

# ESWC 2016



# ESWC 2016

- The 13<sup>th</sup> edition (the 29<sup>th</sup> of May till the 2<sup>nd</sup> of June) at Heraklion, Crete, Greece
- 2 new tracks: Trust and Policy, Smart Cities and GeoSpatial Data
- 47 papers out of 204 submissions
- 15 workshops
- 9 tutorials
- Poster and demo session, PhD symposium

# Tracks

- Linked Data Track
- Machine Learning Track
- Mobile Web, Sensors and Semantic Streams Track
- Natural Language Processing and Information Retrieval Track
- Reasoning Track
- Semantic Data Management, Big Data, Scalability Track
- Services, APIs, Processes and Cloud Computing Track
- Smart Cities, Urban and Geospatial Data Track
- Trust and Privacy Track
- Vocabularies, Schemas, Ontologies Track
- In-Use & Industrial Track

# Vocabularies, Schemas, Ontologies Track

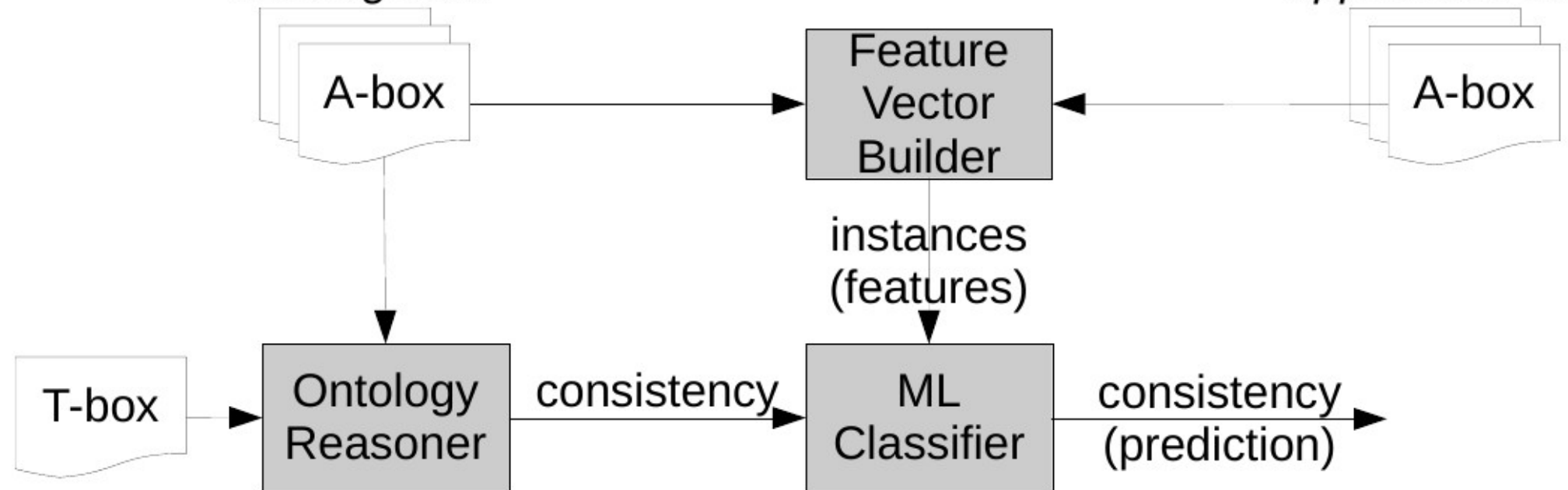
- Alec et al.: An Ontology-Driven Approach for Semantic Annotation of Documents with Specific Concepts.
  - Each document describes an entity of the same domain
  - The goal is to annotate document with concepts
  - SAUPODOC: population of a domain ontology, learning definitions of target concepts based on manual annotations
  - Two domains: destination corpus, film corpus
  - Comparison with SVM, Decision tree (10%-20% improvement)
- Both et al.: Qanary – A Methodology for Vocabulary-Driven Open Question Answering Systems
  - QA system driven by a core QA vocabulary aligned to existing domain ontologies
  - Ontologies aligned using alignment axioms or rules (SPARQL)
- Keet et al.: Test-Driven Development of Ontologies
  - 36 generic test as TBox queries and axioms; implemented as Protégé plugin

# Machine Learning Track

- Paulheim et al.: Fast Approximate A-Box Consistency Checking Using Machine Learning
    - ontology reasoning is a computationally demanding but we can approximate it, e.g. A-Box consistency checking by training a machine learning model which approximates the behaviour of that reasoner for a specific ontology
    - At least 95% accuracy, 2% runtime of a reasoner using decision tree (20 nodes)
    - Validating 293M Microdata documents against schema.org <90 min vs. 18 days by a reasoner
- usable for large number of A-boxes adhering to the same T-box

*training data*

*application data*



# Machine Learning Track

- Ristoski et al.: Enriching Product Ads with Metadata from HTML Annotations
  - Enriching product ads with structured data extracted from thousands of online shops
  - Structured product ads serve as supervision for training feature extraction models (Random Forest, Naive Bayes, SVM, Logistic Regression) working on unstructured product descriptions

# Trust and Privacy Track

- Grzebala et al.: Private Record Linkage: Comparison of Selected Techniques for Name Matching
  - linking based on names while keeping those names encrypted, both on disk and in memory

# Smart Cities, Urban and Geospatial Data Track

- Giannopoulos et al.: Learning to Classify Spatiotextual Entities in Maps
  - To facilitate the annotation process in crowdsourcing map initiatives such as OpenStreetMap
  - Training features: spatial, textual and semantic properties of entities
  - Approaches employed: SVM, kNN, clustering+SVM, clustering+kNN
  - Implemented as JOSM plugin OSMRec for OSM



# Semantic Data Management, Big Data, Scalability Track

- Folz et al.: CyCLaDEs: A Decentralized Cache for Triple Pattern Fragments

- Linked Data Fragment as a trade-off between performance and data availability for querying Linked Data
- LDF clients also caching data during SPARQL query processing but they do not collaborate (decentralized cache hosted by clients is not used)
- CyCLaDEs build a neighborhood cache of clients and it can be used to handle LDF query → significant performance improvement

