

A Context-Aware Approach to Entity Linking

Veselin Stoyanov and James Mayfield
HLTCOE, Johns Hopkins University

Tan Xu and Douglas W. Oard
Univ. of Maryland, College Park

Dawn Lawrie
Loyola University, Maryland

Tim Oates and Tim Finin
Univ. of Maryland, Baltimore County



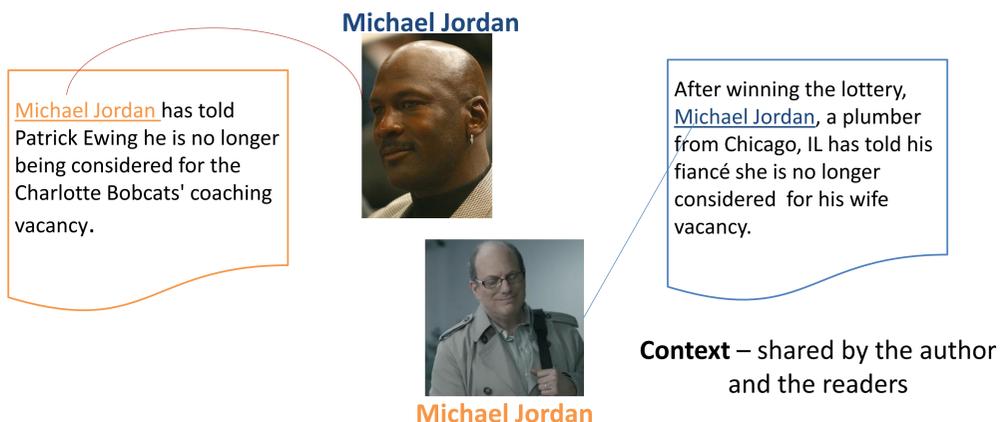
The Importance of Context

Grice's principle for collaborative communication:

- Maxim of Quantity: a contribution should be as informative as is required, but no more.
- Maxim of Manner: one should avoid ambiguity.

Thus, communications are only possible when the author and the audience **share a discourse context**, and entity mentions should be unambiguous in this shared context.

Newswire/Blog Communications



Context-Aware Linker of Entities (CALE)

We propose a new system architecture that explicitly models the notion of context.

Design principles:

1. **Shared context** should be modeled explicitly.
2. Most entity linking should be **trivial** in the shared context.
3. Discourse context implies **joint inference**.

The shared context depends on factors such as the type of communication, the author, and the intended audience.

What is a Context?

For practical purposes, we model a context as a weighted set of entities.

- The time and place of the communication
- The topic of the communication

Our system deals with three types of context:

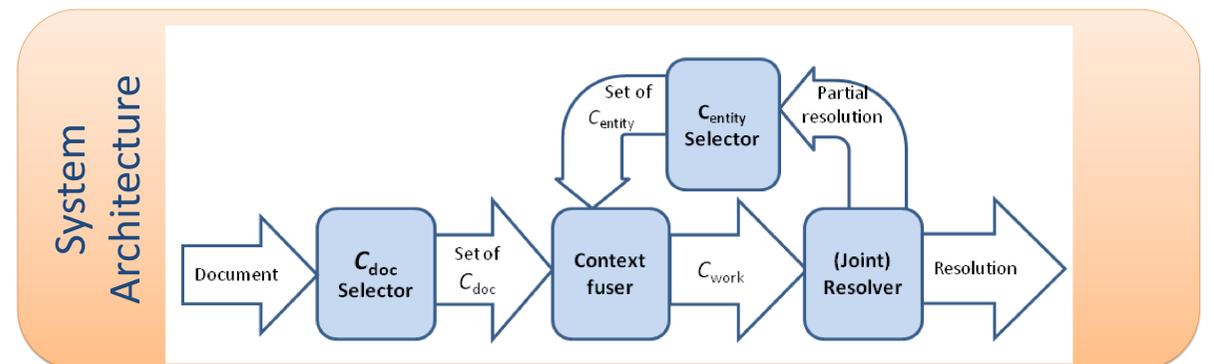
- 1) C_{work} (working context) – the context with which the system is currently resolving mentions.
- 2) C_{doc} (document context) – context triggered by a specific document.
- 3) C_{entity} (entity context) – context associated with a given KB entity.

We utilize *context triggers*: functions that produce a C_{doc} or C_{entity} a given a document or an entity, respectively. For example:

- A NYT trigger that invokes a context of the most popular entities in the New York Times
- A topic trigger that invokes a context of the most popular entities given a topic (e.g., statistics)
- An entity trigger that returns a context of the entities associated with, for example, Jim Halpert.

A context can depend on:

- The intended audience for a communication



Evaluation

We have implemented a prototype system and evaluated it on two tasks with promising results:

- (1) Linking entity mentions in the 2010 TAC knowledge Base Population collection.
- (2) Linking e-mail entity mentions in the Enron collection.