



EVALITA 2011

Evaluation of NLP and Speech Tools for Italian

EVALITA 2011

The News People Search Task: Evaluating Cross-document Coreference Resolution of Named Person Entities in Italian News

L. Bentivogli, A. Marchetti, E. Pianta





Outline

- Introduction and Related Work
- Dataset
- Evaluation and results
- Conclusion and future works



Outline

- Introduction and Related Work
- Dataset
- Evaluation and results
- Conclusion and future works



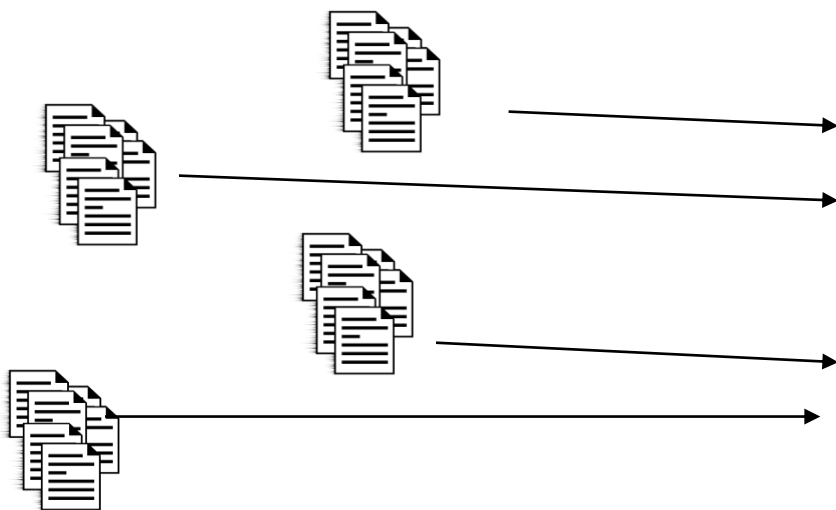
The NePS Task

- The News People Search Task aims at evaluating cross-document coreference resolution of named person entities in Italian news



The NePS Task

- The News People Search Task aims at evaluating cross-document coreference resolution of named person entities in Italian news
- Person Cross-document coreference resolution: recognizing when different documents are referring to the same person



Paolo Rossi
(Comedian)



Paolo Rossi
(Football Player)



The NePS Task

- Cross-document coreference Involves two problematic aspects:

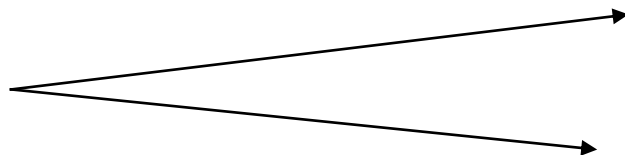


The NePS Task

- Cross-document coreference Involves two problematic aspects:

1. Ambiguity: different people having the same name

Paolo Rossi



Paolo Rossi
(Comedian)



Paolo Rossi
(Football Player)

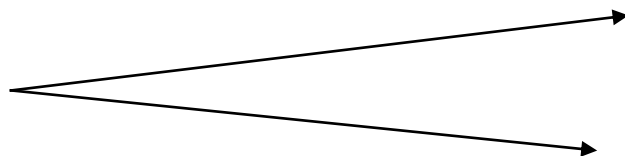


The NePS Task

- Cross-document coreference Involves two problematic aspects:

1. Ambiguity: different people having the same name

Paolo Rossi



Paolo Rossi
(Comedian)



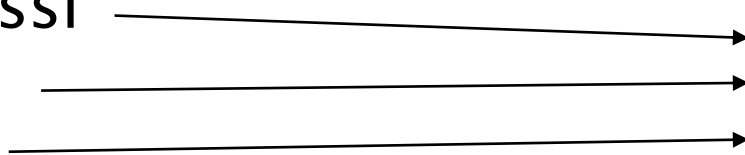
Paolo Rossi
(Football Player)

2. Name variation: different names can refer to the same person

Paolo Rossi

P. Rossi

Pablito



Paolo Rossi
(Football Player)



The NePS Task

- Input: a set of newspaper articles containing at least one mention of an ambiguous person name (Group Name) or of one of its possible variants



The NePS Task

- Input: a set of newspaper articles containing at least one mention of an ambiguous person name (Group Name) or of one of its possible variants

complete name
(First Name –
Last Name)



The NePS Task

- Input: a set of newspaper articles containing at least one mention of an ambiguous person name (Group Name) or of one of its possible variants

complete name
(First Name –
Last Name)

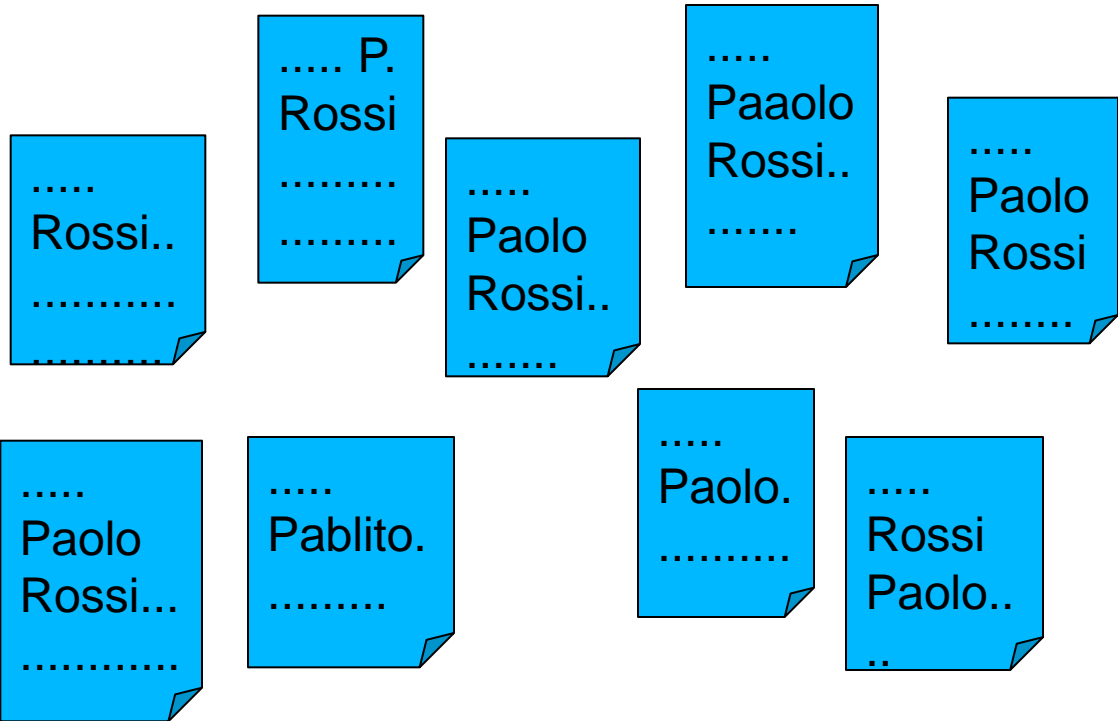
- Output: For each Group Name, systems should return the clustering of the documents
 - each cluster contains all and only those documents that refer to the same individual



The NePS Task

Group Name: **Paolo Rossi**

INPUT

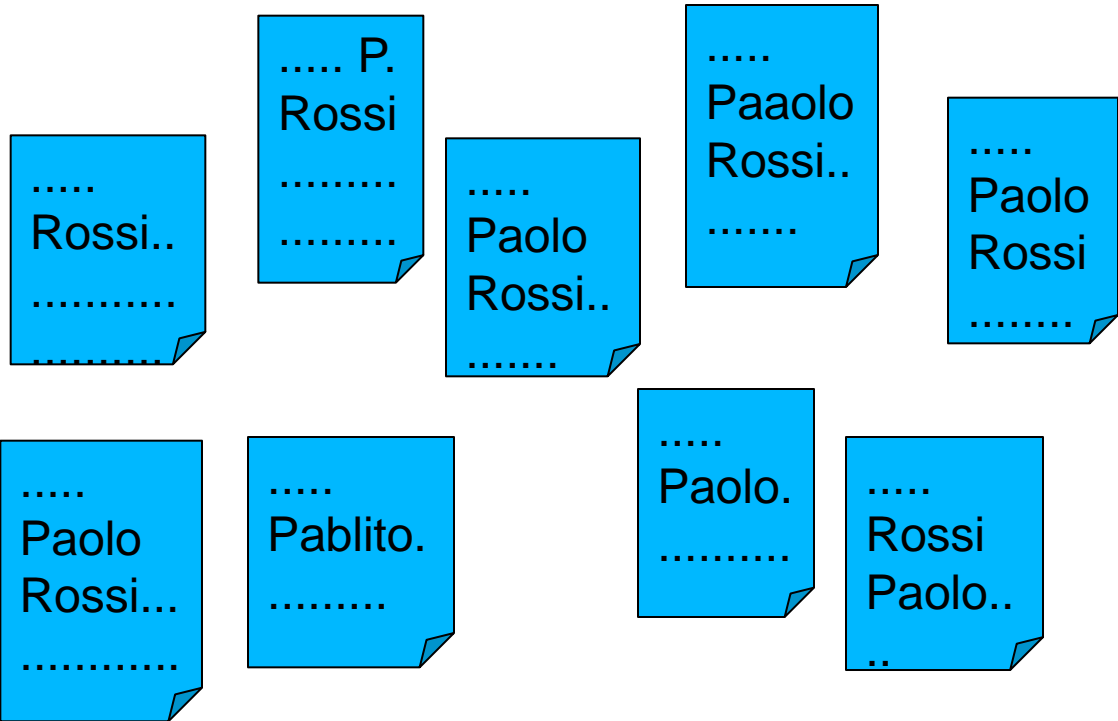




The NePS Task

Group Name: **Paolo Rossi**

OUTPUT



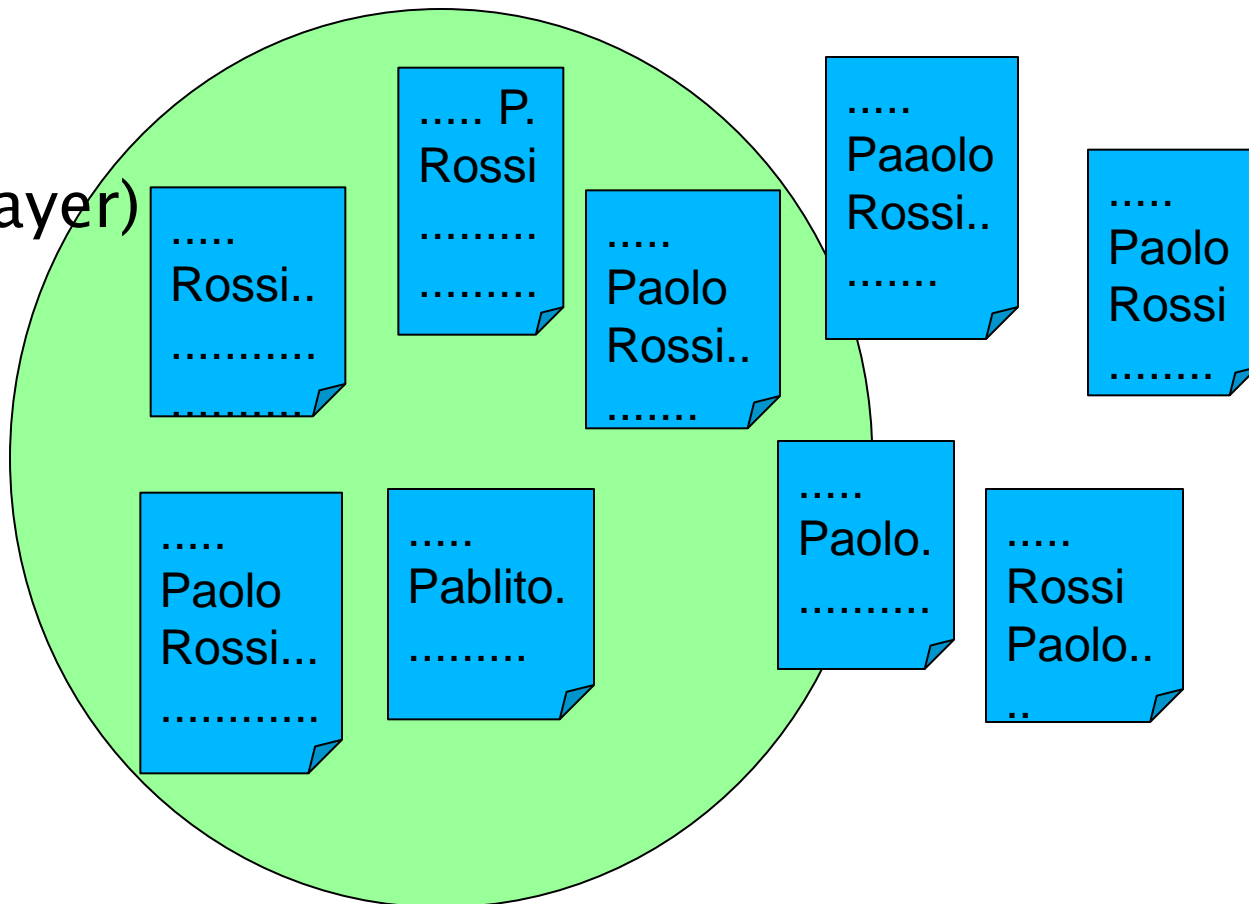


The NePS Task

Group Name: **Paolo Rossi**

Cluster:
Paolo Rossi
(football player)

OUTPUT



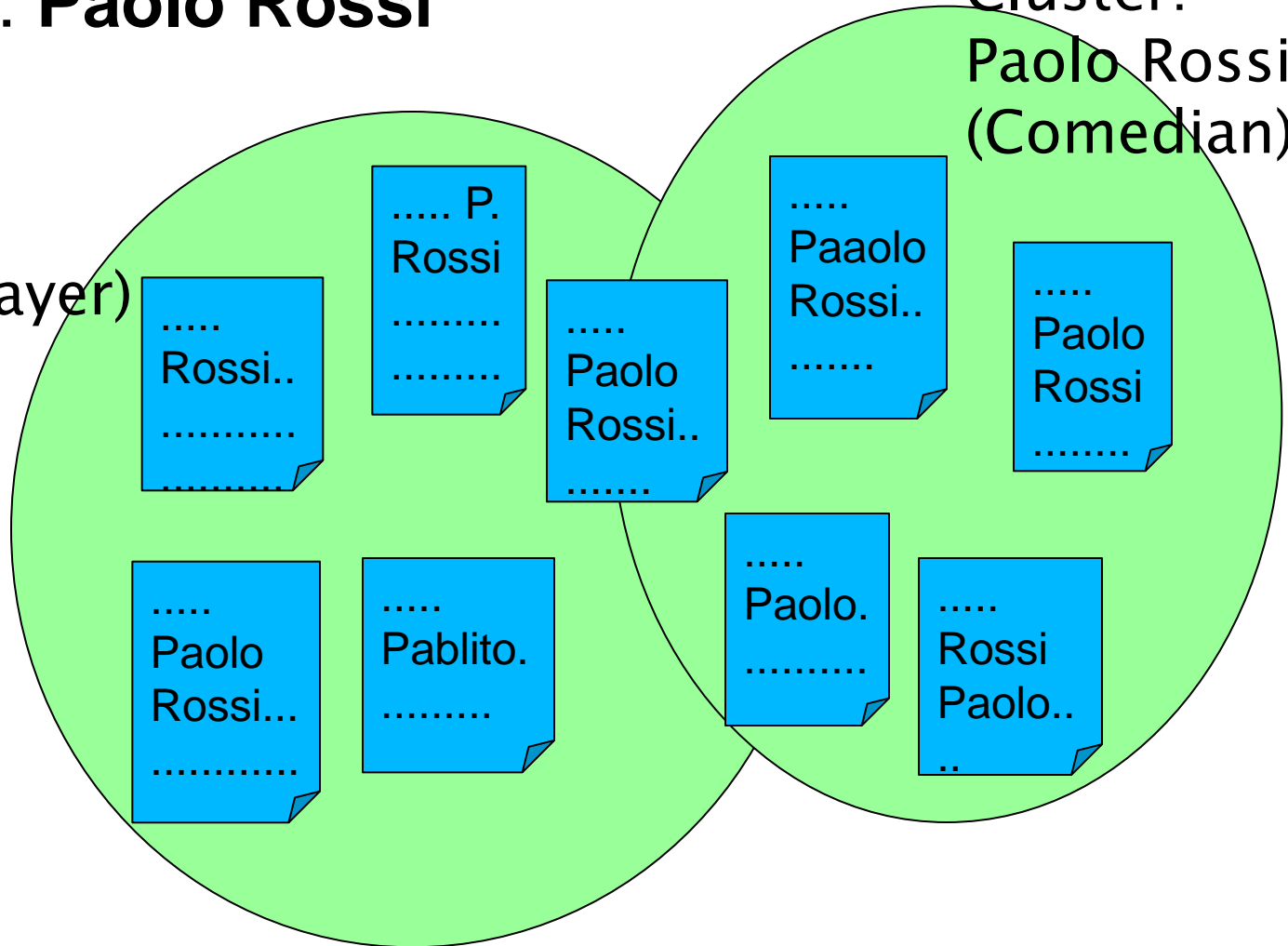


The NePS Task

Group Name: **Paolo Rossi**

Cluster:
Paolo Rossi
(football player)

Cluster:
Paolo Rossi
(Comedian)



OUTPUT



Related Work

- Cross-document coreference represents a central component for NLP applications:
 - Multi document summarization
 - Question answering
 - Information extraction
 - Entity detection and tracking
 - Knowledge base population from texts



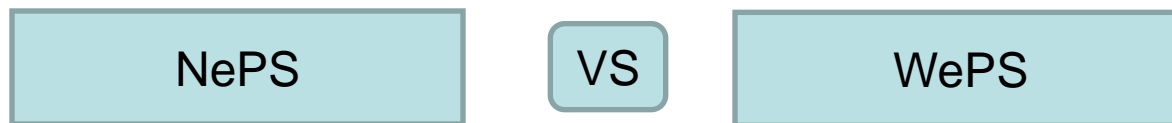
Related Work

- Recent evaluation Campaigns addressed English Language:
 - ACE 2008 Cross–Document (XDOC)
http://projects ldc.upenn.edu/ace/docs/ACE08_XDOC_1.6.pdf
 - Web People Search (WePS), three evaluation campaigns (2007, 2009, 2010) <http://nlp.uned.es/weps/>
- NePS is structured along the same lines as WePS:



Related Work

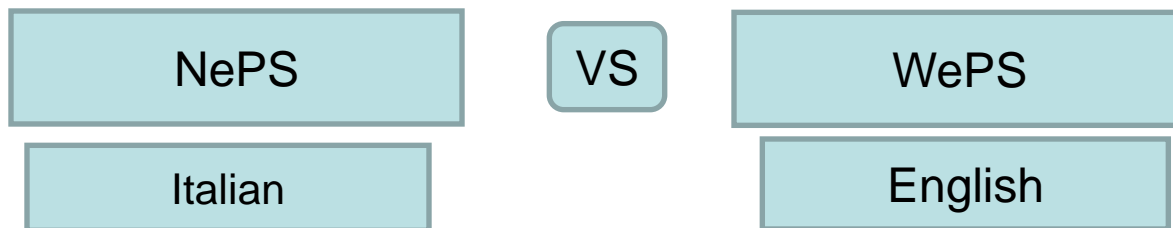
- Recent evaluation Campaigns addressed English Language:
 - ACE 2008 Cross–Document (XDOC)
http://projects.ldc.upenn.edu/ace/docs/ACE08_XDOC_1.6.pdf
 - Web People Search (WePS), three evaluation campaigns (2007, 2009, 2010) <http://nlp.uned.es/weps/>
- NePS is structured along the same lines as WePS:





Related Work

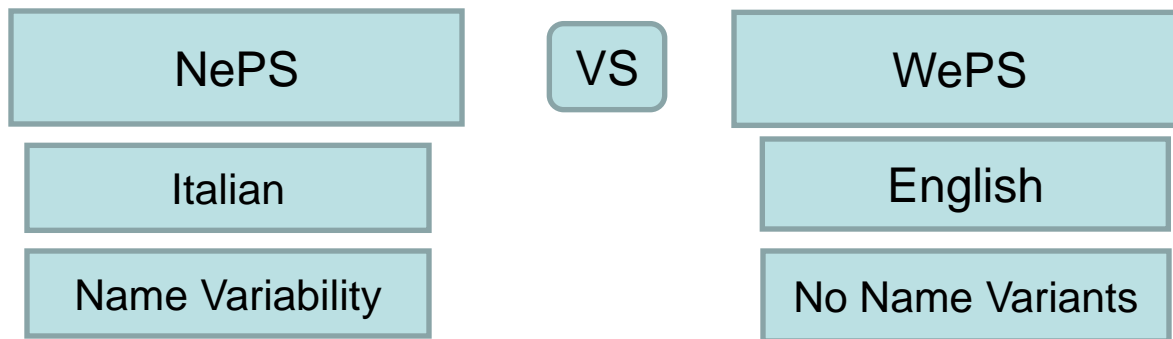
- Recent evaluation Campaigns addressed English Language:
 - ACE 2008 Cross–Document (XDOC)
http://projects.ldc.upenn.edu/ace/docs/ACE08_XDOC_1.6.pdf
 - Web People Search (WePS), three evaluation campaigns (2007, 2009, 2010) <http://nlp.uned.es/weps/>
- NePS is structured along the same lines as WePS:





Related Work

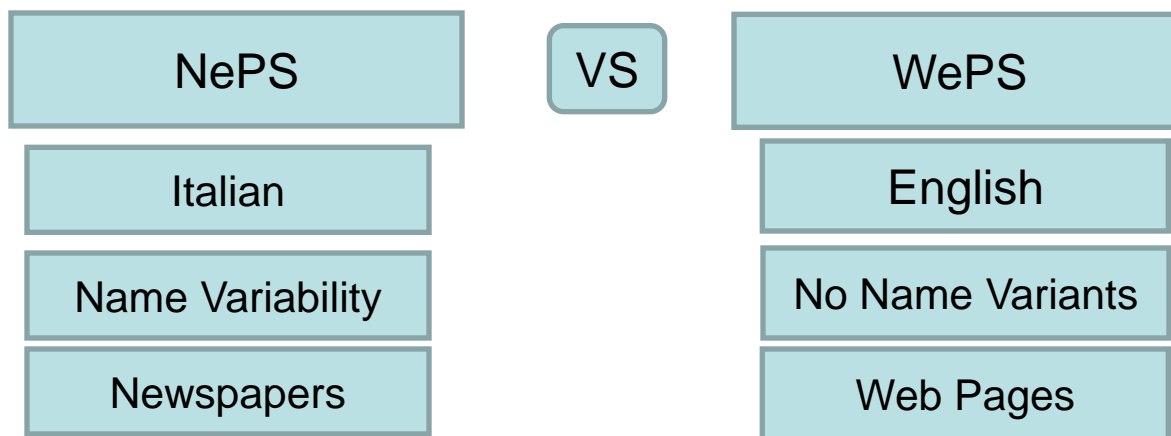
- Recent evaluation Campaigns addressed English Language:
 - ACE 2008 Cross–Document (XDOC)
http://projects.ldc.upenn.edu/ace/docs/ACE08_XDOC_1.6.pdf
 - Web People Search (WePS), three evaluation campaigns (2007, 2009, 2010) <http://nlp.uned.es/weps/>
- NePS is structured along the same lines as WePS:





Related Work

- Recent evaluation Campaigns addressed English Language:
 - ACE 2008 Cross–Document (XDOC)
http://projects.ldc.upenn.edu/ace/docs/ACE08_XDOC_1.6.pdf
 - Web People Search (WePS), three evaluation campaigns (2007, 2009, 2010) <http://nlp.uned.es/weps/>
- NePS is structured along the same lines as WePS:





Outline

- Introduction and Related Work
- **Dataset**
- Evaluation and results
- Conclusion and future works



The Dataset: CRIPCO

- **CRIPCO**: Cross-document Italian People Coreference Corpus (Bentivogli *et al.*, 2008)
 - **Corpus Composition**
 - 43,331 documents from the local newspaper “L’Adige”

# Group Names	#Entities (Clusters)	# Documents
208	690	43,331



The Dataset: CRIPCO

- For Evaluation purposes, the dataset was subdivided according to three **ambiguity ranges**:
 - **no ambiguity**: only 1 person carries the Group Name.
 - **medium ambiguity**: from 2 to 3 persons share the same Group Name.
 - **high ambiguity**: more than 3 persons share the same Group Name.



The Dataset: CRIPCO

- Dataset Subdivided by ambiguity range

Ambiguity	# Group Names	# Entities	Average Ambiguity
No	99	99	1
Medium	47	112	2.38
High	62	489	7.89



The Dataset: CRIPCO

- Dataset Subdivided by ambiguity range

Ambiguity	# Group Names	# Entities	Average Ambiguity
No	99	99	1
Medium	47	112	2.38
High	62	489	7.89

- Average Corpus Ambiguity:
 - 3.36 different entities share the same Group Name



The Dataset: CRIPCO

- Dataset Subdivided by ambiguity range

Ambiguity	# Group Names	# Entities	Average Ambiguity
No	99	99	1
Medium	47	112	2.38
High	62	489	7.89

- Average Corpus Ambiguity:
 - 3.36 different entities share the same Group Name
- Dataset has been subdivided in DEV and TEST set and balanced according to corpus ambiguity



System Evaluation

- System results were compared to the human-annotated gold standard
- The evaluation was carried out using the evaluation metrics and the official scorer used for WePS-2 (<http://nlp.uned.es/weps/weps-1/weps1-data>)
- Evaluation Metrics:
 - Extended B-Cubed Precision (BEP) and Recall (BER) combined with F1 measure
- runs were officially ranked according to their B-Cubed F1 score



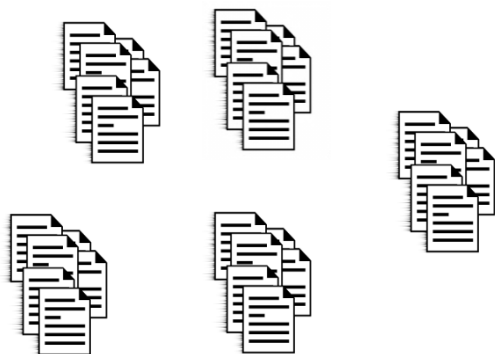
Baseline

ALL-IN-ONE:

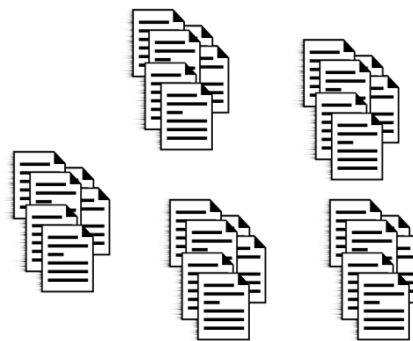
- all the documents for a given Group Name are considered to be related to the same person \ cluster

INPUT

Paolo Rossi



Rubens Barrichello





Baseline

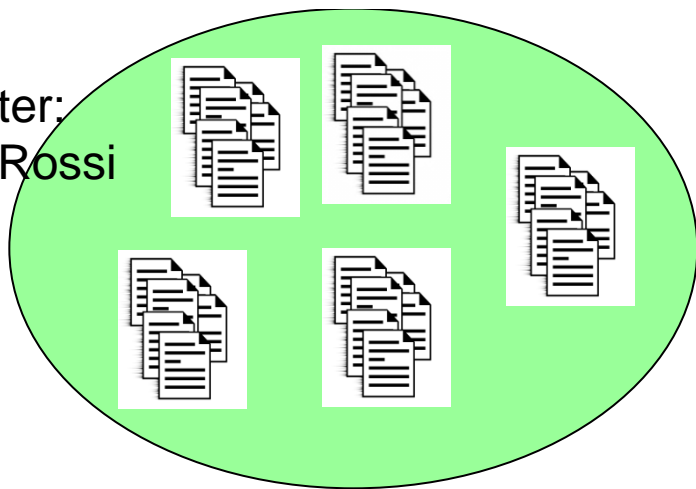
ALL-IN-ONE:

- all the documents for a given Group Name are considered to be related to the same person \ cluster

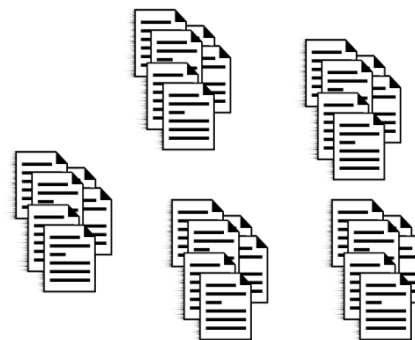
OUTPUT

Paolo Rossi

1 Cluster:
Paolo Rossi



Rubens Barrichello





Baseline

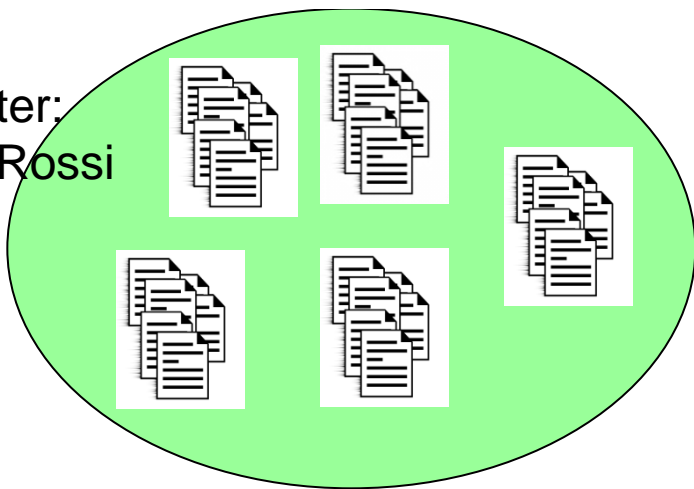
ALL-IN-ONE:

- all the documents for a given Group Name are considered to be related to the same person \ cluster

OUTPUT

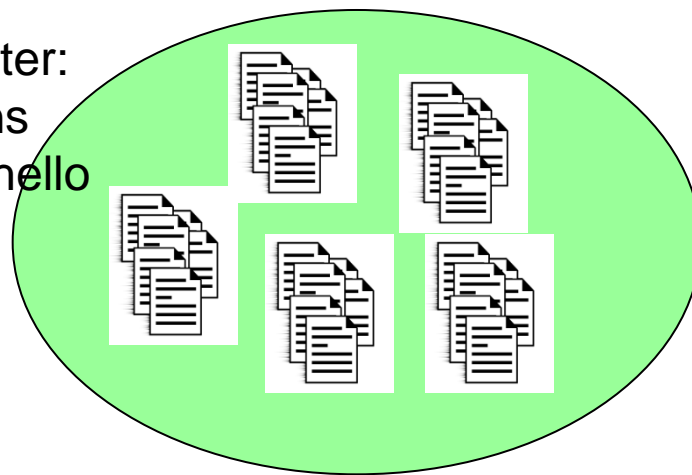
Paolo Rossi

1 Cluster:
Paolo Rossi



Rubens Barrichello

1 Cluster:
Rubens
Barrichello





Outline

- Introduction and Related Work
- Dataset
- **Evaluation and results**
- Conclusion and future works



Evaluation Results

- Five teams registered to the NePS Task and one team participated in it submitting one run
- The system results have been compared to the “ALL_IN_ONE” baseline



Evaluation Results

- Five teams registered to the NePS Task and one team participated in it submitting one run
- The system results have been compared to the “ALL_IN_ONE” baseline

	No ambiguity			Medium ambiguity			High ambiguity			ALL		
	<i>BEP</i>	<i>BER</i>	<i>F1</i>	<i>BEP</i>	<i>BER</i>	<i>F1</i>	<i>BEP</i>	<i>BER</i>	<i>F1</i>	<i>BEP</i>	<i>BER</i>	<i>F1</i>
FBK_0	1.00	0.99	0.99	0.89	0.95	0.92	0.71	0.96	0.82	0.89	0.97	0.93
ALL_IN_ONE	1.00	1.00	1.00	0.86	1.00	0.93	0.56	1.00	0.72	0.84	1.00	0.91



Discussion of Results

- The participating system showed good performances in the Test Set
 - performing similarly to the baseline considering the overall results and no\medium ambiguity Group Names
 - 10 points above it when considering highly ambiguous Group Names (the most interesting and difficult case)



Discussion of Results

- The participating system showed good performances in the Test Set
 - performing similarly to the baseline considering the overall results and no\medium ambiguity Group Names
 - 10 points above it when considering highly ambiguous Group Names (the most interesting and difficult case)
- ALL-IN-ONE baseline is very high for two main reasons:
 - There is a high number of unambiguous names (48/103)
 - Even in the High ambiguous category, the distribution of entities is skewed, i.e. most of the documents refer to one single person



Outline

- Introduction and Related Work
- Dataset
- Evaluation and results
- Conclusion and future works



Conclusion

- The availability of annotated data is crucial for advancing the state of the art in the field
- With the NePS task and the CRIPCO corpus we tried to fill the gap in availability of annotated resources for cross-document coreference resolution for Italian
- The whole corpus, including both the NePS dev and test data, will be freely available from February 2012 at *<http://hlt.fbk.eu/en/CRIPCO>*



EVALITA 2011

Evaluation of NLP and Speech Tools for Italian

Thank You!



References:

- ACE 2008, http://projects.ldc.upenn.edu/ace/docs/ACE08_XDOC_1.6.pdf
- Amigó, E., Gonzalo, J., Artiles, J., and Verdejo F.: A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval* (2008)
- Artiles, J., Gonzalo, J., and Sekine, S.: “The SemEval–2007 WePS Evaluation: Establishing a benchmark for the Web People Search Task”. In: *Proceedings of SemEval–2007, Prague, Czech Republic* (2007)
- Artiles, J., Gonzalo, J., and Sekine, S.: “WePS 2 Evaluation Campaign: Overview of the Web People Search Clustering Task”. In: *Proceedings of WePS 2 Workshop, Madrid, Spain* (2009)
- Artiles J., Borthwick A., Gonzalo J., Sekine S., and Amigó E.: “WePS–3 Evaluation Campaign: Overview of the Web People Search Clustering and Attribute Extraction Tasks”. In: *CLEF 2010 LABs and Workshops Notebook Papers, Padua, Italy* (2010)
- Bentivogli L., Girardi C., and Pianta E.: “Creating a Gold Standard for Person Cross–Document Coreference Resolution in Italian News”. In: *Proceedings of the LREC 2008 Workshop on Resources and Evaluation for Identity Matching, Entity Resolution and Entity Management. Marrakech, Morocco* (2008)