Structuring Job Search via Local Grammars

Job Search Problem
- Job Announcements are Structured Documents
- Real World Web Documents are Still Semantically Flat
- BAG-OF-WORDS Indexing Leads to NOISE
- Local Grammars Generate Structured Forms
- Forms Provide Efficient Information Access

Locating Online Job Offers

Local Grammar that models the context of the word “leader”.

Pipeline of Classification and Transformation
- URLs
  - Preprocessing
  - Correction and Normalization
  - Applying Local Grammars
  - Information Extraction
  - Classification

Pattern extraction:
- Date of publication
- Last date of application
- Employment period
- Company name
- Office: address, e-mail, phone, fax, URL
- Company sector
- Contact details
- Job description
- ...