A Multi-media Approach to Cross-lingual Entity Knowledge Transfer

Di Lu[1], Xiaoman Pan[1], Nima Pourdamghani[2],
Shih-Fu Chang[3], Heng Ji[1], Kevin Knight[2]
[1] Computer Science Department, Rensselaer Polytechnic Institute
{lud2,panx2,jih}@rpi.edu
[2] Information Sciences Institute, University of Southern California
{damghani,knight}@isi.edu
[3] Electrical Engineering Department, Columbia University
sfchang@ee.columbia.edu

Abstract

When a large-scale incident or disaster occurs, there is often a great demand for rapidly developing a system to extract detailed and new information from low-resource languages (LLs). We propose a novel approach to discover comparable documents in high-resource languages (HLs), and project Entity Discovery and Linking results from HLs documents back to LLs. We leverage a wide variety of language-independent forms from multiple data modalities, including image processing (image-to-image retrieval, visual similarity and face recognition) and sound matching. We also propose novel methods to learn entity priors from a large-scale HL corpus and knowledge base. Using Hausa and Chinese as the LLs and English as the HL, experiments show that our approach achieves 36.1% higher Hausa name tagging F-score over a costly supervised model, and 9.4% higher Chinese-to-English Entity Linking accuracy over state-of-the-art.

1 Introduction

In many situations such as disease outbreaks and natural calamities, we often need to develop an Information Extraction (IE) component (e.g., a name tagger) within a very limited time to extract information from low-resource languages (LLs) (e.g., locations where Ebola outbreaks from Hausa documents). The main challenge lies in the lack of labeled data and linguistic processing tools in these languages. A potential solution is to extract and project knowledge from high-resource languages (HLs) to LLs.

A large amount of non-parallel, domain-rich, topically-related comparable corpora naturally exist across LLs and HLs for breaking incidents, such as coordinated news streams (Wang et al., 2007) and code-switching social media (Voss et al., 2014; Barman et al., 2014). However, without effective Machine Translation techniques, even just identifying such data in HLs is not a trivial task. Fortunately many of such comparable documents are presented in multiple data modalities (text, image and video), because press releases with multimedia elements generate up to 77% more views than text-only releases (Newswire, 2011). In fact, they often contain the same or similar images and videos, which are language-independent.

In this paper we propose to use images as a hub to automatically discover comparable corpora. Then we will apply Entity Discovery and Linking (EDL) techniques in HLs to extract entity knowledge, and project results back to LLs by leveraging multi-source multi-media techniques. In the following we will elaborate motivations and detailed methods for two most important EDL components: name tagging and Cross-lingual Entity Linking (CLEL). For CLEL we choose Chinese as the LL and English as HL because Chinese-to-English is one of the few language pairs for which we have ground-truth annotations from official shared tasks (e.g., TAC-KBP (Ji et al., 2015)). Since Chinese name tagging is a well-studied problem, we choose Hausa instead of Chinese as the LL for name tagging experiment, because we can use the ground truth from the DARPA LORELEI program[1] for evaluation.

Entity and Prior Transfer for Name Tagging:
In the first case study, we attempt to use HL extraction results directly to validate and correct names

---

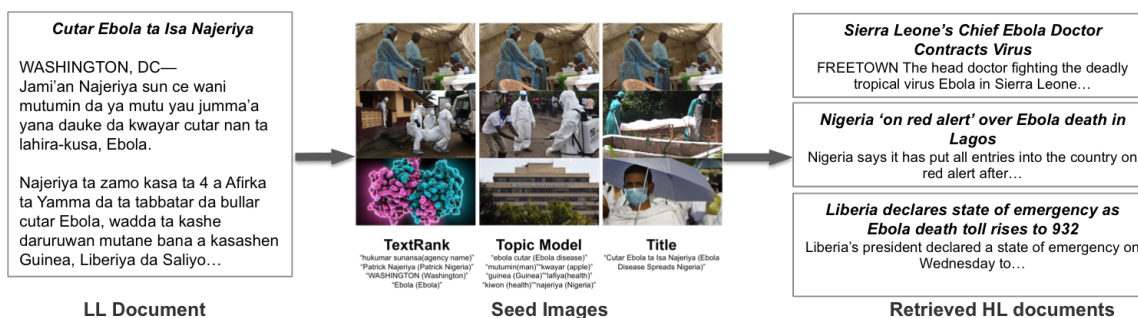[1]http://www.darpa.mil/program/low-resource-languages-for-emergent-incidents

Figure 1: Image Anchored Comparable Corpora Retrieval.

extracted from LLs. For example, in the Hausa document in Figure 1, it would be challenging to identify the location name "*Najeriya*" directly from the Hausa document because it's different from its English counterpart. But since its translation "*Nigeria*" appears in the topically-related English document, we can use it to infer and validate its name boundary.

Even if topically-related documents don't exist in an HL, similar scenarios (e.g., disease outbreaks) and similar activities of the same entity (e.g., meetings among politicians) often repeat over time. Moreover, by running a high-performing HL name tagger on a large amount of documents, we can obtain *entity prior* knowledge which shows the probability of a related name appearing in the same context. For example, if we already know that "*Nigeria*", "*Borno*", "*Goodluck Jonathan*", "*Boko Haram*" are likely to appear, then we could also expect "*Mouhammed Ali Ndume*" and "*Mohammed Adoke*" might be mentioned because they were both important politicians appointed by Goodluck Jonathan to consider opening talks with Boko Haram. Or more generally if we know the LL document is about politics in China in 1990s, we could estimate that famous politicians during that time such as "*Deng Xiaoping*" are likely to appear in the document.

Next we will project these names extracted from HL documents directly to LL documents to identify and verify names. In addition to textual evidence, we check visual similarity to match an HL name with its equivalent in LL. And we apply face recognition techniques to verify person names by image search. This idea matches human knowledge acquisition procedure as well. For example, when a child is watching a cartoon and shifting between versions in two languages, s/he can easily infer translation pairs for the same con-
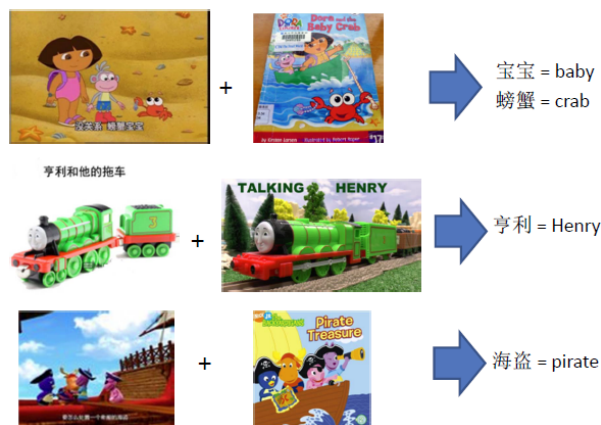

Figure 2: Examples of Cartoons in Chinese (left) and English (right).

cept whose images appear frequently (e.g., "宝宝 (*baby*)" and "螃蟹 (*crab*)" in "*Dora Exploration*", "海盗 (*pirate*)" in "*the Garden Guardians*", and "亨利 (*Henry*)" in "*Thomas Train*"), as illustrated in Figure 2.

Representation and Structured Knowledge Transfer for Entity Linking: Besides data sparsity, another challenge for low-resource language IE lies in the lack of knowledge resources. For example, there are advanced knowledge representation parsing tools available (e.g., Abstract Meaning Representation (AMR) (Banarescu et al., 2013)) and large-scale knowledge bases for English Entity Linking, but not for other languages, including some medium-resource ones such as Chinese. For example, the following documents are both about the event of Pistorius killing his girl friend Reeva:

- **LL document:** 南非残疾运动员皮斯托瑞斯被指控杀害女友瑞娃于其茨瓦内的家中。皮斯托瑞斯是南非著名的残疾人田径选手，有"刀锋战士"之称。...(*The disabled South African sportsman Oscar Pis-*
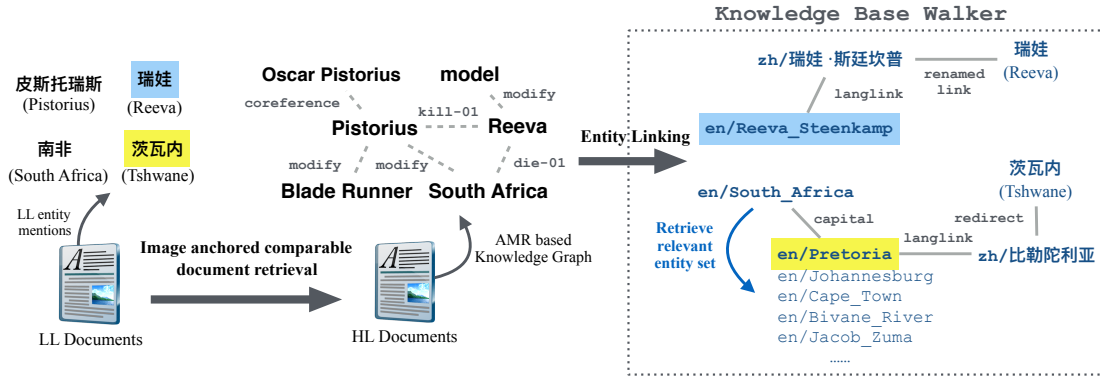
Figure 3: Cross-lingual Knowledge Transfer for Entity Linking.

torius was charged to killing his girl friend Reeva at his home in Tshwane. Pistorius is a famous runner in South Africa, also named as "Blade Runner"...)

- **HL document:** *In the early morning of Thursday, 14 February 2013, "Blade Runner" Oscar Pistorius shot and killed South African model Reeva Steenkamp...*

From the LL documents we may only be able to construct co-occurrence based knowledge graph and thus it's difficult to link rare entity mentions such as "瑞娃 (*Reeva*)" and "茨瓦内 (*Tshwane*)" to an English knowledge base (KB). But if we apply an HL (e.g., English) entity linker, we could construct much richer knowledge graphs from HL documents using deep knowledge representations such as AMR, as shown in Figure 3, and link all entity mentions to the KB accurately. Moreover, if we start to *walk* through the KB, we can easily reach from English related entities to the entities mentioned in LL documents. For example, we can walk from "*South Africa*" to its capital "*Pretoria*" in the KB, which is linked to its LL form "比勒陀利亚" through a language link and then is re-directed to "茨瓦内" mentioned in the LL document through a redirect link. Therefore we can infer that "茨瓦内" should be linked to "*Pretoria*" in the KB.

Compared to most previous cross-lingual projection methods, our approach does not require domain-specific parallel corpora or lexicons, or in fact, any parallel data at all. It also doesn't require any labeled data in LLs. Using Hausa and Chinese as the LLs and English as HL for case study, experiments demonstrate that our approach can achieve 36.1% higher Hausa name tagging over a costly supervised model trained from 337 docu-

ments, and 9.4% higher Chinese-to-English Entity Linking accuracy over a state-of-the-art system.

2 Approach Overview

Figure 4 illustrates the overall framework. It consists of two steps: (1) Apply language-independent key phrase extraction methods on each LL document, then use key phrases as a query to retrieve seed images, and then use the seed images to retrieve matching images, and retrieve HL documents containing these images (Section 3); (2) Extract knowledge from HL documents, and design knowledge transfer methods to refine LL extraction results.
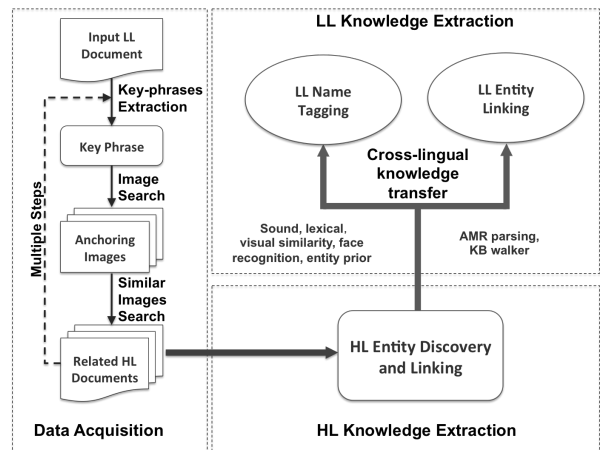

Figure 4: Overall Framework.

We will present two case studies on name tagging (Section 5) and cross-lingual entity linking (CLEL) (Section 6) respectively. Our projection approach consists of a series of non-traditional multi-media multi-source methods based on textual and visual similarity, face recognition, as well as entity priors learned from both unstructured data and structured KB.

56

3 Comparable Corpora Discovery

In this section we will describe the detailed steps of acquiring HL documents for a given LL document via anchoring images. Using a cluster of images as a hub, we attempt to connect topically-related documents in LL and HL. We will walk through each step for the motivating example in Figure 1.

3.1 Key Phrase Extraction

For an LL document (e.g., Figure 1 for the walk-through example), we start by extracting its key phrases using the following three language-independent methods: (1) TextRank (Mihalcea and Tarau, 2004), which is a graph-based ranking model to determine key phrases. (2) Topic modeling based on Latent Dirichlet allocation (LDA) model (Blei et al., 2003), which can generate a small number of key phrases representing the main topics of each document. (3) The title of the document if it's available.

3.2 Seed Image Retrieval

Using the extracted key phrases together as one single query, we apply Google Image Search to retrieve top 15 ranked images as seeds. To reduce the noise introduced by image search, we filter out images smaller than 100×100 pixels because they are unlikely to appear in the main part of web pages. We also filter out an image if its web page contains less than half of the tokens in the query. Figure 1 shows the anchoring images retrieved for the walk-through example.

3.3 HL Document Retrieval

Using each seed image, we apply Google image-to-image search to retrieve more matching images, and then use the TextCat tool (Cavnar et al., 1994) as a language identifier to select HL documents containing these images. It shows three English documents retrieved for the first image in Figure 1.

For related topics, more images may be available in HLs than LLs. To compensate this data sparsity problem, using the HL documents retrieved as a seed set, we repeat the above steps one more time by extracting key phrases from the HL seed set to retrieve more images and gather more HL documents. For example, a Hausa document about "*Arab Spring*" includes protests that happened in Algeria, Bahrain, Iran, Libya, Yemen and Jordan. The HL documents retrieved by LL key phrases and images in the first step missed the detailed information about protests in Iran. However the second step based on key phrases and images from HL successfully retrieved detailed related documents about protests in Iran.

Applying the above multimedia search, we automatically discover domain-rich non-parallel data. Next we will extract facts from HLs and project them to LLs.

4 HL Entity Discovery and Linking

4.1 Name Tagging

After we acquire HL (English in this paper) comparable documents, we apply a state-of-the-art English name tagger (Li et al., 2014) based on structured perceptron to extract names. From the output we filter out uninformative names such as news agencies. If the same name receives multiple types across documents, we use the majority one.

4.2 Entity Linking

We apply a state-of-the-art Abstract Meaning Representation (AMR) parser (Wang et al., 2015a) to generate rich semantic representations. Then we apply an AMR based entity linker (Pan et al., 2015) to link all English entity mentions to the corresponding entities in the English KB. Given a name n_h, this entity linker first constructs a Knowledge Graph $g(n_h)$ with n_h at the hub and leaf nodes obtained from names reachable by AMR graph traversal from n_h. A subset of the leaf nodes are selected as collaborators of n_h. Names connected by AMR conjunction relations are grouped into sets of coherent names. For each name n_h, an initial ranked list of entity candidates $\mathcal{E} = \{e_1, ..., e_M\}$ is generated based on a salience measure (Medelyan and Legg, 2008). Then a Knowledge Graph $g(e_m)$ is generated for each entity candidate e_m in n_h's entity candidate list $\mathcal{E}$. The entity candidates are then re-ranked according to Jaccard Similarity, which computes the similarity between $g(n_h)$ and $g(e_m)$: $J(g(n_h), g(e_m)) = \frac{|g(n_h) \cap g(e_m)|}{|g(n_h) \cup g(e_m)|}$. Finally, the entity candidate with the highest score is selected as the appropriate entity for n_h. Moreover, the Knowledge Graphs of coherent mentions will be merged and linked collectively.

4.3 Entity Prior Acquisition

Given the English entities discovered from the above, we aim to automatically mine related en-

tities to further expand the expected entity set. We use a large English corpus and English knowledge base respectively as follows.

If a name n_h appears frequently in these retrieved English documents,[2] we further mine other related names n'_h which are very likely to appear in the same context as n_h in a large-scale news corpus (we use English Gigaword V5.0 corpus[3] in our experiment). For each pair of names $\langle n'_h, n_h \rangle$, we compute $P(n'_h | n_h)$ based on their co-occurrences in the same sentences. If $P(n'_h | n_h)$ is larger than a threshold,[4] and n'_h is a person name, then we add n'_h into the expected English name set.

Let $\mathcal{E}^0 = \{e_1, ..., e_N\}$ be the set of entities in the KB that all mentions in English documents are linked to. For each $e_i \in \mathcal{E}^0$, we 'walk' one step from it in the KB to retrieve all of its neighbors $\mathcal{N}(e_i)$. We denote the set of neighbor nodes as $\mathcal{E}^1 = \{\mathcal{N}(e_1), ..., \mathcal{N}(e_N)\}$. Then we extend the expected English entity set as $\mathcal{E}^0 \cup \mathcal{E}^1$. Table 1 shows some retrieved neighbors for entity "*Elon Musk*".

Relation	Neighbor
is founder of	SpaceX
is founder of	Tesla Motors
is spouse of	Justine Musk
birth place	Pretoria
alma mater	University of Pennsylvania
parents	Errol Musk
relatives	Kimbal Musk

Table 1: Neighbors of Entity "*Elon Musk*".

5 Knowledge Transfer for Name Tagging

In this section we will present the first case study on name tagging, using English as HL and Hausa as LL.

5.1 Name Projection

After expanding the English expected name set using entity prior, next we will try to carefully select, match and project each expected name (n_h) from English to the one (n_l) in Hausa documents. We scan through every n-gram (n in the order 3, 2, 1) in Hausa documents to see if any of them match an English name based on the following multi-media language-independent low-cost heuristics.

Spelling: If n_h and n_l are identical (e.g., "*Brazil*"), or with an edit distance of one after lower-casing and removing punctuation (e.g., n_h = "*Mogadishu*" and n_l = "*Mugadishu*"), or substring match (n_h = "*Denis Samsonov*" and n_l = "*Samsonov*").

Pronunciation: We check the pronunciations of n_h and n_l based on Soundex (Odell, 1956), Metaphone (Philips, 1990) and NYSIIS (Taft, 1970) algorithms. We consider two codes match if they are exactly the same or one code is a part of the other. If at least two coding systems match between n_h and n_l, we consider they are equivalents.

Visual Similarity: When two names refer to the same entity, they usually share certain visual patterns in their related images. For example, using the textual clues above is not sufficient to find the Hausa equivalent "*Majalisar Dinkin Duniya*" for "*United Nations*", because their pronunciations are quite different. However, Figure 5 shows the images retrieved by "*Majalisar Dinkin Duniya*" and "*United Nations*" are very similar.[5]

We first retrieve top 50 images for each mention using Google image search. Let $\mathcal{I}_h$ and $\mathcal{I}_l$ denote two sets of images retrieved by an n_h and a candidate n_l (e.g., n_h = "*United Nations*" and n_l = "*Majalisar Dinkin Duniya*" in Figure 5), $i_h \in \mathcal{I}_h$ and $i_l \in \mathcal{I}_l$. We apply the Scale-invariant feature transform (SIFT) detector (Lowe, 1999) to count the number of matched key points between two images, $K(i_h, i_l)$, as well as the key points in each image, $P(i_h)$ and $P(i_l)$. SIFT key point is a circular image region with an orientation, which can provide feature description of the object in the image. Key points are maxima/minima of the Difference of Gaussians after the image is convolved with Gaussian filters at different scales. They usually lie in high-contrast regions. Then we define the similarity ($0 \sim 1$) between two phrases as:

$$S(n_h, n_l) = \max_{i_h \in \mathcal{I}_h} \max_{i_l \in \mathcal{I}_l} \frac{K(i_h, i_l)}{\min(P(i_h), P(i_l))} \quad (1)$$

Based on empirical results from a separate small development set, we decide two phrases match if $S(n_h, n_l) > 10\%$. This visual similarity computation method, though seemingly simple, has been

---

[2]for our experiment we choose those that appear more than 10 times

[3]https://catalog.ldc.upenn.edu/LDC2011T07

[4]0.02 in our experiment.

[5]Although existing Machine Translation (MT) tools like Google Translate can correctly translate this example phrase from Hausa to English, here we use it as an example to illustrate the low-resource setting when direct MT is not available.

one of the principal techniques in detecting near-duplicate visual content (Ke et al., 2004).

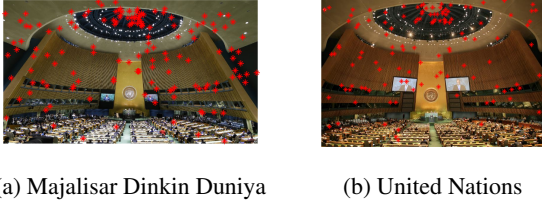(a) Majalisar Dinkin Duniya (b) United Nations

Figure 5: Matched SIFT Key points.

5.2 Person Name Verification through Face Recognition

For each name candidate, we apply Google image search to retrieve top 10 images (examples in Figure 6). If more than 5 images contain and only contain 1-2 faces, we classify the name as a person. We apply face detection technique based on Haar Feature (Viola and Jones, 2001). This technique is a machine learning based approach where a cascade function is trained from a large amount of positive and negative images. In the future we will try other alternative methods using different feature sets such as Histograms of Oriented Gradients (Dalal and Triggs, 2005).

Figure 6: Face Recognition for Validating Person Name 'Nawaz Shariff'.

6 Knowledge Transfer for Entity Linking

In this section we will present the second case study on Entity Linking, using English as HL and Chinese as LL. We choose this language pair because its ground-truth Entity Linking annotations are available through the TAC-KBP program (Ji et al., 2015).

6.1 Baseline LL Entity Linking

We apply a state-of-the-art language-independent cross-lingual entity linking approach (Wang et al., 2015b) to link names from Chinese to an English KB. For each name n, this entity linker uses the cross-lingual surface form dictionary $\langle f, \{e_1, e_2, ..., e_M\}\rangle$, where $\mathcal{E} = \{e_1, e_2, ..., e_M\}$ is the set of entities with surface form f in the KB

according to their properties (e.g., labels, names, aliases), to locate a list of candidate entities $e \in \mathcal{E}$ and compute the importance score by an entropy based approach.

6.2 Representation and Structured Knowledge Transfer

Then for each expected English entity e_h, if there is a cross-lingual link to link it to an LL (Chinese) entry e_l in the KB, we added the title of the LL entry or its redirected/renamed page c_l as its LL translation. In this way we are able to collect a set of pairs of $\langle c_l, e_h \rangle$, where c_l is an expected LL name, and e_h is its corresponding English entity in the KB. For example, in Figure 3, we can collect pairs including "(瑞娃, *Reeva Steenkamp*)", "(瑞娃·斯廷坎普, *Reeva Steenkamp*)", "(茨瓦内, *Pretoria*)" and "(比勒陀利亚, *Pretoria*)". For each mention in an LL document, we then check whether it matches any c_l, if so then use e_h to override the baseline LL Entity Linking result. Table 2 shows some $\langle c_l, e_h \rangle$ pairs with frequency. Our approach not only successfully retrieves translation variants of "*Beijing*" and "*China Central TV*", but also alias and abbreviations.

e_h	e_l	c_l	Freq.
Beijing	北京市	北京(Beijing)	553
		北京市(Beijing City)	227
		燕京(Yanjing)	15
		京师(Jingshi)	3
		北平(Beiping)	2
		首都(Capital)	1
		蓟(Ji)	1
		燕都(Yan capital)	1
China Central TV	中国中央电视台	央视(Central TV)	19
		CCTV	16
		中央电视台(Central TV)	13
		中国央视(China Central TV)	3

Table 2: Representation and Structured Knowledge Transfer for Expected English Entities "*Beijing*" and "*China Central TV*".

7 Experiments

In this section we will evaluate our approach on name tagging and Cross-lingual Entity Linking.

7.1 Data

For name tagging, we randomly select 30 Hausa documents from the DARPA LORELEI program as our test set. It includes 63 person names (PER), 64 organizations (ORG) 225 geo-political entities

(GPE) and locations (LOC). For this test set, in total we retrieved 810 topically-related English documents. We found that 80% names in the ground truth appear at least once in the retrieved English documents, which shows the effectiveness of our image-anchored comparable data discovery method.

For comparison, we trained a supervised Hausa name tagger based on Conditional Random Fields (CRFs) from the remaining 337 labeled documents, using lexical features (character ngrams, adjacent tokens, capitalization, punctuations, numbers and frequency in the training data).

We learn entity priors by running the Stanford name tagger (Manning et al., 2014) on English Gigaword V5.0 corpus.[6] The corpus includes 4.16 billion tokens and 272 million names (8.28 million of which are unique).

For Cross-lingual Entity Linking, we use 30 Chinese documents from the TAC-KBP2015 Chinese-to-English Entity Linking track (Ji et al., 2015) as our test set. It includes 678 persons, 930 geo-political names, 437 organizations and 88 locations. The English KB is derived from BaseKB, a cleaned version of English Freebase. 89.7% of these mentions can be linked to the KB. Using the multi-media approach, we retrieved 235 topically-related English documents.

7.2 Name Tagging Performance

Table 3 shows name tagging performance. We can see that our approach dramatically outperforms the supervised model. We conduct the Wilcoxon Matched-Pairs Signed-Ranks Test on ten folders. The results show that the improvement using visual evidence is significant at a 95% confidence level and the improvement using entity prior is significant at a 99% confidence level. Visual Evidence greatly improves organization tagging because most of them cannot be matched by spelling or pronunciation.

Face detection helps identify many person names missed by the supervised name tagger. For example, in the following sentence, "*Nawaz Shariff*" is mistakenly classified as a location by the supervised model due to the designator "*kasar (country)*" appearing in its left context. Since faces can be detected from all of the top 10 retrieved images (Figure 6), we fix its type to person.

- **Hausa document**: "*Yansanda sun dauki*

*wannan matakin ne kwana daya bayanda PM kasar **Nawaz Shariff** ya fidda sanarwar inda ya bukaci...* (*The Police took this step a day after the PM of the country **Nawaz Shariff** threw out the statement in which he demanded that...*)"

Face detection is also effective to resolve classification ambiguity. For example, the common person name "*Haiyan*" can also be used to refer to the Typhoon in Southeast Asia. Both of our HL and LL name taggers mistakenly label "*Haiyan*" as a person in the following documents:

- **Hausa document**: "*...a yayinda mahaukaciyar guguwar teku da aka lakawa suna **Haiyan** ta fada tsibiran Leyte da Samar.* (*...as the violent typhoon, which has been given the name, **Haiyan**, has swept through the island of Leyte and Samar.*)"

- **Retrieved English comparable document**: "*As **Haiyan** heads west toward Vietnam, the Red Cross is at the forefront of an international effort to provide food, water, shelter and other relief...*"

In contrast using face detection results we successfully remove it based on processing the retrieved images as shown in Figure 7.
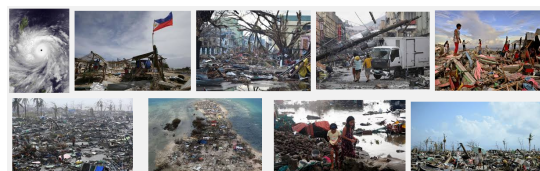

Figure 7: Top 9 Retrieved Images for 'Haiyan'.

Entity priors successfully provide more detailed and richer background knowledge than the comparable English documents. For example, the main topic of one Hausa document is the former president of Nigeria *Olusegun Obasanjo* accusing the current President *Goodluck Jonathan*, and a comment by the former 1990s military administrator of Kano *Bawa Abdullah Wase* is quoted. But *Bawa Abdullah Wase* is not mentioned in any related English documents. However, based on entity priors we observe that "*Bawa Abdullah Wase*" appears frequently in the same contexts as "*Nigeria*" and "*Kano*", and thus we successfully project it back to the Hausa sentence: "*Haka ma **Bawa Abdullahi Wase** ya ce akawai abun dubawa a kalamun*

System	Identification F-score				Classification	Overall
	PER	**ORG**	**LOC**[7]	**ALL**	**Accuracy**	**F-score**
Supervised	36.52	38.64	42.38	40.25	76.84	30.93
Our Approach	**77.69**	**60.00**	**70.55**	**70.59**	**95.00**	**67.06**
Our Approach w/o Visual Evidence	73.77	46.58	70.74	67.98	94.77	64.43
Our Approach w/o Entity Prior	64.91	60.00	70.55	67.59	94.71	64.02

Table 3: Name Tagging Performance (%).

tsohon shugaban kasa kuma tsohon jamiin tsaro. (In the same vein, Bawa Abdullahi Wase said that there were things to take away from the former President's words.)". The impact of entity priors on person names is much more significant than other categories because multiple person entities often co-occur in some certain events or related topics which might not be fully covered in the retrieved English documents. In contrast most expected organizations and locations already exist in the retrieved English documents.

For the same local topic, Hausa documents usually describe more details than English documents, and include more unsalient entities. For example, for the president election in Ivory Coast, a Hausa document mentions the officials of the electoral body such as "*Damana Picasse*": "*Wani wakilin hukumar zaben daga jamiyyar shugaba Gbagbo, Damana Picasse, ya kekketa takardun sakamakon a gaban yan jarida, ya kuma ce ba na halal ba ne. (An official of the electoral body from president Gbagbo's party, Damana Picasse, tore up the result document in front of journalists, and said it is not legal.)*". In contrast, no English comparable documents mention their names. The entity prior method is able to extract many names which appear frequently together with the president name "*Gbagbo*".

7.3 Entity Linking Performance

Table 4 presents the Cross-lingual Entity Linking performance. We can see that our approach significantly outperforms our baseline and the best reported results on the same test set (Ji et al., 2015). Our approach is particularly effective for rare nicknames (e.g., "*C罗*" (C Luo) is used to refer to *Cristiano Ronaldo*) or ambiguous abbreviations (e.g., "*邦联*" (federal) can refer to *Confederate States of America*, 邦联制 (*Confederation*) and many other entities) for which the contexts in LLs are not sufficient for making correct linking decisions due to the lack of rich knowledge rep-

resentation. Our approach produces worse linking results than the baseline for a few cases when the same abbreviation is used to refer to multiple entities in the same document. For example, when "巴" is used to refer to both "巴西 (*Brazil*)" or "巴勒斯坦 (*Palestine*)" in the same document, our approach mistakenly links all mentions to the same entity.
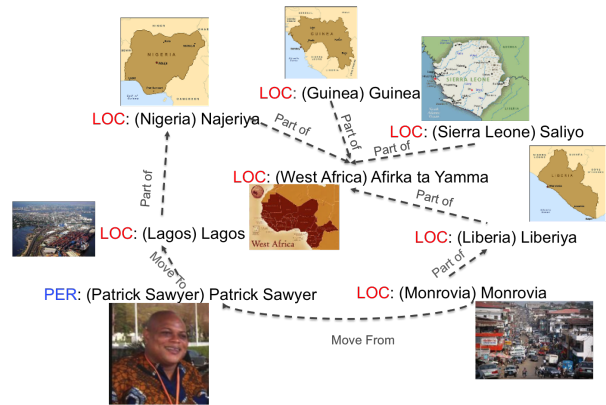

Figure 8: An Example of Cross-lingual Cross-media Knowledge Graph.

7.4 Cross-lingual Cross-media Knowledge Graph

As an end product, our framework will construct cross-lingual cross-media knowledge graphs. An example about the Ebola scenario is presented in Figure 8, including entity nodes extracted from both Hausa (LL) and English (HL), anchored by images; and edges extracted from English.

8 Related Work

Some previous cross-lingual projection methods focused on transferring data/annotation (e.g., (Padó and Lapata, 2009; Kim et al., 2010; Faruqui and Kumar, 2015)), shared feature representation/model (e.g., (McDonald et al., 2011; Kozhevnikov and Titov, 2013; Kozhevnikov and Titov, 2014)), or expectation (e.g., (Wang and Manning, 2014)). Most of them relied on a large

Approach	Overall					Linkable Entities				
	PER	ORG	GPE	LOC	ALL	PER	ORG	GPE	LOC	ALL
Baseline	49.12	60.18	80.97	80.68	66.57	67.27	67.61	81.05	80.68	74.70
State-of-the-art	49.85	64.30	75.38	**96.59**	65.87	68.28	72.24	75.46	**96.59**	73.91
Our Approach	**52.36**	**67.05**	**93.33**	93.18	**74.92**	**71.72**	**75.32**	**93.43**	93.18	**84.06**
Our Approach w/o KB Walker	50.44	67.05	84.41	90.91	70.32	69.09	75.32	84.50	90.91	78.91

Table 4: Cross-lingual Entity Linking Accuracy (%).

amount of parallel data to derive word alignment and translations, which are inadequate for many LLs. In contrast, we do not require any parallel data or bi-lingual lexicon. We introduce new cross-media techniques for projecting HLs to LLs, by inferring projections using domain-rich, non-parallel data automatically discovered by image search and processing. Similar image-mediated approaches have been applied to other tasks such as cross-lingual document retrieval (Funaki and Nakayama, 2015) and bilingual lexicon induction (Bergsma and Van Durme, 2011). Besides visual similarity, their method also relied on distributional similarity computed from a large amount of unlabeled data, which might not be available for some LLs.

Our name projection and validation approaches are similar to other previous work on bi-lingual lexicon induction from non-parallel corpora (e.g., (Fung and Yee, 1998; Rapp, 1999; Shao and Ng, 2004; Munteanu and Marcu, 2005; Sproat et al., 2006; Klementiev and Roth, 2006; Hassan et al., 2007; Udupa et al., 2009; Ji, 2009; Darwish, 2010; Noeman and Madkour, 2010; Bergsma and Van Durme, 2011; Radford et al., ; Irvine and Callison-Burch, 2013; Irvine and Callison-Burch, 2015)) and name translation mining from multi-lingual resources such as Wikipedia (e.g. (Sorg and Cimiano, 2008; Adar et al., 2009; Nabende, 2010; Lin et al., 2011)). We introduce new multimedia evidence such as visual similarity and face recognition for name validation, and also exploit a large amount of monolingual HL data for mining entity priors to expand the expected entity set.

For Cross-lingual Entity Linking, some recent work (Finin et al., 2015) also found cross-lingual coreference resolution can greatly reduce ambiguity. Some other methods also utilized global knowledge in the English KB to improve linking accuracy via quantifying link types (Wang et al., 2015b), computing pointwise mutual information for the Wikipedia categories of consecutive pairs of entities (Sil et al., 2015), or using linking as

feedback to improve name classification (Sil and Yates, 2013; Heinzerling et al., 2015; Besancon et al., 2015; Sil et al., 2015).

9 Conclusions and Future Work

We describe a novel multi-media approach to effectively transfer entity knowledge from high-resource languages to low-resource languages. In the future we will apply visual pattern recognition and concept detection techniques to perform deep content analysis of the retrieved images, so we can do matching and inference on concept/entity level instead of shallow visual similarity. We will also extend anchor image retrieval from document-level into phrase-level or sentence-level to obtain richer background information. Furthermore, we will exploit edge labels while walking through a knowledge base to retrieve more relevant entities. Our long-term goal is to extend this framework to other knowledge extraction and population tasks such as event extraction and slot filling to construct multimedia knowledge bases effectively from multiple languages with low cost.

References

Eytan Adar, Michael Skinner, and Daniel S Weld. 2009. Information arbitrage across multi-lingual

wikipedia. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 94–103.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of Association for Computational Linguistics 2013 Workshop on Linguistic Annotation and Interoperability with Discourse*.

Utsab Barman, Amitava Das, Joachim Wagner, and Jennifer Foster. 2014. Code mixing: A challenge for language identification in the language of social media. In *Proceedings of Conference on Empirical Methods in Natural Language Processing Workshop on Computational Approaches to Code Switching*.

Shane Bergsma and Benjamin Van Durme. 2011. Learning bilingual lexicons using the visual similarity of labeled web images. In *Proceedings of International Joint Conference on Artificial Intelligence*.

Romaric Besancon, Hani Daher, Hervé Le Borgne, Olivier Ferret, Anne-Laure Daquo, and Adrian Popescu. 2015. Cea list participation at tac edl english diagnostic task. In *Proceedings of the Text Analysis Conference*.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3(993-1022).

William B Cavnar, John M Trenkle, et al. 1994. N-gram-based text categorization. In *Proceedings of SDAIR1994*.

Navneet Dalal and Bill Triggs. 2005. Histograms of oriented gradients for human detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

Kareem Darwish. 2010. Transliteration mining with phonetic conflation and iterative training. In *Proceedings of the 2010 Named Entities Workshop*.

Manaal Faruqui and Shankar Kumar. 2015. Multilingual open relation extraction using cross-lingual projection. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics*.

Tim Finin, Dawn Lawrie, Paul McNamee, James Mayfield, Douglas Oard, Nanyun Peng, Ning Gao, Yiu-Chang Lin, Josh MacLin, and Tim Dowd. 2015. HLTCOE participation in TAC KBP 2015: Cold start and TEDL. In *Proceedings of the Text Analysis Conference*.

Ruka Funaki and Hideki Nakayama. 2015. Image-mediated learning for zero-shot cross-lingual document retrieval. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*.

Pascale Fung and Lo Yuen Yee. 1998. An ir approach for translating new words from nonparallel and comparable texts. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*.

Ahmed Hassan, Haytham Fahmy, and Hany Hassan. 2007. Improving named entity translation by exploiting comparable and parallel corpora. *AMML07*.

Benjamin Heinzerling, Alex Judea, and Michael Strube. 2015. Hits at tac kbp 2015: Entity discovery and linking, and event nugget detection. In *Proceedings of the Text Analysis Conference*.

Ann Irvine and Chris Callison-Burch. 2013. Combining bilingual and comparable corpora for low resource machine translation. In *Proc. WMT*.

Ann Irvine and Chris Callison-Burch. 2015. Discriminative Bilingual Lexicon Induction. *Computational Linguistics*, 1(1).

Heng Ji, Joel Nothman, Ben Hachey, and Radu Florian. 2015. Overview of tac-kbp2015 tri-lingual entity discovery and linking. In *Proceedings of the Text Analysis Conference*.

Heng Ji. 2009. Mining name translations from comparable corpora by creating bilingual information networks. In *Proceedings of the 2nd Workshop on Building and Using Comparable Corpora: from Parallel to Non-parallel Corpora*.

Yan Ke, Rahul Sukthankar, Larry Huston, Yan Ke, and Rahul Sukthankar. 2004. Efficient near-duplicate detection and sub-image retrieval. In *Proceedings of ACM International Conference on Multimedia*.

Seokhwan Kim, Minwoo Jeong, Jonghoon Lee, and Gary Geunbae Lee. 2010. A cross-lingual annotation projection approach for relation detection. In *Proceedings of the 23rd International Conference on Computational Linguistics*.

Alexandre Klementiev and Dan Roth. 2006. Weakly supervised named entity transliteration and discovery from multilingual comparable corpora. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*.

Mikhail Kozhevnikov and Ivan Titov. 2013. Cross-lingual transfer of semantic role labeling models. In *Proceedings of Association for Computational Linguistics*.

Mikhail Kozhevnikov and Ivan Titov. 2014. Cross-lingual model transfer using feature representation projection. In *Proceedings of Association for Computational Linguistics*.

Qi Li, Heng Ji, Yu Hong, and Sujian Li. 2014. Constructing information networks using one single model. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*.

Wen-Pin Lin, Matthew Snover, and Heng Ji. 2011. Unsupervised language-independent name translation mining from wikipedia infoboxes. In *Proceedings of the First workshop on Unsupervised Learning in NLP*, pages 43–52. Association for Computational Linguistics.

David G Lowe. 1999. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on Computer vision*.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *In Proceedings of Association for Computational Linguistics*.

Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*.

Olena Medelyan and Catherine Legg. 2008. Integrating cyc and wikipedia: Folksonomy meets rigorously defined common-sense. In *Wikipedia and Artificial Intelligence: An Evolving Synergy, Papers from the 2008 AAAI Workshop*.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into texts. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*.

Dragos Stefan Munteanu and Daniel Marcu. 2005. Improving machine translation performance by exploiting non-parallel corpora. *Computational Linguistics*, 31(4)(477–504).

Peter Nabende. 2010. Mining transliterations from wikipedia using pair hmms. In *Proceedings of the 2010 Named Entities Workshop*, pages 76–80. Association for Computational Linguistics.

PR Newswire. 2011. Earned media evolved. In *White Paper*.

Sara Noeman and Amgad Madkour. 2010. Language independent transliteration mining system using finite state automata framework. In *Proceedings of the 2010 Named Entities Workshop*, pages 57–61. Association for Computational Linguistics.

Margaret King Odell. 1956. *The Profit in Records Management*. Systems (New York).

Sebastian Padó and Mirella Lapata. 2009. Crosslingual annotation projection for semantic roles. *Journal of Artificial Intelligence Research*, 36:307–340.

Xiaoman Pan, Taylor Cassidy, Ulf Hermjakob, Heng Ji, and Kevin Knight. 2015. Unsupervised entity linking with abstract meaning representation. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics–Human Language Technologies*.

Lawrence Philips. 1990. Hanging on the metaphone. *Computer Language*, 7(12).

Will Radford, Xavier Carreras, and James Henderson. Named entity recognition with document-specific KB tag gazetteers.

Reinhard Rapp. 1999. Automatic identification of word translations from unrelated English and German corpora. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 519–526.

Li Shao and Hwee Tou Ng. 2004. Mining new word translations from comparable corpora. In *Proceedings of the 20th international conference on Computational Linguistics*, page 618.

Avirup Sil and Alexander Yates. 2013. Re-ranking for joint named-entity recognition and linking. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*.

Avirup Sil, Georgiana Dinu, and Radu Florian. 2015. The ibm systems for trilingual entity discovery and linking at tac 2015. In *Proceedings of the Text Analysis Conference*.

Philipp Sorg and Philipp Cimiano. 2008. Enriching the crosslingual link structure of wikipedia-a classification-based approach. In *Proceedings of the AAAI 2008 Workshop on Wikipedia and Artifical Intelligence*.

Richard Sproat, Tao Tao, and ChengXiang Zhai. 2006. Named entity transliteration with comparable corpora. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*.

Robert L Taft. 1970. *Name Search Techniques*. New York State Identification and Intelligence System, Albany, New York, US.

Raghavendra Udupa, K Saravanan, A Kumaran, and Jagadeesh Jagarlamudi. 2009. Mint: A method for effective and scalable mining of named entity transliterations from large comparable corpora. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 799–807.

Paul Viola and Michael Jones. 2001. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

Clare R Voss, Stephen Tratz, Jamal Laoudi, and Douglas M Briesch. 2014. Finding romanized arabic dialect in code-mixed tweets. In *Proceedings of Language Resources and Evaluation Conference*.

Mengqiu Wang and Christopher D Manning. 2014. Cross-lingual projected expectation regularization for weakly supervised learning. *Transactions of the Association for Computational Linguistics*.

Xuanhui Wang, ChengXiang Zhai, Xiao Hu, and Richard Sproat. 2007. Mining correlated bursty topic patterns from coordinated text streams. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Chuan Wang, Nianwen Xue, and Sameer Pradhan. 2015a. Boosting Transition-based AMR Parsing with Refined Actions and Auxiliary Analyzers. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*.

Han Wang, Jin Guang Zheng, Xiaogang Ma, Peter Fox, and Heng Ji. 2015b. Language and domain independent entity linking with quantified collective validation. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*.