Evaluation Results Regarding the Different Entity Linking Challenges (see Sec. 2.2 in Paper)

Explanations:
Challenge 1: Known surface form, known entity.
Challenge 2: Unknown surface form, known entity.
 Challenge 2a: The mention is a substring of an existing surface form of the entity to be linked to.
 Challenge 2b: An existing surface form of the entity to be linked to is substring of the mention.
 Challenge 2c: The mention is an abbreviation of an existing surface form of the entity to be linked to.
 Challenge 2d : Otherwise.
Challenge 3: Known surface form, unknown entity.
Challenge 4: Unknown surface form, unknown entity.

Number of (not unique) annotations (mention+entity) found in news:

Challenge 1	204,675,773	
Challenge 2a	18,547	
Challenge 2b	92,283	192,832
Challenge 2c	3,860	
Challenge 2d	78,142	
Challenge 3	149,610 (active: 149,366)	
Challenge 4	194,745 (active: 193,647)	
Sum	205,225,526	

Regarding Challenge 2:
In order to examine for Challenge 2 how similar the mentions are w.r.t. the existing surface forms of the KG entities to which the mentions should link, we determine the **minimum normalized edit distance** between each mention in the set of annotations with Challenge 2d and each of the surface forms of the respective target entity. Hereby, we get:

Mean:	0.573
Standard deviation:	0.299

 We can state that on average the mention differs from the lexically nearest surface form in half of the characters, which makes a detection of new surface forms non-trivial.

Number of unique entities/sfforms/(entity,sfform) found in news:

	unique entities		unique sfforms		unique tupel (entity,sfform)	
Challenge 1	958,842		1,712,875		1,824,983	
Challenge 2a	709		713		713	
Challenge 2b	3,714		3,949		3,950	
Challenge 2c	16	7,394	16	7,746	16	7,750
Challenge 2d	2,955		3,068		3,071	
Challenge 3	1,720 (active: 1707)		2,619		2,632	
Challenge 4	4,054 (active: 3937)		4,397		4,398	
Total (!= SUM)	966,125		1,727,806		1,840,256	