

# Modeling Concept-Attribute Structure

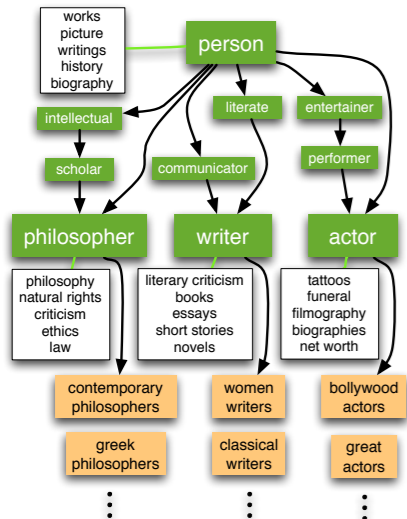
Joseph Reisinger<sup>1</sup> and Marius Pasca<sup>2</sup>  
 The University of Texas at Austin<sup>1</sup> Google<sup>2</sup>

## Abstract

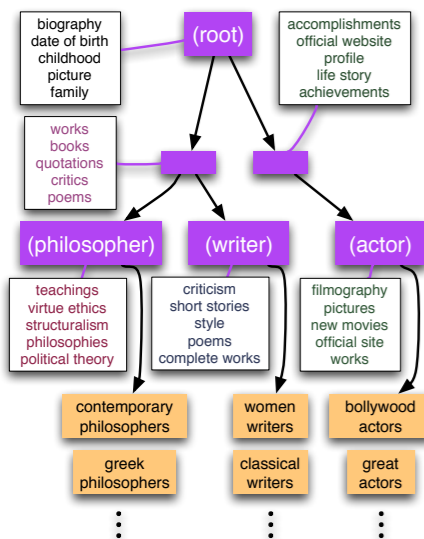
We apply hierarchical Latent Dirichlet Allocation (hLDA) to the problem of ontology annotation; automatically extending WORDNET with new concepts and annotating existing concepts with generic property fields, or attributes. The resulting annotations are evaluated along two dimensions: (1) the precision of the ranked lists of attributes at each concept, and (2) the specificity of the attribute assignments to WORDNET concepts. We find that hLDA and several variants outperform previously proposed heuristic methods, significantly improving the resulting annotation quality, and confirming the ability of topic models to reduce extraction noise.

## Models

### fixed (LDA+wordnet)



### learned (nCRP)



## Labeled Attribute Sets

antineoplastic agents	mechanism of action, solubility, extravasation, contraindications
book publishers	adaptation, scientific name, adaptations, online dictionary, definition
federal agencies	castle, pay banding, locality pay, history, careers, secretary
mammals	digestive system, habitat, life cycle, respiratory system, reproduction
scientific journals	journal, impact factor, definition, archive, ranking, process, picture
shipwrecks	survivors, shipwreck, story, route, sinking, salvage, passenger list
social issues	health risks, cause and effect, definition, cartoons, meaning
special diets	definition, meaning, history, symptoms, low fat recipes, vitamins
turkish cities	population, history, climate, maps, weather, tourism, sightseeing
turtles	respiratory system, life cycle, sickness, habitat, drawing, predators
tyrants	autobiography, early life, childhood, mausoleum, bibliography
vulnerabilities	definition, history, list, different types, prevention, tutorial, statistics
writers	family crest, coat of arms, clan, family tree, bibliography, tartan

## Results

### (1) Attribute re-ranking / noise-filtering / smoothing precision

$$p_{lda}(w|w_d) = \sum_c p(w|c, \eta) p(c|w_d, \alpha)$$

$$p(w|w_d) = p_{lda}(w|w_d) p_{base}(w|w_d) \quad p_{base}(w|w_d) \stackrel{\text{def}}{=} \theta^r(w, w_d)$$

### (2) Concept assignment precision

