My Thesis Topic

By
Toshihiro Kuboi

Genealogical Relation Extraction

Make a computer extract genealogical relation information from plain text automatically.

Genealogy relation - parents, children, etc. relationship between persons

Plain Text - non-structured text such as articles on Wikipedia or Novels

Why is the topic intersting?

Because it is very difficult to make a computer to read and understand text as we do!

- Named Entity Recognition:
  - The computer needs to be able to tell nouns referring to persons apart from those referring to non-persons.
- Anaphora Resolution:
  - The computer needs to be able to determine who a personal pronoun is referring to.
- Synonym resolution:
  - There are many terms to describe genealogical relationships.

Related Works

Previous works only tried to extract genealogical information from structured data such as census data.

Few work has covered the extraction from plain text.

There are works which tried to extract named entity relation information from plain text.

Related Works

- (Brin, 1998) proposed DIPRE, which creates information extraction pattern for author-title relation based on information extracted from the web, containing seed words.
- (Agichtein and Gravano, 2000) proposed Snowball, which is based on DIPRE, but uses more flexible pattern matching strategy.
- (Etzioni et al., 2005) proposed KnowItAll, which classifies its own training examples using domain-independent extraction patterns to generate candidate facts, which are then automatically tested for plausibility using PMI statistics.

Any Questions?