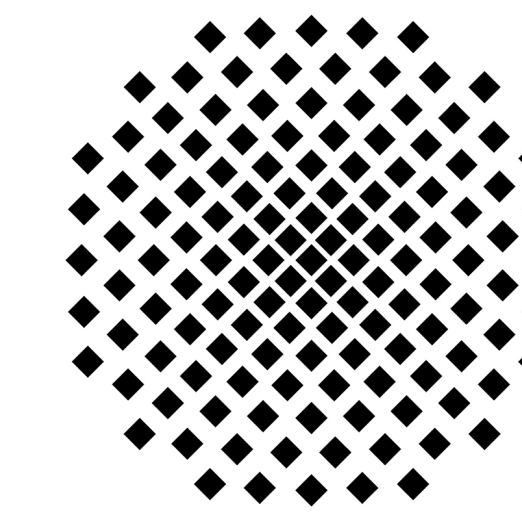


Visualization, Search, and Error Analysis for Coreference Annotations

Markus Gärtner, Anders Björkelund, Gregor Thiele, Wolfgang Seeker and Jonas Kuhn
 Institut für Maschinelle Sprachverarbeitung, University of Stuttgart
 firstname.lastname@ims.uni-stuttgart.de



University of Stuttgart
 Germany

EXPLORATION VIEWS

Text View

Presenter Text (bn/cnn/01/cnn_0140); part 000 [predicted/-]

Highlight Type Background

#begin document (bn/cnn/01/cnn_0140); part 000
 Two months after [a bomb] blasted [its hull], [the "USS Cole"] is back in the United States. [The damaged ship] was carried from Yemen aboard a Norwegian transport ship. 17 sailors were killed when [the ship] was attacked by suicide bombers in the port of Aden. [A temporary patch] has already been made to cover [the 40] by 40 hole in [the ship's hull]. [It] will be welded to [the ship] before [it] is unloaded from [the carrier]. Repairs in Mississippi are expected to cost more than \$ 150 million and last a year at a ship - building facility in Pascagoula, Mississippi.

- Textual representation with color markup
- Customizable formatting and text properties
- Filter out singletons or clusters that are of no interest

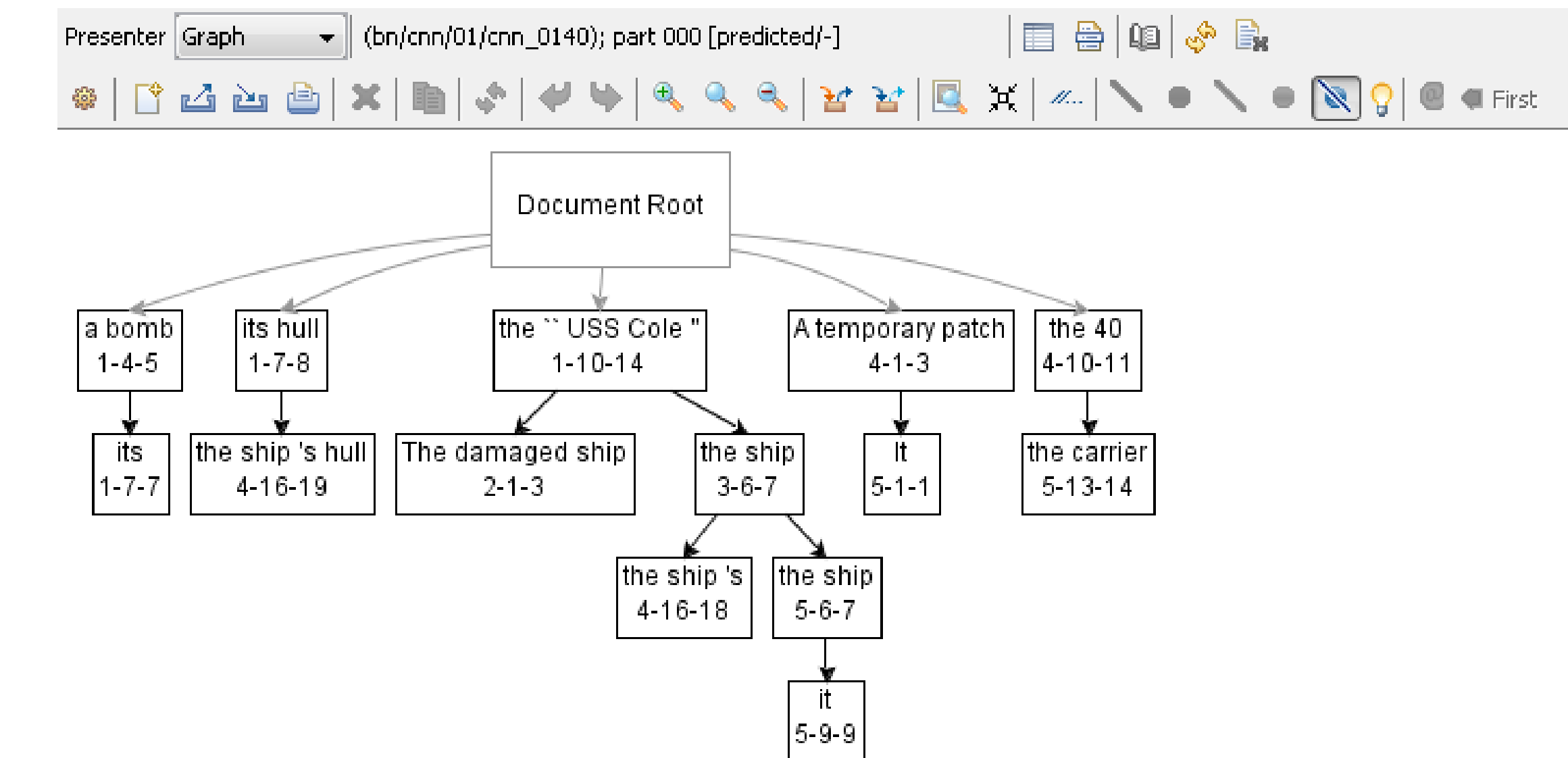
Entity Grid View

Presenter Entity-Grid (bn/cnn/01/cnn_0140); part 000 [predicted/-]

	a bomb	the ship 's hull	the " USS Cole "	A temporary patch	the carrier
1	[a bomb - Common,its - Pronoun]	[its hull - Common]	[the " USS Cole " - Common]		
2			[The damaged ship - Common]		
3			[the ship - Common]		
4		[the ship 's hull - Common]	[the ship 's - Common]	[A temporary patch - Common]	[the 40 - Common]
5			[the ship - Common,it - Pronoun]	[It - Pronoun]	[the carrier - Common]
6					

- Tabular view inspired by Barzilay and Lapata's *Entity Grid* [1]
- Lists entities as columns and sentences as rows
- Cell contents customizable via *Label Patterns*
- Graphical error summary when used for error analysis

Tree View

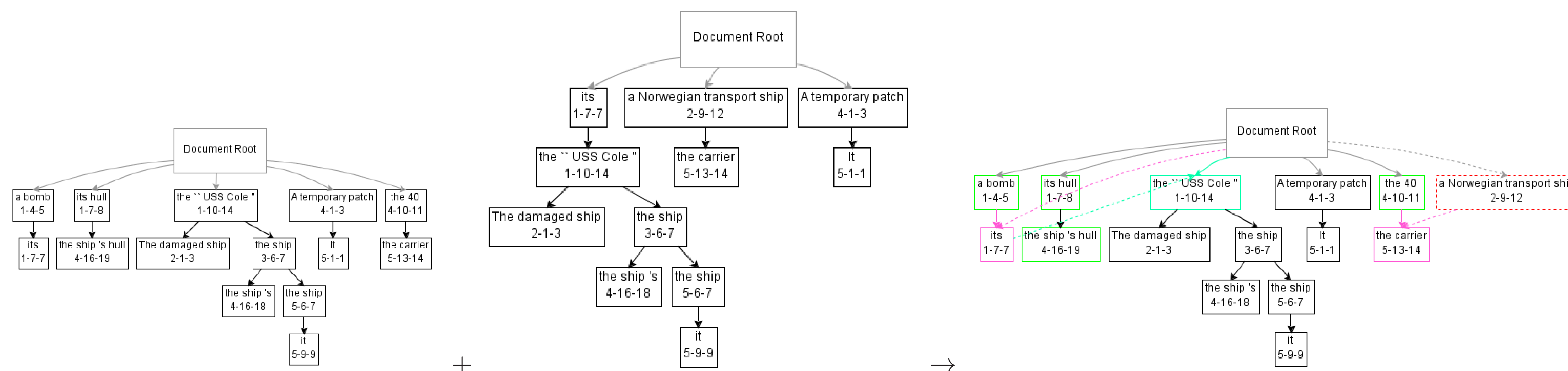


- Displays clusters as subtrees of a virtual document root
- Node and edge text customizable via *Label Patterns*
- Export for several formats (svg, png, xml, ...)

Both the **entity grid** and **tree views** support *Label Patterns* to customize text content. Those patterns are strings that define the format according to which a mention will be displayed. They allow various properties of a mention to be used as label text (e.g. `$form$` extracts the full surface form of a mention, whereas `#form#` would only extract the surface form of the head word of a mention). Switching between exploration views requires only a simple click and selections or filtering of mentions can be preserved when switching to another view.

ERROR ANALYSIS

Trees representing (a) system output, (b) correct annotation, and (c) a merged version of the other two with highlighting of their individual differences (i.e. the errors made in the prediction)



(a) system output

(b) correct

(c) merged (solid edges are system output, dashed nodes and edges are present in the correct version, but not in the system output)

- Assign multiple stand-off coreference annotations per data set
- Select up to two annotations for comparison in the exploration views
- Fine-grained analysis with 5 different error types (*false positive*, *false negative*, *foreign antecedent*, *invented antecedent*, *invalid cluster root*)
- Specialized visualizations available for grid and tree views
- Experimental prototype of a quantitative error breakdown (needs further work)

SEARCH

Example search query for cataphoric pronouns and corresponding result overview

2 groups - 25 matches

Root
 Property(Type) = Pronoun
 Head-Property(form) <=>

Transitive

Property(Type) != Pronoun
 Property(Type) <=>

Name	Common
I	3
you	1
it	2
me	3
You	1
he	1

- Interfaces with the built-in search engine of ICARUS [2]
- Enables searches on sets of documents
- Express queries graphically or in plain text
- Inspect search results with any of the 3 available exploration views

SUMMARY

- Multiple exploration views for coreference annotations
- Interactive graphical interface integrated in ICARUS [2]
- Fine-grained comparison of different data sets
- Supports various levels of user expertise
- Highly customizable visualization
- Java-based, platform independent, requires no installation



The latest version can be found here:
<http://www.ims.uni-stuttgart.de/data/icarus.html>

REFERENCES

- [1] Regina Barzilay and Mirella Lapata. Modeling Local Coherence: An Entity-Based Approach. *Computational Linguistics*, 34(1):1–34, 2008.
- [2] Markus Gärtner, Gregor Thiele, Wolfgang Seeker, Anders Björkelund, and Jonas Kuhn. ICARUS – An Extensible Graphical Search Tool for Dependency Treebanks. In *Proc. of ACL: System Demonstrations*, pages 55–60, Sofia, Bulgaria, August 2013. ACL.