



InPhO: A System for Collaboratively Populating and Extending a Dynamic Ontology

http://inpho.cogs.indiana.edu/



Mathias Niepert, Cameron Buckner, Jaimie Murdock, Colin Allen
Indiana University, Bloomington, IN, USA



ABSTRACT

InPhO is a system that combines statistical text processing, information extraction, human expert feedback, and logic programming to populate and extend a dynamic ontology for the field of philosophy. Integrated in the editorial workflow of the Stanford Encyclopedia of Philosophy (SEP), it will provide important metadata features such as automated generation of cross-references, semantic search, and ontology driven conceptual navigation.

BACKGROUND

The Stanford Encyclopedia of Philosophy (SEP; <http://plato.stanford.edu>)

- Dynamic, peer-managed reference work with state-of-the-art editorial process using web interfaces
- *Dynamic*: new and revised entries come online each month
- *Substantial and complex*: More than 9 million words of sophisticated humanities content
- *Expert-driven*: More than 1,100 professional philosophers serve as authors and editors

OBJECTIVES

- High quality, expert-reviewed metadata
- Faceted and Semantic Search (“Who were Plato’s students?”, “What philosophical ideas were they concerned with?”)
- Automatic generation of cross-references for SEP entries
- Ontology-driven conceptual navigation and visualization of the underlying ontology

METHODS

The Indiana Philosophy Ontology (InPhO; <http://inpho.cogs.indiana.edu>) is a *dynamic ontology* constructed in a 3-step iterative process which was bootstrapped by creating a small, hand-built formal ontology using the subject-area structure of the SEP and other resources.

1. Statistical methods are run over the entries in the SEP to identify likely relationships among terms. In addition, relations are populated using external sources (e.g., Wikipedia and genealogy projects)
2. Feedback from authors, editors and other experts is used to assess the “uncertain” results of step 1. This feedback is then stored in a knowledge base
3. Logic programming is used to put the pieces of knowledge together to form a global, populated ontology

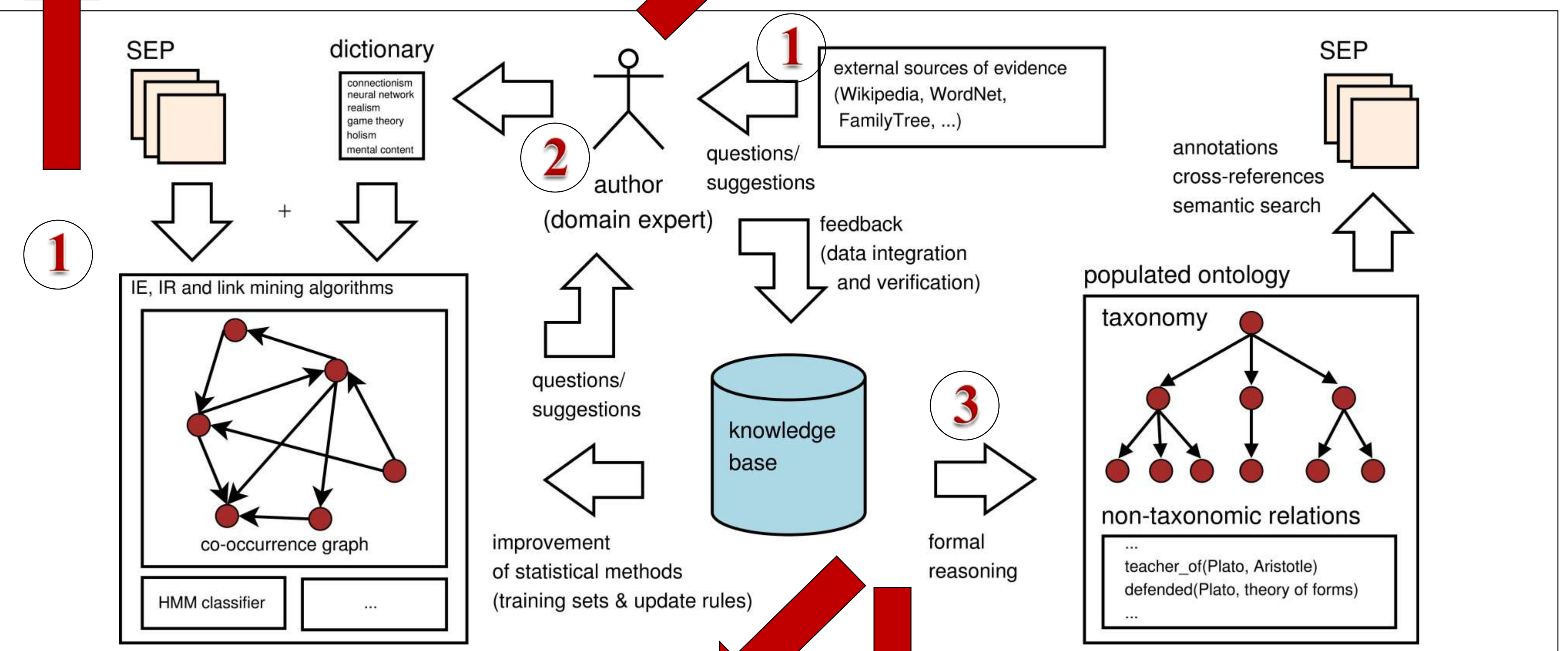
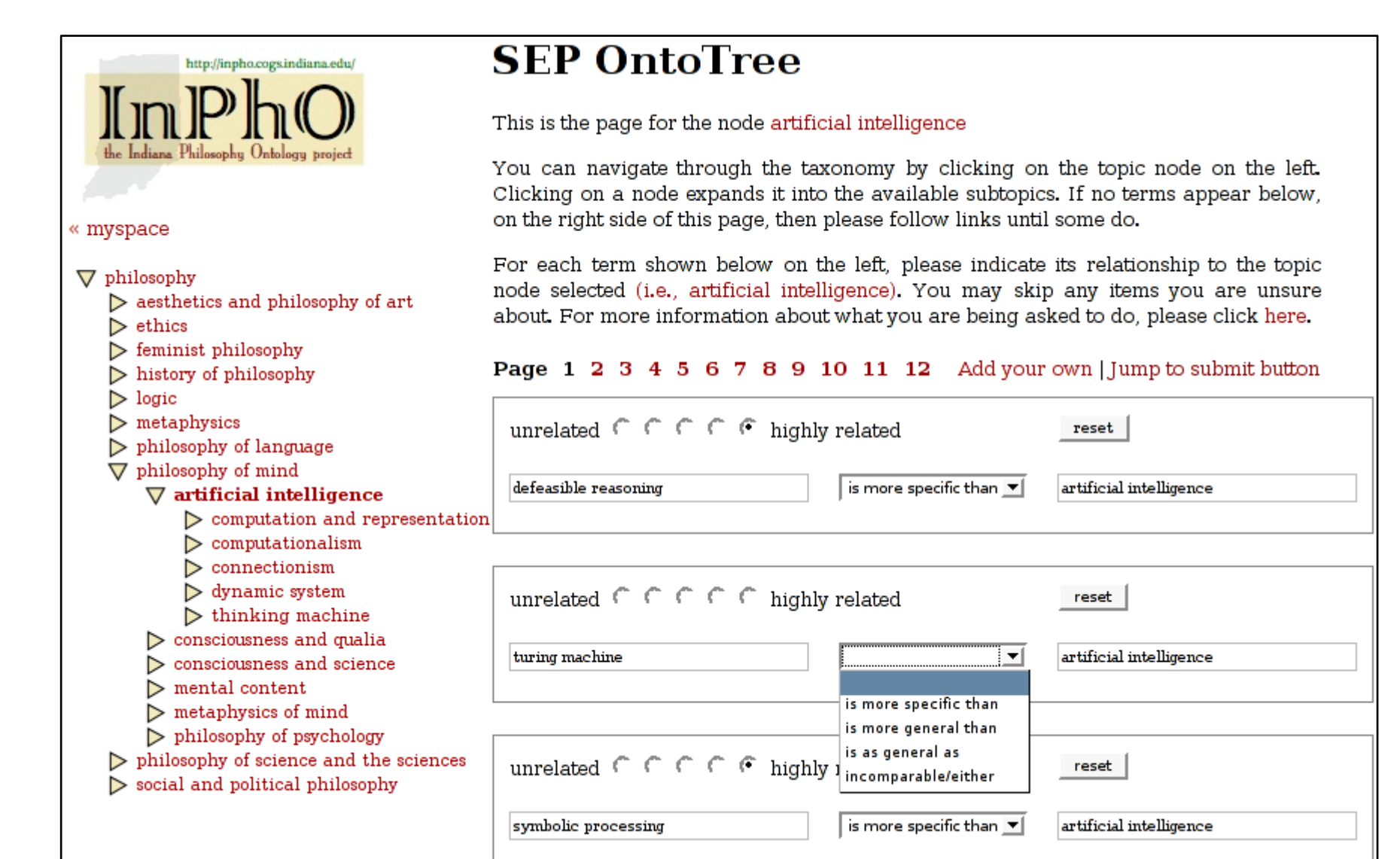
1. STATISTICAL TEXT PROCESSING

We use documents of the SEP supplemented by other sources such as Wikipedia and philosophy genealogy projects to derive candidate instances of the relations of our formal ontology.

- A domain-specific dictionary (documents, philosophers and philosophical ideas) is maintained semi-automatically
- Using the content of the SEP entries, we build a *co-occurrence graph* and use a combination of semantic similarity and node-entropy measures to create candidate instances for hyponymy and hypernymy relationships between terms in the dictionary
- Pattern matching algorithms are applied to Wikipedia entries and other semi-structured sources to derive instances for non-taxonomic relations (e.g., is-teacher-of)

2. EXPERT FEEDBACK INTERFACES

- “Uncertain” recommendations derived in step 1 are presented to authors and editors for verification and integration
- This is done for taxonomic *and* non-taxonomic relations (e.g., is-student-of)
- No knowledge in ontology design is assumed
- Interfaces will be part of the editorial process of the Stanford Encyclopedia of Philosophy



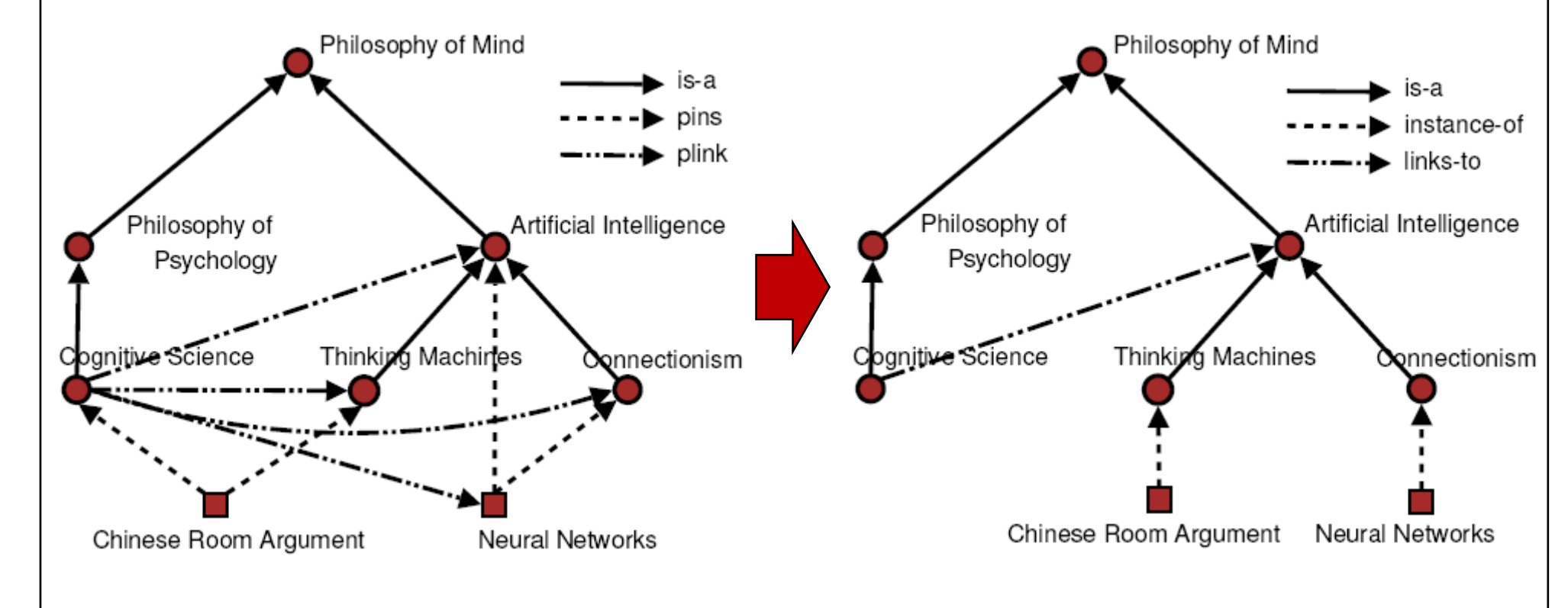
3. ANSWER SET PROGRAMMING

Answer set programming is a non-monotonic logic programming paradigm. Every answer set program consists of three parts:

- *Signature*: predicate symbols (e.g., is-a) and set of objects (here: terms referring to ideas in philosophy)
- *Declaration*: Set of expert feedback facts (e.g., more-specific(Neural Network, Connectionism)) and the facts given by the existing ontological structure
- *Regular Part* (set of rules). *Examples*:
 - more-specific(X, Y) :- more-general(Y, X).
 - possible-instance(X, Y) :- highly-related(X, Y), more-specific(X, Y), class(Y), not class(X).
- Conflicting feedback is possible and modeled using predicate ic (inconsistent):
 - ic(X, Y) :- ms(X, Y), mg(X, Y).

3. ANSWER SET PROGRAMMING (continued)

- Can be used to model “semantic links” between incomparable ideas
- Rules minimize amount of non-taxonomic, semantic links while maintaining reachability of highly related ideas in the taxonomy



ACKNOWLEDGEMENTS

This research has been funded by Indiana University under the grant “New Frontiers in the Arts and Humanities” and by Digital Humanities Start-up Grant HD-50203-07 from the National Endowment for the Humanities.