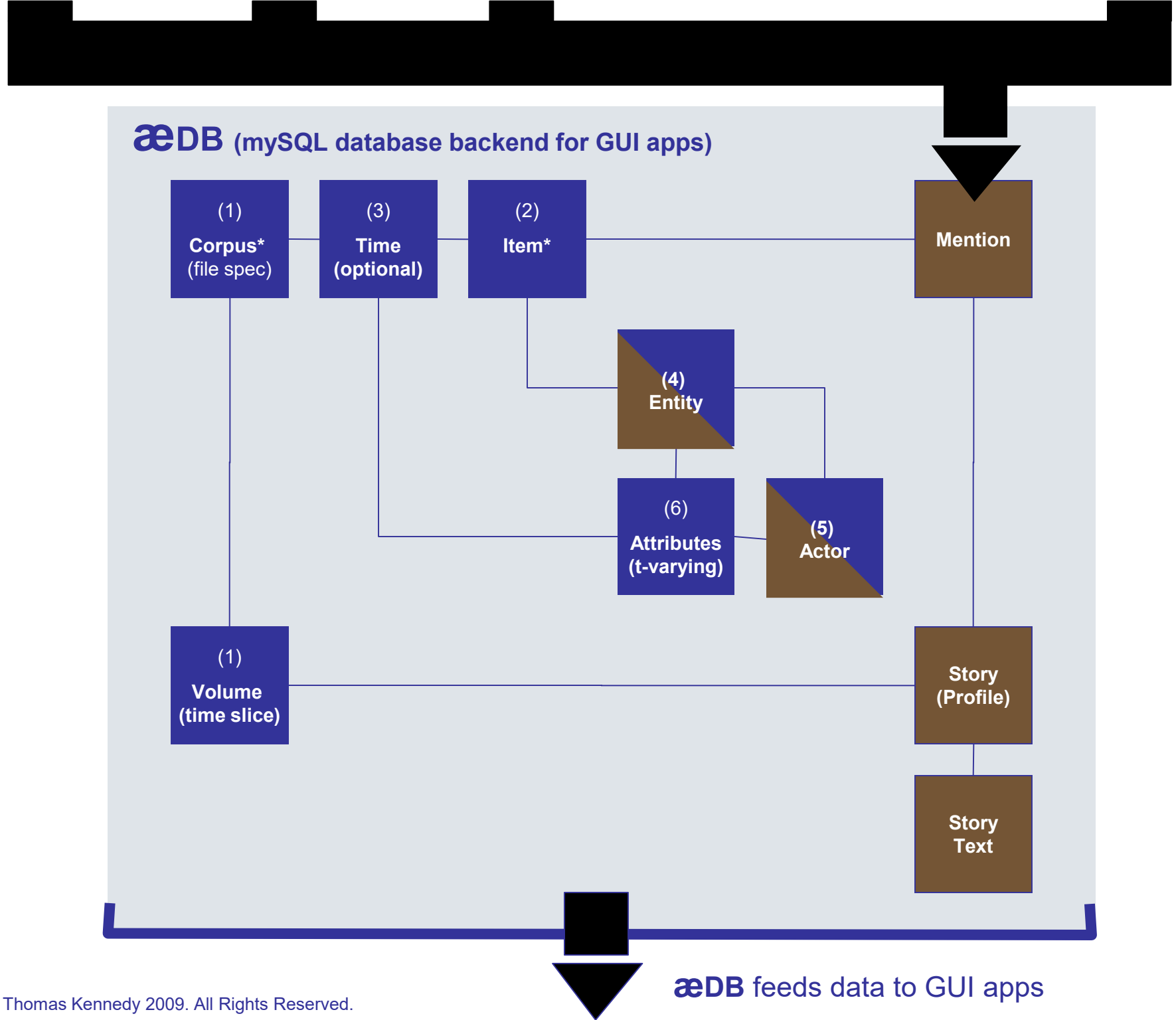
graph meta-data (.gmd)

dynamic graphs, data set (.net, .dat)









Script language at first, but user-friendly GUI apps coming…

*items: write and run æ scripts using a GUI wizard …

item manager (separate window)

Create and manage items by class (category)
- text-patterns to search for
- timing restrictions (if any)

*seer: use network data views to browse and tag (code) corpus item mentions …

Graph Browser

- association results shown in graph form with controls for …
- time/frame: |< < > >| (frame snapshot advance)
- animation: ■ ▶ • (smooth transitions – record = export)
- layout style: several "styles" (layout algorithms) offered in dropdown
- view: zoom in / out, rotate (2 axes), 3d pan

- selecting (mouse clicking on) graph elements (nodes and links) queries aeDB to populate the other interface elements with the data that supports the association map (graph) being displayed
  (1) clicking a node puts a list of all the texts in which it is mentioned in the "hits" box
  (2) text viewer
  (3) entity / actor profiles

Text Viewer / Tagger

- text of selected "hit" in scrollable text window with item mentions highlighted for easy viewing

Controls for …
- moving to next / previous mention
- moving to next / previous text in hits
- browsing text (scroll bars, mouse wheel, etc.)
- cut / copy / paste
- selecting limited range of font / size options

Hits (texts)

clickable list of texts in which the selected graph item(s) appear (much like Google)

Metadata (optional)

If desired, background on each hit text is also given from the story profile—things like publisher, author, length, date, online source, etc.

mention tags (coding)

view and manage mention "tags" -- e.g., aspects of mention tone or meaning
- delete or modify auto-coding
- add human coding

To try it out on your own project, email me with subject “ae mailing list”

Need easy instructions for …

- downloading æ code and tutorial materials (script & corpus)
- downloading visualization software and viewing output
- running the tutorial script on its corpus
- getting æ output into statistical analysis applications
- preparing a corpus for a new æ project
- adapting the tutorial materials script to write a new æ script
- running an æ script on a corpus prepared for it