HITS' Monolingual and Cross-lingual Entity Linking System at TAC 2012: A Joint Approach

Angela Fahrni, Thierry Göckel and **Michael Strube**
Heidelberg Institute for Theoretical Studies
Schloss-Wolfsbrunnenweg 35
69118 Heidelberg, Germany
{Angela.Fahrni|Thierry.Goeckel|Michael.Strube}@h-its.org

Abstract

This paper presents HITS' system for monolingual and cross-lingual entity linking at TAC 2012. We propose a joint system for entity disambiguation, recognition of NILs and clustering using Markov Logic. The proposed model (1) is *global*, i.e. a group of mentions in a text is disambiguated in one single step combining various global and local features, and (2) performs disambiguation, unknown entity detection and clustering *jointly*. The model for all languages is exclusively trained on English Wikipedia articles.

The results achieved in the TAC monolingual and cross-lingual entity linking tasks show that our approach is competitive: our best English run achieves 8.5 percent points above median, while we outperformed all other participating systems in the Chinese cross-lingual subtask. The results for the Spanish subtask are lower due to a bug. Our unofficial Spanish results (after fixing the bug) are close to the ones of the best system.

1 Introduction

HITS participated in the English monolingual and the Chinese and Spanish cross-lingual entity linking tasks at TAC 2012 with a novel approach (see also Fahrni and Strube (2012)). While we already participated in the Chinese cross-lingual entity linking task in 2011 (Fahrni et al., 2012), we participated in the two other tasks for the first time.

Entity linking involves three subtasks:

1. **Entity Disambiguation** is the task of identifying the corresponding entry in a predefined knowledge base (*KB*) for proper nouns that refer to persons, places or organizations (*query terms*).
2. **Recognition of NILs** is the task of detecting query terms that do not refer to a known entity, which is part of the predefined KB, but to an unknown entity.
3. **Clustering of NILs** is the task of grouping query terms with no corresponding entry in the KB so that all query terms in a cluster refer to the same unknown entity.

Previously, these three subtasks have been approached in a cascaded way (Ji et al., 2011). We propose a novel joint approach that performs entity disambiguation, recognition of NILs and clustering jointly using Markov Logic. Mentions are not only clustered if they refer to unknown entities, but also if they refer to a known entity. The main motivation behind this approach is that disambiguation and clustering are interleaved and can support each other. While disambiguation models relations between entities and mentions, clustering focuses on modeling relations between mentions.

Our model is not optimized for TAC data. We propose a general disambiguation and clustering system that uses few features and considers common and proper nouns (in Fahrni and Strube (2012), we apply our model to a different dataset where we disambiguate not only proper nouns but also common nouns). Given the TAC testing data, we do not disambiguate only the query term, but all mentions that influence the decision for the query term given our features. The model is trained on 500 English Wikipedia articles and then applied to English, Spanish and Chinese data. No TAC training data has been used to train the model.

The mapping strategy for the two cross-lingual entity linking tasks is the same as last year. We map the articles in the Chinese and Spanish Wikipedia to entries in the English Wikipedia using interlanguage links beforehand. The disambiguation for Chinese and Spanish queries is done with respect to this mapped index.

The remainder of the paper is organized as follows. In Section 2 we discuss related work. Our approach is presented in Section 3, while the experiments are analyzed in Section 4.

2 Related Work

Most systems, including our last year's system (Fahrni et al., 2012), approach entity disambiguation and recognition of NILs (Bunescu and Paşca, 2006; Dredze et al., 2010) or disambiguation and clustering (Ji et al., 2011) in a cascaded way. Monahan et al. (2011) interleave entity linking and clustering, but they do not approach the two tasks jointly: after disambiguation, mentions are clustered. Then each cluster is assigned an entity in the knowledge base if there exists a corresponding one. Dai et al. (2011) perform entity disambiguation and recognition of the NILs jointly using Markov Logic. In contrast to us, they do not cluster mentions and focus on one specific type of mentions in the biological domain, namely mentions that refer to genes.

Global disambiguation approaches are another strand of work that is similar to ours. While early work often uses local classifiers or rankers that select an entity for each mention independently (Csomai and Mihalcea, 2008; Milne and Witten, 2008; Dredze et al., 2010), recently various global approaches have been proposed. Kulkarni et al. (2009) propose a method that maximizes local context-concept compatibility and global concept coherence. Fahrni et al. (2011) use a graph-based approach and select the best combination of entities given the graph structure. Han and Sun (2012) use a generative model integrating topic coherence (one topic per document) and local context compatibility. Ratinov et al. (2011) describe a two pass method and use the output of the first pass as input for the second one. While all these approaches use a limited number of global features, we integrate various global features and also learn their weights.

While previous disambiguation approaches are mainly evaluated on one single language, recently released multilingual evaluation data sets (e.g. NTCIR 9[1], TAC 2011 and Mayfield et al. (2011)) allow to evaluate systems on several languages and to focus on portability across them.

The most prominent research line for sense induction are distributional approaches (Schütze, 1998). Pedersen (2006) gives an overview over state-of-the art techniques. Recently, the efficiency problem caused by the number of necessary comparisons has been addressed (Singh et al., 2011). While Rao et al. (2010) apply streaming clustering, Wick et al. (2012) propose a discriminative hierarchical model and partition entities into trees of latent sub-entities. None of these approaches for clustering also disambiguate entities at the same time.

3 Approach

This section describes our joint approach of entity disambiguation, recognition of NILs and clustering using Markov Logic Networks and explains how we extend the system for the cross-lingual tasks.

3.1 Markov Logic Networks

Markov Logic (ML) combines first-order logic with probabilities (Domingos and Lowd, 2009). A Markov Logic Network (MLN) consists of a set of pairs (F_i, w_i), where F_i is a first-order formula and $w_i \in \mathbb{R}$ is a weight associated with the formula F_i. It builds a template for constructing a Markov Network given a set of constants C. This Markov Network contains a binary node for each possible grounding for each predicate of the Markov Logic Network. If the grounding of the predicate is true the value of this binary node is 1, otherwise 0. In addition, it contains one feature[2] for each ground formula. If a ground formula is true, the feature for this ground formula has the value 1, otherwise 0. The weight of the feature is given by w_i.

[1] `http://ntcir.nii.ac.jp/CrossLink`

[2] Note that *feature* is used differently in this section than in the rest of the paper.

The probability distribution for the ground Markov Network is represented by

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_i w_i n_i(x)\right)$$

where $n_i(x)$ is the number of true groundings of F_i in x. The normalization factor Z is the partition function.

To learn the weights for the formulas and to perform MAP inference, we use *thebeast*.[3] *thebeast* employs cutting plane inference (Riedel, 2008). To learn the weights we use a perceptron.

3.2 Disambiguation and Clustering with MLNs

In order to define how disambiguation, recognition of NILs and clustering interact, we use hard constraints. While we describe these constraints in this section, we explain the features in the next one.

Table 1 shows all used predicates and formulas. Each formula is associated with a positive or negative weight. While the weight – except for hard constraints – is learnt from training data, the polarity of the weights is set manually. In the following, we indicate the polarity by the $+$ or $-$ in front of each formula. For some formulas, the final weight consists of a learned weight w multiplied by a score s (e.g. the prior probability). In these cases, the final weight for a formula does not just depend on the respective formula, but also on the instantiation, e.g. a specific mention and candidate entity. We indicate such combined weights by the term $w \cdot s$, while w refers to cases where the formula is exclusively weighed by the learned weight. M denotes all mentions and E_m refers to all candidate entities of a mention m.

3.2.1 The Core of the Approach

Entity disambiguation and clustering are two different ways to deal with lexical ambiguities. The two tasks focus on different relations:

- Entity disambiguation models the relation between mentions and entities in a knowledge base. In order to solve lexical ambiguities, mentions are linked to entities in a given knowledge base.
- Entity clustering models the relation between mentions. Mentions are clustered, so that all mentions in a group refer to the same entity.

In the task of recognizing NILs the same relation is considered as in entity disambiguation, i.e. the relation between mentions and entities. Whereas in entity disambiguation the question is to which entity a mention given its context refers to, the task of recognizing unknown entities is to determine if such an entity relation exists for a given mention at all.

To approach entity disambiguation, recognition and clustering of NILs with ML, we define a hidden predicate for each relation we are interested in. The predicate *hasEntity(Mention, Entity)* models the relation between mentions and entities in the knowledge base (Table 1, p1). To ensure that each mention refers to at most one entity, a hard cardinality constraint is defined: for each mention the predicate *hasEntity* is true at most once. This constraint allows us to do joint disambiguation and recognition of NILs (Table 1, f1).

To model if two mentions refer to the same entity, the predicate *hasSameEntity(Mention, Mention)* is used (Table 1, p2). It is true for all mention pairs that refer to the same entity, independently of the fact whether the referred entity exists in the knowledge base or not. This clustering relation is *transitive* and *symmetric* (Table 1, f2, f3).

In order to perform joint disambiguation and clustering, we need to define how the mention-entity relation (disambiguation, recognition of NILs) and the clustering relation are interrelated (Table 1, f4, f5). Given that two mentions refer to the same entity in the knowledge base, they belong to the same cluster (f5). On the other hand, if two mentions are part of the same cluster and one of them refers to an entity in the knowledge base, the other mention in the cluster has to refer to the same entity (f4). Note that two mentions can also be in the same cluster without referring to an entity in the knowledge base.

3.3 Features

All features have a corresponding predicate which is part of at least one formula (see Table 1).

3.3.1 Local Features

Local features involve one single mention and its candidate entities.

[3] `http://code.google.com/p/thebeast`

Prior probability (p3, f8) The prior probability is defined as the probability that a mention m refers to an entity e. To estimate this probability, all internal hyperlinks are extracted from the English, Chinese and Spanish Wikipedia dumps. For each linked mention m, it is counted how many times it links to a particular Wikipedia page, i.e. entity. This count is normalized by the number of times mention m is linked in Wikipedia.

Relatedness (p4, f9, f12) This feature reflects the average pairwise relatedness of a candidate entity for a mention to the context. The pairwise relatedness measure considers the incoming link structure in Wikipedia and is calculated in the same way as proposed by Milne and Witten (2008).

Co-occurrence probability of two entities (p5, f13) Given a candidate entity for a mention, the co-occurrence probability to the candidates of all other mentions is calculated. The minimum (if possible non-zero) value is taken as a score. The co-occurrence probability is calculated as described in Fahrni et al. (2012).

Local context similarity (p6, f10) The local context similarity measures how similar the current local context C_m – consisting of seven words before and after the mention – is to the local contexts for that entity in Wikipedia. For each mention in Wikipedia that is linked to a certain Wikipedia page e we extract the surrounding words T_e using the same context definition as above. We then calculate the local context similarity ($sim(e,m)$) for a candidate entity e of a mention m via

$$sim(e,m) = \frac{1}{|C_m|} \sum_{c \in C_m} s(c, T_e)$$

where the first term is used for normalization and $s(c, T_e)$ denotes the frequency of c in T_e divided by the number of times c appears in the context of all entities in Wikipedia.[4]

String edit distance (p7, f11) This feature accounts for the difference between the mention string m used in the text and the preferred name p for a candidate entity of m. We assume that the Wikipedia article title and the titles of its redirects are preferred names for an entity. To measure the distance between preferred names and the mention in the text, we calculate the edit distance[5] and normalize it by the length of the longer string. If there exists more than one preferred term for an entity, we take the minimum distance. This feature indicates a negative relation between a candidate entity and a mention. The more distant a preferred name is from a mention, the less likely it is that the mention refers to this entity.

3.3.2 Global Features

In contrast to local features, global features involve more than one mention. From a disambiguation perspective, these features define which mentions are disambiguated jointly.

Shared lemma (p10, f14) The one sense per discourse assumption states that one mention string is used to refer to one sense, i.e. in our case to one entity, in one discourse (Gale et al., 1992). For each document, we extract all mentions with the same lemma and the inverse distance in sentences between the two. The bigger the inverse distance is, the closer the two mentions are to each other and the more likely it is that they refer to the same entity.

Head match (p8, f6) The one entity per discourse assumption often applies to mentions which are substrings of each other and share the same syntactic head lemma. We extract all these pairs and the inverse distance between the respective mentions.

Acronyms (p8, f6) In texts, especially in newspaper texts, acronyms are often introduced by the pattern *full name (acronym)*. We extract all these mention pairs, where one mention is the full name and the other one the acronym.[6]

Partial string match (p9, f7) If two mentions are person names and one is a substring of the other, we assume that they refer to the same entity with a certain probability. We extract all these pairs and the inverse distance between the respective mentions. In

[4] We take its logarithm.

[5] We use the Lingpipe implementation (`http://alias-i.com/lingpipe/`).

[6] In our Wikipedia training data, acronyms are relatively rare. Hence it is difficult to learn a weight for the acronym feature. As it is similar to the head match feature, we use the same predicate and weight for the two features.

Predicates

Hidden predicates	
p1	$hasEntity(m, e)$
p2	$hasSameEntity(m, n)$
Predicates realizing Wikipedia Miner features	
p3	$hasCommonness(m, e, s)$
p4	$hasRelatedness(m, e, s)$
p5	$hasCoocProb(m, e, s)$
Additional predicates involving one mention and one entity	
p6	$hasContextSimilarity(m, e, s)$
p7	$hasStringDistance(m, e, s)$
Predicates involving two mentions (intradocument)	
p8	$isSubStringHeadMatch(m, n, s)$
p9	$isPartialStringMatch(m, n, s)$
p10	$haveSameLemma(m, n, s)$
Predicates involving two mentions (cross-document)	
p11	$shareNgram(m, n, s)$

Formulas

Hard constraints			
f1			$\forall m \in M:\ \lvert\{e \in E : hasEntity(m, e)\}\rvert \leq 1$
f2			$\forall m, n \in M:\ m \neq n \land hasSameEntity(m, n) \rightarrow hasSameEntity(n, m)$
f3			$\forall m, n, l \in M:\ m \neq n \land m \neq l \land n \neq l$ $\land\ hasSameEntity(m, n) \land hasSameEntity(n, l) \rightarrow hasSameEntity(m, l)$
f4			$\forall m, n \in M:\ m \neq n \land hasSameEntity(m, n) \land hasEntity(m, e)$ $\rightarrow hasEntity(n, e)$
f5			$\forall m, n \in M:\ m \neq n \land hasEntity(m, e) \land hasEntity(n, e)$ $\rightarrow hasSameEntity(m, n)$
Formulas with learned weights			
f6	$+$	$(w \cdot s)$	$\forall m, n \in M\ \forall e \in E_m:\ m \neq n \land isSubStringHeadMatch(m, n, s)$ $\rightarrow hasEntity(m, e) \land hasEntity(n, e)$
f7	$+$	$(w \cdot s)$	$\forall m, n \in M\ \forall e \in E_m:\ m \neq n \land isPartialStringMatch(m, n, s)$ $\rightarrow hasEntity(m, e) \land hasEntity(n, e)$
f8	$+$	$(w \cdot s)$	$\forall m \in M\ \forall e \in E_m:\ hasCommonness(m, e, s)$ $\rightarrow hasEntity(m, e)$
f9	$+$	$(w \cdot s)$	$\forall m \in M\ \forall e \in E_m:\ hasRelatedness(m, e, s) \rightarrow hasEntity(m, e)$
f10	$+$	$(w \cdot s)$	$\forall m \in M\ \forall e \in E_m:\ hasContextSimilarity(m, e, s)$ $\rightarrow hasEntity(m, e)$
f11	$-$	$(w \cdot s)$	$\forall m \in M\ \forall e \in E_m:\ hasStringDistance(m, e, s)$ $\rightarrow hasEntity(m, e)$
f12	$-$	(w)	$\forall m \in M\ \forall e \in E_m:\ hasRelatedness(m, e, s) \land s = 0$ $\rightarrow hasEntity(m, e)$
f13	$-$	$(w \cdot (1.0 - s))$	$\forall m \in M\ \forall e \in E_m:\ hasCoocProbability(m, e, s)$ $\rightarrow hasEntity(m, e)$
f14	$+$	$(w \cdot s)$	$\forall m, n \in M:\ m \neq n \land hasSameString(m, n, s)$ $\rightarrow hasSameEntity(m, n)$
f15	$+$	$(w \cdot s)$	$\forall m, n \in M:\ m \neq n \land shareNgram(m, n, s)$ $\rightarrow hasSameEntity(m, n)$

Table 1: Predicates and formulas used for entity disambiguation and clustering (m, n, l represent mentions, M sets of mentions, e an entity, E all entities, E_m all candidate entities for mention m and s scores)

order to decide if an English mention is a name of a person, we use the *CoNLL* gender list (Bergsma and Lin, 2006). If a mention is part of the list and mainly associated with the male or female gender, we consider it as a person name. In Spanish, we check if one of the candidate entities for a mention refers to a person according to the TAC KB. For Chinese, we use a list of Chinese names and their English equivalents we extracted from *Baidu Baike*. If the English equivalent refers to a person according to the TAC KB, we consider the mention as a person name.

Cross-document n-gram feature (p11, f15) In contrast to the previous features, this one is a cross-document feature. The assumption is that we work with a document collection. We extract all mention pairs with the same lemma but coming from two different documents. For each of these mentions, we extract all n-grams that include the respective mention and that consist of nouns and adjectives. If the two mentions share at least one of these n-grams, we consider them as referring to the same entity and add as score the number of shared n-grams.

3.4 From a Monolingual to a Cross-lingual System

In order to extend our monolingual system for the cross-lingual task, we pursue the same mapping strategy as last year (Fahrni et al., 2012). We map the Chinese and Spanish Wikipedia articles to the English articles using interlanguage links[7] and disambiguate Chinese and Spanish mentions directly with respect to these mapped articles. If a Chinese or Spanish Wikipedia article has no corresponding English page in Wikipedia, we keep the original article, but assign it a new ID. The advantage of doing the mapping before linking is that we can use the English link structure – which is richer as for example the Chinese one – to calculate relatedness. We also did some experiments with a list of Chinese and English equivalents we extracted from *Baidu Baike*[8], a Chinese encyclopedia. As the performance went down due to noise, we did not use this list for our submissions to identify candidate entities, but just for feature *f7*.

Apart from the lexicon, the cross-lingual system only differs from the monolingual system regarding a relatedness feature. Experiments have shown that formula *f12* is not strong in the cross-lingual scenario, as the relatedness calculations behind this feature is not as strong in the cross-lingual case. We therefore removed formula *12* in the cross-lingual subtasks and added another relatedness measure, which performed better (*p5, f13*).

Both the monolingual and the cross-lingual system are trained on 500 English Wikipedia articles.

4 Experiments

4.1 Processing TAC queries

The system proposed in Section 3 is a general purpose disambiguation and clustering system that is not designed for the TAC scenario. To suit the TAC scenario, some modifications need to be done.

In the proposed system, mentions are disambiguated with respect to the English Wikipedia. In order to suit the TAC evaluation, we need to map our IDs to the ones of the TAC KB. The mapping is done by comparing the article titles and the ones of their corresponding redirects with the entry names of the TAC KB. In case of ties, we select the entry whose description has the highest cosine similarity to the Wikipedia article. In total, we mapped 790,963 entries.

While in the TAC scenario the focus lies on a few selected query terms, we disambiguate all mentions in a text. To process the TAC testing data, we proceed as follows:

1. **Text cleaning:** HTML tags and noise (e.g. in Web documents) are removed.
2. **Preprocessing:** We tokenize the texts and perform POS tagging and parsing. For English and Chinese, we use *Stanford CoreNLP*[9], for Spanish, we use *FreeLing* (Padró and Stanilovsky, 2012).
2. **Mention detection and feature extraction:** In this step mentions are identified and the features

[7]We use the following Wikipedia dumps: English (2012/01/04), Chinese (2012/08/22), Spanish (2012/07/28), German (2012/01/16), Italian (2012/01/26), Dutch (2012/01/19).

[8]`http://baike.baidu.com`

[9]`http://nlp.stanford.edu/software/corenlp.shtml`

are extracted. To identify candidates for the mentions we use a lexicon. To create this lexicon we extract all anchors, article titles and titles from redirects from the English, Chinese and Spanish Wikipedia dumps. In order to reduce the noise of anchors in English, we just consider an anchor if it is used at least two times to refer to the respective Wikipedia article. We did not set this constraint in Chinese and Spanish, as these Wikipedia dumps are smaller.

2. **Query term identification:** In order to decide which identified mention is a query term, string similarity between the identified mentions and the query terms for a document is calculated. For each query term we select the mention with the highest similarity. In case of ties, we select the mention with the smallest difference regarding the offsets. Due to the cleaning and preprocessing, the original offset information is lost, which is why the chosen mention does not always have to be the intended one.

3. **Inference:** We use *thebeast* (Riedel, 2008). We disambiguate and cluster the query terms as well as all identified mentions that influence the decision for the query terms given our features directly or transitively. As the clustering can cross document boundaries, we process the mentions from documents that can be in the same cluster (given our features) at the same time.

4. **Postprocessing:** For some runs (see Table 2) we ignore the clusters produced by the inference step and use string matching to cluster the query terms that refer to the same entity. In the following this strategy is called *cluster_strm*. For all other runs, we use the cluster produced during inference. The query terms which are in no cluster are grouped using string matching as a back-off strategy (*cluster_mln*).[10]

4.2 Results

HITS participated in all three entity linking tasks. In total, we submitted four runs for the English, two runs for the Chinese and four runs for the Spanish entity linking task. The differences between the runs are summarized in table 2.

Table 3 shows the results for all runs. Except for the Chinese subtask where string match is a strong baseline for clustering, the induced clusters are better than the one produced by simple string match: run *HITS1* and *HITS2* as well as *HITS3*, *HITS3** and *HITS4* (English and Spanish) only differ regarding the postprocessing.

The results show that our approach is competitive. While the results for our best English run are between median and best scores, we outperform all other participating systems in the Chinese cross-lingual entity linking task. Our official results in the Spanish cross-lingual entity linking task are low due to problems with the post-processing. We solved the problems (a small bug) and present the new numbers for Spanish (HITS3*), too. The Spanish (unofficial) results are close to the ones of the best system.

A more detailed analysis of our system's results shows that our system performs well regarding persons and organizations across all languages, while the scores for GPEs are lower. The current system does not include specific features for different types of named entities. The performance for GPEs could be improved by adding more specific features.

5 Conclusions

HITS participated with a novel approach in the monolingual and Chinese and Spanish cross-lingual subtasks. We propose a disambiguation system that approaches entity disambiguation, recognition of unknown entities and clustering jointly using Markov Logic Networks. We trained all models for all languages on 500 English Wikipedia articles and did not use any TAC data to train the system. In Fahrni and Strube (2012) we apply disambiguate not only to proper nouns but also to common nouns and show that our joint approach can serve as a general purpose disambiguation system.

Acknowledgments. We would like to thank Mathias Niepert for the advice on MLNs and *thebeast* and our colleagues Ye Lin and Jie Cai for help with regard to Chinese. This work has been partially funded by the European Commission through the CoSyne project FP7-ICT-4-248531 and the Klaus Tschira Foundation.

[10] The induced clusters are also affected by the fact that some mentions are NILs according to the TAC KB, but not according to the Wikipedia dumps we use.

Run ID	Features / Formulas	Postprocessing	Resources
English Entity Linking Task			
HITS1	full model	cluster_strm	*CoNLL* gender data
HITS2	full model	cluster_mln	*CoNLL* gender data
HITS3	full model, without predicate *p5* and formula *f13*	cluster_strm	*CoNLL* gender data
HITS4	full model, without predicate *p5* and formula *f13*	cluster_mln	*CoNLL* gender data
Chinese Entity Linking Task			
HITS1	EN: full model, without predicate *p5* and formula *f13* ZH: full model, without formula *f12*	cluster_strm	List from *Baidu Baike*
HITS2	EN: full model, without predicate *p5* and formula *f13* ZH: full model, without formula *f12*	cluster_mln	List from *Baidu Baike*
Spanish Entity Linking Task			
HITS1	EN: full model, without predicate *p5* and formula *f13* ES: full model, without predicate *p5* and formulas *f13* and *f12*	cluster_mln	
HITS2	EN: full model, without predicate *p5* and formula *f13* ES: full model, without predicate *p5* and formulas *f13* and *f12*	cluster_stm	
HITS3	EN: full model ES: full model, without formula *f12*	cluster_mln	
HITS4	EN: full model ES: full model, without formula *f12*	cluster_stm	
HITS3*	Unofficial run, same as HITS3, but a bug is fixed	cluster_mln, fixed bug	

Table 2: Description of the different runs of HITS for the monolingual and cross-lingual entity linking tasks at TAC 2012

Run	**Micr.**	**B^3 P**	**B^3 R**	**B^3 F1**	**B^{3+} P**	**B^{3+} R**	**B^{3+} F1**
English Entity Linking Task							
Best							0.730
Median							0.536
HITS2	0.718	0.751	0.932	0.832	0.572	0.678	0.621
HITS1	0.718	0.625	0.938	0.750	0.465	0.683	0.553
HITS4	0.679	0.750	0.929	0.830	0.557	0.641	0.596
HITS3	0.679	0.615	0.933	0.741	0.443	0.644	0.525
Chinese Cross-lingual Entity Task							
Best							0.740
HITS1	0.843	0.863	0.811	0.836	0.738	0.742	0.740
HITS2	0.843	0.882	0.794	0.836	0.753	0.727	0.740
Spanish Cross-lingual Entity Task							
Best							0.641
HITS3	0.707	0.648	0.880	0.746	0.464	0.638	0.538
HITS4	0.707	0.530	0.910	0.670	0.359	0.662	0.465
HITS1	0.674	0.658	0.900	0.760	0.472	0.621	0.536
HITS2	0.674	0.510	0.918	0.656	0.340	0.634	0.443
HITS3*	0.707	0.904	0.830	0.866	0.660	0.612	0.635

Table 3: HITS' performance compared to the best and median scores in the monolingual and cross-lingual entity linking tasks

References

Shane Bergsma and Dekang Lin. 2006. Bootstrapping path-based pronoun resolution. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics,* Sydney, Australia, 17–21 July 2006, pages 33–40.

Razvan Bunescu and Marius Paşca. 2006. Using encyclopedic knowledge for named entity disambiguation. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics,* Trento, Italy, 3–7 April 2006, pages 9–16.

Andras Csomai and Rada Mihalcea. 2008. Linking documents to encyclopedic knowledge. *IEEE Intelligent Systems*, 23(5):34–41.

Hong-Jie Dai, Richard Tzong-Han Tsai, and Wen-Lian Hsu. 2011. Entity disambiguation using a Markov-Logic network. In *Proceedings of the 5th International Joint Conference on Natural Language Processing,* Chiang Mai, Thailand, 8–13 November 2011, pages 846–855.

Pedro Domingos and Daniel Lowd. 2009. *Markov Logic: An Interface Layer for Artificial Intelligence*. Morgan Claypool Publishers.

Mark Dredze, Paul McNamee, Delip Rao, Adam Gerber, and Tim Finin. 2010. Entity disambiguation for knowledge base population. In *Proceedings of the 23rd International Conference on Computational Linguistics,* Beijing, China, 23–27 August 2010, pages 277–285.

Angela Fahrni and Michael Strube. 2012. Jointly disambiguating and clustering concepts and entities with Markov logic. In *Proceedings of the 24th International Conference on Computational Linguistics,* Mumbai, India, 8–15 December 2012.

Angela Fahrni, Vivi Nastase, and Michael Strube. 2011. HITS' graph-based system at the NTCIR-9 cross-lingual link discovery task. In *Proceedings of the 9th NTCIR Workshop Meeting,* Tokyo, Japan, 6-9 December 2011, pages 473–480.

Angela Fahrni, Vivi Nastase, and Michael Strube. 2012. HITS' cross-lingual entity linking system at TAC 2011: One model for all languages. In *Proceedings of the Text Analysis Conference,* National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 14–15 November 2011.

William A. Gale, Kenneth W. Church, and David Yarowsky. 1992. One sense per discourse. In *Proceedings of the DARPA Speech and Natural Language Workshop,* New York, N.Y., 23-26 February 1992, pages 233–237.

Xianpei Han and Le Sun. 2012. An entity-topic model for entity linking. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics,* Jeju Island, Korea, 8–14 July 2012, pages 105–115.

Heng Ji, Ralph Grishman, and Hoa Dang. 2011. Overview of the TAC 2011 knowledge base population track. In *Proceedings of the Text Analysis Conference,* National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 14–15 November 2011.

Sayali Kulkarni, Amit Singh, Ganesh Ramakrishnan, and Soumen Chakrabarti. 2009. Collective annotation of Wikipedia entities in web text. In *Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* Paris, France, 28 June – 1 July 2009, pages 457–466.

James Mayfield, Dawn Lawrie, Paul McNamee, and Douglas W. Oard. 2011. Building a cross-language entity linking collection in 21 languages. In *Proceedings of the 2011 Conference on Multilingual and Multimodal Information Access Evaluation,* Amsterdam, The Netherlands, 19-22 September 2011.

David Milne and Ian H. Witten. 2008. Learning to link with Wikipedia. In *Proceedings of the ACM 17th Conference on Information and Knowledge Management (CIKM 2008),* Napa Valley, Cal., USA, 26–30 October 2008, pages 1046–1055.

Sean Monahan, John Lehmann, Timothy Nyberg, Jesse Plymale, and Arnold Jung. 2011. Cross-lingual cross-document coreference with entity linking. In *Proceedings of the Text Analysis Conference,* National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 14–15 November 2011.

Lluís Padró and Evgeny Stanilovsky. 2012. FreeLing 3.0: towards a wider multilinguality. In *Proceedings of the 8th International Conference on Language Resources and Evaluation,* Istanbul, Turkey, 21–27 May 2012.

Ted Pedersen. 2006. Unsupervised corpus-based methods for WSD. In E. Agirre and P.G. Edmonds, editors, *Word Sense Disambiguation: Algorithms and Applications*, pages 133–166. Springer, Heidelberg, Germany.

Delip Rao, Paul McNamee, and Mark Dredze. 2010. Streaming cross document entity coreference resolution. In *Proceedings of Coling 2010: Poster Volume,* Beijing, China, 23–27 August 2010, pages 1050–1058.

Lev Ratinov, Dan Roth, Doug Downey, and Mike Anderson. 2011. Local and global algorithms for disambiguation to Wikipedia. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics,* Portland, Oreg., USA, 19–24 June 2011, pages 1375–1384.

Sebastian Riedel. 2008. Improving the accuracy and efficiency of MAP inference for Markov logic. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence,* Helsinki, Finland, 9–12 July 2008, pages 468–475.

Hinrich Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.

Sameer Singh, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. 2011. Large-scale cross-document coreference using distributed inference and hierarchical models. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics,* Portland, Oreg., USA, 19–24 June 2011, pages 793–803.

Michael Wick, Sameer Singh, and Andrew McCallum. 2012. A discriminative hierarchical model for fast coreference at large scale. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics,* Jeju Island, Korea, 8–14 July 2012, pages 379–388.