Cross-Lingual Image Search on the Web

Kobi Reiter, Stephen Soderland, and **Oren Etzioni**

Turing Center
Dept. of Computer Science and Engineering
University of Washington, Seattle, WA 98195 USA
{dbgalur , soderlan , etzioni} @cs.washington.edu

Abstract

Most people locate images on the Web by querying image search engines such as Google's. The images are tagged by the words in their "vicinity", which limits the ability of a searcher to retrieve them. Although images are universal, an English searcher will fail to find images tagged in Chinese, and a Spanish searcher will fail to find images tagged in English. Cross-lingual homonyms cause problems as well. For example, the Hungarian word for tooth is 'fog', which makes it challenging for Hungarian searchers to find images of teeth on the Web.

To solve these and related problems we introduce the PANIMAGES cross-lingual image search engine.[1] PANIMAGES enables searchers to translate and disambiguate their queries before sending them to Google. PANIMAGES utilizes several machine-readable dictionaries composed into a graph that records relationships between words and word senses across different languages. The graph enables PANIMAGES to infer translations by composing information garnered from multiple dictionaries.

PANIMAGES can translate over 437,000 words and supports queries in 50 languages. Our experiments show that for queries in languages with a limited Web presence (ranging from Dutch or Norwegian to Lithuanian or Telugu), PANIMAGES increases the number of correct images by 75% in the first 15 pages (270 results), while increasing precision by 27%.

1 Introduction

The Web has emerged as a rich source of images that serve a wide range of purposes from children adorning their homework with pictures to anthropologists studying cultural nuances. Most people find images on the Web by querying an image search engine such as Google's (available at `images.google.com`).

[1] `cs.washington.edu/research/panimages`

Google collects images as part of its crawl of the Web and tags them with the words that appear in their vicinity on the crawled HTML documents and links. It is not surprising that most of the tags are in "major" languages such as English. So while images are universal, most of them can be found through Google only if you can query in the "right" language.

More broadly, monolingual image search engines face the following challenges:

- **Limited Resource Languages** - The lower the Web presence of a language, the fewer hits a speaker of that language gets from a query. A query for 'grenivka' (Slovenian for 'grapefruit') produces only 24 results, of which only 9 are images of grapefruits. Yet translating the query into English produces tens of thousands of images with high precision.
- **Cross-Cultural Images** - Results of an image search may vary considerably depending on the language of the query term. Translating the query 'baby' or 'food' into Chinese, Arabic, or Zulu allows an interesting cultural comparison.
- **Cross-Lingual Masking** - A word in one language is often a homonym for an unrelated word in another language. Relevant results can be swamped by results for the unrelated word. The Hungarian word for tooth happens to be 'fog'; the only way to get images of teeth rather than misty weather is to query with a translation that doesn't suffer from cross-lingual masking.
- **Word Sense Ambiguity** - Searching for an image that corresponds to a minor sense of a word is problematic. Most results for the query 'spring' are images of flowers and trees in bloom. If a user wants images of flexible coils or of bubbling fountains, the most effective queries are translations of this sense of 'spring' into languages where that word is not ambiguous.

We present PANIMAGES, an implemented system that allows a monolingual user to select from any of 50 input languages, automatically looks up word-sense specific translations into over 900 languages, and lets the user control which translations are sent to an image search engine. Figure 1 shows the system architecture. At compile time, PANIMAGES merges information from multiple wiktionaries and open-source dictionaries into a translation graph as described in

Section 2. At run time, PANIMAGES accepts a query from a user, presents the user with possible translations found in the translation graph, then sends the translations selected by the user to Google's image search as described in Section 3. We present statistics on the translation graph as well as PANIMAGES translation accuracy and coverage in Section 4, showing that PANIMAGES increases both recall and precision for queries in languages that do not have high Web presence. Section 5 discusses related work, followed by conclusions and future work in Section 6.

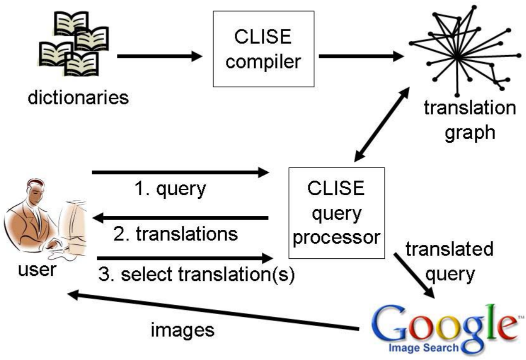

Figure 1: Architecture: The PANIMAGES compiler creates a translation graph from multiple dictionaries. The query processor takes a user query and presents a set of translations. The user selects the desired translation(s), which PANIMAGES sends to Google Image Search.

2 Cross-Lingual Translation Graph

PANIMAGES's knowledge is encoded in a novel lexical resource called the *translation graph*. This section describes the meaning of the graph, how it is constructed from multiple dictionaries, and how it is used by PANIMAGES.

2.1 The Translation Graph

Nodes in the graph are ordered pairs (w, l) where w is a word in a language l. An edge in the graph between (w_1, l_1) and (w_2, l_2) represents the belief that w_2 is a translation into l_2 of a particular word sense of the word w_1. The edge is labeled by an integer denoting an ID for that word sense. Thus, an edge is a triple of the form: $((w_1, l_1), s, (w_2, l_2))$. Word sense translation is symmetric, thus the edges are undirected. Word sense translation is transitive, thus paths through the graph represent correct translations so long as all the edges on the path share a single word sense. Paths through the graph enable PANIMAGES to identify translations that are absent from any of its source dictionaries.

Figure 2 shows a portion of a translation graph for two senses of 'spring' in English. The graph also shows edges from a French dictionary for the words 'printemps' (spring season) and 'ressort' (flexible spring).

PANIMAGES builds the translation graph incrementally based on entries from multiple, independent dictionaries as described in detail in Section 2.2. As edges are added based on entries from a new dictionary, some of the new word sense IDs are redundant because they are equivalent to word senses already in the graph from another dictionary. For example, PANIMAGES assigns a new word sense ID to the French dictionary entry for 'printemps', yet another to the Spanish dictionary entry for 'primavera', and so forth (see labels '1' and '3' in Figure 2). We refer to this phenomenon as *sense inflation*.

Sense inflation would severely limit the utility of the translation graph, so we have developed a mechanism for merging duplicate word senses automatically. PANIMAGES computes the probability $prob(s_i = s_j)$ that a pair of distinct IDs s_i and s_j refer to the same word sense (see Section 2.2 for the details). Thus, PANIMAGES determines that word sense ID '3' on edges from 'printemps' has a high probability of being equivalent to ID '1'.

The graph in Figure 2 also illustrates the power of paths. For example, the path from the Basque word 'udaherri' to its Maori translation 'koanga' is along edges with the same word sense, since ID '1' is equivalent to ID '3'. Thus, PANIMAGES infers that 'udaherri' can be translated as 'koanga', even though this translation is absent from the graph's source dictionaries—the graph representation facilitates the computation of novel translations.

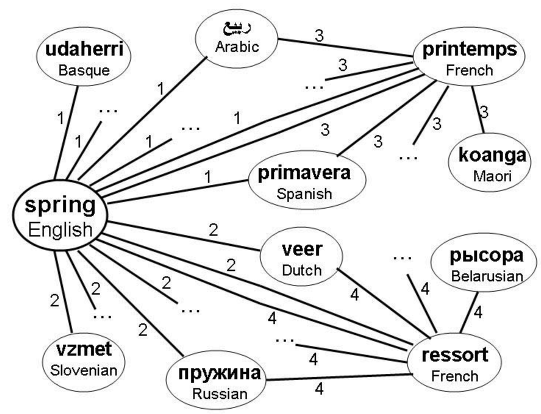

Figure 2: A fragment of a translation graph for two senses of the English word 'spring'. Edges with the label '1' or '3' are for spring in the sense of a season; edges labeled '2' or '4' are for the flexible coil sense. This graph shows translation entries from an English dictionary merged with translation entries from a French dictionary.

The following section discusses building the graph, focusing on the algorithm for merging word senses originating from different dictionaries.

2.2 Building the Translation Graph

PANIMAGES builds the translation graph in two stages: first adding entries from individual machine-readable dictionaries, then resolving word senses across entries from separate dictionaries. We make two assumptions about the dictionaries. 1) Entries in the dictionary distinguish separate word senses of a given word, and 2) the dictionary has lists of foreign translations for distinct senses of the word. Each list of translations gives us a set of words in multiple languages that are translations of each other in the same word sense.

We implemented the translation graph as a relational database. Each row in a *Translation table* represents an edge in the graph, while each row in the *Word sense equivalence table* represents the probability, $prob(s_i = s_j)$, that two word sense IDs s_i and s_j are equivalent.

As PANIMAGES adds to the translation graph from each new entry in a dictionary, it assigns a new, unique word sense ID for each word sense in that entry. Thus, edges for translations of the season 'spring' from the English dictionary have one word sense ID, edges for translations of the flexible coil 'spring' have a different word sense ID, and so forth.

Figures 3 and 4 give a schematic illustration of how PANIMAGES merges entries from multiple dictionaries. Figure 3 shows graph edges from an entry for the word E from an English dictionary that gives translations into French, German, Hungarian, and Spanish. PANIMAGES assigns the word sense ID 1 for these edges. This figure also shows edges from an entry for word R from a Russian dictionary, which in this case has translations into Chinese, English, German, Hungarian, and Latvian. These edges are assigned word sense ID 2.

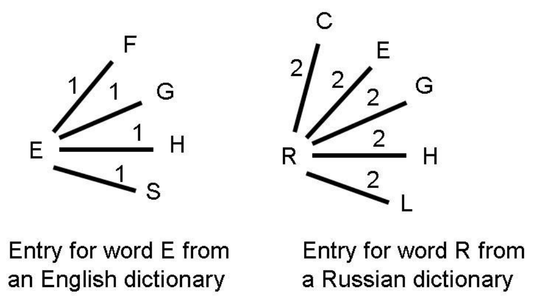

Figure 3: Schematic diagram of edges from an entry for the word E from an English dictionary and edges from an entry for the word R from a Russian dictionary.

Figure 4 shows the situation after both sets of edges have been added to the translation graph. The overlap of sense 1 with respect to sense 2 provides evidence that these IDs refer to the same word sense. The hypothesis is that both dictionaries are incomplete, and that the entry for word E should have included C, L, and R along with F, G, H, and S. Likewise the entry for word R should have included F and S along with C, E, G, H, and L. Under this hypothesis, all of the nodes in Figure 4 are translations of each other, even though no single dictionary lists all of these translation pairs.

Should we compute the overlap as $\frac{|intersection(1,2)|}{|nodes(1)|} = 0.60$ or as $\frac{|intersection(1,2)|}{|nodes(2)|} = 0.50$? We believe that taking the maximum of the intersections is most effective as illustrated by the following scenario. Suppose that there are 40 nodes with word sense ID s_i from one dictionary and only 5 nodes with word sense ID s_j from another dictionary, but that 4 out of 5 of the nodes with s_j are in the intersection of s_i and s_j. From the point of view of s_i the overlap is only 4/40 or 0.10; from the point of view of s_j the overlap is 4/5 or 0.80. It seems most likely that these are actually equivalent word senses, but that the second dictionary simply has lower coverage.

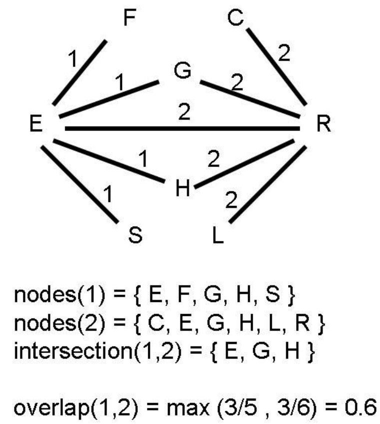

Figure 4: After the entries from Figure 3 have both been added to the graph, the nodes with word sense ID 1 overlaps with the set of nodes for word sense ID 2. The proportion of overlapping nodes gives evidence that the two word senses may be equivalent.

PANIMAGES determines the probability that two word sense IDs s_i and s_j are equivalent as follows:

- A word sense is equivalent to itself: $prob(s, s) = 1$.
- If s_i and s_j are alternate word senses from the same entry in a dictionary, then they are assumed to be distinct: $prob(s_i = s_j) = 0$.
- If word senses s_i and s_j have at least k intersecting nodes, then set the probability by equation 1 below.
- In all other cases, the probability is undefined.

PANIMAGES estimates the probability that s_i and s_j are equivalent word senses by the following equation.

If $|nodes(s_i) \cap nodes(s_j)| \geq k$, then $prob(s_i = s_j) =$

$$max(\frac{|nodes(s_i) \cap nodes(s_j)|}{|nodes(s_i)| + m}, \frac{|nodes(s_i) \cap nodes(s_j)|}{|nodes(s_j)| + m}) \quad (1)$$

where nodes(s) is the set of nodes that have edges labeled by word sense ID s, m is a smoothing factor and k is a sense intersection threshold.

As an example of computing the probability of word sense IDs being duplicates, consider the situation where all entries of 'spring' from an English dictionary are already in the translation graph with word sense ID s_1 for the season and s_2 for the flexible coil sense. PANIMAGES is now adding entries from a French dictionary and has assigned word sense s_3 for links from the word 'printemps'.

The dictionary entry for 'printemps' has 11 translations, giving 12 nodes for s_3 including (printemps,French). 8 of these 12 nodes have edges labeled with word sense ID s_1 – these are the overlapping translations of 'printemps' with translations from the English season 'spring'. Since s_1 has 56 nodes, $prob(s_1 = s_3) = max(\frac{8}{56+1}, \frac{8}{12+1}) = 0.62$ if the smoothing term is set to 1.

The only node in $nodes(s_3)$ with an edge labeled s_2 (flexible coil) is the ambiguous (spring,English). If the intersection threshold k is set to 1, then $prob(s_2 = s_3) = 0.08$ with $k \geq 2$ this probability is left undefined.

The computations in equation 1 can be made efficiently even with a large translation graph from a large number of dictionaries. The set of nodes whose word sense IDs have a non-zero intersection with given s_i is constant in the size of the translation graph – it is simply the set of translations given in an entry from a particular dictionary.

2.3 Computing Translation Probabilities

Given the translation graph coupled with the word sense equivalence probabilities, PANIMAGES can compute the probability that a particular word is a translation of another in a particular word sense. First, we show how to compute the probability of a single translation path. Then, we show how we combine evidence across multiple paths.

Consider a single path P that connects node n_1 to n_k, where n_i is the word w_i in language l_i and the ith edge has word sense s_i. Let $pathProb(n_1, n_k, s, P)$ be the probability that (w_1, l_1) is a correct translation of (w_k, l_k) in word sense s, given a path P connecting these nodes.

The simple case is where the path is of length 1. If s is the same sense ID as s_1, then the probability is simply 1.0, otherwise it is the probability that the two senses are equivalent:

$$pathProb(n_1, n_2, s, P) = prob(s = s_1) \qquad (2)$$

Where the path P has more than one edge, the path probability is reduced by $prob(s_i = s_{i+1})$ whenever the word sense ID changes along the path. If the desired sense s is not found on the path, we need to factor in the probability that s is equivalent to at least one sense s_i on the path, which we approximate by the maximum of $prob(s = s_i)$ over all s_i.

We make the simplifying assumption that sense-equivalence probabilities are mutually independent, which enables us to state the following heuristic formula for paths of length greater than one (*i.e.*, $|P| > 1$):

$$pathProb(n_1, n_k, s, P) = \max_{i=1\ldots|P|}(prob(s = s_i)) \times \prod_{i=1\ldots|P|-1} prob(s_i = s_{i+1}) \qquad (3)$$

Frequently, there are multiple, distinct paths from one node to another in the translation graph, so we need to combine the evidence from each of these paths to arrive at the desired translation probability between words. We define two paths from n_1 to n_k to be distinct if the set of unique word sense IDs on each path is different. For example, all paths from (udaherri,Basque) to (koanga,Maori) in Figure 2 are derived from the same two dictionary entries and contain ID 1 and ID 3. This means that there is only one distinct path between these two nodes.

We use the standard Noisy-Or model to combine evidence. The basic intuition is that translation is correct unless every one of the translation paths fails to maintain the desired sense s. We multiply the probability of failure for each path. We then subtract that probability from one to get the probability of correct translation.

Then, the probability that n_1 is a correct translation of n_k in word sense s is:

$$prob(n_1, n_k, s) = 1 - \prod_{P \in distinctP} (1 - pathProb(n_1, n_k, s, P)) \qquad (4)$$

where $distinctP$ is a set of distinct paths from n_1 to n_k.

We now turn to discussion of a practical application that uses a translation graph for cross-lingual image search.

3 Image Search with a Translation Graph

This section describes PANIMAGES, a system for cross-lingual image search that uses a translation graph built automatically from open-source dictionaries. PANIMAGES accepts user queries in any of 50 languages, presents possible translations in over 900 languages, then sends translated queries to Google's image search (http://images.google.com).

3.1 Open-Source Dictionaries

We found dictionaries from the Wiktionary Project (www.wiktionary.org) to be particularly useful – these are open source 'wiki' dictionaries created by Web volunteers, with separate wiktionaries for 171 languages. A wiktionary for language l_i has entries for words in l_i that may include multiple word senses for a word. Some wiktionary entries have a list of translations of an l_i word into multiple languages, which is generally broken into separate lists of translations for different word senses. The wiktionaries for English and French have the greatest coverage – The English wiktionary has about 19,500 English words that have translations, and the French wiktionary has about 12,700 French words that have translations. Each wiktionary is a joint effort by many Web volunteers, so the number of translations varies greatly across wiktionaries and varies for different words and word senses within a wiktionary. Wiktionaries continue to grow in coverage – we downloaded these wiktionaries in July, 2006.

As an example, the English wiktionary entry for 'spring' has definitions for 8 word senses, of which 6 have lists of translations: spring in the sense of to start to exist, to jump or leap, a season, a fountain of water, a device made of flexible material, and a rope on a boat. There are 55 translations in 48 languages for the season sense of spring, 33 translations of the flexible coil sense of spring, and so forth.

The French wiktionary has 3 word senses for the word 'printemps' (spring) of which only one sense (the season) has a list of translations, with 11 translations in 9 languages.

Another open-source dictionary that we used is an Esperanto dictionary (http://purl.org/net/voko/revo/), which also has translations into numerous languages for most of its entries. The Esperanto dictionary has 16 translations for 'printempo' in 13 languages, including English 'spring' in the sense of a season.

Parsing the entries for the English, French, and Esperanto dictionaries was not entirely straightforward. The format of entries was different for each dictionary, and was not always consistent within dictionaries. In particular, the foreign language names were not standardized. For example, Albanian

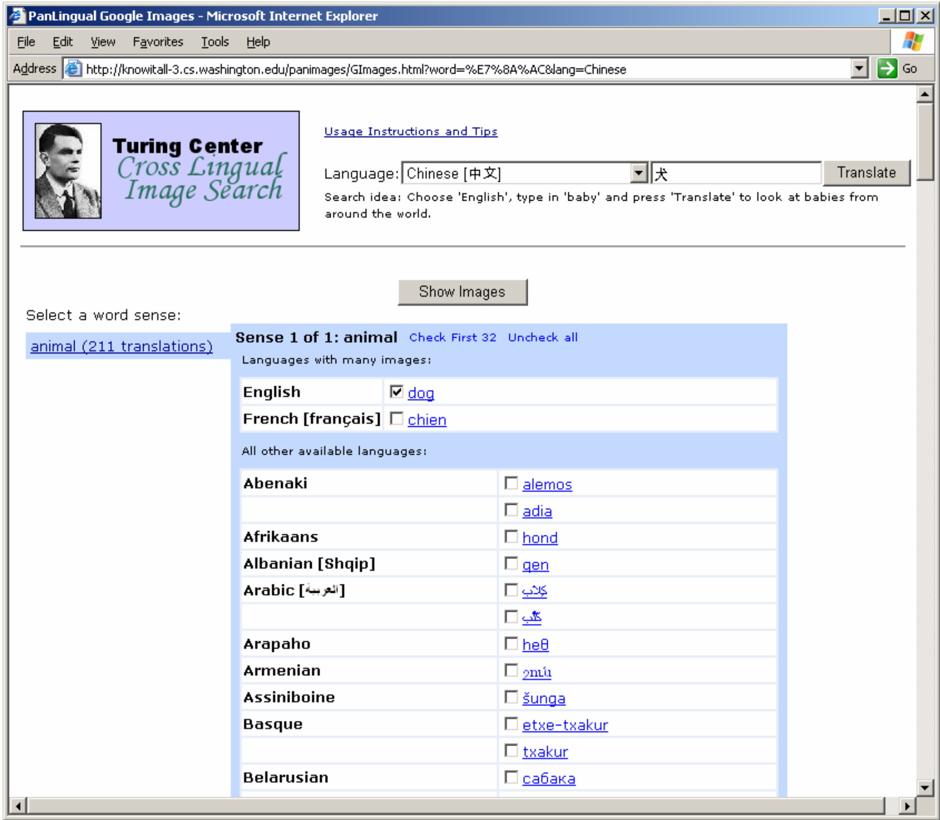

Figure 5: A snapshot of the PANIMAGES interface after a user has asked for translations of the Chinese word for dog. This produces 211 translations, with the English word selected as a default.

was indicated by any of the following: Albania, Albanian, Albanian(Gheg), als, or sq. We found it necessary to create tables that gave the normalized form of the variant language names.

Using Bilingual Dictionaries

The method for computing word sense equivalence that is discussed in 2.2 relies on having translations for each word into multiple languages. If we use bilingual dictionaries to augment our translation graph, equation 1 is not effective. Instead, we can infer word sense equivalence from local cliques of 3 nodes in the graph.

Suppose we are adding to the translation graph in Figure 2 from a Vietnamese-English dictionary. The entry for the word ‘xuân’ has the English translation ‘spring’, but no clue as to the intended word sense of spring. PANIMAGES adds a new node (xuân,Vietnamese) with an edge to (spring,English). If a Vietnamese-French dictionary translates ‘xuân’ as ‘printemps’, we have strong evidence that ‘xuân’, ‘spring’, and ‘printemps’ are all mutual translations in the sense of the season of spring.

In general, such cliques of three nodes gives a high probability that the three word senses are equivalent. We are exploring extensions to PANIMAGES to compute word sense equivalences from multiple bilingual dictionaries, based on cliques in the graph.

3.2 Interface Design

The PANIMAGES graphical user interface allows a user to enter a search query in any of k source languages ($k = 50$ in the present implementation). PANIMAGES presents translations of the query term, presenting multiple sets of translations if there are multiple word senses of the term. The user selects one or more translations, and PANIMAGES sends this as a query to Google Images.

Finding Translations:

PANIMAGES looks up the node $(w_i,\ l_i)$ in the translation graph that corresponds to the query word and language, then follows edges in the graph to create one or more sets of nodes (w_k, l_k) where w_k is a translation into l_k for a particular sense of w_i. For each word sense, PANIMAGES follows paths of any length as long as the probability that the word sense has not changed according to Equation 4 is above a threshold.

In the example in Figure 2, the English word “spring” has edges with two word sense IDs, and thus has two sets of translations. The first set includes all nodes reachable by edges labeled with word sense 1, including (printemps, French), which has further edges with word sense 3. The second set is all nodes reachable by edges labeled with word sense 2 or 4.

Presenting Translations to the User:

PANIMAGES presents these sets of translations and allows the user to select one or more translation to be sent to Google

Images. Figure 5 shows the translations resulting from the Chinese word for dog.

As a practical consideration, PANIMAGES defaults to selecting translations in a language with high Web presence: an English translation for all source languages but English, and a French translation for English queries. The user may add or remove any of the translation-language pairs to the query before clicking on *Show Images*. Another option is to click on a single translation to send that as a query.

We were surprised to find that creating a Google Images query from a large number of translations does not improve the results. The more translations, the greater chance of ambiguity in one of the translated words or accidental homonyms with another language.

PANIMAGES finds over 100 translations of the English word "horse". Combining multiple translations in a single query gives poor results. Translations such as Gaelic ("each"), Faroese ("ross"), Hungarian ("ló"), or Romanian ("cal") produce only false hits. The Bulgarian ("кон") is ambiguous and returns images of religious icons as well as horses. The default translation (cheval,French) returns over 200,000 images of horses with high precision.

Handling Word Senses:
PANIMAGES lists each distinct word sense along with a gloss if available and the number of translations for this word sense. The user can click on a word sense to see the list of translations for that sense. PANIMAGES presents the word senses with the largest number of translations first, and selects this as the default word sense.

We felt that it would be confusing rather than helpful to include glosses in multiple languages (e.g. a gloss in French for entries from a French dictionary), so we display glosses in English only. Ideally, we would automatically translate the gloss into the user's language, although we have not tackled that problem.

Handling Gaps in Coverage:
There is a wide variety in the coverage of different source languages that a user can select in PANIMAGES. English, German, French, Russian, and Dutch have about 30,000 nodes in our translation graph, while Kurdish, Lithuanian and Latvian have just over 1,000 nodes. PANIMAGES uses automatic word completion to help a user navigate the space of words covered by a given input language. As a user begins to type in a word, a drop-down list appears of up to 20 completions of the characters entered so far. While this is more useful for alphabetic languages, it is also helpful for languages such as Chinese to show possible multi-character words.

There are further issues we have yet to resolve when a language has alternate orthography for the same word, such as optional diacritical marks in Hebrew or Arabic, or alternate character sets in Japanese or Serbian. The dictionary entries do not always match the form found commonly on Web pages.

4 Experimental Results

We now present statistics on a translation graph that PANIMAGES built from three dictionaries, along with recall and precision from a sample of image search queries over this translation graph. PANIMAGES built this translation graph from the English and French wiktionaries and an Esperanto dictionary.

4.1 Graph Statistics

The translation graph was built from over 437,000 words found in the three source dictionaries, giving a total of 656,031 translations.[2] Table 1 shows the number of languages in this translation graph and the distribution of the number of nodes per language. Of the 962 languages with nodes in the graph, three languages have over 30,000 nodes and an additional 53 languages have over 1,000 nodes. The majority of languages have 10 or fewer nodes.

Coverage of Translation Graph	
Number of languages	Number of words
3	30,000 - 33,200
53	1,000 - 30,000
73	100 - 1,000
157	10 - 100
274	2 - 10
402	1
962	1 - 33,200

Table 1: The distribution of the 962 languages represented in a translation graph built from English, French, and Esperanto dictionaries. 6% of the languages have at least 1,000 nodes in the graph, while 70% of the languages have 10 or fewer nodes.

Building a translation graph from a combination of the English wiktionary, French wiktionary, and Esperanto dictionary provided more translations than any of these dictionaries alone. Adding the French wiktionary and Esperanto dictionary to the English wiktionary, increased the number of English words in the translation graph from 19,500 to over 33,000. The coverage of French was similarly increased from 12,700 words to 29,400, and coverage of Esperanto from 23,000 words to 26,000.

There is a high degree of ambiguity for many common words in the graph, particularly for English words. Entries in the English wiktionary have an average of 1.46 word senses. We are only counting the word senses that have lists of translations, and ignoring word senses that did not produce edges in the translation graph. Over 100 English words have 10 or more word senses (e.g. ball, box, dirty, rock, key). 21% of the English words have translations for at least two word senses.

Merging equivalent word senses is a definite gain from a user's point of view. A common word such as the node (bicycle,English) has six edges with separate word sense IDs: one from the English dictionary entry, three from French words with translations as bicycle, and two from Esperanto words. After word sense merging, PANIMAGES presents only two word senses to the user: the noun and verb sense of bicycle.

[2] A word often has multiple translations.

We did preliminary evaluations of the precision and recall gain from inferences using Equations 1 through 4. We took a random set of 1,000 English words and considered Hebrew or Russian translations of the most common word sense of the English word. We chose these language pairs because we could easily find a bilingual speaker for English-Hebrew and English-Russian. We counted the number of English words with a direct edge to at least one Russian or Hebrew word, and counted the number of additional English words that have inferred translations. We also examined the precision (translation accuracy) of the additional translations.

Smoothing in Equation 1 is set to 1, the sense intersection threshold is set to 2, and we only consider paths with a maximum of 3 word sense IDs and where the translation probability is > 0.5. Table 2 shows the results for these two language pairs.

	Direct	Inferred translations		
	edges	Gain	Pct	Precision
English-Hebrew	65	41	63%	0.83
English-Russian	365	113	31%	0.68

Table 2: Translations of 1,000 random English words into Hebrew or Russian. 65 English words had direct translations into Hebrew, and 365 English words had direct translations into Russian. Graph inference found translations into Hebrew for an additional 41 words (63% increase) and translations into Russian for an additional 113 words (31% increase). 83% of the new translations in Hebrew and 68% of the new Russian translations maintain the original word sense.

Error analysis shows that Equations 1 through 4 are behaving as desired. The errors arise when our source dictionaries have entries that mix together multiple senses of a word. This causes a shift in word sense. We are exploring methods to detect when a sense ID needs to be split — when graph analysis shows that a subset of nodes linked to that ID have different translation behavior than another subset of nodes.

4.2 Image Retrieval Performance

We also evaluated coverage and precision of image search for non-English queries, comparing the results of sending the non-English query directly to Google Image search with the results of sending the default PANIMAGES translation instead. To generate our test set of words, we selected 10 arbitrary concepts that are associated with distinctive images, 6 nouns (ant, clown, fig, lake, sky, train), 2 verbs (eat, run), and two adjectives (happy, tired). For each of these, we selected 10 random translations of the English word from the translation graph. We restricted this to translations in one of PANIMAGES's 50 input languages. 29% of the random words were from 7 languages with relatively high Web presence (French, German, Spanish, Italian, Russian, Chinese, and Japanese). The remaining words are from 33 relatively minor languages ranging from Danish and Dutch to Telugu and Lithuanian.

5 Related Work

There has been considerable research on cross-lingual information retrieval (CLIR) in the past decade, prompted in part

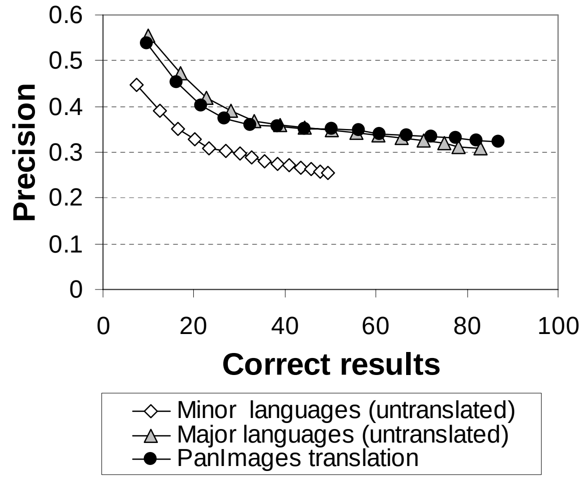

Figure 6: Image search results for 100 random non-English words and for the default PanImages translation. Results from 7 major languages that account for 29% of the non-English words are indistinguishable from PanImage results. For the remaining 33 minor languages, PanImages increases correct results by 75% on the first 15 pages (270 results), while also increasing precision by 27%. Total results increased 57-fold for the minor languages, from 33,000 to 1,856,000.

by the TREC conferences [Harman, 1996], the NTCIR conferences in Asia [Kishida *et al.*, 2004], and a series of SIGIR CLIR workshops [Gey *et al.*, 2006]. Surveys of CLIR research may be found in [Oard, 1997] and [Kishida, 2005].

In contrast to PANIMAGES, this research has focused on a small number of language pairs, much of it building systems that must be adapted to one language pair at a time. Examples of this type of system translate English queries into Spanish, French, or German [Buckley *et al.*, 2000] or translate English queries into Chinese [Xu, Fraser, & Weischedel, 2001].

While early CLIR systems typically relied on bilingual dictionaries [Hull & Grefenstette, 1996; Davis & Dunning, 1996], corpus-based methods or hybrid methods soon outstripped purely dictionary-based systems [Yang *et al.*, 1998]. Methods that derive word-translations from parallel text include [Gale & Church, 1991; Fung, 1995; Melamed, 1997; Franz, McCarly, & Zhu, 2001]. There are also hybrid systems [Ballesteros & Croft, 1998] that use corpus-based techniques to disambiguate translations provided by bilingual dictionaries.

The main drawback of using bilingual dictionaries alone is word sense ambiguity. A single term in the source language is typically translated into multiple terms in the target language, mixing different word senses. Combining information from multiple bilingual dictionaries only exacerbates this problem: translating from language l_1 into l_2 and then translating each of the possible l_2 translations into a third language l_3, quickly leads to an explosion of translations.

Commercial search engines such as Google (http://google.com), French Yahoo (http://fr.yahoo.com) and German Yahoo (http://de.yahoo.com), offer query translation capability for only a handful of languages. For example, French and German Yahoo automatically translate query terms into any of several major languages using Systran (http://www.systran.com) and translate the resulting Web pages.

In contrast, PANIMAGES translates between a large number of languages, and infers word-sense-preserving translations that are not found in any single dictionary.

6 Conclusions and Future Work

This paper introduces PANIMAGES, a fully-implemented cross-lingual image search system for the Web. While cross-lingual search systems have been available before, we believe that PANIMAGES has the broadest language coverage to date. In addition, PANIMAGES's coverage will continue to grow as it accumulates more source dictionaries.

PANIMAGES ingests multiple, independently-authored translation dictionaries to create the *translation graph*—a data structure that represents the translation relationships between words and their senses. Naive construction of the graph from multiple dictionaries results in *sense inflation*— equivalent word senses drawn from different dictionaries appear to be distinct. We have described a probabilistic method for solving this problem and for inferring translations from the graph.

PANIMAGES solves several difficulties in Web image search including both translation and word-sense disambiguation. Our preliminary results show that, for queries in languages that do not have high Web presence, PANIMAGES increases the total number of results 57-fold (from 33,000 to 1,856,000). PANIMAGES increases the number of correct images by 75% on the first 15 pages (270 results), while increasing precision by 27%.

In future work, we plan to apply the translation mechanism underlying PANIMAGES to other tasks including the translation of tags in social tagging systems such as *del.icio.us*, and in on-line games such as the Von Ahn's "ESP game" [von Ahn & Dabbish, 2004].

Acknowledgments

This research was supported in part by NSF grants IIS-0535284 and IIS-0312988, DARPA contract NBCHD030010, ONR grant N00014-02-1-0324 as well as gifts from Google, and carried out at the University of Washington's Turing Center. We thanks Ethan Phelps-Goodman, Doug Downey, and Jonathan Pool for helpful comments.

References

[Ballesteros & Croft, 1998] Ballesteros, L., and Croft, W. B. 1998. Resolving ambiguity for cross-language retrieval. In *Proceedings of ACM SIGIR 1998*, 64–71.

[Buckley *et al.*, 2000] Buckley, C.; Mitra, M.; Walz, J. A.; and Cardie, C. 2000. Using clustering and superconcepts within SMART: TREC 6. *Information Processing and Management* 36(1):109–131.

[Davis & Dunning, 1996] Davis, M., and Dunning, T. 1996. A TREC evaluation of query translation methods for multi-lingual text retrieval. In *Proceedings of TREC-4*.

[Franz, McCarly, & Zhu, 2001] Franz, M.; McCarly, S.; and Zhu, W. 2001. English-Chinese information retrieval at IBM. In *Proceedings of TREC 2001*.

[Fung, 1995] Fung, P. 1995. A pattern matching method for finding noun and proper noun translations from noisy parallel corpora. In *Proceedings of ACL-1995*.

[Gale & Church, 1991] Gale, W., and Church, K. 1991. A Program for Aligning Sentences in Bilingual Corpora. In *Proceedings of ACL-1991*.

[Gey *et al.*, 2006] Gey, F.; Kando, N.; Lin, C.-Y.; and Peters, C. 2006. New directions in multilingual information access: Introduction to the workshop at SIGIR 2006. In *Proceedings of Workshop on New Directions in Multilingual Information Access at SIGIR 2006*.

[Harman, 1996] Harman, D. 1996. Overview of the Fourth Text REtrieval Conference (TREC-4). In *Proceedings of TREC-4*.

[Hull & Grefenstette, 1996] Hull, D., and Grefenstette, G. 1996. Querying across languages: a dictionary-based approach to multilingual information retrieval. In *Proceedings of ACM SIGIR 1996*, 49–57.

[Kishida *et al.*, 2004] Kishida, K.; Chen, K.; Lee, S.; Chen, H.; Kando, N.; Kuriyama, K.; Myaeng, S.; and Eguchi, K. 2004. Cross-language information retrieval (CLIR) task at the NTCIR workshop 3. *ACM SIGIR Forum* 38(1):17–20.

[Kishida, 2005] Kishida, K. 2005. Technical issues of cross-language information retrieval: a review. *Information Processing and Management* 41:433–455.

[Melamed, 1997] Melamed, I. 1997. A Word-to-Word Model of Translational Equivalence. In *Proceedings of ACL-1997 and EACL-1997*, 490–497.

[Oard, 1997] Oard, D. 1997. Cross-language text retrieval research in the USA. In *Proceedings of 3rd DELOS Workshop*.

[von Ahn & Dabbish, 2004] von Ahn, L., and Dabbish, L. 2004. Labeling images with a computer game. In *ACM CHI Notes 2004*.

[Xu, Fraser, & Weischedel, 2001] Xu, J.; Fraser, A.; and Weischedel, R. M. 2001. TREC 2001 cross-lingual retrieval at BBN. In *Proceedings of TREC 2001*.

[Yang *et al.*, 1998] Yang, Y.; Carbonell, J. G.; Brown, R. D.; and Frederking, R. E. 1998. Translingual information retrieval: Learning from bilingual corpora. *Artificial Intelligence* 103(1-2):323–345.