The University Of Sheffield.

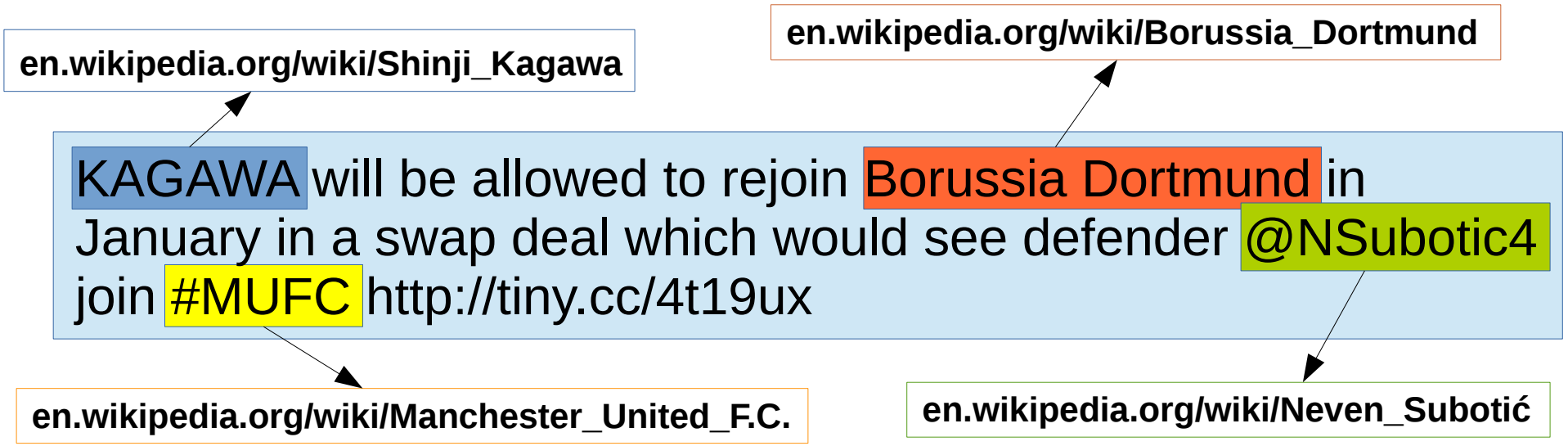
Using @Twitter Conventions to Improve #LOD-based Named Entity Disambiguation

Genevieve Gorrell, Johann Petrak, and Kalina Bontcheva

The University Of Sheffield.

Named Entity Recognition and Disambiguation (NERD)

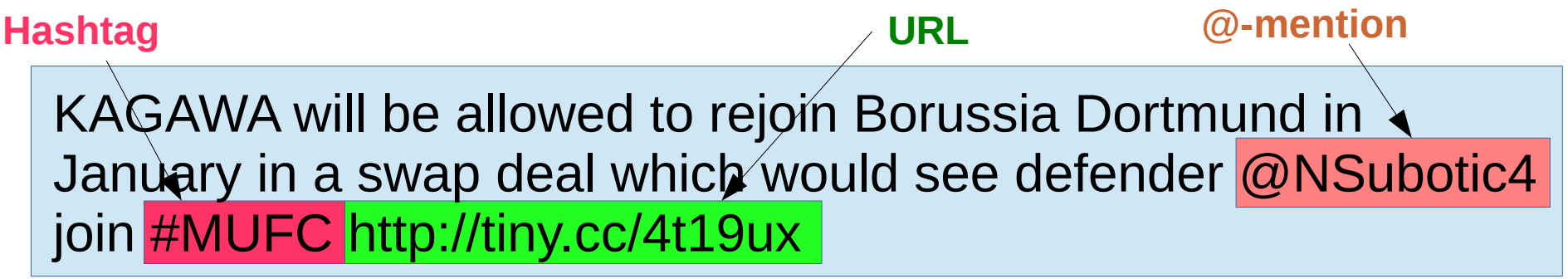
- Finding and linking entities against unique referents in a database such as Wikipedia is an enabler for information extraction applications and a popular task



The University Of Sheffield.

NERD in Tweets

- Tweets are particularly hard
  - Short messages provide little contextual information
- However Twitter also has extra information that we can use, such as hashtags, user @-mentions and hyperlinks
- **In this work we evaluate the improvement obtainable by using extra content from these Twitter *expansions***



We need ...

- A NERD system in which to evaluate the contribution of Twitter expansion
  - Introducing YODIE (Yet another Open Data Information Extraction system!)
- A NERD tweet corpus
  - Existing corpora weren't suitable at the time
    - MSM corpus now also provides NERD-annotated tweets
  - We have created and shared a new corpus
    - Available here: https://gate.ac.uk/applications/yodie.html
    - Selected from news/financial and climate change domains

The University Of Sheffield.

New Tweet NERD Corpus

	Tweets	Total NEs	URLs	Hashtags	@mentions
Total	794	681	504 (236)	359 (188)	334 (316)
Training	397	257	242 (112)	172 (88)	167 (157)
Test	397	424	262 (124)	187 (100)	167 (159)

- Manually annotated by three annotators and expert-adjudicated
  - Unanimous agreement between annotators for 89% of entities
- 252 person annotations, 309 location annotations, 347 organization annotations, 218 nil annotations
- Bracketed figures above show number of expansions that were successful at the time the work was done

YODIE

- YODIE is an NERD system developed over several diverse EU projects that aims to:
  - Develop reusable, extensible open source GATE components for named entity linking
  - Port rapidly to new domains
  - Achieve competitive performance on different input types
- YODIE is suitable for this work because:
  - It is a typical approach
  - Performance is comparable with existing SoA systems

YODIE Overview

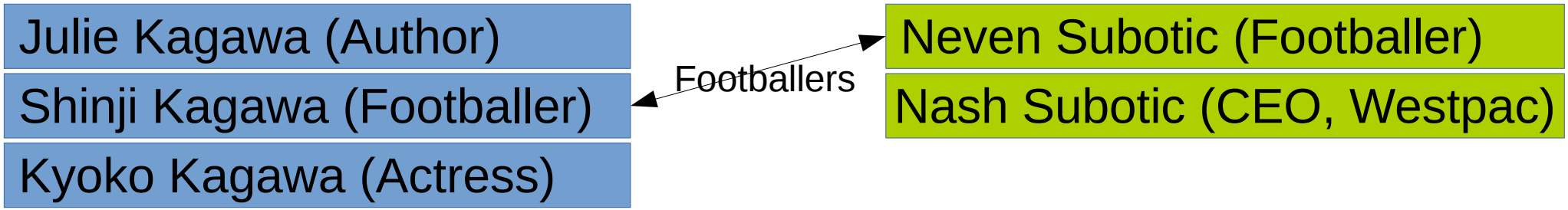
- Candidates are mapped to labels gathered from aliases, link text etc.
- These phrases and terms are then located in the text
  - Label-based high recall “NER” stage (similar to Han et al., 2011)
  - Avoids the NER performance bottleneck
  - Makes no attempt to detect nils—however nils aren't part of this work
- Candidates are evaluated using scores
- Scores, among other things, used as features for ML-based (SVM) disambiguation step that evaluates candidates
- Final two steps are where tweet expansion has scope for impact!

The University Of Sheffield.

YODIE Scores—Structural

- Pairwise structural semantic approach assigns higher score to related candidates
- Different ways of deciding if candidates are related

KAGAWA will be allowed to rejoin Borussia Dortmund in January in a swap deal which would see defender @NSubotic4



The University Of Sheffield.

Structural—Relatedness (Milne & Witten 2008)

- Wikipedia-based
  - A is set of Wikipedia pages that link to a
  - B is set of Wikipedia pages that link to b
  - W is all pages in Wikipedia

$$r(a,b) = \frac{\log(max(|A|,|B|)) - \log(|A \cap B|)}{\log(|W|) - \log(min(|A|,|B|))}$$

The University Of Sheffield.

Structural—LOD-based similarity

- DBpedia-based (higher quality but sparser)
  - C_c is current candidate
  - n is number of mentions in context window (entire tweet)
  - m is number of candidates on the entity
  - R is count of relationships between candidates in DBpedia
  - w is inverse square of degrees of separation in relationship
  - d is distance between mentions in characters

$$s(C_c) = \sum_{i=1}^{n} \left(\frac{\sum_{j=0}^{m} w\, R(C_c, C_{ij})}{d_{(c,i)}} \right)$$

YODIE Scores—Text-based

- Text-based Similarity
  - Term-document matrix constructed over DBpedia abstracts
    - TF-IDF transformed to up-weight important words
  - Tweet text and candidate DBpedia text data BoW vectors multiplied with T-D matrix to expand with related terms
  - Resulting vectors cosined together to obtain similarity score

The University Of Sheffield.

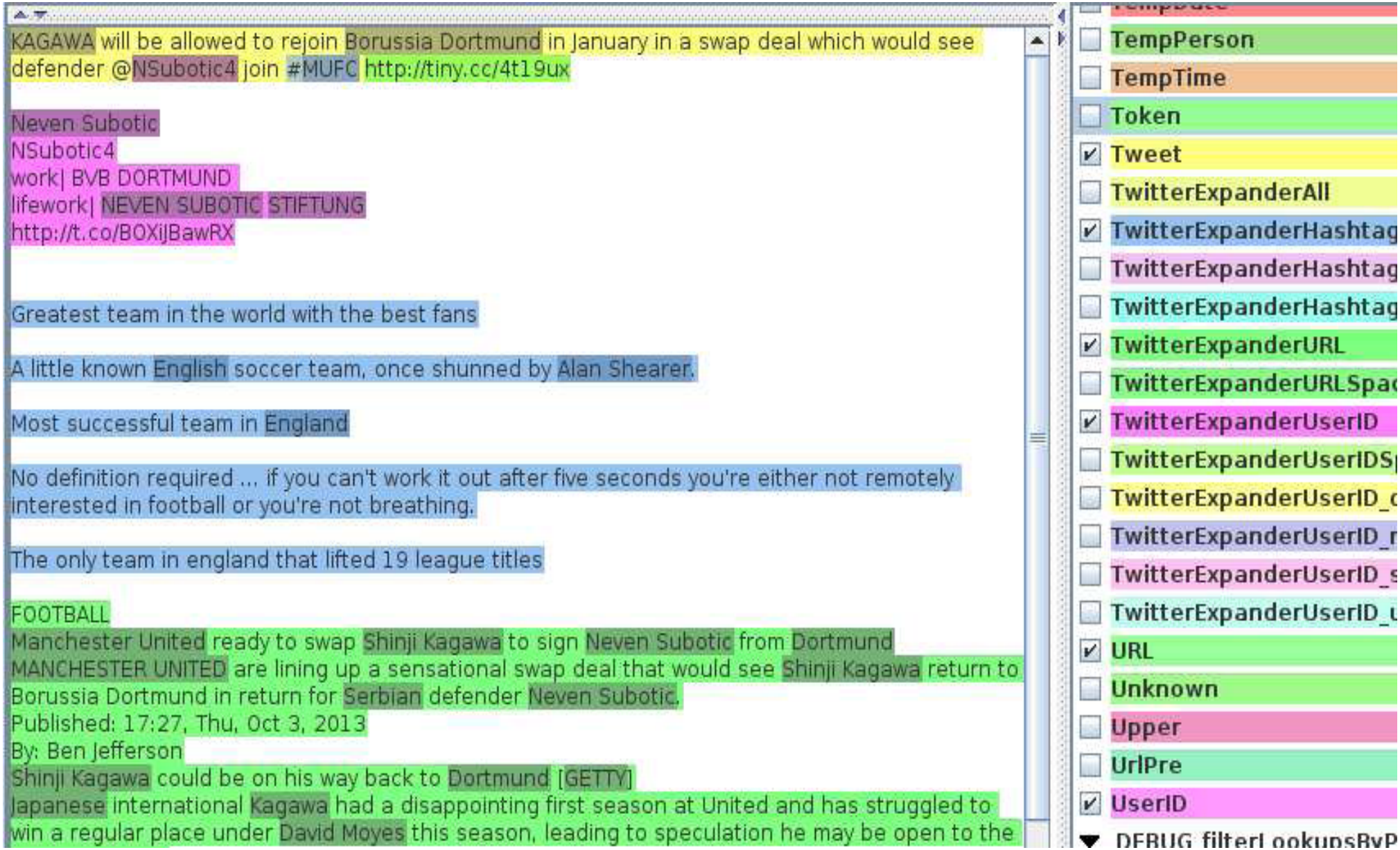
YODIE Disambiguation

- SVM-based candidate selection trained on disambiguation corpora to determine probability of a candidate being correct
- Where no candidate is considered likely to be correct, all are rejected—the spot is considered spurious
- Where multiple candidates are accepted, probability is used to select the best

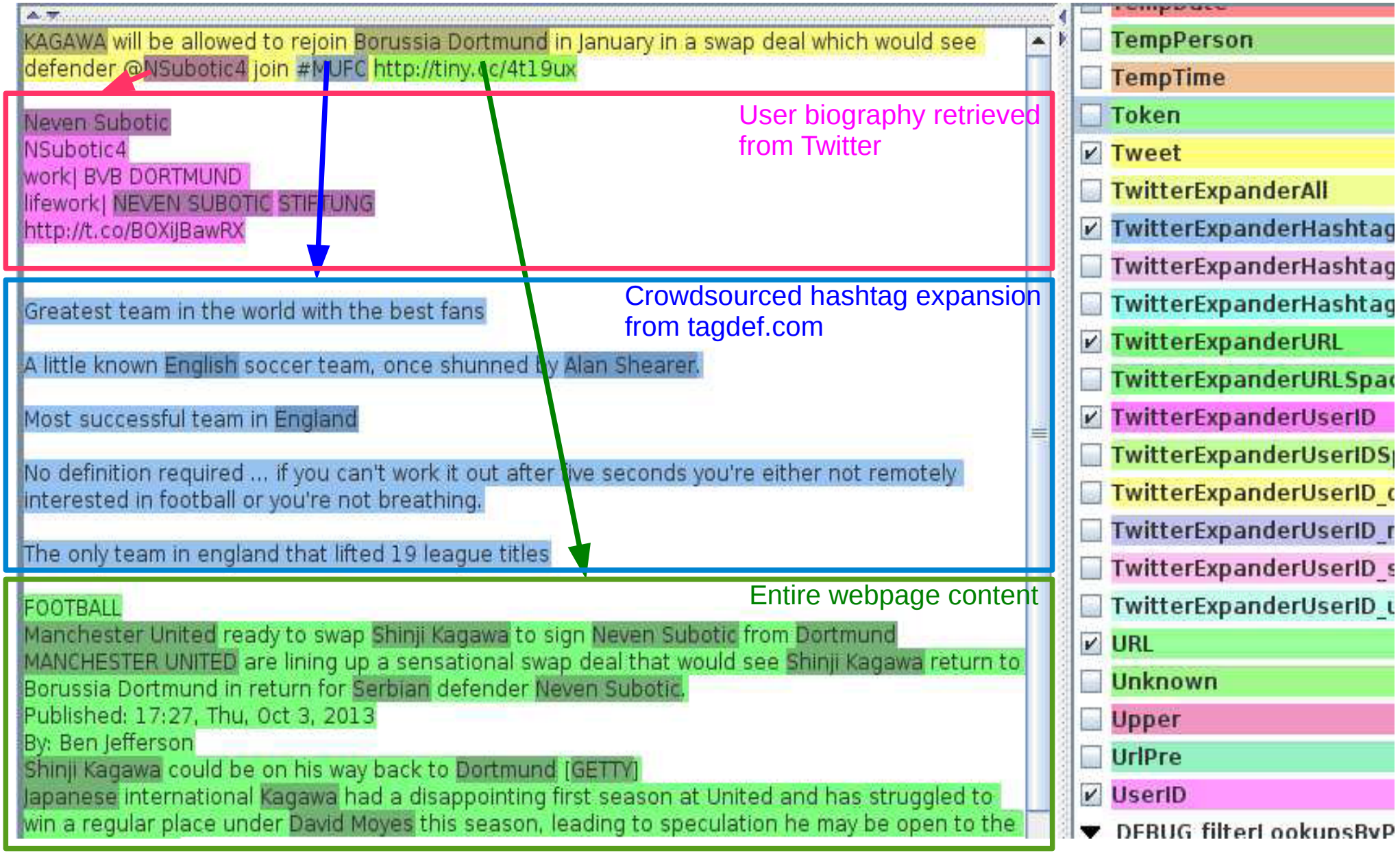
Tweet Expansions Studied

- Hashtag definitions are retrieved from https://tagdef.com
- URL expansion retrieves entire of page content
- @-mention expansion retrieves user biography

GATE GUI Screenshot with Expansion Examples



GATE GUI Screenshot with Expansion Examples



The University Of Sheffield.

How Expansions Make a Difference

- Additional material gives structural scores more entities to relate
- Additional material gives textual scores more relevant text to operate across
- Additionally, any entities found in a biography are added as candidates to the @-mention it was expanded from (hereafter referred to as “back-projection”)
- Disambiguation stages adds further intelligence

Experimental Conditions

- **Base**—A baseline system with no Twitter expansion
- **Id**—Includes @-mention expansion only
- **Url**—Includes URL expansion only
- **Hash**—Includes hashtag expansion only
- **Id + Proj**—Includes @-mention expansion with back-projection
- **Id + Proj + Url**—Try excluding hashtag expansion since later results suggest limited utility
- **All**—Everything altogether

Baselines

- Picking a candidate at random: F1 = 0.229
- Picking candidate with most URI mentions in Wikipedia: F1 = 0.521
  - URI frequency in Wikipedia is a good prior and a strong baseline, hard to beat!
  - Other scores however contribute different kinds of information and can increase the score in conjunction with the prior even if they are lower

The University Of Sheffield.

Structural Scores—LOD-based

	Precision	Recall	F1
Base	**0.416**	0.267	0.326
Id	0.399	**0.276**	0.326
URL	0.414	0.272	**0.328**
#	0.385	0.253	0.305
Id w/proj	0.318	0.266	0.313
Id w/proj + URL	0.373	0.269	0.312
All	0.373	0.260	0.306

- LOD-based structural similarity is a strong but sparse metric—had hoped for a better result!

The University Of Sheffield.

Structural Scores—Relatedness

	Precision	Recall	F1
Base	0.236	0.244	0.240
Id	**0.253**	0.272	**0.262**
Url	0.236	0.242	0.239
Hash	0.235	0.241	0.238
Id +Proj	0.244	0.269	0.256
Id + Proj + Url	0.249	**0.276**	**0.262**
All	0.250	0.266	0.258

- Though a weaker score than LOD-based, Twitter expansion has more impact here
  - @-mentions seem particularly helpful

The University Of Sheffield.

Text-based—Abstracts Only

	Precision	Recall	F1
Base	0.201	0.421	0.272
Id	0.208	0.436	0.282
Url	0.194	0.407	0.263
Hash	0.204	0.427	0.276
Id +Proj	**0.217**	**0.463**	**0.295**
Id + Proj + Url	0.212	0.454	0.289
All	0.216	0.461	0.294

- URL seems to make it worse but everything else helps
- Particularly, back-projection improves recall

The University Of Sheffield.

Text-based—Abstracts Plus Other Textual Fields in DBpedia

	Precision	Recall	F1
Base	0.221	0.379	0.279
Id	0.226	0.389	0.286
Url	0.234	0.402	0.296
Hash	0.234	0.401	0.295
Id +Proj	0.235	0.414	0.300
Id + Proj + Url	0.247	0.434	0.315
All	**0.253**	**0.446**	**0.323**

- Including, for each candidate, all textual fields found in DBpedia, gives most obvious improvement of scores considered
- Expansions all seem to help to some degree

The University Of Sheffield.

Including the Disambiguation Stage

	Precision	Recall	F1	Accuracy
Base	0.442	0.550	0.490	0.550
Id	0.444	0.557	0.494	0.557
Id + Proj	0.444	0.642	0.525	0.642**
Url	0.452	0.568	0.504	0.568*
Hash	0.446	0.559	0.496	0.559*
Id + Proj + Url	0.452	**0.660**	0.536	**0.660****
All	**0.495**	0.623	**0.552**	0.623**

- Individual score results vary from low to somewhat substantial, but combining together with SVM stage increases scope
- Significant improvements in accuracy of 7% are obtainable, and 6% in F1 (significance not calculated for F1)

The University Of Sheffield.

Contextualizing the Result

	Prec	Recall	F1
YODIE (Base)	0.44	0.55	0.49
YODIE (Exp)	0.50	0.62	**0.55**
Aida 2014	**0.59**	0.38	0.46
Lupedia	0.50	0.24	0.32
Spotlight	0.09	0.51	0.15
TagMe	0.10	**0.67**	0.17
TextRazor	0.19	0.44	0.26
Zemanta	0.48	0.56	0.52

- Table above shows that this magnitude of improvement is substantial on the scale of performances obtained by competitive systems, repositioning YODIE
  - (This isn't a competition though! The other systems don't use tweet expansion)

Contextualizing the Result

- Improvement obtainable depends on characteristics of corpus
  - How many expansions in it? How many expansions succeed?
- If every tweet had a successful expansion ..
  - Hashtag—accuracy gain 0.04
  - URL—accuracy gain 0.06
  - @-mention—accuracy gain 0.23
- Different tweet selections have different characteristics
  - E.g. news tweets often have a URL

The University Of Sheffield.

Thanks for Listening!

Talk references:

Basave, Amparo Elizabeth Cano, et al. "Making Sense of Microposts (# MSM2013) Concept Extraction Challenge." # MSM. 2013.

Finin, Tim, et al. "Annotating named entities in Twitter data with crowdsourcing." Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Association for Computational Linguistics, 2010.

Han, Xianpei, Le Sun, and Jun Zhao. "Collective entity linking in web text: a graph-based method." Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 2011.

Meij, Edgar, Wouter Weerkamp, and Maarten de Rijke. "Adding semantics to microblog posts." Proceedings of the fifth ACM international conference on Web search and data mining. ACM, 2012.

Ritter, Alan, Sam Clark, and Oren Etzioni. "Named entity recognition in tweets: an experimental study." Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011.

Witten, I., and David Milne. "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links." Proceeding of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy, AAAI Press, Chicago, USA. 2008.