Towards an Automated Approach to Extract and Compare Fictional Networks: An Evaluation

Marcello Trovati and James Brady

School of Computing and Mathematics
University of Derby

Outline

- During this presentation I will discuss
  - What it is we set out to do
  - Our approach
  - Some examples of algorithms and look up tables which we used/created
  - Our evaluation/conclusions
  - Further research/development

Introduction

This paper is concerned with extraction of text from a fictional work leading to categorisation and understanding of relations between characters and the types of relations such as friendly and/or hostile ones.

Our Approach

- Generally
- Collocation and Character Identification
- Relation Extraction
- Identification of Relation Types

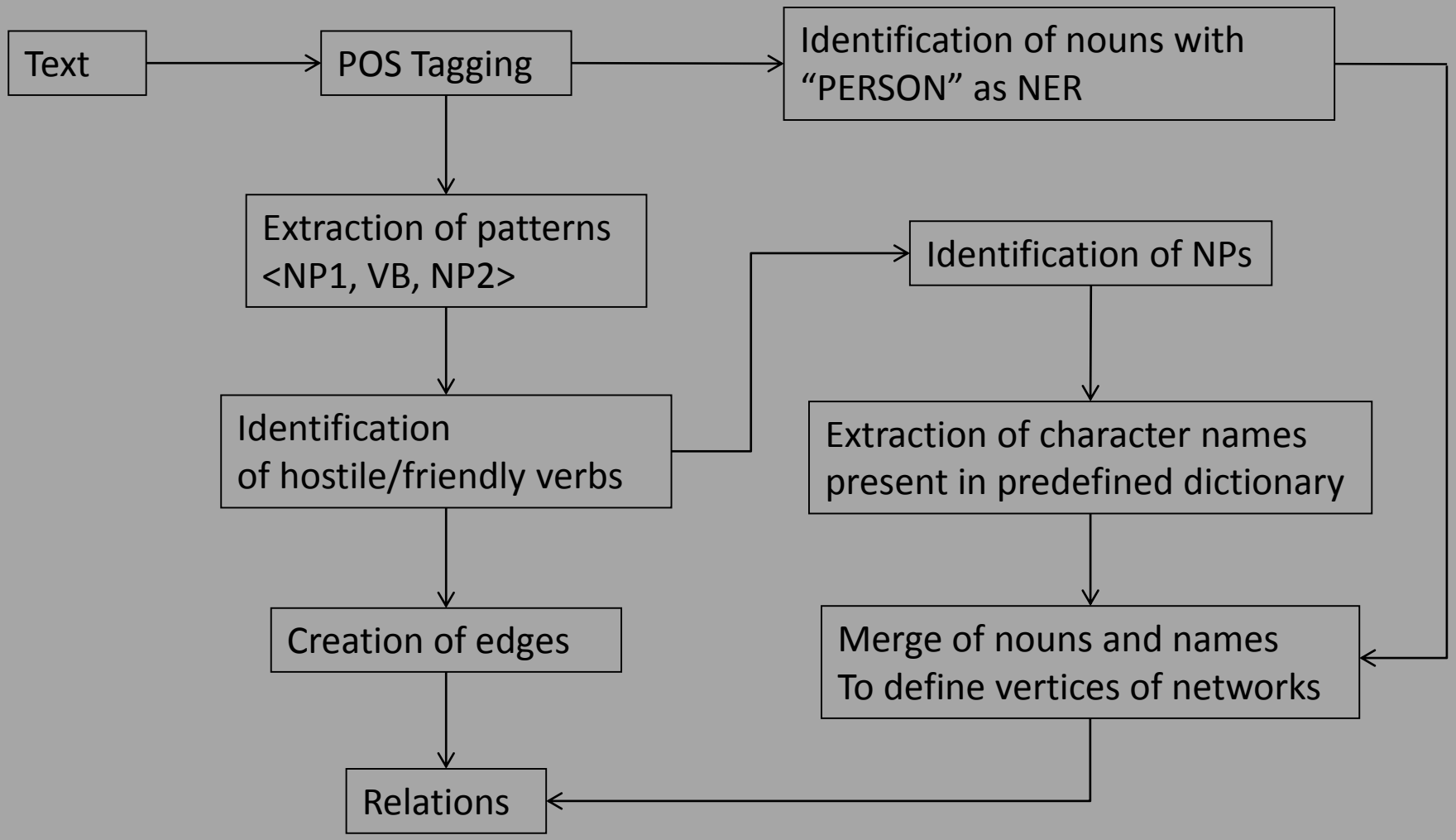
A selection of friendly and hostile verbs from a constructed dictionary

Hostile Verbs	Friendly Verbs
kill	accommodate
hate	admire
injure	aid
harm	approve
attack	cherish
annihilate	cooperate
asphyxiate	collaborate
assassinate	cuddle
crucify	esteem
drown	fondle
eradicate	glorify
erase	idolise
execute	kiss
exterminate	love

Text Patterns for Term Extraction

Tag Pattern
Adj Noun
Noun Noun
Adj Adj Noun
Adj Noun Noun
Noun Adj Noun
Noun Noun Noun
Noun Prep Noun

Relation Extraction Algorithm



Identification of the relation types

- Friendly
- Hostile
- Other/Unknown

Type Aggregation Algorithm

Remove any relation that has not yet been recognised as either friendly or hostile (i.e. unknown) and group the remaining ones into a set, L

Let $l_1, \ldots, l_j$ be its (disjoint) subgroups whose elements *are* the triples referring to the same two characters. Thus,

$$\bigcup_{i=1}^{j} l_i = L, and \sum_{i=1}^{j} |l_i| = |L|$$

$$where\ |\cdot|\ refers\ to\ the\ cardinality\ of\ a\ set$$

for all $l_i, 1 \leq i \leq j$ **do**

assign the most frequently occurring type to the relation between the two characters

end for

Evaluation

- 89% precision
- 68% recall
- 55% of correct extractions had correct type
- 9% reduction in average path length

Discussion

- Clustering properties
- General dynamical properties
- Identification of general topological features

Conclusion

- Initial steps of automated extraction
- Limited capabilities

- Future Research
  - Wider set of relations
  - Directed networks
  - Improvements to type aggregation
  - Similarity measure

Thank you for your attention and especially for still being here at the very end of the very last day!