Schema Linking

Juan Carlos Rosito Cuellar

Universita' degli studi di Milano-Bicocca

16-4-2019

objectives

- Make a recommendation of schema semantical annotation and the schema linking of a table.
  - Fetch and preprocess the headers of the web table.
  - Make a surrogate model that can annotate all possible entities of each header, within different ontology's entities.
  - Get all the possible decision trees that comprise all the possible entities of the web table.
  - Get the unique Entities, types and relationships that maximize the similarity of the Web table.
  - Considerate how the knowledge graph evolves in time

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Introduction

- Query Input
- Recommender of Vocabulary Term
  - Features for ranking
  - Learning to Ranking
- Query Output

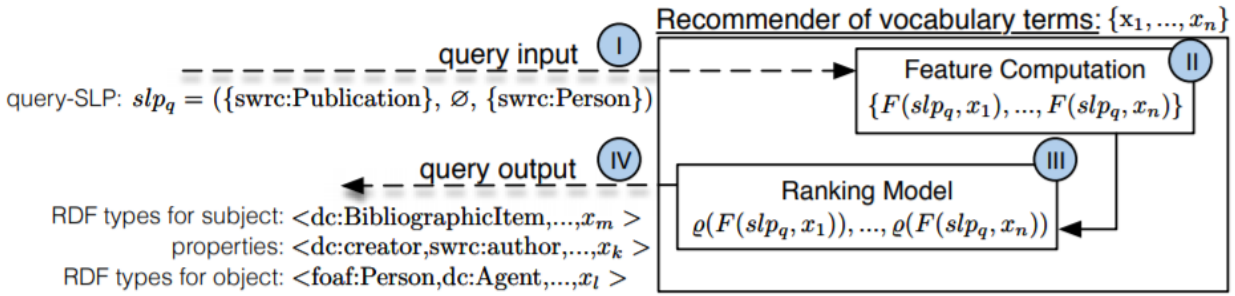

Figure 1: Schema of the process of the query

Introduction: Definitions

Variable	Definition
$\mathbb{V}$	Set of all vocabularies on the LOD cloud
$\mathbb{T}$	Set of all RDF types from all vocabularies in $\mathbb{V}$
$\mathbb{P}$	Set of all properties all vocabularies in $\mathbb{V}$
slp	A schema-level pattern with $slp = (sts, ps, ots)$
sts	Subject type set with $sts \in \mathcal{P}(\mathbb{T})$: RDF types describing a resource in subject position of a triple
ots	Object type set with $ots \in \mathcal{P}(\mathbb{T})$: RDF types describing a resource in object position of a triple
ps	Property set with $ps \in \mathcal{P}(\mathbb{P})$: properties interlinking resources of types in sts and ots
$\mathbb{DS}$	The set of datasets that are published on the LOD cloud
G	A graph representing a dataset such that $G \in \mathbb{DS}$
(s, p, o, c)	An RDF quadruple consisting of a subject, property, object, and a context URI where G can be found

Figure 2: Tabular overview of the variables

Introduction:

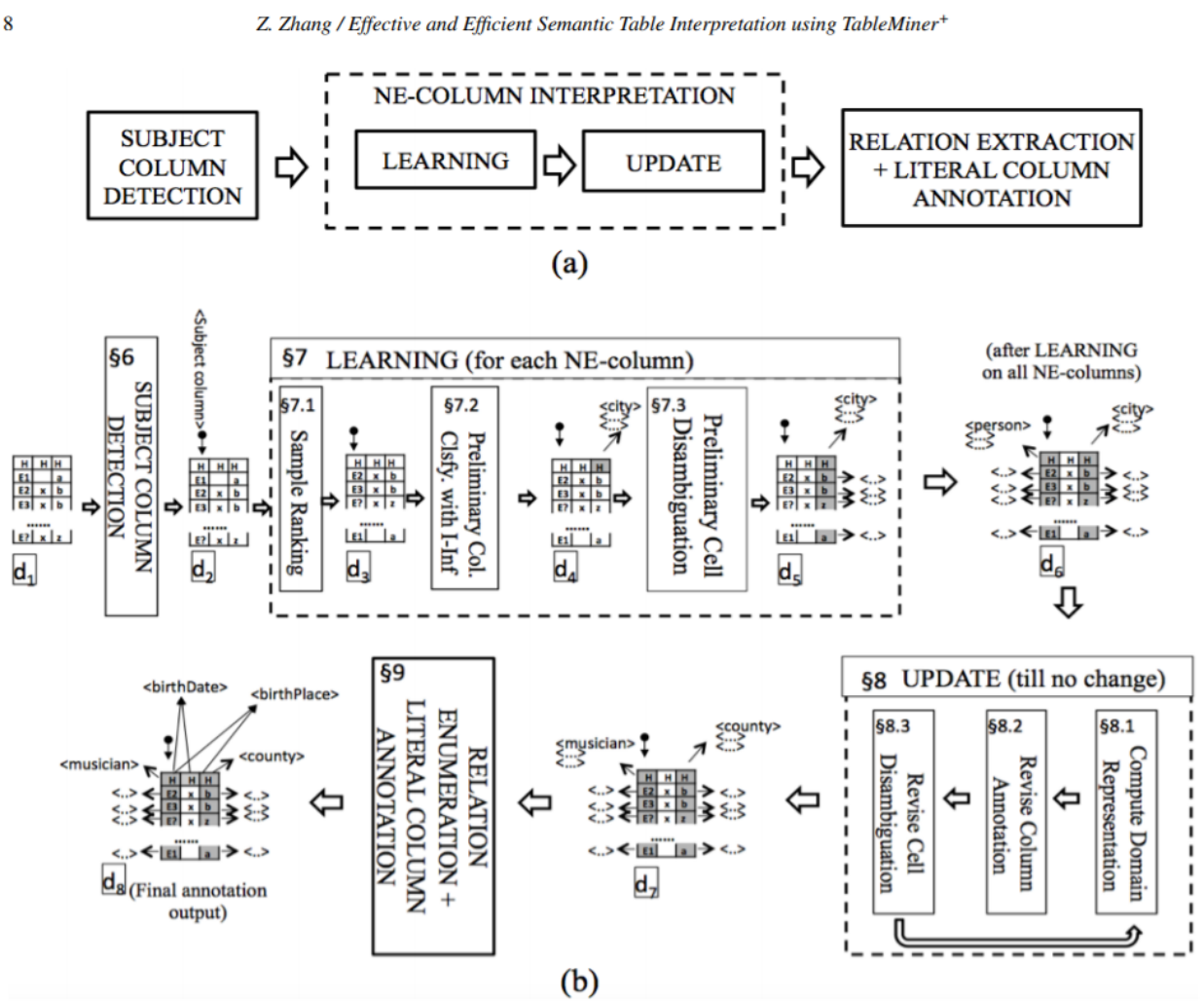

Fig. 2. The overview of TableMiner+. (a) a high-level architecture diagram; (b) detailed architecture with input/output. d - table data. Grey colour indicates annotated table elements. Angle brackets indicates annotations. Inside a table: H - header. E, a, b, x, z - content cells

Figure 3: TableMiner design by Z.Zhang

Introduction

Introduction

Dataset ABSTAT and ontologies

Conclusion