

# Coreference Resolution for Offline Answer Extraction using Instances



W.A.R. Bergen, Norway  
September 29<sup>th</sup> 2005

Lonneke van der Plas and Jori Mur  
Humanities Computing  
University of Groningen



# Overview

- Question Answering (QA)
- Off-line Answer Extraction
- Coreference Resolution using Instances
- Evaluation
- Results
- Conclusion
- Future Work



# Question Answering

- Task: Find an answer in a corpus to a question posed in a natural language
- Typical questions:
  - Where was Nelson Mandela born?
  - How many inhabitants has Brazil?
  - What is the capital of Russia?



# Question Answering

- Corpus: Newspaper text
- Example of text containing the answer
  - Where was Nelson Mandela born?

In a recent television broadcast BBC commentator Brian Walden argued that **Nelson Mandela**, born in the black homeland of **Transkei** (...)



# Offline Answer Extraction

Extract answers **before** question answering process has started

Abbreviation	Function
Age	Inhabitants
Capital	Location
Currency	Nobelprizewinners
Date of Birth	Place of Birth
Date of Death	Place of Death
Founders	Died age



# Offline Answer Extraction

Extract Answers by looking at frequently occurring patterns

- [name], born in [answer]
- [name], who was born in [answer]
- [name] was born in [answer]



# Offline Answer Extraction

Extract Answers by looking at frequently occurring patterns

- [name], born in [answer]
- [name], who was born in [answer]
- [name] was born in [answer]
- [name]'s new book contains autobiographical elements. **The writer** was born in [answer]



# Coreference Resolution

- **Todd Martin** was the opponent of the quiet **Ivanisevic** during the finals yesterday. The American, who defeated the local hero **Boris Becker** a day earlier, had no chance to beat the **26-year old Croatian**.
- **[name]**, 26 years old



# Coreference Resolution using Instances

- Van Gogh **ISA** painter
- Ivanisevic **ISA** Croatian
- George Bush **ISA** president of the US
- **Apposition** relation: George Bush, the president of the US, arrived yesterday
- **Predicate** relation: Van Gogh is a famous Dutch painter



# Coreference Resolution

- Todd Martin was the opponent of the quiet **Ivanisevic** during the finals yesterday. The American, who defeated the local hero Boris Becker a day earlier, had no chance to beat the **26-year old Croatian**.



# Extracting more facts

## 1. Adjusting Patterns

[answer]-year old [name] -> [answer]-year old [noun]



# Extracting more facts

## 1. Adjusting Patterns

[answer]-year old [name] -> [answer]-year old [noun]

## 2 Selecting name

DocId

Ivanisevic ...  
Boris Becker ...  
26-year old Croatian

Boris Becker



# Extracting more facts

## 1. Adjusting Patterns

[answer]-year old [name] -> [answer]-year old [noun]

## 2 Selecting name



## 3 Check if name occurs in combination with predicate in ISA-database

✗ Boris Becker ISA Croation  
Ivanisevic ISA Croation



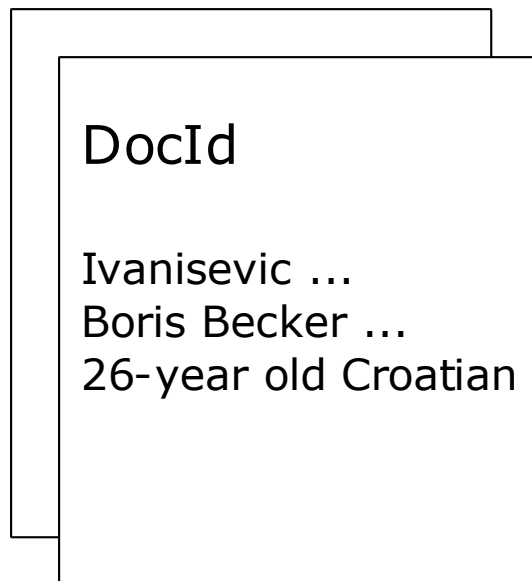


# Extracting more facts

1. Adjusting Patterns

[answer]-year old [name] -> [answer]-year old [noun]

2 Selecting name



3 Check if name occurs in combination with predicate in ISA-database

✗ Boris Becker ISA Croation  
Ivanisevic ISA Croation



4 Store name and answer (age) as a table entry

Ivanisevic	26	DocId



# Evaluation

- Baseline: no ISA-filter -> antecedent is first preceding person name
- How many more facts are extracted compared to original table
- Estimate precision
  - How many correct facts
  - How many incorrect facts



# Categories

Abbreviation

Age

Capital

Currency

Date of Birth

Date of Death

Founders

Function

Inhabitants

Location

Nobelprizewinners

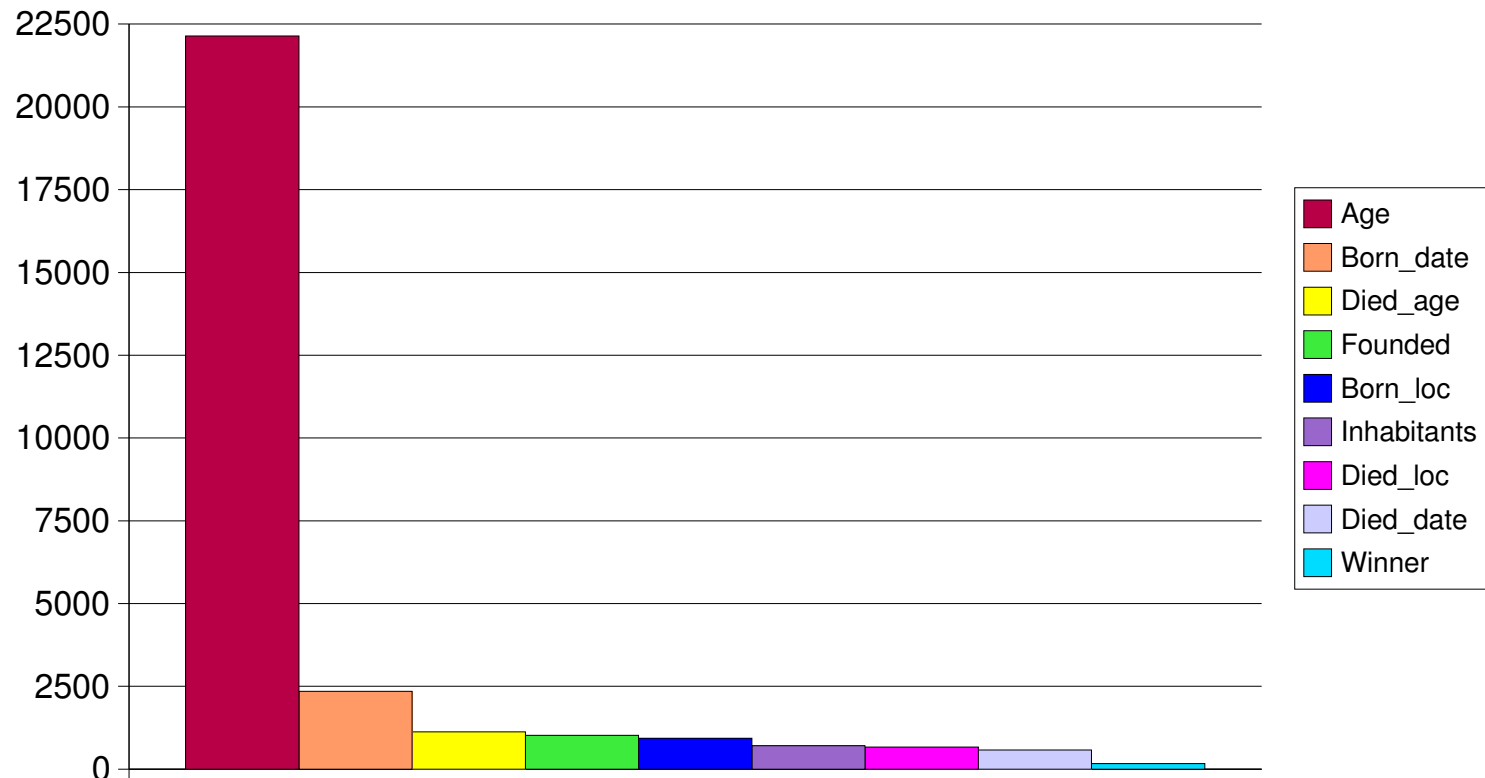
Place of Birth

Place of Death

Died Age



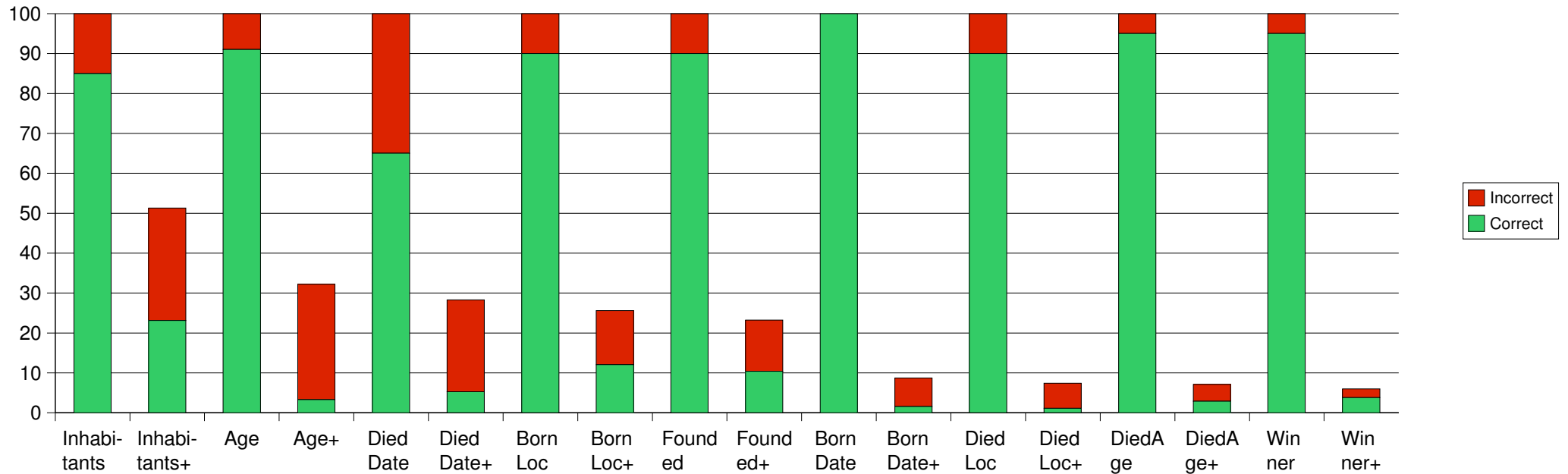
# Size of tables





# Results

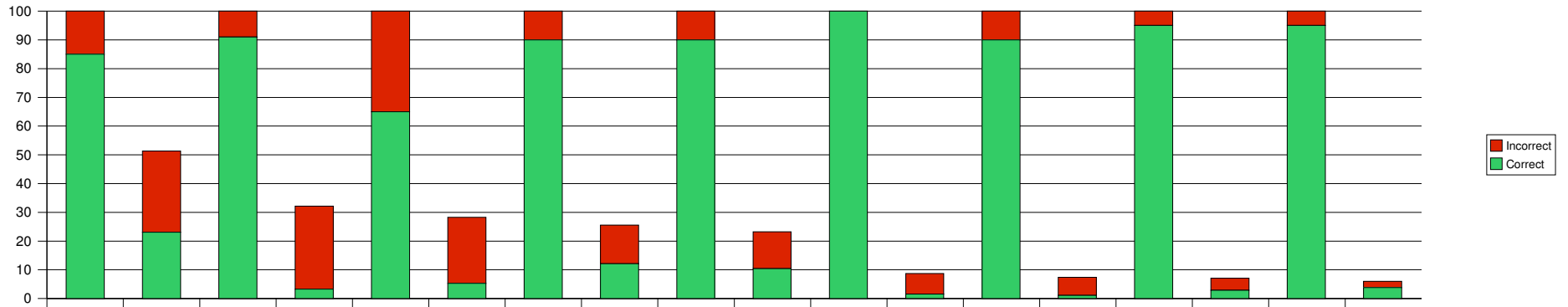
### Coreference without Filter



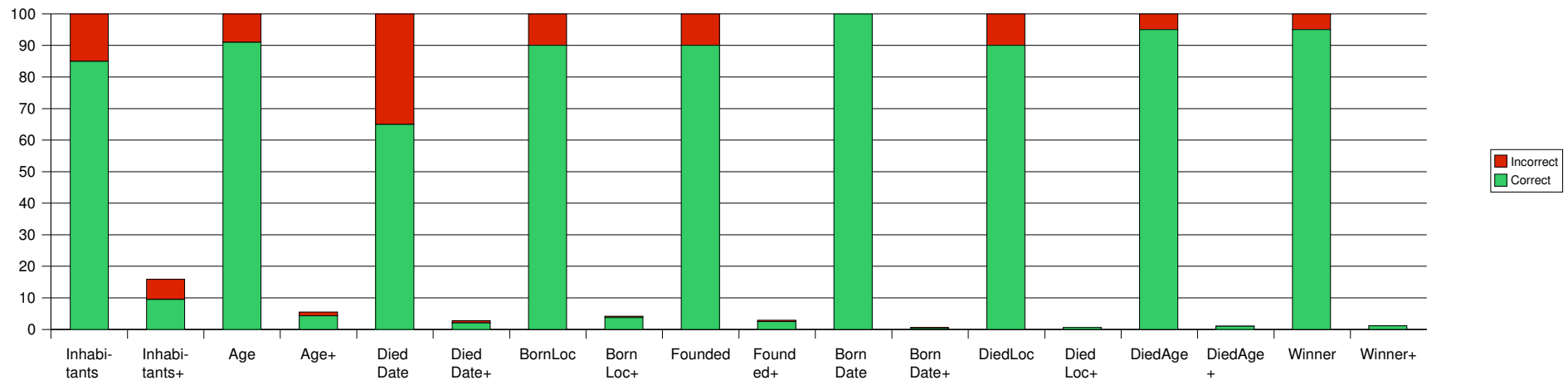


# Results

Coreference without Filter



Coreference with Filter





# Conclusion

- More facts are extracted using Coreference Resolution
- It hurts recall to apply filter, but increases precision a lot
- Age is exception



# Future work

- How will it affect question answering?
  - Without filter better for QA?
- Is it useful to extend it to more categories?
- Should we make use of the syntactic information?
- Is it helpful to look at similar words?
- Could it improve existing techniques for coreference resolution?