Dependency Path Based Relation Extraction

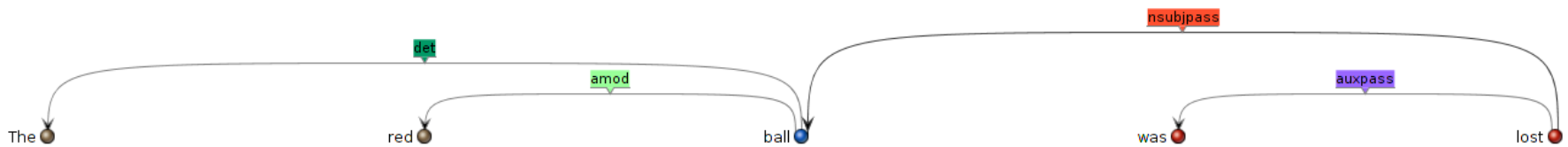
Peculiarity of Numerical Relations

- Analyzing a number of sentences expressing numerical relations lead to several insights as already discussed.
- **Keywords** We can expect presence of certain keywords that might help in identifying relations.
- **Modifiers** A large number of false positives stem out of mentions where a change in the numerical attribute is mentioned.

Dependency Path Based Relation Extraction

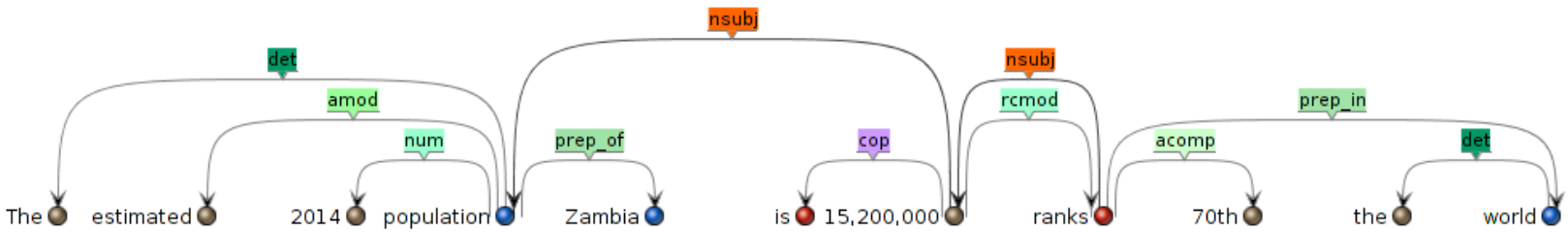
Dependencies

- Dependencies: Grammatical relation between two words, governer and dependent.
- "The red ball was lost"
-
  - **amod(ball,3,red,2)** "Red" is an adjective for "ball"
  - **det(ball,3,The,1)** "the" is a determiner of "ball"
  - **nsubjpass(lost,5,ball,3)** "ball is the subject of lost"
  - **auxpass(lost,5,was,4)** "was is an auxiliary of lost"



Dependency Path Based Relation Extraction

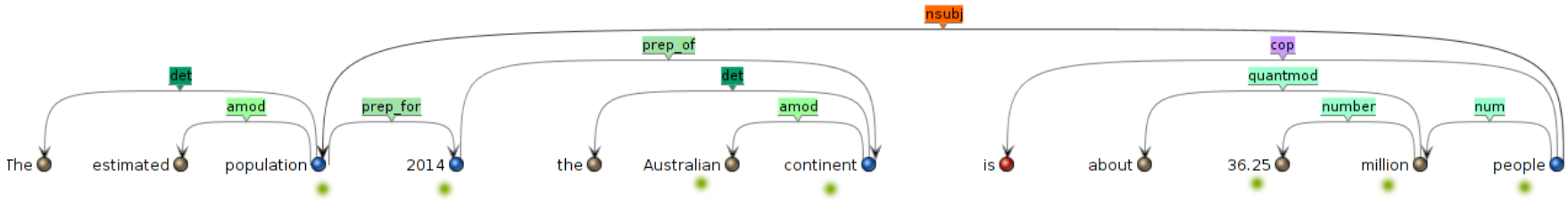
- Given a Country-Number pair, extract the shortest undirected path between them in the dependency graph.



- Path(Zambia - 15,200,000) = {Zambia, population, 15,200,000}
- For a match, the path:
  - Should have one of the keywords
  - Should not have a modifier

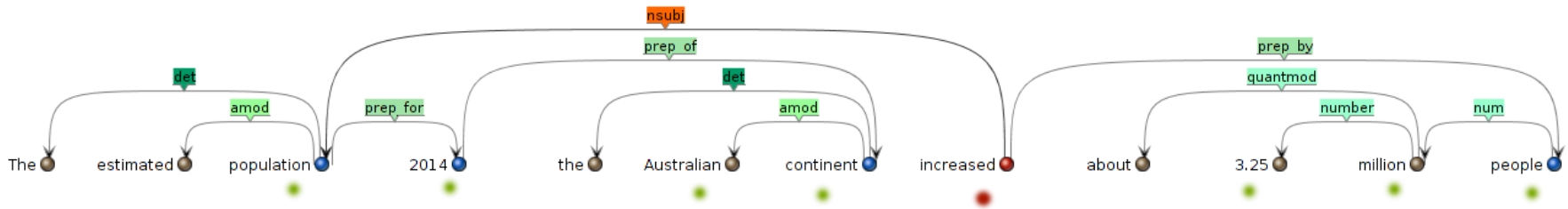
Dependency Path Based Relation Extraction

Example



Path has a keyword, and no modifier

Extracted



Path has a keyword: people, but also a modifier

Not Extracted

- The extractor was applied to 30 sentences expressing 23 different relations.
-

	Relations Present	Relations not Present (False positives)
Extracted	16	17
Not Extracted	7	N/A

  - Precision: 48.4%
  - Recall: 69.6%
- The precision will increase further on applying unit based pruning.