

ContactFiller: Extracting Structured Entity Profiles from Unstructured Sources

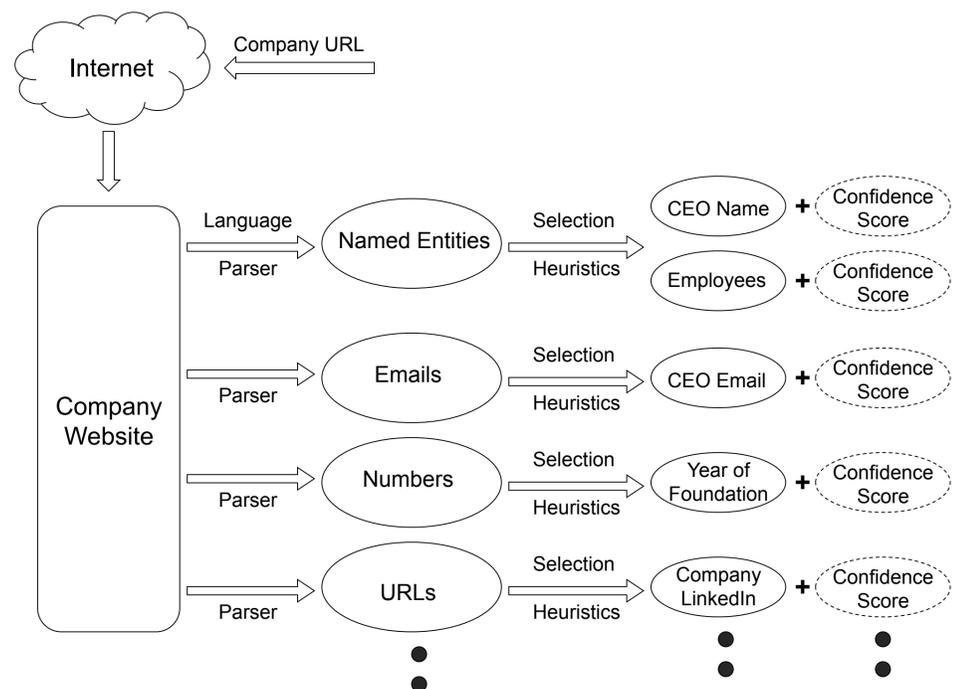
PROBLEM / MARKET NEED

- Information is key in the modern commercial landscape and often we have very little time to do due diligence about a prospective client or business contact before or after a meeting.
- There is a need for a rapid information extraction system that takes a set of information about a person or business and searches the Web and other unstructured data sources for more specific information about that entity.
- While there are existing solutions that allow business cards to be scanned and the information contained in them to be extracted, there are not tools that fully exploit online information resources in a targeted way.

- Technologies involved in this system:
 - Web crawler explores massive collections of web pages for potentially relevant information. The extractors allow for the targets of retrieval to be meta keywords which lead to specific content.
 - Named entity recognition (NER) is applied over unstructured text sources in order to disambiguate term references.
 - Conflict resolution is applied with a selection of metrics which generate confidence scores for each candidate.

TECHNOLOGY SOLUTION

- In order to address end users objectives for information browsing, we have developed a software tool for assisting in extracting web-based company information.
- The objective is to retrieve a list of selected fields of a named entity or an organisation. For example, for a given company, the tool will retrieve properties such as
 - Name of CEO
 - CEO's email address
 - Employee names
 - Year of foundation
 - Social media handles
- We propose three aspects to a solution for this objective:
 - Natural language processing technologies that can find points of information from unstructured sources.
 - Technologies that can be used to verify that information from a range of sources actually relates to the same person.
 - Techniques to aggregate possibly conflicting information into a single record.



THE FUTURE OF THIS TECHNOLOGY SOLUTION

- The ContactFiller project is currently tailored to retrieve a list of selected information of a company, but it is designed as a general approach to extract structured entity profiles from unstructured sources. This technology can be easily extended to retrieve information from diverse unstructured sources.

RESEARCH TEAM

Dr Jing Su, UCD

Dr Oisín Boydell, UCD

Dr Brian MacNamee, UCD