Capturing and Classifying Ontology Evolution in News Media Archives

Albert Weichselbraun*, Arno Scharl and Wei Liu

* Vienna University of Economics and Business Administration
Department of Information Systems and Operations
Augasse 2-6, 1090 Vienna

albert.weichselbraun@wu-wien.ac.at

September 2nd, 2008

Agenda

Problem & Motivation

- domain knowledge evolves continually
  $\longrightarrow$ most real world ontologies *do* change
- Stojanovic et al.: *Ontology evolution* process of adaptation of an ontology
  - to arisen changes
  - maintaining consistency (ontology + artifacts)
- two research projects (AVALON, RAVEN)

Data Driven Ontology Change

Stojanovic et al. Δ

(i) explicit, usage driven changes

(ii) implicit, data-driven changes

This work focuses on *data-driven* changes.
$\longrightarrow$ observe changes in a domain

Requirements

ontology analysis tool

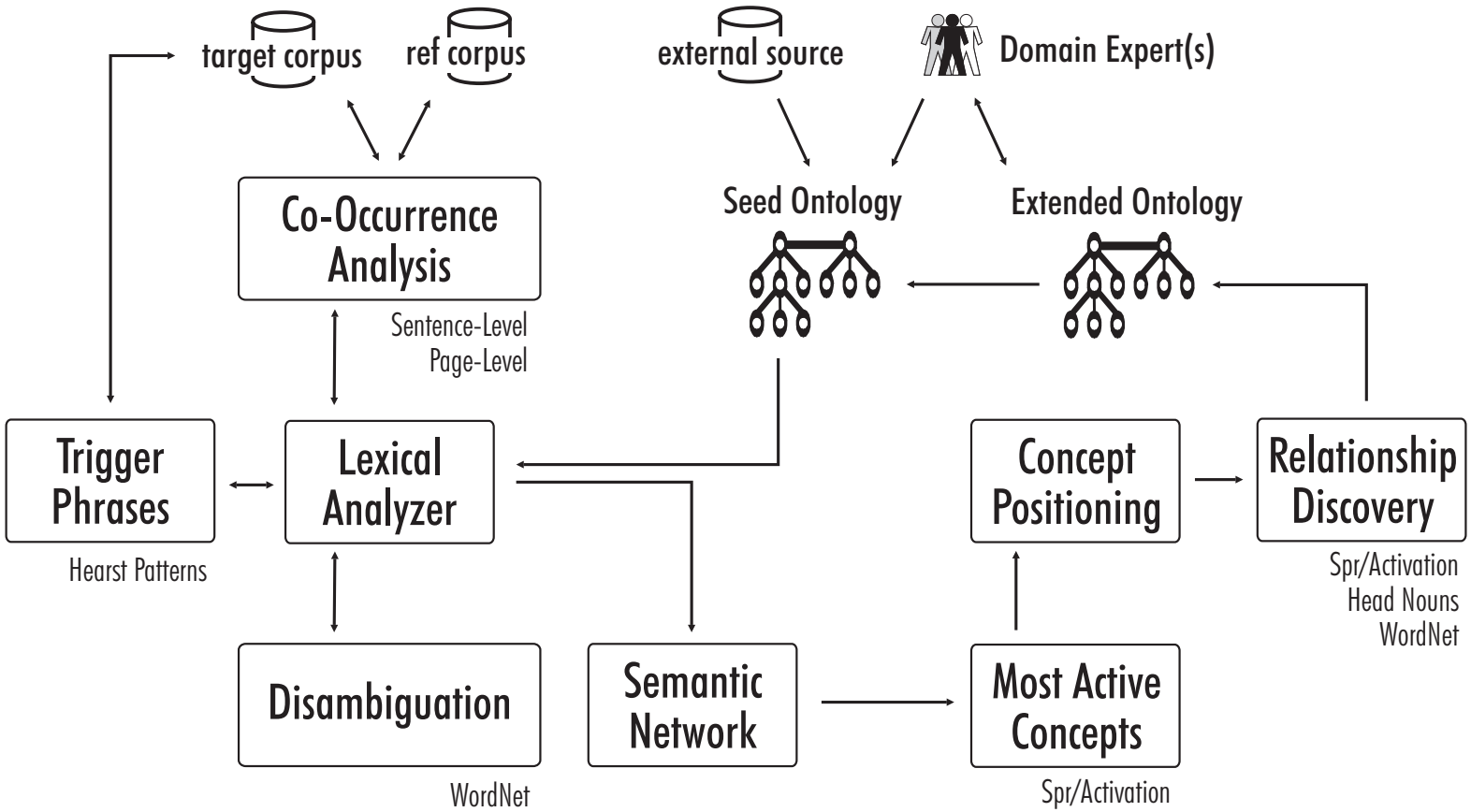
- standardized process to track changes in the domain
  $\longrightarrow$ ontology learning
  - less laborious
  - no inter-/intra personal variations
  - lightweight ontologies
- well defined and *volatile* domain

Sampling

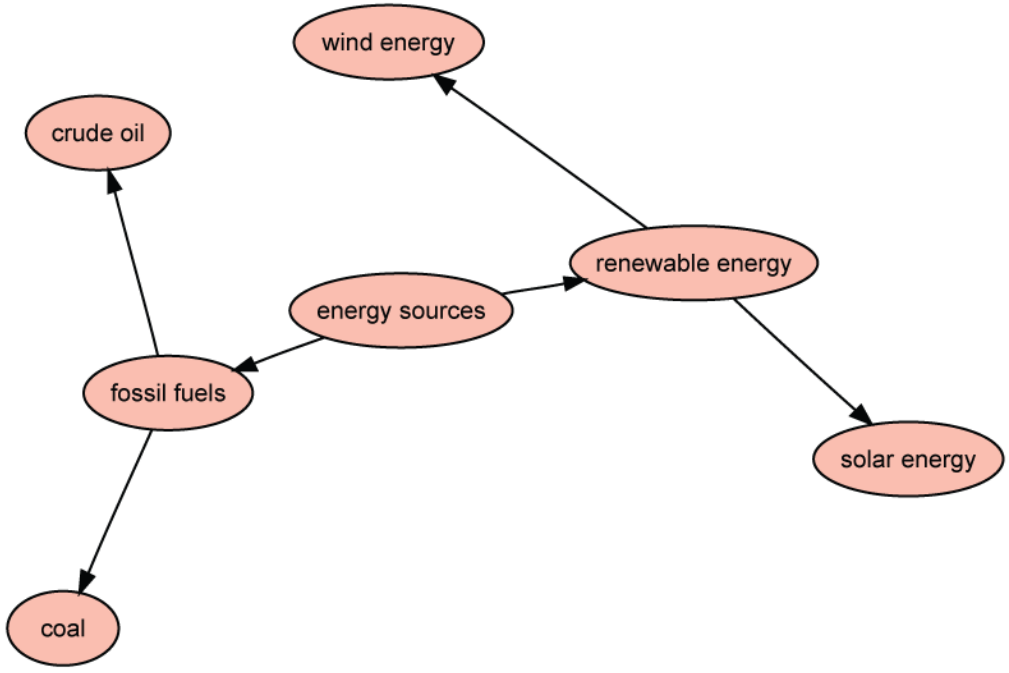
well defined domain

- media coverage on energy sources
- data repository: webLyzard - sample based mirroring
  - 156 news media sites from five English-speaking countries
  - weekly mirrors; from November 2005 to August 2006

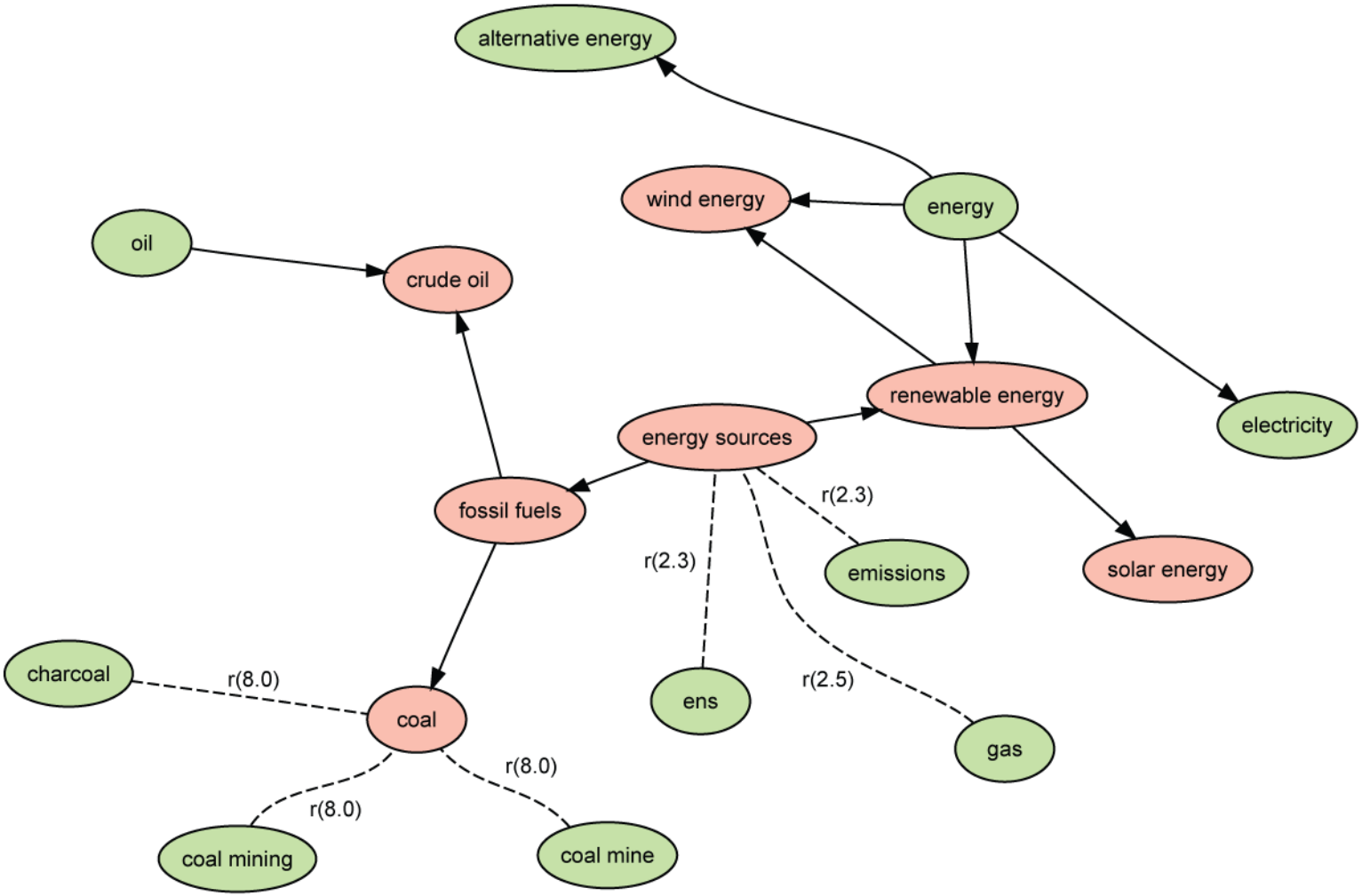
Ontology Learning



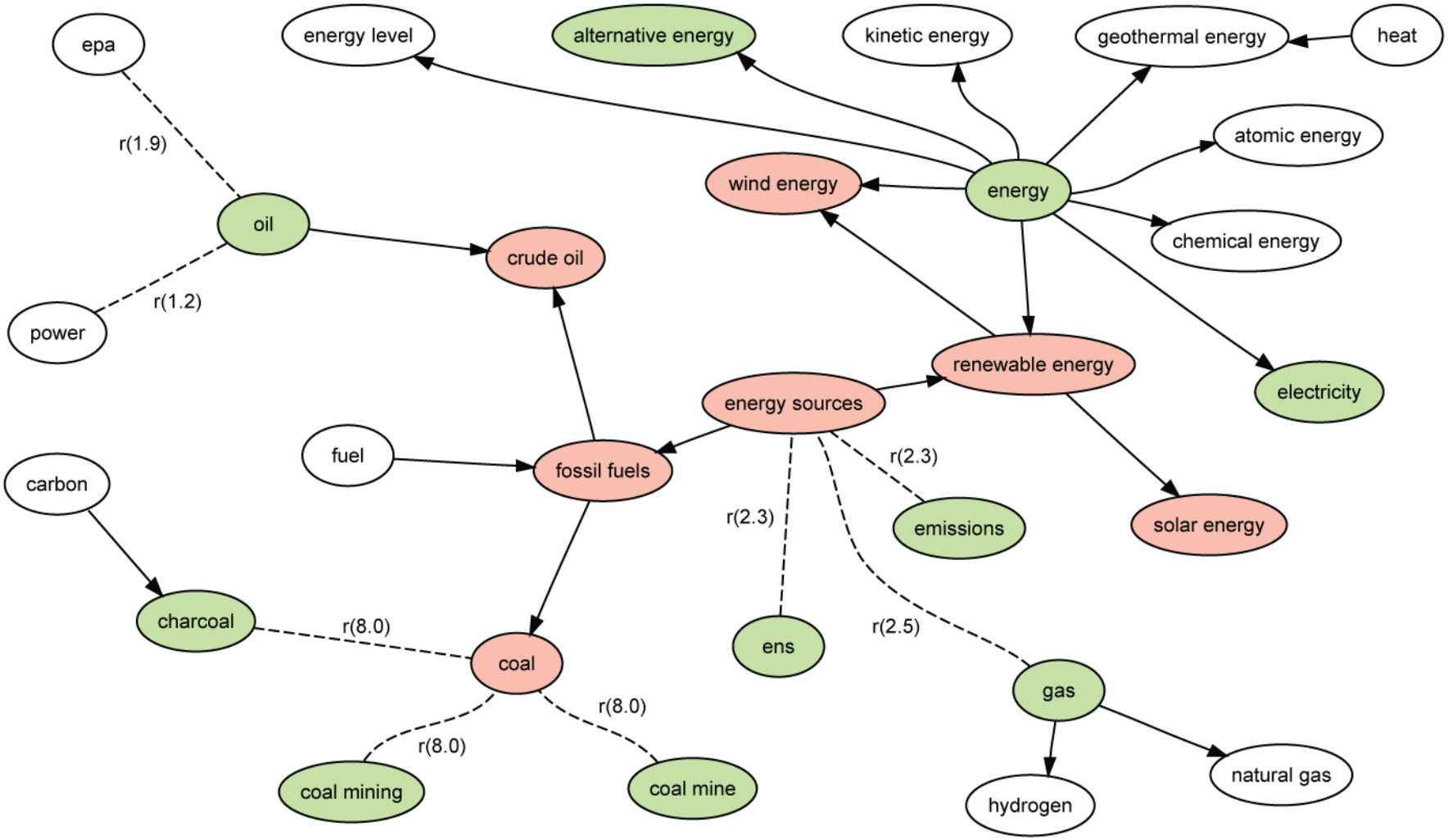
Ontology Learning



Ontology Learning



Ontology Learning



Limitations

- detect changes to the *domain language*, but not changes of the conceptualization
- not *one* authoritative usage, but *averages* (e.g. `alternative energy`)
- handling of salience, limited disambiguation
- very coarse handling of relation types (hierarchical)

Ontology Evolution

domain terminology

- **core domain terminology** comprises frequently used concepts; constantly included into the domain's ontology
- **extended domain terminology** additional domain concepts; lower relevance/importance; used for special topics within the domain (e.g. `nuclear power`,); not as universally used as the core domain terminology
- **peripheral terminology** is used documents; does not carry important domain concepts; not included in the domain ontology

Ontology Evolution

domain relations

- **core domain relations** featuring essential relations between core domain vocabulary,
- **extended domain relations** comprising relations to extended domain vocabulary as well as non-essential relations between the core vocabulary, and
- **Peripheral domain relations** which do not carry enough weight to be included into the ontology.

influenced by: scope, granularity, etc.

A Small Example

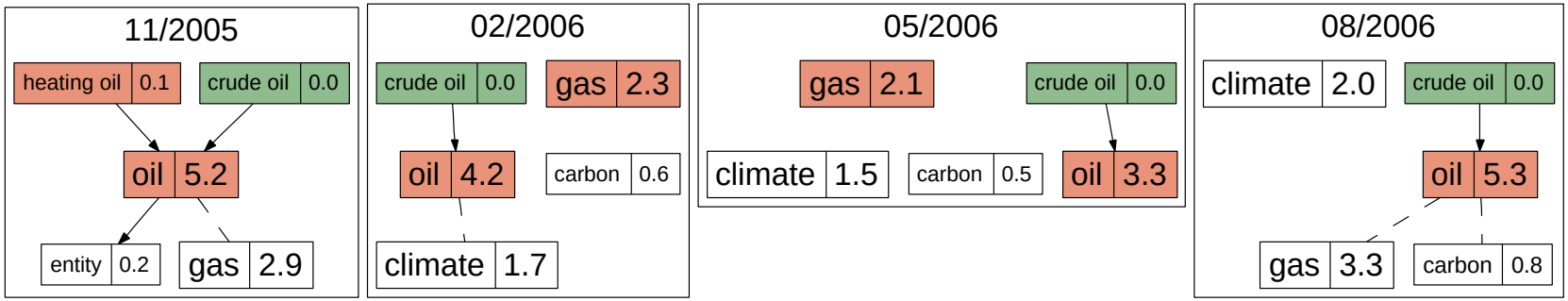


Figure: Evolution of the concept "oil" from November 2005 to August 2006.

Evolution Patterns

Terminology

- *Changes in a term's importance*; focus of media coverage shifts
- *Change of the assigned concept*
  - Change in *term focus* `oil`
  - Change in *term assignment* `fuel, storage`
  - Change in *context* `Sri Lanka, Maldives`

Visualization

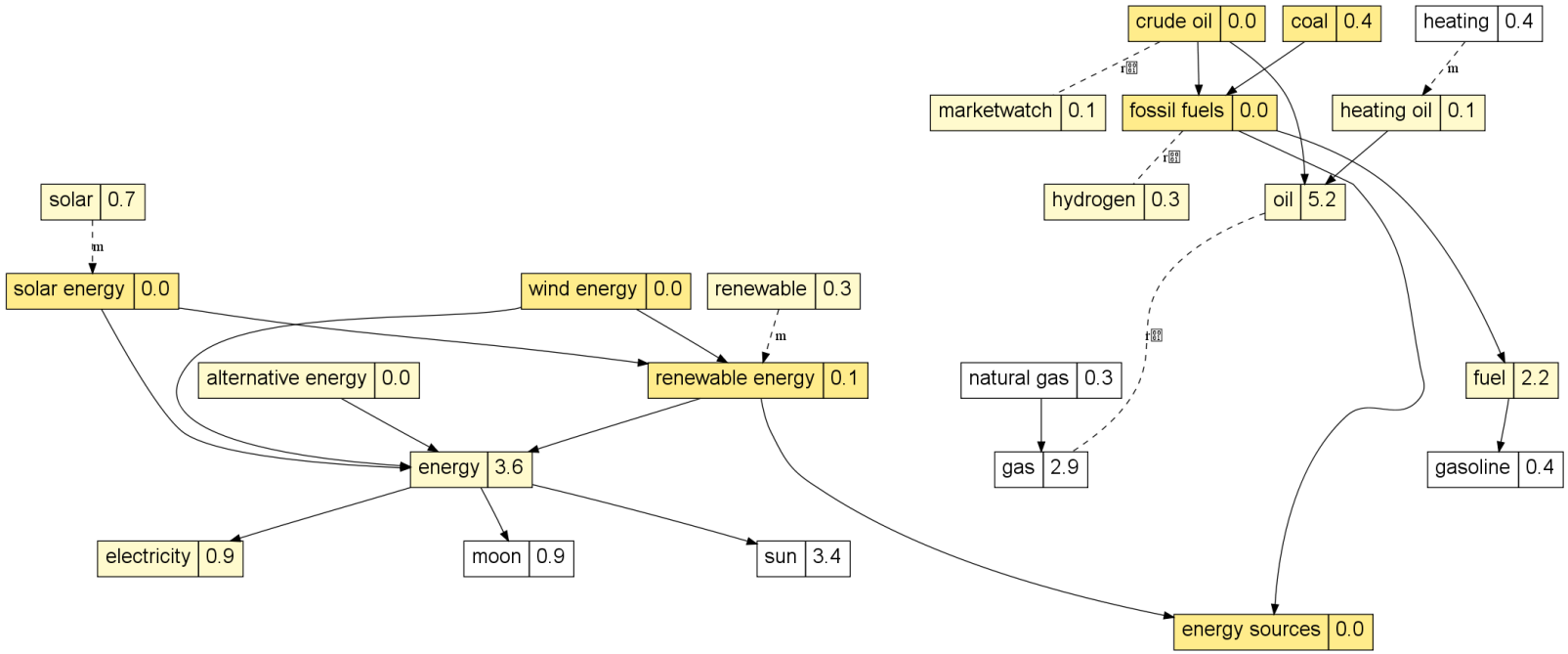


Figure: Extended Ontology (November 2005)

Visualization

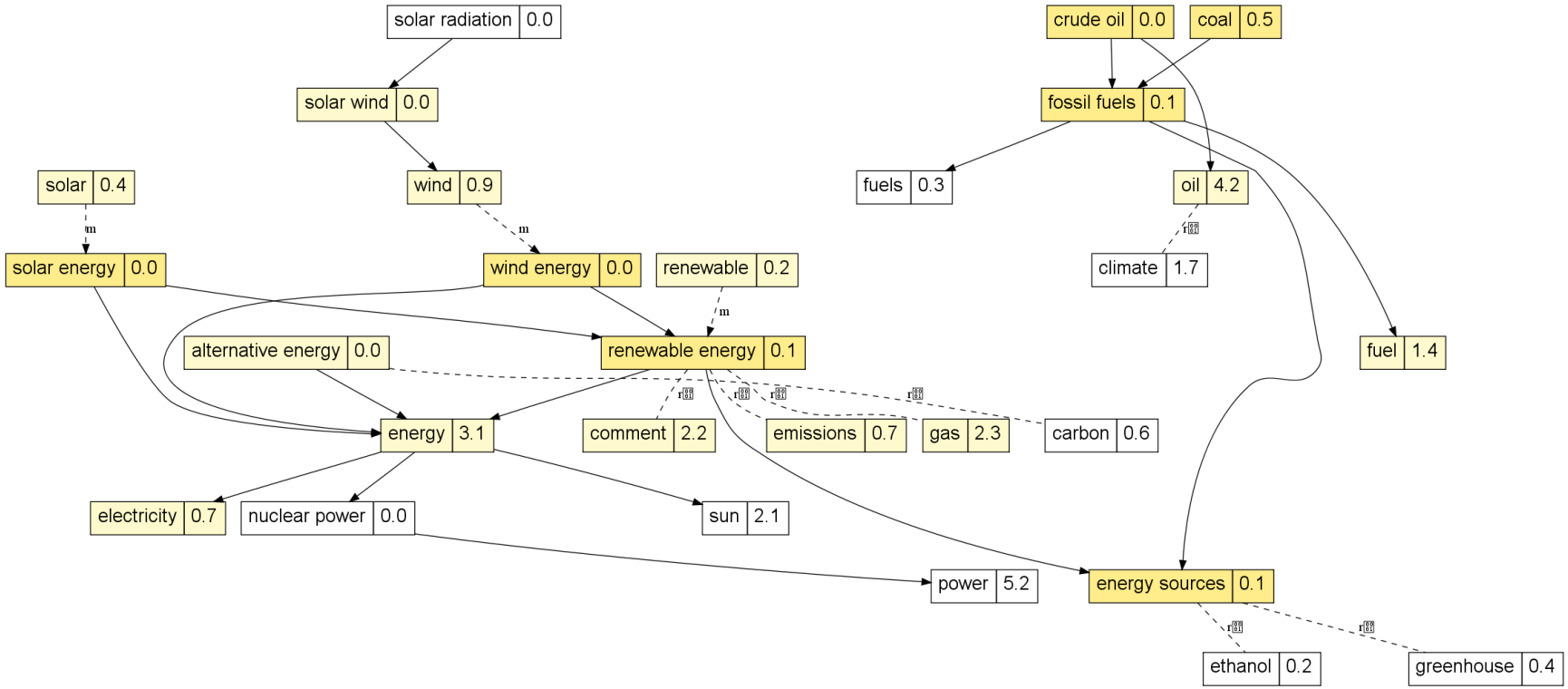


Figure: Extended Ontology (February 2006)

Visualization

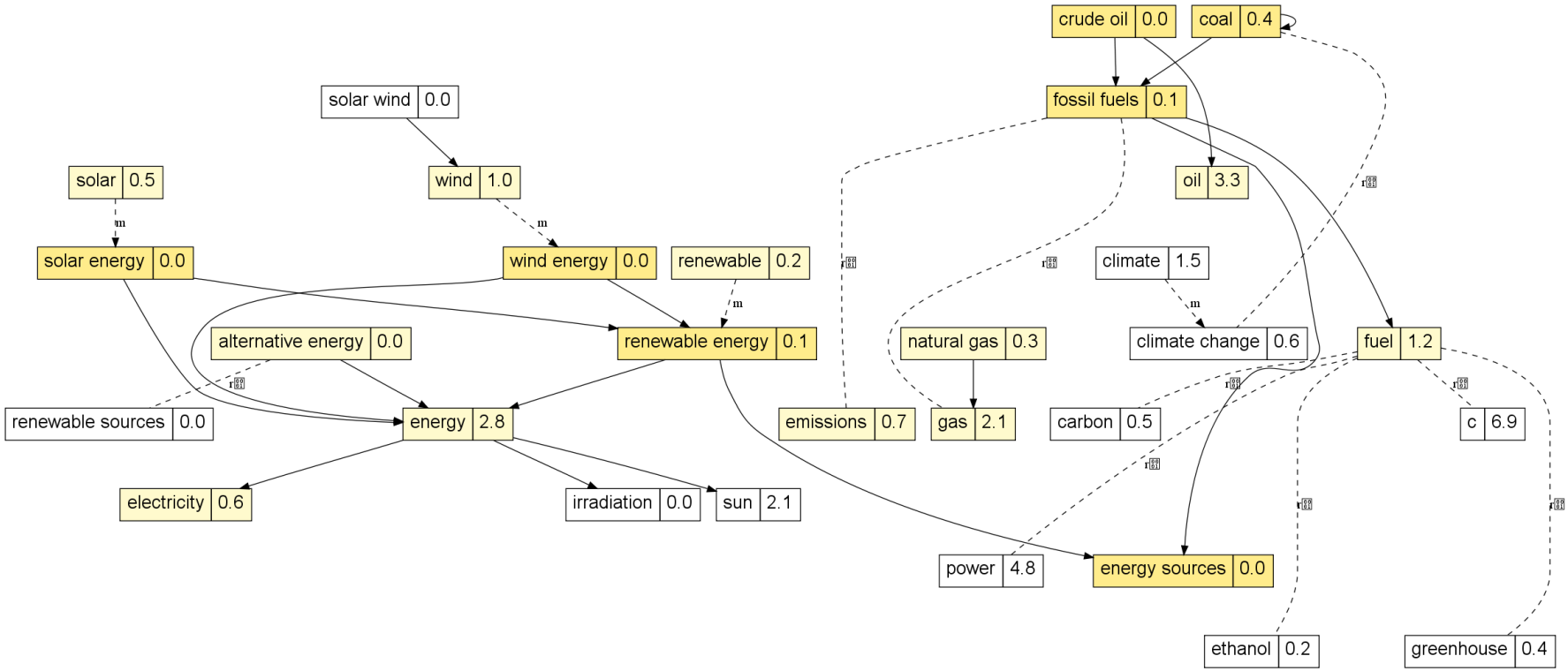


Figure: Extended Ontology (May 2006)

Visualization

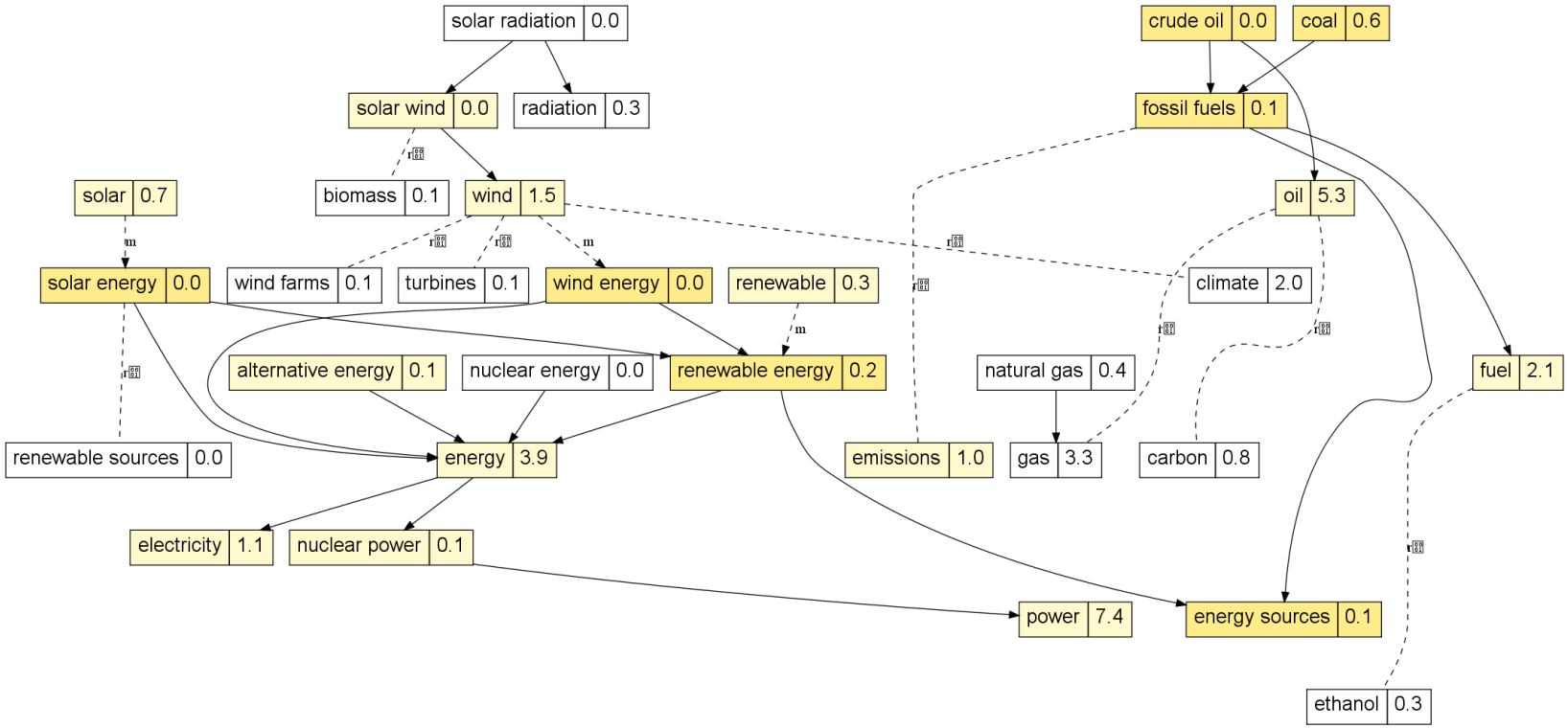


Figure: Extended Ontology (August 2006)

Conclusions

- system for tracking changes in domain ontologies
- visualization
- empirical study (online media)
  - three levels of domain concepts and relations (core, extended and peripheral)
  - observed changes to a term's importance and meaning

Outlook

- tight integration with the *Media Watch on Climate Change*
- formalization of changes to the ontology
  $\longrightarrow$ temporal reasoning
- improvements to the ontology learning component
  - relation type detection
  - user feedback (Δ community versus domain experts)