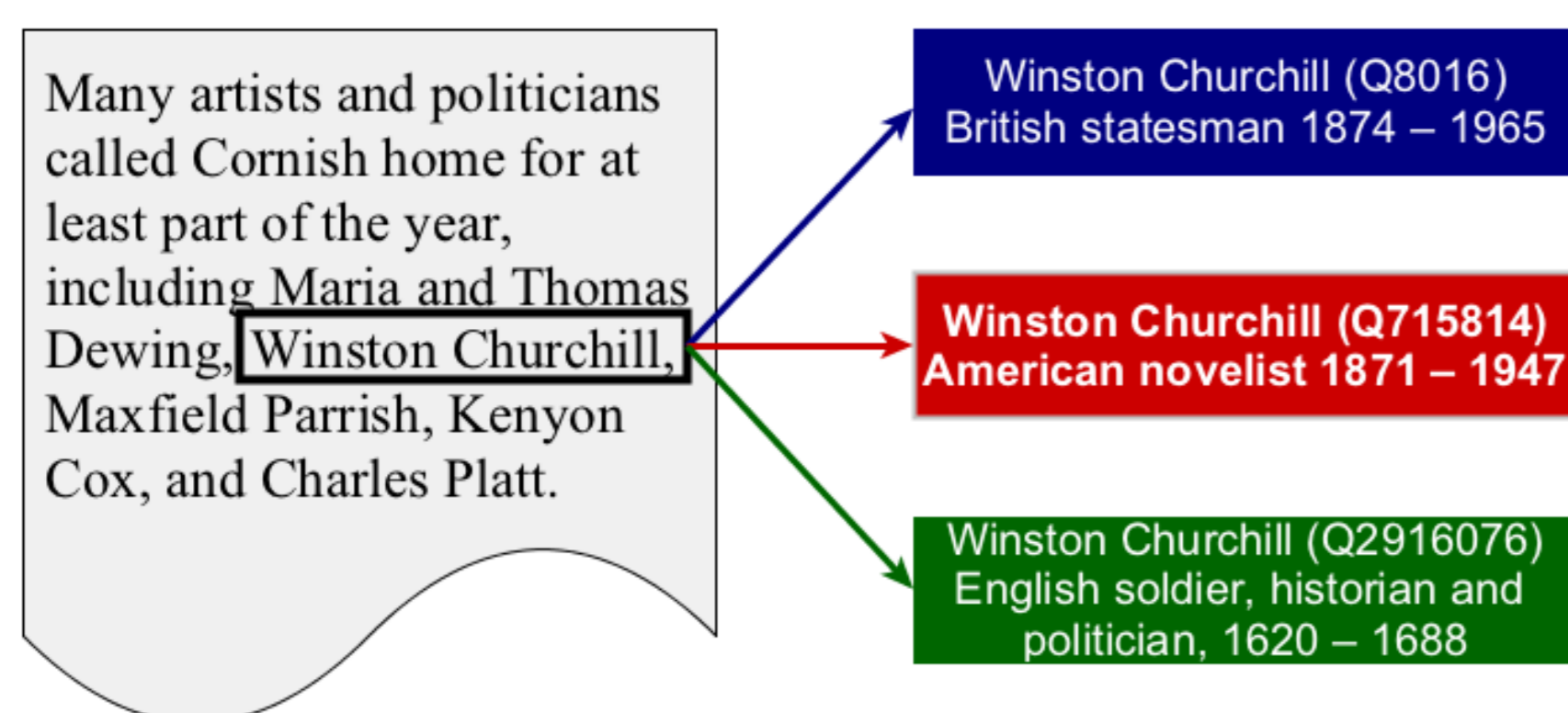


Person Name Linking

Named Entity Disambiguation and Linking to a knowledge base

Goal: uniquely assign a real-world entity to a person mention in text document

Problem: person mentions can be ambiguous



SocNNEL

State-of-the-art approaches

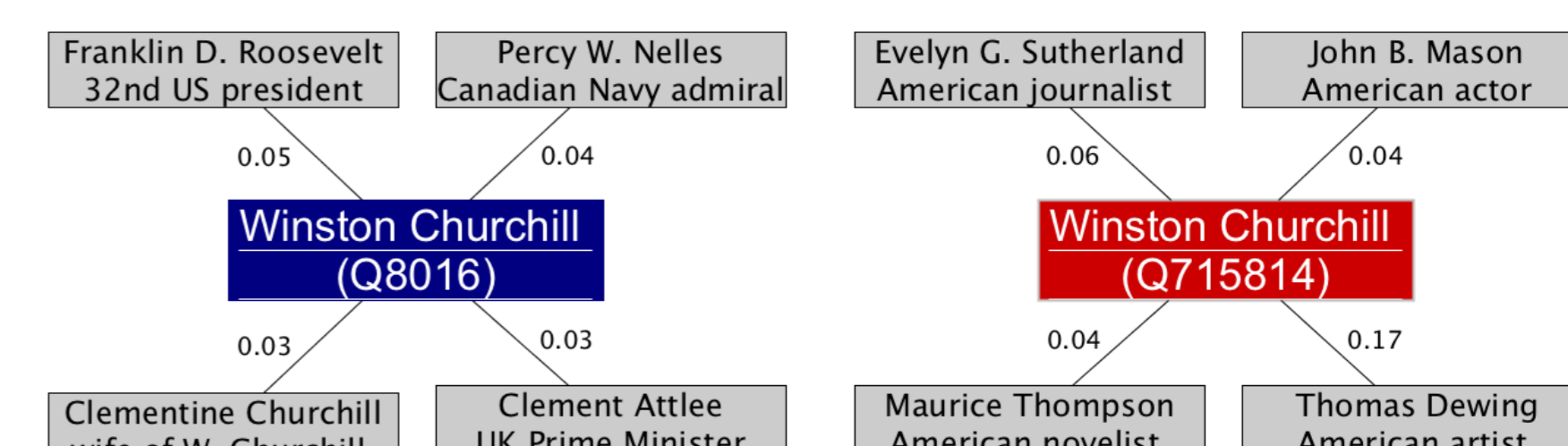
- focus on context features that describe a person to compare them to external concepts/topics
- methods: topic models, PageRank or random walks on networks
- are often language dependent and tailored towards specific domains

Social Network Named Entity Linking:

- only uses neighboring person mentions, information from Wikidata and the Wikipedia Social Network^[2]
- language independent, reliable, simple

Wikipedia Social Network (WSN)^[2]

- extracted from the English Wikipedia using Wikilinks and person information in Wikidata
- nodes represent persons
- edges have a relationship strength (r)
- relationship strength is based on the number of co-occurrences and distances between mentions



Disambiguation Model

Disambiguation by comparison of entity candidates to uniquely identified persons

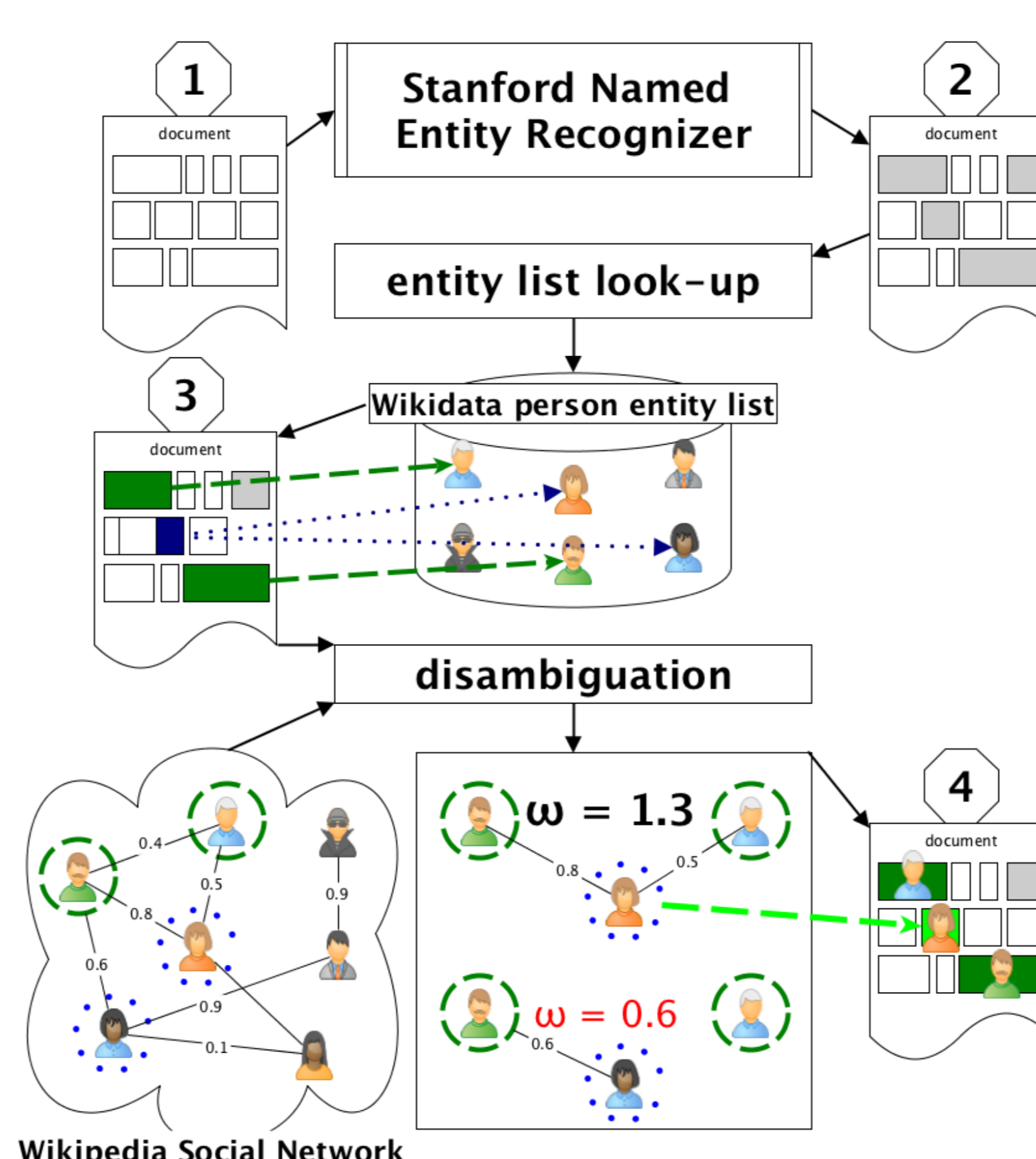
1. Person Name Extraction

Stanford NER

2. Entity List Look-Up

matching the mention to Wikidata labels and alternative names:

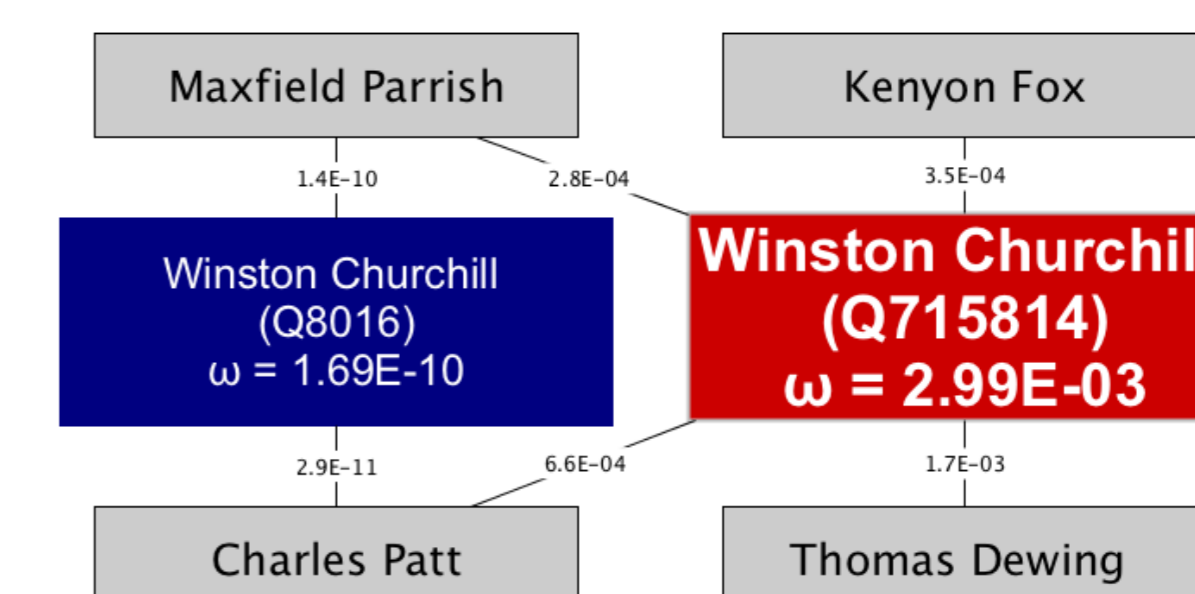
- no match (grey)
- seed sp : one match (green)
- ambiguous mention m : multiple matches (blue), each match for m is called an entity candidate $c \in C_m$



3. WSN based Disambiguation

- selecting the candidate c with the strongest relation to the neighborhood N consisting of seed persons sp .
- weight function $\omega : (c, N) \rightarrow \mathbb{R}$ for the neighborhood relation between a candidate $c \in C_m$ for a person mention m and a neighborhood N of seed persons sp in the document d :

$$\omega(c, N) := \sum_{sp \in N} r(c, sp)$$



Experiments

Dataset: base 1: lowest Wikidata id
1.5 M person wikilinks base 2: random selection

	SocNNEL		base ₁		base ₂	
	P	A	P	A	P	A
all mentions	96.4	94.1	89.0	88.9	78.5	78.4
seed persons		98.7		98.7		98.7
ambiguous	89.8	82.4	63.0		24.2	

Table 1: Accuracy (A) and Precision (P) in %.

Number of candidates

- number of candidates influences the precision
- maximum number of candidates should be 10 - 40

Evaluation

Dataset: AIDA CoNLL-YAGO testset b (984 person mentions)

	SocNNEL		base ₁		base ₂	
	A	P	A	P	A	P
all m	84.4	94	74.2	74.3	62	61.9
truth possible	86.2	95.7	75.8	75.8	63.2	63.2
ambiguous m	66.8	87.6	43.1	43.1	14.5	14.5
truth possible	68	88.4	43.9	43.9	14.7	14.7

Table 2: Accuracy (A) and Precision (P) in % on the AIDA CoNLL-YAGO testset-b.

- comparable to state-of-the-art approaches

Conclusion

- disambiguation model that employs a mention's neighborhood and the Wikipedia Social Network.
- measure for the relationship between a person and its neighborhood
- comparable to state-of-the-art approaches, simple, reliable and easily adoptable to other document types and languages

Ongoing Work

- different named entity types (e.g., locations, organizations)
- refining the model to cover more mentions and become independent of seeds persons
- limiting the number of candidates per mention

References

- [1] J. Geiß, M. Gertz: **With a Little Help from my Neighbors: Person Name Linking Using the Wikipedia Social Network**, *Wiki Workshop*, WWW 2016
- [2] J. Geiß, A. Spitz and M. Gertz: **Beyond Friendships and Followers: The Wikipedia Social Network**, *ASONAM* 2015

Contact

Dr. Johanna Geiß
geiss@informatik.uni-heidelberg.de
<http://dbs.ifi.uni-heidelberg.de/>

