UniNE at CLEF 2018: Author Masking

Mirco Kocher and Jacques Savoy

{Mirco.Kocher, Jacques.Savoy}@unine.ch

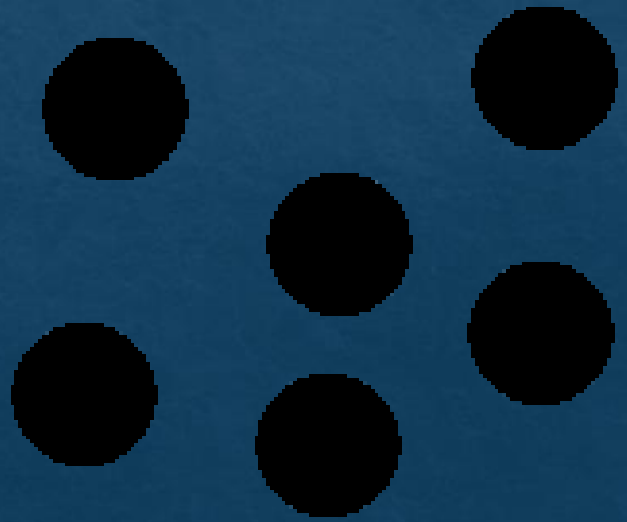
University of Neuchâtel, Switzerland

Author Masking

- Text should not match writing style of its author
- Safe, if author can not be verified
- Sensible, if masked text is inconspicuous
- Sound, if text retains its meaning

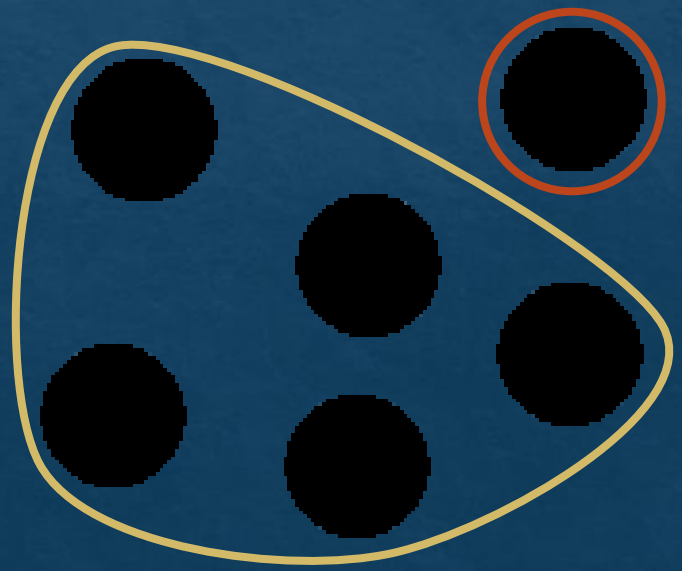
Author Masking

- Text should not match writing style of its author
- Safe, if author can not be verified
- Sensible, if masked text is inconspicuous
- Sound, if text retains its meaning



Author Masking

- Text should not match writing style of its author
- Safe, if author can not be verified
- Sensible, if masked text is inconspicuous
- Sound, if text retains its meaning

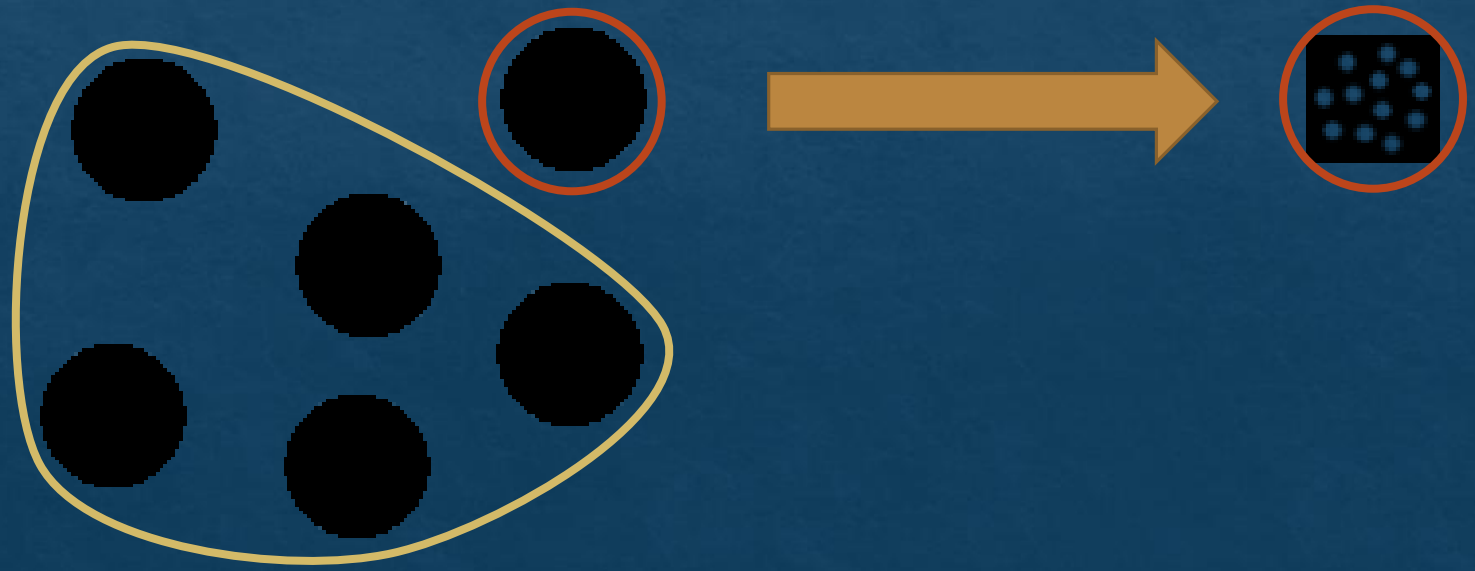


original, text to mask

same, comparable texts

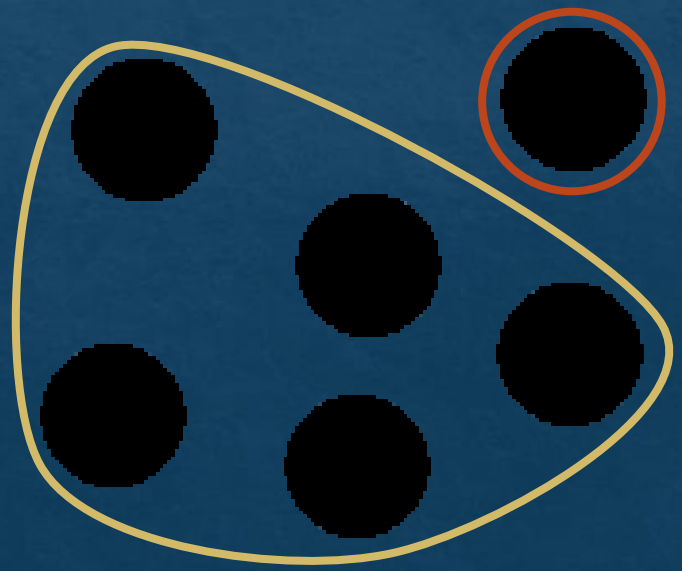
Author Masking

- Text should not match writing style of its author
- Safe, if author can not be verified
- Sensible, if masked text is inconspicuous
- Sound, if text retains its meaning



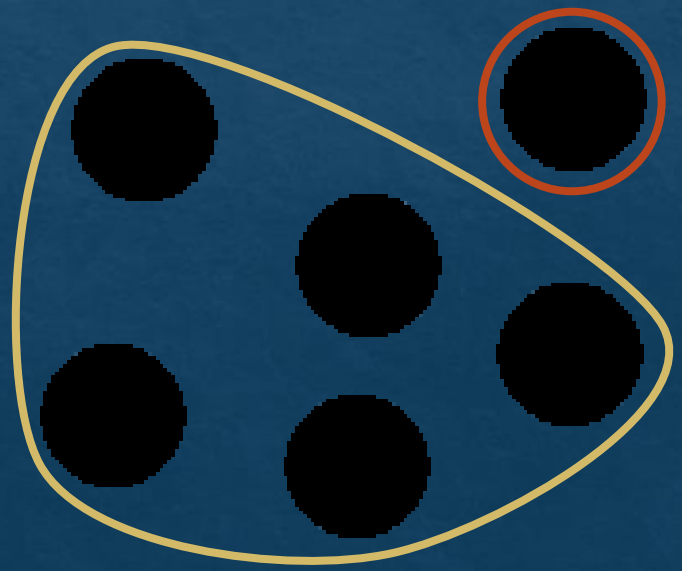
Approach

- Focus on sensibleness and soundness
- Only increase safety if high probability of correctness
- Rule based approach to attach frequency features
- Bag of Words



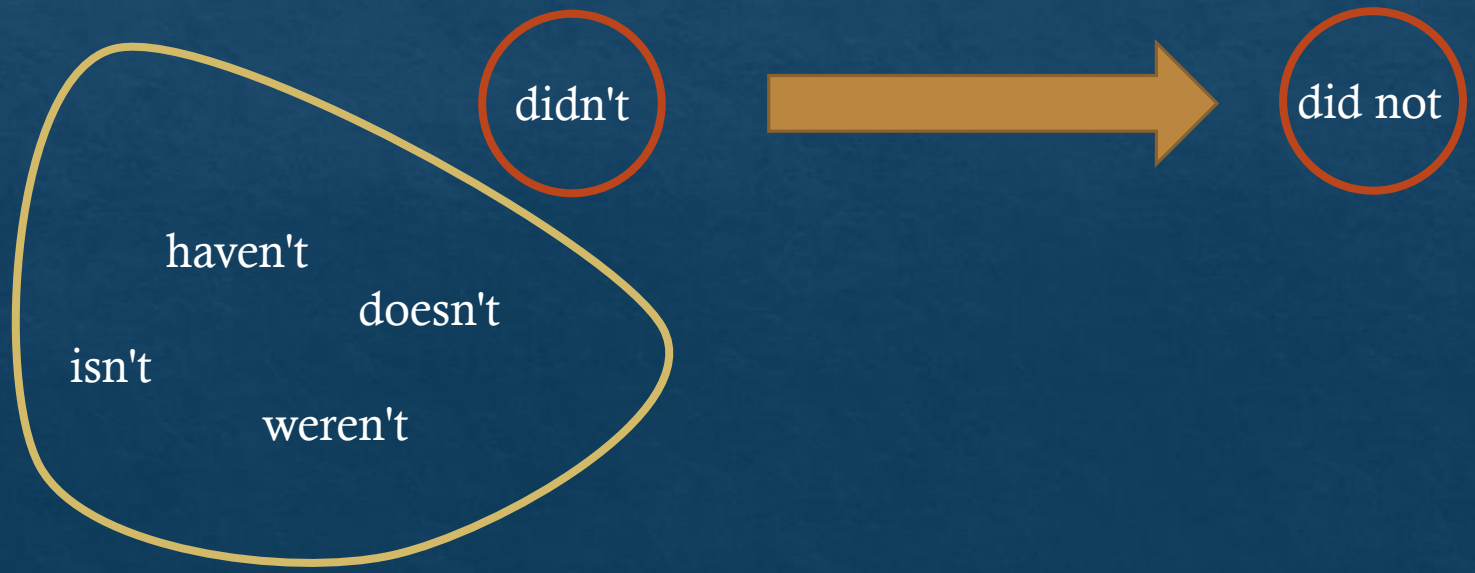
Approach

- Compare frequency features
- Make text from original dissimilar from same
- Decrease common features from same
- Increase sparse features from same



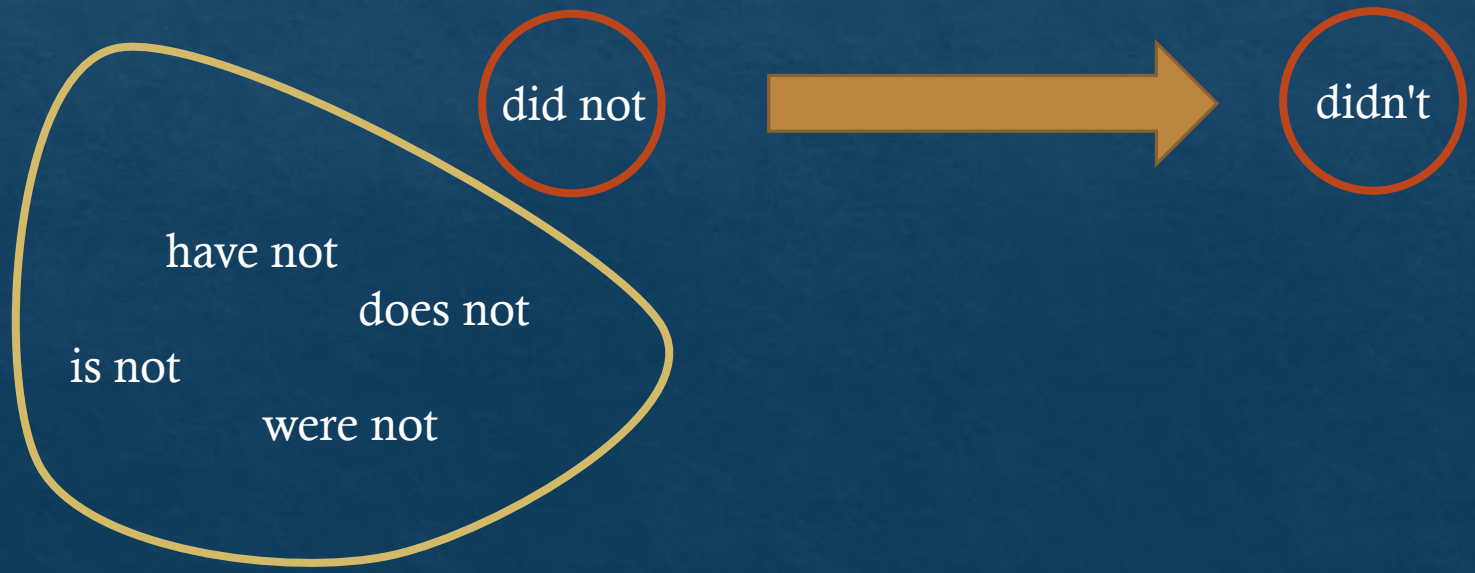
Example Rules

- Contractions in same
- Use full form in masked original



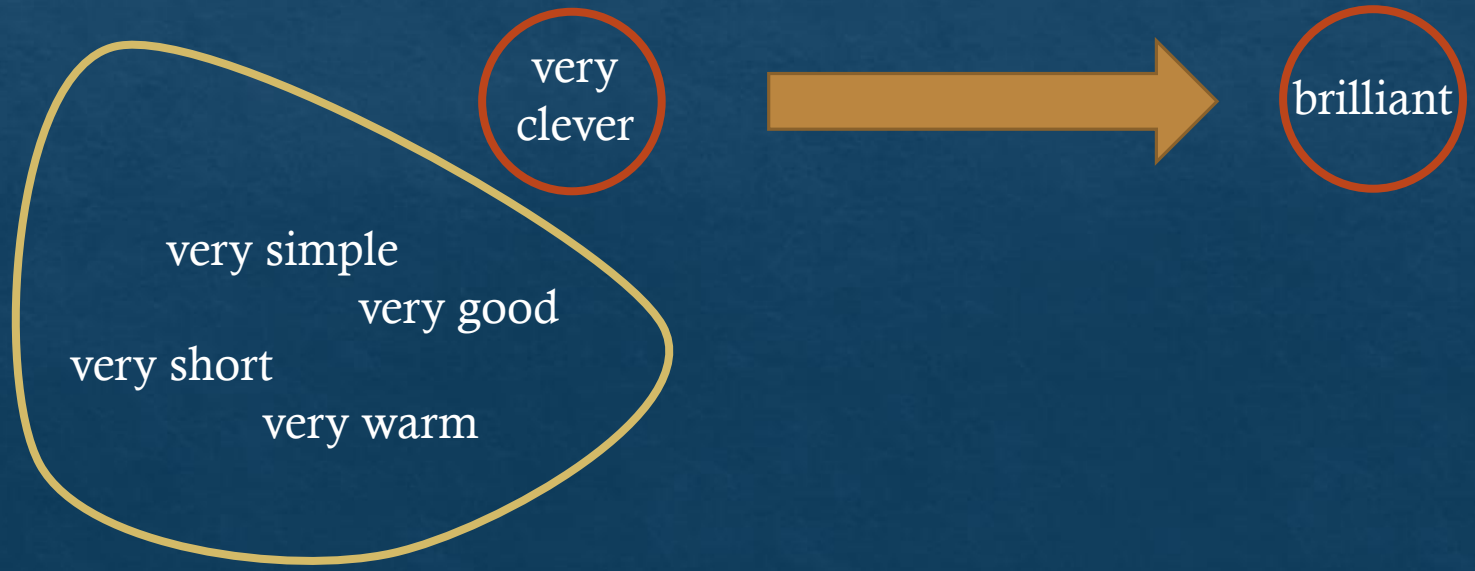
Example Rules

- Full form in same
- Use contractions in masked original



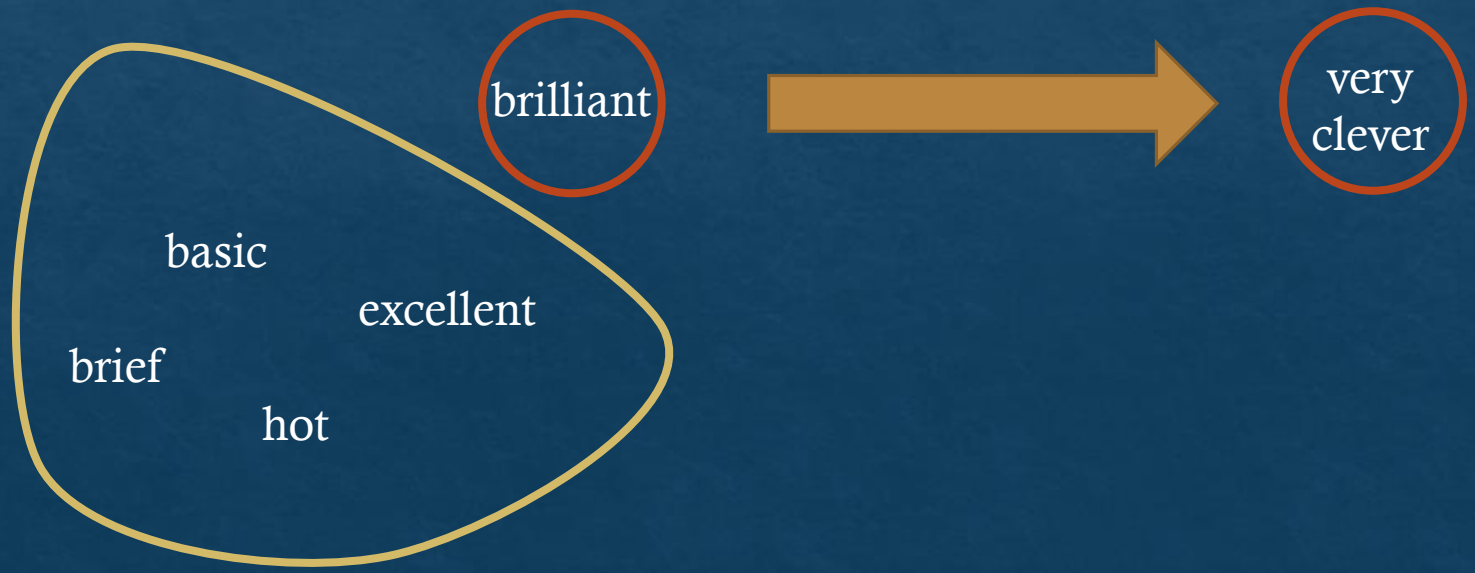
Example Rules

- Overused " very "
- Use synonym



Example Rules

- Underused " very "
- Replace synonym



Example Rules

- Decrease frequency of " in "
  - " in order to " -> " to "
- Decrease frequency of " the " or " of "
  - " the X of the Y " -> " the Y X "
- Change frequency of " and "
  - "X, Y, and Z" <-> "X, Y, as well as Z"
- Synonyms
  - "However" <-> "On the contrary"
  - "Actually" <-> "In fact"

Example Rules

- Decrease frequency of "! "
  - "! " -> ". "
- Increase frequency of "! "
  - "! " -> "!! " or "!!! "
- Increase frequency of "? "
  - "? " -> "?? " or "??? "

Probabilistic Rules

- Introduce spelling errors from repeated characters
  - e.g., "missing" -> "mising"
  - e.g., "following" –> "folllowing”
- Randomly only for 5%

Results

Team	Safe	Sensible	Sound

Results

Team	Safe	Sensible	Sound
2017 Ca.	474		
2016 Mi.	466		
2018 Ra.	355		
2016 Ke.	297		
2017 Ba.	291		
2016 Ma.	209		
2018 Ko.	**108**		

Results

Team	Safe	Sensible	Sound
	world↓		
2017 Ca.	474		
2016 Mi.	466		
2018 Ra.	355		
2016 Ke.	297		
2017 Ba.	291		
2016 Ma.	209		
2018 Ko.	**108**		

Results

Team	Safe				Sensible	Sound
	acc	rec	imp	world↓		
2017 Ca.	-0.12	-0.21	0.39	474		
2016 Mi.	-0.13	-0.24	0.45	466		
2018 Ra.	-0.10	-0.19	0.37	355		
2016 Ke.	-0.10	-0.18	0.37	297		
2017 Ba.	-0.07	-0.13	0.25	291		
2016 Ma.	-0.06	-0.10	0.21	209		
2018 Ko.	**-0.04**	**-0.08**	**0.16**	**108**		

Results

Team	Safe							Sensible	Sound
	AUC	c@1	final	acc	rec	imp	world↓		
2017 Ca.	-0.12	-0.08	-0.10	-0.12	-0.21	0.39	474		
2016 Mi.	-0.13	-0.10	-0.11	-0.13	-0.24	0.45	466		
2018 Ra.	-0.11	-0.08	-0.09	-0.10	-0.19	0.37	355		
2016 Ke.	-0.09	-0.07	-0.08	-0.10	-0.18	0.37	297		
2017 Ba.	-0.06	-0.05	-0.06	-0.07	-0.13	0.25	291		
2016 Ma.	-0.05	-0.04	-0.04	-0.06	-0.10	0.21	209		
2018 Ko.	**-0.12**	**-0.11**	**-0.08**	**-0.04**	**-0.08**	**0.16**	**108**		

Results

Team	Safe				Sensible	Sound
	acc	rec	imp	world↓		
2017 Ca.	-0.12	-0.21	0.39	474		
2016 Mi.	-0.13	-0.24	0.45	466		
2018 Ra.	-0.10	-0.19	0.37	355		
2016 Ke.	-0.10	-0.18	0.37	297		
2017 Ba.	-0.07	-0.13	0.25	291		
2016 Ma.	-0.06	-0.10	0.21	209		
2018 Ko.	**-0.04**	**-0.08**	**0.16**	**108**		

Results

Team	Safe	Sensible	Sound
2017 Ca.	474		
2016 Mi.	466		
2018 Ra.	355		
2016 Ke.	297		
2017 Ba.	291		
2016 Ma.	209		
2018 Ko.	**108**		

Results

Team	Safe	Sensible	Sound
2017 Ca.	474	4.0	3.0
2016 Mi.	466	2.5	3.0
2018 Ra.	355	3.0	2.0
2016 Ke.	297	5.0	3.0
2017 Ba.	291	1.5	2.0
2016 Ma.	209	4.0	3.0
2018 Ko.	**108**	**2.0**	**1.5**

Conclusion

- Retain text meaning and remain inconspicuous
- Use light technique to attack author masking
- Safety slightly increased after obfuscation
- More alterations should be included
- Broader applicability needed

Thank you for your attention

Mirco Kocher and Jacques Savoy

{Mirco.Kocher, Jacques.Savoy}@unine.ch

University of Neuchâtel, Switzerland