Across-Document Neighborhood Expansion for Candidate Retrieval
Laura Dietz dietz@cs.umass.edu and Jeffrey Dalton jdalton@cs.umass.edu

Abstract
Last year’s competition demonstrated that the NER context contains important information that should not be ignored in entity linking. State-of-the-art approaches anchor on unambiguous entities, look for overlap in categories, or approximate a joint model of candidate assignments, after Wikipedia candidates have been selected. Current candidate approaches, such as anchor text maps, are effective but may lead to very large candidate sets to be examined. UMass has two objectives for our TAC submission. First, we use cross-document context information to perform entity neighborhood expansion and estimate the importance of entity context using corpus-wide information. Second, we use probabilistic information retrieval that incorporates the neighborhood information to generate a ranked candidate set in a single step. The result is a small candidate set that even for less than 50 candidates contains the true answer in 95% of the cases, allowing for time intensive re-ranking methods!

Motivating Example: Relevance of NERs
Motivating Example: Relevance of NERs

Example Query: ABC shot the TV drama “Lost” in Australia.
Candidates:
- Australian Broadcasting Corporation Television
- American Broadcast Central

“Australia” is an unambiguous entity
But: “Australia” is not relevant for American Broadcast Central.
Danger to lead to the wrong conclusion.

Results as Candidate Retrieval
Neighborhood Expansion retrieves the true entity at high cutoff rates
MRR 0.75 (versus 0.72) 95% recall at rank 45
Small candidate set allows for time intensive re-ranking methods!

Results as Entity Linking System 2012
micro-avg precision B^3+F1

Results as Entity Linking System 2011
micro-avg precision B^3+F1

Joint Neighborhood Assignment Models
For each NER span: Assuming candidate set is retrieved
Goal: find joint assignment that maximize likelihood

Candidate Retrieval Model
Mention t, name variants v, sentences s, NER spans e
component weights k, relevance weights φ

Candidate Retrieval with Neighborhood Expansion
Neighborhood expansion estimates reliability for disambiguating the query mention.
No candidate set necessary! joint assignment model is optimized during candidate retrieval!