Using Semantic Technologies to Construct Research Topicology

Gandhi, Sunil
Peshave, Akshay
Rao, Raghavendra

Introduction

- Motivation: Well connected graph of Authors, Publications and Topics find numerous applications in areas of information mining, pattern analysis and data visualisation.
- Proposal: To create triples representing the above associations and with inferred data using Semantic Web Technology.
- Related Works: Semantic Web Conference Corpus, Google Scholar , PubZone

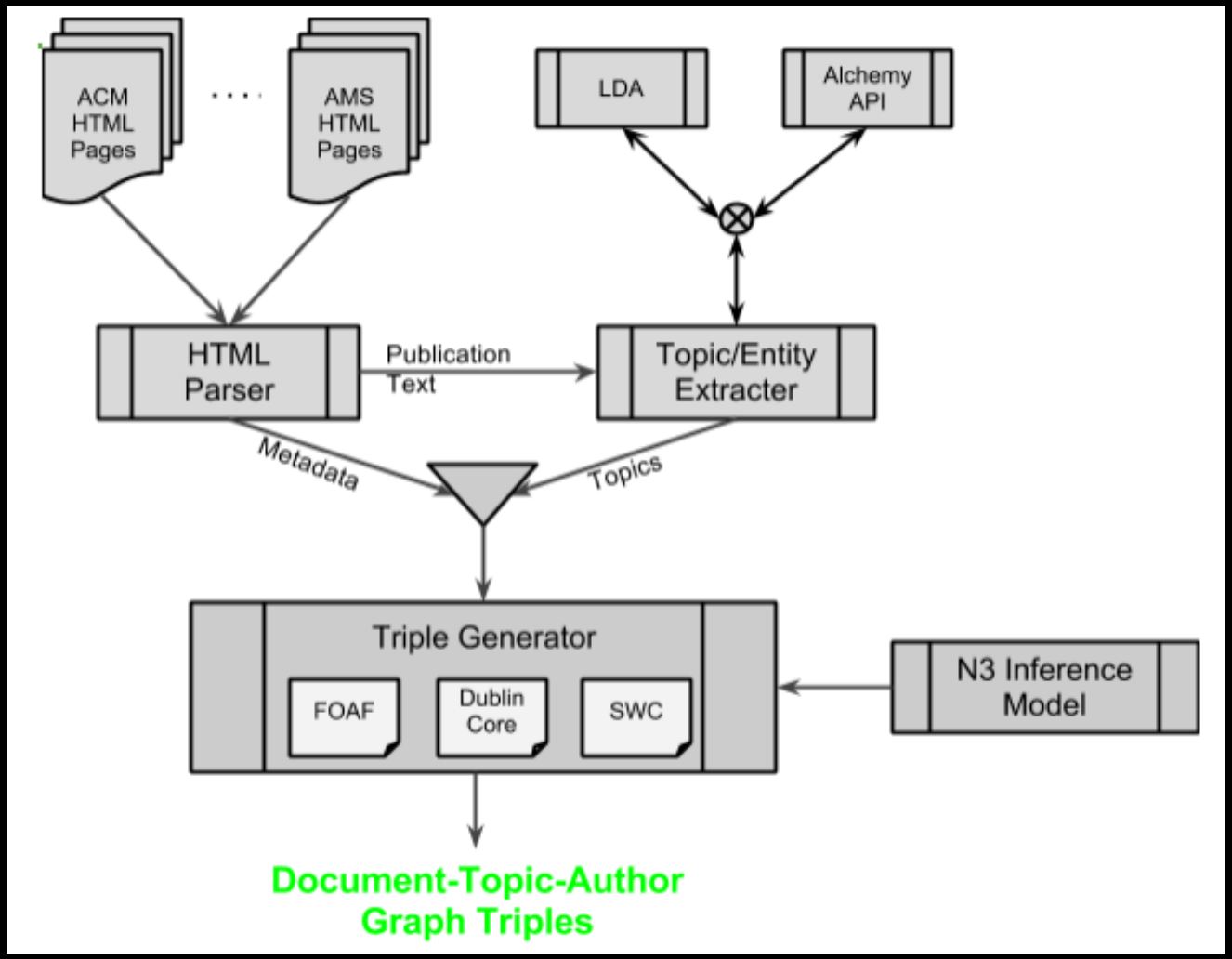
Proposed Approach

Data store: Publication/conference portal ACM

Topic/Entity extraction: LDA (Latent Dirichlet Allocation) libraries and/or Alchemy API

Ontologies: Dublin Core, FOAF and SWF ontology for storing the triples describing authors, topics and documents.

Workflow Diagram



References

[1] Google [Online]. Availabe: www.google.com

[2] Semantic Web Conference Corpus [Online]. Available: http://data.semanticweb.org/

[3] Google Scholar [Online]. Available: http://scholar.google.com/

[4] PubZone [Online]: Available: http://www.pubzone.org/index.do

[5] Latent Dirchlet Allocation [Online]: Availaible: http://www.cs.princeton.edu/~blei/papers/BleiNgJordan2003.pdf

[6] AlchemyAPI [Online]: Availaible: http://www.alchemyapi.com/

[7] Dublin Core Metadata Initiative [Online]: Availaible: http://dublincore.org/

[8] Friend-of-a-Friend Ontology [Online]: Availaible: http://xmlns.com/foaf/spec/

[9] Semantic Web Conference Ontology [Online]: Availaible: http://data.semanticweb.org/ns/swc/swc_2009-05-09.html