Theory and Method for

Studying Meaning and Related Social Structures

Introducing “Association Engine” (Æ),
A Practical Tool for Relational Content Analysis

2010 Academy of Management Meetings | Montreal
Content Analysis in Organizational Research:
Techniques and Applications

Outline

Theory meets Method

- Cool research questions
- Meaning and social structure
- New (research) realities

Anatomy of a Project /Introducing Æ

Relational Content Analysis Made Easy

- Key Inputs
- Simple interface
- Robust technologies

Conclusion

- Community, standards matter!

Cool topics being addressed by linking macro entities to patterns of association

Published	**Social and political movements:** How do fringe interests become mainstream issues?	Grass-fed meat, dairy—activists found a collection of codes seen as ground for new identity, practices.[1]
	Markets: how do erstwhile misfits become new categories?	Workstations—Media co-mentions of rivals reveal category, prototype.[2]
In process	**Corporate reputation:** how do criteria change and emerge?	Green—seen by finding overlap in antagonists' association of attributes.[3]
	Fields, Forms & Genres: how do cross-category blends get naturalized?	Nanotech—repeated patterns of association among patent classes.[4]

1. Weber et al. 2008 (ASQ)
2. Kennedy 2008 (ASR)
3. Kennedy & Chok
4. Lo & Kennedy

Repeated patterns of association: a window to dynamics of social realities

Meaning, Structure

Because meaning and structure are both seen in relations among entities/ideas, they are dually constituted (Mohr and Duquenne 1997, Mohr 1998)

Networks, Structure

Recognition of social realities (sometimes also legitimate) emerges from repeated associations that link actors, entities or ideas (Emirbayer 1997, Tilly 2008, Phillips et al. 2000, Powell et al. 2005, Kennedy 2008)

But this type of work is laborious and risky because:

1. the data is hard to get and code;
2. finding relations requires RDB and tricky queries;
3. theory and method are new to most reviewers

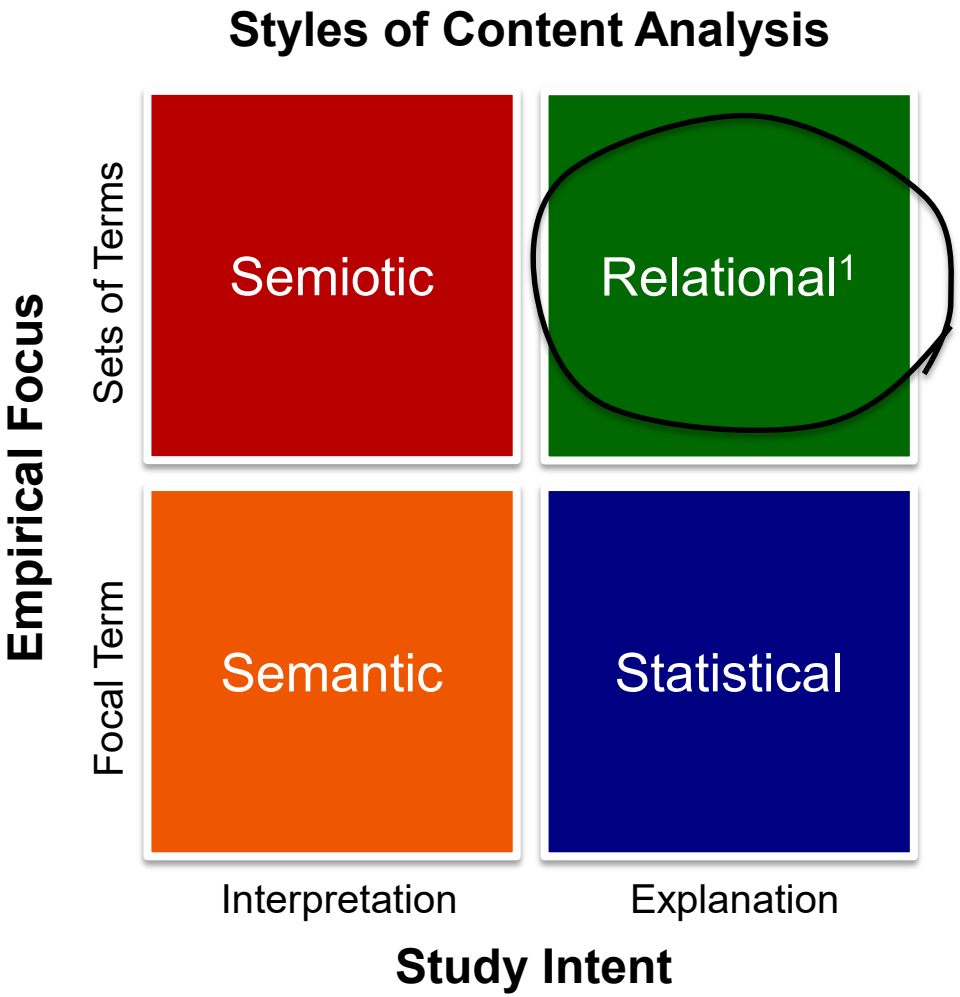
Process, Network Dynamics

One can trace the process of field-level change (Hardy & Maguire 2008) by studying the dynamics of networks of changing patterns of association between entities, ideas, etc., (Owen-Smith & Powell 2008) and vice versa

Still hard, but the flood of web content = big opportunities for content analysis

	Before	*Opportunity*
Coding	Human coding	Computer-assisted
Scale	Hundreds	Hundreds of thousands+
Sources	– Media, PR	– Also Blogs, RSS feeds
	– Reporters, Analysts	– Commenters, too
	– Published catalogs	– Make-your-own!
Lag	Year or more	Months?
Periods	Usually years	Months, weeks, days, …
Product	Qualitative themes	Semantic networks

Of the styles of content analysis, this calls for *relational* content analysis*



1. Carley and Palmquist 1992

Anatomy of a project: Besides a good question, you need 3 key inputs

Corpus

A collection of text relevant for your topic & research Q

- Can be Lexis-Nexis® or Factiva downloads or RSS feeds from blogs

Fuzzy Ontology

Terms *potentially* relevant to "facets" of meaning[1]

- Instances, attributes, synonyms, antonyms
- Parents, children, neighbors
- Producers (sources)

Like an imprecise dictionary (aka "loose" ontology)

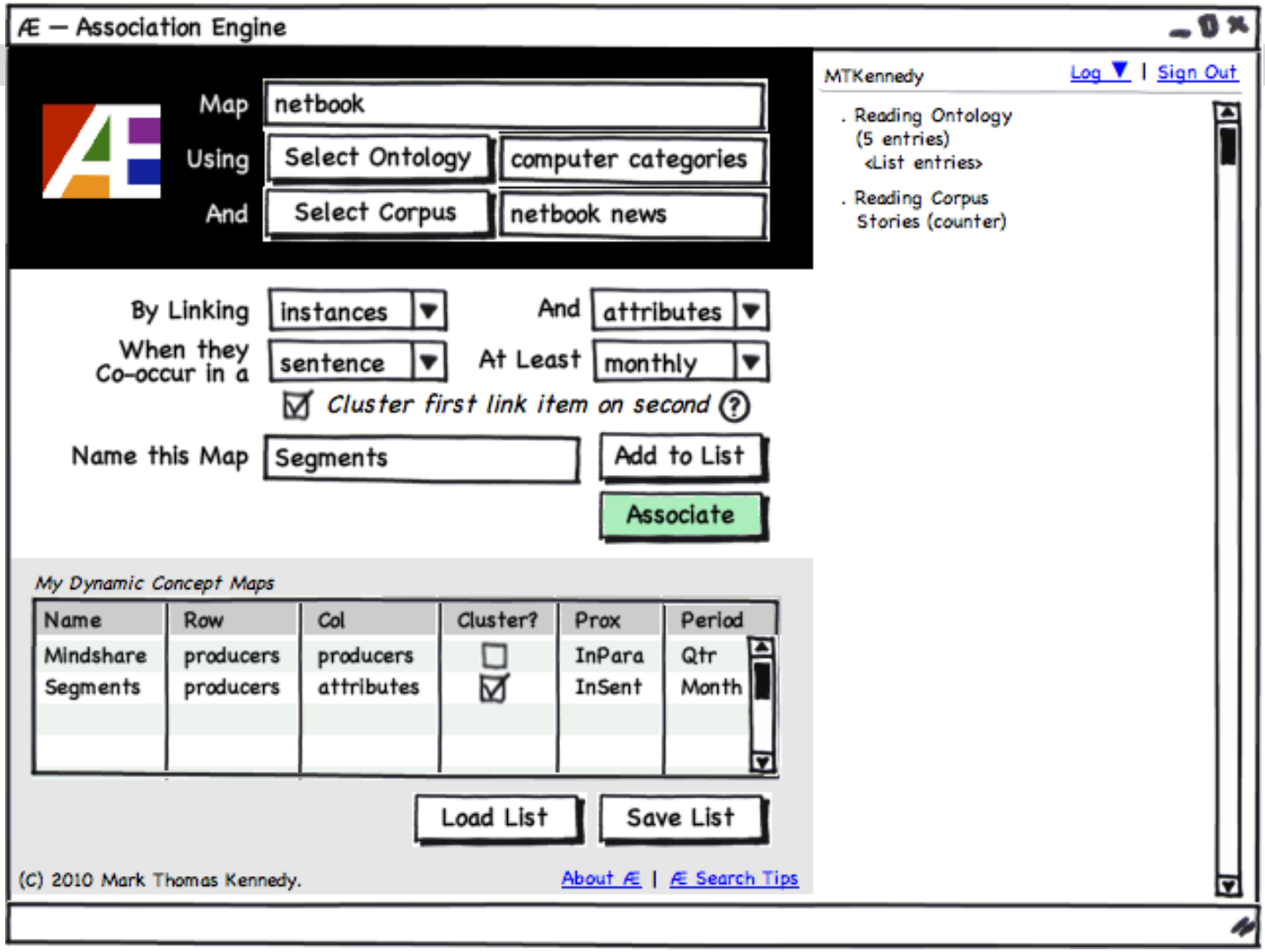
Recipes

Facets to associate (e.g., instances, attributes)

Update frequency (period) and when to link (proximity)

1. To recall these, I use the acronym "PCASPIAN"

Interface (netbook example)

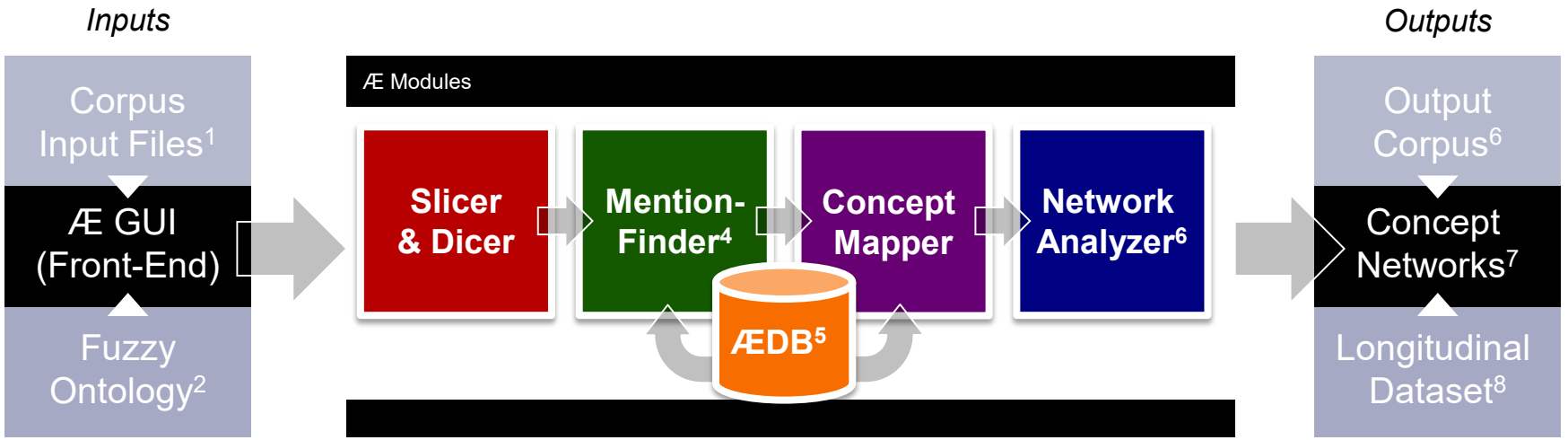


Here's what a fuzzy ontology looks like

```
<ontology>
  <source>MTK</source>
  <name>computers</name>
  <entry>
    <term>netbook</term>
    <definition>a small, light-weight low cost laptop computer with long battery life; often used for Internet access and wireless
      communication</definition>
    <partofspeech>noun</partofspeech>
    <facets>
      <attributes>
        lightweight, inexpensive, wireless,
          <aliases>small, <phrase>smaller than a laptop</phrase></aliases>
        <phrase>long battery life</phrase>
      </attributes>
      <producers>
        Dell, Sylvania, Acer, MSI, Lenovo, Samsung, HP, Toshiba, <pattern>Asus|ASUS</pattern>
      </producers>
      <instances>HP 2140 Mini-Note, Samsung NC10, Dell Inspiron Mini 9, MSI Wind, U100-279US, Asus Eee PC 1000HA, Lenovo
      Ideapad S10, HP Mini 1000 XP Edition, Acer Aspire One AOA150-1126, Asus Eee PC 900HA, Sylvania GNET28001SO, Meso
      </instances>
      <synonyms>laptop, mini-laptop, notebook, mini-notebook, sub-notebook, ultraportable</synonyms>
      <antonyms>mainframe, workstation, server</antonyms>
      <parents>
        laptop, notebook, <aliases>personal computer, PC</aliases>
      </parents>
      <neighbors>smartphone, tablet</neighbors>
    </facets>
  </entry>
  <entry>personal computer
      <definition>a small digital computer based on a microprocessor and designed to be used by one person at a time</definition>
      <attributes>personal, small, microprocessor, desktop, speed, size, memory, storage</attributes>
      <synonyms>PC, desktop</synonyms>
    </entry>
</ontology>
```

Æ builds on robust open source libraries for search, database, network analysis

Æ is an "association engine'', a search utility that produces dynamic concept networks from 3 inputs:
(1) a collection of relevant texts from LexisNexis, Factiva, or RSS/Atom feeds from blogs or websites (corpus);
(2) a set of terms *potentially* relevant to the concept's meaning ("fuzzy ontology"); and
(3) rule(s) for when to treat co-mentions of these terms as links ("association rules").



Notes

1. From any of the following:
   - LexisNexis® (HTML),
   - Factiva (XML), or
   - the web (RSS/Atom feeds).
2. .æfo file containing terms, definitions, synonyms, etc., or commands for fetching them automatically.
3. Builds on **Lucene**, an open source library (Java) used in a wide range of production search applications.
4. Builds on **Lucene and MySQL**, an open source implementation of SQL. For table structure, see Æ Functional Specification, section 2.4.
5. Builds on **JUNG**, an open source library (in Java) for social network analysis.
6. Corpus input files converted and combined into one .æc file for later re-use as desired.
7. As Pajek .net file for analysis, visualization, exchange.
8. Output (.dat) file contains time-varying measures of concept network structure, position of relevant terms.

Anatomy of a Project: 3 Key Outputs

Visualize

Æ produces dynamic concept networks (DCNs),

- Graph models of meaning and structure
- Visualize / animate to observe changing patterns of association over time

Analyze

Get flat-file datasets for quantitative analysis (one row per for each node-period observation)

Explore
** planned **

Use DCNs to browse the corpus

- Clicking nodes fetches texts that mention that term
- Clicking a link fetches texts that co-mention the linked terms

Conclusion: Sober Excitement

Vision Enable study of difficult questions (dynamics), but cut time, expense, risk—increase research productivity

Fitness … Environmental factors giving content analysis new zip

- Explosion of web content—ripe for analysis
- Open source, esp. for search engine development
- Advances in theory that link macro-level entities to recognition of patterns in lower-level relations

Or Just Deviation? Take-off will also require: theory, identity, community, mutual recognition, complementary roles

- Hazard-rate reg.+ institutions / ecology = take-off!
- New content analysis + _____ = ???

Other ideas?