

Translating Queries into Snippets for Improved Query Expansion

Stefan Riezler and Yi Liu and Alexander Vasserman

Google Inc.

1600 Amphitheatre Parkway

Mountain View, CA 94043

{riezler,yliu,avasserm}@google.com

Abstract

User logs of search engines have recently been applied successfully to improve various aspects of web search quality. In this paper, we will apply pairs of user queries and snippets of clicked results to train a machine translation model to bridge the “lexical gap” between query and document space. We show that the combination of a query-to-snippet translation model with a large n-gram language model trained on queries achieves improved *contextual query expansion* compared to a system based on term correlations.

1 Introduction

In recent years, user logs of search engines have attracted considerable attention in research on query clustering, query suggestions, query expansion, or general web search. Besides the sheer size of these data sets, the main attraction of user logs lies in the possibility to capitalize on users’ input, either in form of user-generated query reformulations, or in form of user clicks on presented search results. However noisy, sparse, incomplete, and volatile these data may be, recent research has presented impressive results that are based on simply taking the majority vote of user clicks as a signal for the relevance of results.

In this paper we will apply user logs to the problem of the “word mismatch” or “lexical chasm” (Berger et al., 2000) between user queries and documents. The standard solution to this problem, query expansion, attempts to overcome this

mismatch in query and document vocabularies by adding terms with similar statistical properties to those in the original query. This will increase the chances of matching words in relevant documents and also decrease the ambiguity of the overall query that is inherent to natural language. A successful approach to this problem is local feedback, or pseudo-relevance feedback (Xu and Croft, 1996), where expansion terms are extracted from the top-most documents that were retrieved in an initial retrieval round. Because of irrelevant results in the initial retrieval, caused by ambiguous terms or retrieval errors, this technique may cause expansion by unrelated terms, leading to query drift. Furthermore, the requirement of two retrieval steps is computationally expensive.

Several approaches have been presented that deploy user query logs to remedy these problems. One set of approaches focuses on user reformulations of queries that differ only in one segment (Jones et al., 2006; Fonseca et al., 2005; Huang et al., 2003). Such segments are then identified as candidate expansion terms, and filtered by various signals such as cooccurrence in similar sessions or log-likelihood ratio of original and expansion phrases. Other approaches focus on the relation of queries and retrieval results, either by deploying the graph induced by queries and user clicks in calculating query similarity (Beeferman and Berger, 2000; Wen et al., 2002; Baeza-Yates and Tiberi, 2007), or by leveraging top results from past queries to provide greater context in finding related queries (Raghavan and Sever, 1995; Fitzpatrick and Dent, 1997; Sahami and Heilman, 2006). Cui et al. (2002) present an all together different way to deploy user clickthrough data by extracting expansion terms directly from clicked results. They claim significant improvements over

the local feedback technique of Xu and Croft (1996).

Cui et al.’s (2002) work is the closest to ours. We follow their approach in extracting expansion terms directly from clicked results, however, with a focus on high precision of query expansion. While expansion from the domain of document terms has the advantage that expansion terms are guaranteed to be in the search domain, expansion precision may suffer from the noisy and indirect “approval” of retrieval results by user clicks. Thus expansion terms from the document domain are more likely to be generalizations, specifications, or otherwise related terms, than terms extracted from query substitutions that resemble synonyms more closely. Furthermore, if the model that learns to correlate document terms to query terms is required to ignore context in order to generalize, finding appropriate expansions for ambiguous query terms is difficult.

Our approach is to look at the “word mismatch” problem as a problem of translating from a source language of queries into a target language of documents, represented as snippets. Since both queries and snippets are arguably natural language, statistical machine translation technology (SMT) is readily applicable to this task. In previous work, this has been done successfully for question answering tasks (Riezler et al., 2007; Soricut and Brill, 2006; Echihabi and Marcu, 2003; Berger et al., 2000), but not for web search in general. Cui et al.’s (2002) model is to our knowledge the first to deploy query-document relations for direct extraction of expansion terms for general web retrieval. Our SMT approach has two main advantages over Cui et al.’s model: Firstly, Cui et al.’s model relates document terms to query terms by using simple term frequency counts in session data, without considering smoothing techniques. Our approach deploys a sophisticated machine learning approach to word alignment, including smoothing techniques, to map query phrases to snippet phrases. Secondly, Cui et al.’s model only indirectly uses context information to disambiguate expansion terms. This is done by calculating the relationship of an expansion term to the whole query by multiplying its contributions to all query terms. In our SMT approach, contextual disambiguation is done by deploying an n-gram language model trained on queries to decide about the appropriateness of an expansion term in the con-

text of the rest of the query terms. As shown in an experimental evaluation, together the orthogonal information sources of a translation model and a language model provide significantly better contextual query expansion than Cui et al.’s (2002) correlation-based approach.

In the following, we recapitulate the essentials of Cui et al.’s (2002) model, and contrast it with our SMT-based query expansion system. Furthermore, we will present a detailed comparison of the two systems on a real-world query expansion task.

2 Query-Document Term Correlations

The query expansion model of Cui et al. (2002) is based on the principle that if queries containing one term often lead to the selection of documents containing another term, then a strong relationship between the two terms is assumed. Query terms and document terms are linked via clicked documents in user sessions. Formally, Cui et al. (2002) compute the following probability distribution of document words w^d given query words w^q from counts over clicked documents D :

$$P(w^d|w^q) = \sum_D P(w^d|D)P(D|w^q) \quad (1)$$

The first term in the righthandside of equation 1 is a normalized *tfidf* weight of the the document term in the clicked document, and the second term is the relative cooccurrence of document and query term in sessions.

Since equation 1 calculates expansion probabilities for each term separately, Cui et al. (2002) introduce the following cohesion formula that respects the whole query Q by aggregating the expansion probabilities for each query term:

$$CoWeight_Q(w^d) = \ln\left(\prod_{w^q \in Q} P(w^d|w^q) + 1\right) \quad (2)$$

In contrast to local feedback techniques (Xu and Croft, 1996), Cui et al.’s algorithm allows to precompute term correlations offline by collecting counts from query logs. This reliance on pure frequency counting is both a blessing and a curse: On the one hand it allows for efficient non-iterative estimation, on the other hand it makes the implicit assumption that data sparsity will be overcome by counting from huge datasets. The only attempt at smoothing that is made in this approach is a recurrence to words in query context, using equation 2, when equation 1 assigns zero probability to unseen pairs.

3 Query-Snippet Translation

The SMT system deployed in our approach is an implementation of the alignment template approach of Och and Ney (Och and Ney, 2004). The basic features of the model consist of a translation model and a language model which go back to the noisy channel formulation of machine translation in Brown et al. (1993). Their “fundamental equation of machine translation” defines the job of a translation system as finding the English string \hat{e} that is a translation of a foreign string \mathbf{f} such that

$$\begin{aligned}\hat{e} &= \arg \max_e P(\mathbf{e}|\mathbf{f}) \\ &= \arg \max_e P(\mathbf{f}|\mathbf{e})P(\mathbf{e})\end{aligned}\quad (3)$$

Equation 3 allows for a separation of a language model $P(\mathbf{e})$, and a translation model $P(\mathbf{f}|\mathbf{e})$. Och and Ney (2004) reformulate equation 3 as a linear combination of feature functions $h_m(\mathbf{e}, \mathbf{f})$ and weights λ_m , including feature functions for translation models $h_i(\mathbf{e}, \mathbf{f}) = P(\mathbf{f}|\mathbf{e})$ and language models $h_j(\mathbf{e}) = P(\mathbf{e})$:

$$\hat{e} = \arg \max_e \sum_{m=1}^M \lambda_m h_m(\mathbf{e}, \mathbf{f}) \quad (4)$$

The translation model used in our approach is based on the sequence of alignment models described in Och and Ney (2003). The relationship of translation model and alignment model for source language string $\mathbf{f} = f_1^J$ and target string $\mathbf{e} = e_1^I$ is via a hidden variable describing an alignment mapping from source position j to target position a_j :

$$P(f_1^J | e_1^I) = \sum_{a_1^J} P(f_1^J, a_1^J | e_1^I) \quad (5)$$

The alignment a_1^J contains so-called null-word alignments $a_j = 0$ that align source words to the empty word. The different alignment models described in Och and Ney (2003) each parameterize equation 5 differently so as to capture different properties of source and target mappings. All models are based on estimating parameters θ by maximizing the likelihood of training data consisting of sentence-aligned, but not word-aligned strings $\{(\mathbf{f}_s, \mathbf{e}_s) : s = 1, \dots, S\}$. Since each sentence pair is linked by a hidden alignment variable $\mathbf{a} = a_1^J$, the optimal $\hat{\theta}$ is found using unlabeled-data log-likelihood estimation techniques such as

the EM algorithm (Dempster et al., 1977):

$$\hat{\theta} = \arg \max_{\theta} \prod_{s=1}^S \sum_{\mathbf{a}} p_{\theta}(\mathbf{f}_s, \mathbf{a} | \mathbf{e}_s) \quad (6)$$

The final translation model is calculated from relative frequencies of phrases, i.e. consecutive sequences of words occurring in text. Phrases are extracted via various heuristics as larger blocks of aligned words from best word alignments, as described in Och and Ney (2004).

Language modeling in our approach deploys an n-gram language model that assigns the following probability to a string w_1^L of words (see Brants et al. (2007)):

$$P(w_1^L) = \prod_{i=1}^L P(w_i | w_1^{i-1}) \quad (7)$$

$$\approx \prod_{i=1}^L P(w_i | w_{i-n+1}^{i-1}) \quad (8)$$

Estimation of n-gram probabilities is done by counting relative frequencies of n-grams in a corpus of user queries. Remedies against sparse data problems are achieved by various smoothing techniques, as described in Brants et al. (2007).

For applications of the system to translate unseen queries, a standard dynamic-programming beam-search decoder (Och and Ney, 2004) that tightly integrates translation model and language model is used. Expansion terms are taken from those terms in the 5-best translations of the query that have not been seen in the original query string.

In our opinion, the advantages of using an alignment-based translation model to correlate document terms with query terms, instead of relying on a term frequency counts as in equation 1, are as follows. The formalization of translation models as involving a hidden alignment variable allows us to induce a probability distribution that assigns some probability of being translated into a target word to every source word. This is a crucial step towards solving the problem of the “lexical gap” described above. Furthermore, various additional smoothing techniques are employed in alignment to avoid overfitting and improved coping with rare words (see Och and Ney (2003)). Lastly, estimation of hidden-variable models can be based on the well-defined framework of statistical estimation via the EM algorithm.

Similar arguments hold for the language model: N-gram language modeling is a well-understood

	sentence pairs	source words	target words
tokens	3 billion	8 billion	25 billion
avg. length	-	2.6	8.3

Table 1: Statistics of query-snippet training data for translation model.

problem, with a host of well-proven smoothing techniques to avoid data sparsity problems (see Brants et al. (2007).)

In combination, translation model and language model provide orthogonal sources of information to the overall translation quality. While the translation model induces a smooth probability distribution that relates source to target words, the language model deploys probabilities of target language strings to assess the adequacy of a target word as a *translation in context*. Reliance on ordering information of the words in the context of a source word is a huge advantage over the bag-of-words aggregation of context information in Cui et al’s (2002) model. Furthermore, in the SMT model used in our approach, translation model and language model are efficiently integrated in a beam-search decoder.

In our application of SMT to query expansion, queries are considered as source language sentences and snippets of clicked result documents as target sentences. A parallel corpus of sentence-aligned data is created by pairing each query with each snippet of its clicked results. Further adjustments to system parameters were applied in order to adapt the training procedure to this special data set. For example, in order to account for the difference in sentence length between queries and snippets, we set the null-word probability to 0.9. This allows us to improve precision of alignment of noisy data by concentrating the alignment to a small number of key words. Furthermore, extraction of phrases in our approach is restricted to the intersection of alignments from both translation directions, thus favoring precision over recall also in phrase extraction. The only major adjustment of the language model to the special case of query-snippet translation is the fact that we train our n-gram model on queries taken from user logs, instead of on standard English text.

1-grams	2-grams	3-grams
9 million	1.5 billion	5 billion

Table 2: Statistics of unique query n-grams in language model.

	items w/ agreement	disagreements included
# items	102	125
mean item score	0.333	0.279
95% conf. int.	[0.216, 0.451]	[0.176, 0.381]

Table 3: Comparison of SMT-based expansion with correlation-based expansion on 7-point Likert-type scale.

4 Experimental Evaluation

4.1 Data

The training data for the translation model and the correlation-based model consist of pairs of queries and snippets for clicked results taken from anonymized query logs. Using snippets instead of full documents makes iterative training feasible and also reduces noise considerably. This parallel corpus of query-snippet pairs is fed into a standard SMT training pipeline (modulo the adjustments to word and phrase alignment discussed above). The parallel corpus consists of 3 billion query-snippet pairs that are input to training of word and phrase alignment models. The resulting phrase translation table that builds the basis of the translation model consists 700 million query-snippet phrase translations. A collection of data statistics for the training data is shown in table 1.

The language model used in our experiment is a trigram language model trained on English queries in user logs. N-grams were cut off at a minimum frequency of 4. Data statistics for resulting unique n-grams are shown in table 2.

4.2 Experimental Comparison

Our experimental setup for query expansion deploys a real-world search engine, google.com, for a comparison of expansions from the SMT-based system and the correlation-based system. The experimental evaluation was done as direct comparison of search results for queries where both experimental systems suggested expansion terms. Since expansions from both experimental systems are done on top of the same underlying search engine, this allows us to abstract away from interactions with the underlying system. The queries used for evaluation were extracted randomly from 3+ word

query	SMT-based expansions	corr-based expansions	score
applying U.S. passport	passport - visa	applying - home	-1.0
configure debian to use dhcp	debian - linux configure - install	configure - configuring	-1.0
how many episodes of 30 rock?	episodes - season episodes - series	how many episodes - tv many episodes - wikipedia	-0.83
lampasas county sheriff department	department - office	department - home sheriff - office	-0.83
weakerthans cat virtue chords	chords - guitar chords - lyrics chords - tab	cat - tabs chords - tabs	-0.83
Henry VIII Menu Portland, Maine	menu - restaurant menu - restaurants	portland - six menu - england	1.3
ladybug birthday parties	parties - ideas parties - party	ladybug - kids	1.3
political cartoon calvin coolidge	cartoon - cartoons	political cartoon - encyclopedia	1.3
top ten dining, vancouver	dining - restaurants	dining vancouver - 10	1.3
international communication in veterinary medicine	communication - communications communication - skills	international communication - college	1.3

Table 4: SMT-based versus correlation-based expansions with mean item score.

queries in user logs in order to allow the systems to deploy context information for expansion.

In order to evaluate Cui et al.’s (2002) correlation-based system in this setup, we required the system to assign expansion terms to particular query terms. This could be achieved by using a linear interpolation of scores in equation 2 and equation 1. Equation 1 thus introduces a preference for a particular query term to the whole-query score calculated by equation 2. Our reimplementation uses unigram and bigram phrases in queries and expansions. Furthermore, we use *Okapi BM25* instead of *tfidf* in the calculation of equation 1 (see Robertson et al. (1998)).

Query expansion for the SMT-based system is done by extracting terms introduced in the 5-best list of query translations as expansion terms for the respective query terms.

The evaluation was performed by three independent raters. The raters were given task-specific rating guidelines, and were shown queries and 10-best search results from both systems, anonymized, and presented randomly on left or right sides. The raters’ task was to evaluate the results on a 7-point Likert-type¹ scale, defined as:

-1.5: much worse

-1.0: worse

-0.5: slightly worse

¹Likert’s (1932) original rating system is a 5-point scale using integer scores 1 through 5. Our system uses average scores over three raters for each item, and uses a 7-point instead of a 5-point scale. See Dawes (2008) on the comparability of 5-, 7-, or 10-point scales.

0: about the same

0.5: slightly better

1.0: better

1.5: much better

Results on 125 queries where both systems suggested expansion terms are shown in table 3. For each query, rating scores are averaged over the scores assigned by three raters. The overall mean item score for a comparison of SMT-based expansion against correlation-based expansion was 0.333 for 102 items with rater agreement, and 0.279 for 125 items including rater disagreements. All result differences are statistically significant.

Examples for SMT-based and correlation-based expansions are given in table 4. The first five examples are losses for the SMT-based system. In the first example, *passport* is replaced by the related, but not synonymous term *visa* in the SMT-based expansion. The second example is a loss for SMT-based expansion because of a replacement of the specific term *debian* by the more general term *linux*. The correlation-based expansions *tv 30 rock* in the third example, *lampasas county sheriff home* in the fourth example, and *weakerthans tabs* in the fifth example directly hit the title of relevant web pages, while the SMT-based expansion terms do not improve retrieval results. However, even from these negative examples it becomes apparent that the SMT-based expansion terms are clearly related to the query terms, and for a majority cases this has a positive effect. Such examples are shown in

(herbs , herbs) (for , for) (chronic , chronic) (constipation , constipation)
(herbs , herb) (for , for) (chronic , chronic) (constipation , constipation)
(herbs , remedies) (for , for) (chronic , chronic) (constipation , constipation)
(herbs , medicine) (for , for) (chronic , chronic) (constipation , constipation)
(herbs , supplements) (for , for) (chronic , chronic) (constipation , constipation)
(herbs , herbs) (for , for) (mexican , mexican) (cooking , cooking)
(herbs , herbs) (for , for) (cooking , cooking) (mexican , mexican)
(herbs , herbs) (for , for) (mexican , mexican) (cooking , food)
(mexican , mexican) (herbs , herbs) (for , for) (cooking , cooking)
(herbs , spices) (for , for) (mexican , mexican) (cooking , cooking)

Table 5: Unique 5-best phrase-level translations of queries *herbs for chronic constipation* and *herbs for mexican cooking*.

query terms	n-best expansions		
herbs	com	treatment	encyclopedia
chronic	interpret	treating	com
constipation	interpret	treating	com
herbs for	medicinal	support	women
for chronic	com	gold	encyclopedia
chronic constipation	interpret	treating	
herbs	cooks	recipes	com
mexican	recipes	com	cooks
cooking	cooks	recipes	com
herbs for	medicinal	women	support
for mexican	cooks	com	allrecipes

Table 6: Correlation-based expansions for queries *herbs for chronic constipation* and *herbs for mexican cooking*.

the second set of expansions. SMT-based expansions such as *henry viii restaurant portland, maine*, or *ladybug birthday ideas*, or *top ten restaurants, vancouver* achieve a change in retrieval results that does not result in a query drift, but rather in improved retrieval results. In contrast, the terms introduced by the correlation-based system are either only vaguely related or noise.

5 Discussion

We attribute the experimental result of a significant preference for SMT-based expansions over correlation-based expansions to the fruitful combination of translation model and language model provided by the SMT system. The SMT approach can be viewed as a combined system that proposes candidate expansion via the translation model, and filters them by the language model. Thus we may find a certain amount of non-sensical expansion candidates at the phrase translation level. This can be seen from inspecting table 7 which shows the most probable phrase translations that are applicable to the queries *herbs for chronic constipation* and *herbs for mexican cooking*. The phrase table includes identity translations and closely related terms as most probable translations for nearly every phrase, however, it also clearly includes noisy

and non-related terms. More importantly, an extraction of expansion terms from the phrase table alone would not allow to choose the appropriate term for the given query context. This can be attained by combining the phrase translations with a language model: As shown in table 5, the 5-best translations of the full queries attain a proper disambiguation of the senses of *herbs* by replacing the term by *remedies*, *medicine*, and *supplements* for the first query, and with *spices* for the second query. Expansion terms highlighted in bold face.

The fact that the most probable translation for the whole query mostly is the identity translation can be seen as a feature, not as a bug, of the SMT-based approach: By the option to prefer identity translations or word reorderings over translations of source words, the SMT model effectively can choose not to generate any expansion terms. This will happen if none of the candidate phrase translations fit with high enough probability into the context of the whole query, as assessed by the language model.

In contrast to the SMT model, the correlation-based model cannot fall back onto the ordering information of the language model, but aggregates information for the whole query from a bag-of-words of query terms. Table 6 shows the top three

correlation-based expansion terms assigned to unigrams and bigrams in the queries *herbs for chronic constipation* and *herbs for mexican cooking*. Expansion terms are chosen by overall highest weight and shown in bold face. Relevant expansion terms such as *treatment* or *recipes* that would disambiguate the meaning of *herbs* are in fact proposed by the correlation-based model, however, the cohesion score also promotes terms such as *interpret* or *com* as best whole-query expansions, thus leading to query drift.

6 Conclusion

We presented an approach to *contextual query expansion* that deploys natural language technology in form of statistical machine translation. The key idea of our approach is to consider the problem of the “lexical gap” between queries and documents from a linguistic point of view, and attempt to bridge this gap by translating from the query language into the document language. Using search engine user logs, we could extract large amounts of parallel data of queries and snippets from clicked documents. These data were used to train an alignment-based translation model, and an n-gram based language model. The same data were used to train a reimplementation of Cui et al.’s (2002) term-correlation based query expansion system. An experimental comparison of the two systems showed a considerable preference for SMT-based expansions over correlation-based expansion. Our explanation for this result is the fruitful combination of the orthogonal information sources from translation model and language model. While in the SMT approach expansion candidates proposed by the translation model are effectively filtered by ordering information on the query context from the language model, the correlation-based approach resorts to an inferior bag-of-word aggregation of scores for the whole query. Furthermore, each component of the SMT model takes great care to avoid sparse data problems by various sophisticated smoothing techniques. In contrast, the correlation-based model relies on pure counts of term frequencies.

An interesting task for future work is to dissect the contributions of translation model and language model, for example, by combining a correlation-based system with a language model filter. The challenge here is a proper integration of n-gram lookup into correlation-based expansion.

References

- Baeza-Yates, Ricardo and Alessandro Tiberi. 2007. Extracting semantic relations from query logs. In *Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD’07)*, San Jose, CA.
- Beeferman, Doug and Adam Berger. 2000. Agglomerative clustering of a search engine query log. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’00)*, Boston, MA.
- Berger, Adam L., Rich Caruana, David Cohn, Dayne Freitag, and Vibhu Mittal. 2000. Bridging the lexical chasm: Statistical approaches to answer-finding. In *Proceedings of SIGIR’00*, Athens, Greece.
- Brants, Thorsten, Ashok C. Popat, Peng Xu, Franz J. Och, and Jeffrey Dean. 2007. Large language models in machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP’07)*, Prague, Czech Republic.
- Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.
- Cui, Hang, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma. 2002. Probabilistic query expansion using query logs. In *Proceedings of WWW 2002*, Honolulu, Hawaii.
- Dawes, John. 2008. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1):61–77.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(B):1–38.
- Echihabi, Abdessamad and Daniel Marcu. 2003. A noisy-channel approach to question answering. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL’03)*, Sapporo, Japan.
- Fitzpatrick, Larry and Mei Dent. 1997. Automatic feedback using past queries: Social searching? In *Proceedings of SIGIR’97*, Philadelphia, PA.
- Fonseca, Bruno M., Paulo Golgher, Bruno Possas, Berthier Ribeiro-Neto, and Nivio Ziviani. 2005. Concept-based interactive query expansion. In *Proceedings of the 14th Conference on Information and Knowledge Management (CIKM’05)*, Bremen, Germany.

Huang, Chien-Kang, Lee-Feng Chien, and Yen-Jen Oyang. 2003. Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology*, 54(7):638–649.

Jones, Rosie, Benjamin Rey, Omid Madani, and Wiley Greiner. 2006. Generating query substitutions. In *Proceedings of the 15th International World Wide Web conference (WWW'06)*, Edinburgh, Scotland.

Likert, Rensis. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 140:5–55.

Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

Och, Franz Josef and Hermann Ney. 2004. The alignment template approach to statistical machine translation. *Computational Linguistics*, 30(4):417–449.

Raghavan, Vijay V. and Hayri Sever. 1995. On the reuse of past optimal queries. In *Proceedings of SIGIR'95*, Seattle, WA.

Riezler, Stefan, Alexander Vasserman, Ioannis Tsochantaridis, Vibhu Mittal, and Yi Liu. 2007. Statistical machine translation for query expansion in answer retrieval. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL'07)*, Prague, Czech Republic.

Robertson, Stephen E., Steve Walker, and Micheline Hancock-Beaulieu. 1998. Okapi at TREC-7. In *Proceedings of the Seventh Text REtrieval Conference (TREC-7)*, Gaithersburg, MD.

Sahami, Mehran and Timothy D. Heilman. 2006. A web-based kernel function for measuring the similarity of short text snippets. In *Proceedings of the 15th International World Wide Web conference (WWW'06)*, Edinburgh, Scotland.

Soricut, Radu and Eric Brill. 2006. Automatic question answering using the web: Beyond the factoid. *Journal of Information Retrieval - Special Issue on Web Information Retrieval*, 9:191–206.

Wen, Ji-Rong, Jian-Yun Nie, and Hong-Jiang Zhang. 2002. Query clustering using user logs. *ACM Transactions on Information Systems*, 20(1):59–81.

Xu, Jinxi and W. Bruce Croft. 1996. Query expansion using local and global document analysis. In *Proceedings of SIGIR'96*, Zurich, Switzerland.

herbs	herbs herbal medicinal spices supplements remedies
herbs for	herbs for herbs herbs and with herbs
herbs for chronic	herbs for chronic and herbs for chronic herbs for
for	for
for chronic	for chronic chronic of chronic
for chronic constipation	for chronic constipation chronic constipation for constipation
chronic	chronic acute patients treatment
chronic constipation	chronic constipation of chronic constipation with chronic constipation
constipation	constipation bowel common symptoms
for mexican	for mexican mexican the mexican of mexican
for mexican cooking	mexican food mexican food and mexican glossary
mexican	mexican mexico the mexican
mexican cooking	mexican cooking mexican food mexican cooking
cooking	cooking culinary recipes cook food recipe

Table 7: Phrase translations applicable to source strings *herbs for chronic constipation* and *herbs for mexican cooking*.