Using Part-of-speech Patterns to Reduce Query Ambiguity

James Allan and Hema Raghavan
Center for Intelligent Information Retrieval
Department of Computer Science
University of Massachusetts
Amherst, MA 01003
{allan,hema}@cs.umass.edu

ABSTRACT

Query ambiguity is a generally recognized problem, particularly in Web environments where queries are commonly only one or two words in length. In this study, we explore one technique that finds commonly occurring patterns of parts of speech near a one-word query and allows them to be transformed into clarification questions. We use a technique derived from statistical language modeling to show that the clarification queries will reduce ambiguity much of the time, and often quite substantially.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*linguistic processing*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*query formulation, search process*

Keywords

Clarity, part of speech, query ambiguity

1. INTRODUCTION

A generally acknowledged issue in information retrieval, particularly with Web search engines, is that users provide very short queries [15] that are sometimes very ambiguous. A classic example of this problem is the meaning of "Java" as a query: is the searcher's interest in coffee, a programming language, or tourism in Indonesia? There is quite a range of ways in which ambiguity can appear. In the "java" example, it is based on subject matter. A query could also be ambiguous because the user's underlying task is unclear (e.g., buy Java, use Java, introduction to Java), because the type of user is unknown (e.g., expert or novice), or because the style of querying is not clear (e.g., obtain facts, an overview, exhaustive coverage, or summary).

In this study, we investigate a particular technique for resolving ambiguity that is motivated by task-level ambiguity. How do the query words relate to other words in the text?

What roles can they assume and which is the user interested in? For example, given a query "boat", the user might be interested in:

- things a boat does (floats, anchors, overturns)

- types of boats (whaling, fishing, red)

- ways to boat (quickly, safely)

- things to do with boats (drive them, paint them, sink them)

The intent is for the system to present a list of clarifying options to the searcher that can be selected if needed. Choosing one of those options focuses the query by including only those documents that use the query words in the right manner.

Ultimately, these ideas should be evaluated by field-testing a system that incorporates them. In this study we are laying the groundwork for such a system. Our focus is a novel evaluation to demonstrate that this approach generates useful clarifications of an ambiguous query. In the next section, we sketch some prior work toward coping with ambiguity. In Section 4 we discuss the approach that we use to find the clarifying options; the actual implementation is described in Section 3. We describe our evaluation technique in Section 5 and present results showing the value of this approach in Section 6. We conclude in Section 7 by describing where this work is headed.

2. RELATED WORK

There has been substantial work investigating issues surrounding word sense disambiguation [16], a type of query ambiguity that arises regularly in information retrieval. That type of ambiguity is often resolved implicitly when queries are long enough—the additional words provide sufficient context to clear up confusion—but is still a critical problem when queries are short [20]. Users of Web search engines generally provide short queries [15], and we are focusing on that situation in this study.

There is a range of information retrieval interface ideas that attempt to help the user deal with ambiguous queries. Sanderson, Lawrie, and Croft have been working on concept hierarchies that provide a hierarchical map of words and their relationships [19, 21, 11, 17]. A user can navigate the hierarchy to find the sense that is relevant. To date that research has not directly investigated whether the hierarchies help with disambiguation, looking instead at their ability to affect recall and precision.

Clustering the retrieved document set is another way to deal with ambiguity: ideally documents covering different senses of a word will be placed in different clusters. Much work has been done on clustering, either investigating the clustering directly [1, 2, 14] or exploring issues related to clustering and interactive search [13, 18, 23]. The Web service Northern Light[1] classifies returned documents into a set of labeled clusters, showing the clusters as well as the top ranked documents. All of these clustering techniques group documents by topic rather than by the way that the query word is used.

A limited number of studies have been done that directly evaluate the effectiveness of query reformulation when searching. It has been shown that query reformulation can improve the effectiveness of a query [5], though the focus was on the cognitive burden it places on the searcher.

This work was inspired by Tables 8 and 9 of a paper by Church et al. [7]. That work describes how statistical occurrence patterns of words in text can be used to find lexically interesting items. One of those is to look for verbs that occur near a particular word in an interesting way (measured by mutual information), allowing someone to find "what does a boat do?" or "what do you typically do with *food* and *water*?" We felt that if a query "boat" were used, then a possible clarification might be that question, and that documents containing the appropriate pattern (viz., boat as a subject and a verb "interestingly" near it) would be appropriate matches.

AltaVista[2] at one point provided a service called Live-Topics [22] that showed the inter-relationships between query terms and other terms in the corpus. This map was derived from the corpus (and human corrections) and provided another form of a "hyperindex" to allow the user to select words to improve a query.

The HiB system [4] allows the user to clarify (refine) the query by offering well-formed phrases in which the query appears. The user's query is mapped into a "hyperindex" by this approach, that shows relationships between words based on their occurrence in the titles of retrieved documents. The motivation for this work is similar to ours, but requires a much more elaborate infrastructure. They do not appear to have evaluated the value of the technique for query disambiguation. Grefenstette also suggests a system similar to ours [12], but also has no evaluation of its effectiveness.

Anick and Tipirneni [3] developed a query refinement technique based on calculating the lexical dispersion of words occurring in the top-ranked documents. Their work is similar to this study in that the candidates are identified by part-of-speech patterns at index time, but they use a substantially reduced set of patterns ("adjective? noun+"), they present patterns to the user on the assumption that the top ranked documents are a good source for them, and they are more focussed on query expansion than disambiguation (though the difference is often subtle). We believe that some aspects of their approach (e.g., using dispersion to find good candidates) may be valuable for our work, but we have not yet investigated those ideas.

3. EXPERIMENTAL SETUP

All experiments in this study were conducted using about

Tag	Part-Of-Speech
NN	Common noun (singular)
NP	Proper noun
NNS	Common noun (plural)
VB	Verb (present tense)
VBN	Verb(Past participle)
VBD	Verb(Past tense)
JJ	Adjective
AT	Determiner
CC	Conjunction
IN	Preposition
CD	Number

Figure 1: Some of the common part of speech tags and their corresponding parts of speech in the Brown corpus.

178,000 documents from the *trec5-fbis*, the FBIS subset of the fifth TREC volume. These are documents from FBIS dated 1994. We also used a second corpus, *tdt2*, which includes the English news stories from the TDT-2 collection, amounting to approximately 40,000 news stories from newswire and broadcast news sources. For stories from an audio source, the closed caption was used rather than speech recognition output.

Our indexing and retrieval was done using V3.2 of the CIIR's InQuery retrieval system [6]. This system incorporates fast and reasonably accurate part of speech tagging using JTAG [24]. JTAG uses the Brown tag-set (some tags of which are listed in Figure 1).

When documents were indexed, each word was indexed along with its part of speech. Proximity operators allow a particular part of speech to be selected for a word, as well as restrictions on which parts of speech occurred nearby. For example, #1(boat VBD) means the word "boat" immediately followed by the past participle of a verb (e.g., "boat sank"). The query #0(boat NN) means the word "boat" used *as* a noun rather than as another part of speech. The combined query,

$$\texttt{\#3(\#0(boat NN) VBZ)}$$

means "boat" as a noun followed within three words by the present third person singular form a verb (e.g., "boat sinks" or "boat quickly sinks"). This type of construction allows us to build up quite complex restrictions on query words and nearby parts of speech.

4. APPROACH

Given a corpus that has parts of speech indexed and that provides flexible query support such as that outlined above, we can attempt query disambiguation as follows. We (1) analyze the text to find patterns of parts of speech that occur frequently near the query word, (2) map the patterns to questions that reflect the patterns, (3) present the questions to the user as options to consider, and (4) re-run the retrieval based on the disambiguated question(s) the searcher selects.

In this section, we sketch how this approach works in a functional system. We show several examples that illustrate the value of this approach and suggest that it will be effective. In Section 5 we empirically evaluate the effectiveness of

this approach toward reducing ambiguity. Our longer term goal is to implement these ideas in a fully functioning system to get user feedback.

4.1 Finding patterns

The patterns that we are looking for center around a single query word and incorporate a small number of words on either side. For example, given the query word *boat* we might look for the pattern "an adjective followed by the noun *boat*." In the trec5-fbis corpus, we would find adjectives such as *Haitian*, *fishing*, *wooden*, *speed*, and so on. If we find a sufficient number of occurrences of that pattern (regardless of how many different adjectives there are), we would note that "JJ boat:NN" is an interesting pattern ("JJ" is the part of speech code for an adjective, and "NN" denotes a singular noun).

Looking for all such patterns near a query word would be prohibitively expensive to carry out at search time. For that reason, we enhanced InQuery's indexing stage to extract all patterns centered around non-stopwords so that we could use it at run-time, and organized them by that keyword. To make this process slightly less expensive we used a 50% random sample of the corpus. For example, Figure 2 shows a sample of patterns and counts that appear in the 80,000 document subset of the trec5-fbis corpus.

We found by observation that any pattern that occurs more than 50 times in the corpus is "interesting" and ignored all others (a pattern might occur less than 50 times with particular query word, but many more than 50 times overall). Given the large number of patterns, making this decision in that way does not seem unreasonable, though it would ultimately be preferable to use a measure such as mutual information to select and keep the more interesting patterns [7].

4.2 Patterns to questions

The patterns are interesting, but they are not appropriate for displaying to a user. Instead, we convert them into questions that disambiguate the query. For example, here are some patterns and corresponding questions for the query word *party*:

JJ party:NN	varieties of a *party*
NP party:NP	names of a *party*
NN IN party:NN NNS	things done with a *party*
VB party:NN NN	things done to a *party*

For each pattern that has a corresponding question, we index the pattern by the query word and keep track of the number of times the pattern occurs.

Given a user's single-word query, all patterns including that query word are obtained from the index, and the list of questions is constructed. Patterns that correspond to the same question are collapsed into a single question.

The questions are presented ranked in order of the most frequently occurring first. (The evaluation in Section 6 suggests that questions with many matches are not much less ambiguous than the original query, so it may be preferable to rank them by an approximation of ambiguity reduction.)

A partial list of patterns and corresponding questions is presented in Figure 3. Note that some language processing could be useful to handle variants of queries (e.g., to change "varieties of mouse" into "varieties of mice").

We wish to make it clear that the mapping from patterns

Pattern	Question
Q:noun VB	things that QQ does
Q:noun VBD	things that QQ did
Q:noun HV verb	things that QQ did
VBD Q:NN	things that happened to QQ
Q:VBN NN(S)	things that were QQ
Q:VBN TOIN AT noun	things one can QQ to
adv Q:verb	ways to be QQ
Q:VBG TO VB	things you could be QQ to do
adj Q:noun	varieties of QQ
Q:noun BER adj	varieties of QQ
Q:noun BER verb	things done to QQ
Q:noun BER noun	things that QQ are

Figure 3: Partial list of patterns recognized by the query clarification index component, along with the questions corresponding to those patterns. In the pattern, "Q:pos" means the query word occurring in that part of speech. The codes "noun," "verb," and "adj" are macros that indicate all varieties of nouns that the tagger identifies. In the Question column, QQ would be replaced by the query word.

to questions was created by hand. We found that a surprisingly small set of questions could accommodate most patterns, but it is most likely that some questions are absent. We are currently exploring better approaches for generating this mapping.

It is likely that the clarifying questions themselves may not be sufficient to help the user recognize the distinction. To help with that, we present the user which a list of several sample phrases that match the corresponding patterns. For example, for the query "*boat*", we might get (examples taken from our corpora):

> Are you interested in varieties of boats?
> > *Haitian boat, makeshift boat, wooden boat, Cuban boat, big boat, …*
>
> Are you interested in things a boat did?
> > *boat followed, boat capsized, boat sank, boat hit, boat had, …*
>
> Are you interested in things done with a boat?
> > *rescue the boat, rock the boat, miss the boat, …*

This sort of "keyword in context" list provides more information about the meaning of the clarification question, and does it in a corpus-specific way. That is, on a different corpus, the clarification questions chosen might not be the same, and the sample phrases including the query word would also be different.

4.3 Running the clarification question

When the user selects one of the query questions, the system then transforms it into an appropriate query and provides a new list of matching documents. The InQuery query is a logical "or" of all patterns that correspond to the selected query. For example, for the question "Are you interested in things done to a boat?", the following query could be generated

Instance	Count	Examples
JJ NN/NP IN		
JJ party:NP IN	2864	Democratic Party of, Communist Party of
JJ trade:NN IN	675	free trade with, foreign trade with
JJ tax:NN IN	175	new tax of, residential tax of
JJ war:NN IN	779	nuclear war in, psychological war to
VB NN/NNS NN		
VB tax:NN NN	114	improve tax collection, ensure tax fairness
VB president:NP NP	563	assassinate President Reina, meet President Husni
VB party:NN NN	149	strengthen party membership, promote party style
VB government:NN NNS	190	attend government sessions, include government officials
TO VB NN/NNS		
TO trade:VB NN	103	to trade protectionism, to trade fleece
TO finance:VB NN	110	to finance income, to finance pension
TO transport:VB NNS	36	to transport laborers, to transport weapons
TO address:VB NNS	32	to address rallies, to address imbalances
NNS/NN/NP NNS/NN/NP VB/VBD/VBG/VBZ		
NP citizens:NNS VBG	53	US citizens residing, African citizens living
NP ticket:NN VBZ	3	Samper-De la Calle ticket clears
NP clinton:NP VBD	647	Bill Clinton began, Bill Clinton met, President Clinton expressed
NP airlines:NP VBD	9	World Airlines said, Lufthansa Airlines opened

Figure 2: Some patterns that occur frequently in the trec5-fbis corpus, and selected instances of the patterns with a query word. The instances list the pattern with the query word inserted, the number of times that pattern occurs in the 80,000 document subset of the corpus, and then one or two examples of the pattern in the text.

```
#or( #1( #0(boat NN) BER verb )
     #1( #0(boat NN) NV BEN verb )
     #1( HVZ VBN #0(boat NN) )
     … )
```

In practice, the patterns generally need to be relaxed slightly to find matches that are very close to the patterns. We have observed that allowing a few extra words improves the recall of useful documents, though we have not carried out a formal study of that effect. The above query is transformed into,

```
#or( #3( #0(boat NN) BER verb )
     #3( #0(boat NN) NV BEN verb )
     #3( HVZ VBN #0(boat NN) )
     … )
```

5. EVALUATION

We are interested in whether the clarifying questions result in a more focused set of documents in response to the searcher's query. One way to measure that would be to ask someone to judge each set of returned documents to decide whether they were more or less focused than the original set. Such a process is time consuming and unwieldy.

Instead, we will use a measure of query clarity based on vocabulary distribution in both sets of retrieved documents [8, 9]. If one query has more "clarity" than another, it is less ambiguous, and perhaps more useful for retrieval.

5.1 Query clarity

Clarity is defined [8] as the Kullback-Liebler (KL) divergence between the collection and the query:

$$\text{clarity}(Q) \equiv \sum_{w \in V} P(w|Q) \log_2 \frac{P(w|Q)}{P(w)}$$

where V is the set of all terms in the collection (e.g., trec5-fbis or tdt2), Q represents the query, and $P(w)$ represents the probability of the word occurring in a document in the collection. To estimate the probability distribution of words given the query (the "query language model"), we use,

$$P(w|Q) \cong \sum_{d \in R} P(w|d)P(d|Q)$$

where R is the set of documents retrieved in response to the query, and $P(w|d) = \lambda P_{ML}(w|d) + (1 - \lambda)P(w)$, a linear combination of the corpus probability and the maximum likelihood estimate based on the document. We set $\lambda = 0.6$.

What the clarity measure does is compare the language model (probability distribution of words) of the corpus with the language model generated by a query. The way it is calculated means that if the distributions are identical, the clarity will be zero, and as they become more and more different, the value will rise. As a result, if the set of retrieved documents has roughly the same coverage as the entire corpus, the query that generated them will have a low clarity value—i.e., it is an ambiguous query.

It has also been shown that there is a strong correlation between clarity and the performance of the query as measured by average precision [10], so it is possible to predict, to some extent, the performance of a query on a collection without relevance information.

5.2 Using clarity

To determine whether the clarification questions are providing any focusing of the results, we compare the clarity of the original query word with that of the modified query. Each query Q_i has a set of clarifying questions $Q_{i,1}$ through

Q_{i,n_Q} where n_Q varies by query and indicates the number of questions whose patterns occurred more than 50 times in the corpus.

We compute the following measures to compare the effectiveness of the system.

- The number of times that the clarification question is clearer than the original query.

$$\sum_{j=1}^{n_Q} I(\text{clarity}(Q_{i,j}) > \text{clarity}(Q_i))$$

- The increase in clarity moving from the original query to the clarification question. If the number is negative or near zero, it means that the questions were not an improvement.

$$\sum_{j=1}^{n_Q} \text{clarity}(Q_{i,j}) - \text{clarity}(Q_i)$$

In addition, both measures are averaged across all clarification questions (per-question average, also called a pooled average), and weighted by the original queries (per-query average). The latter is important because n_Q can vary dramatically between queries and we do not want queries with few clarification possibilities to be overshadowed by the others.

6. RESULTS

To evaluate the disambiguation potential of this technique, we choose 50 one-word queries using a mix of several different ways:

- We used 25 common one word queries from web logs. This we found from multiple sources as follows: (1) words that were listed as being common single word queries using a Word Tracker[3] report that lists commonly used query terms mined from a large log of metasearch queries; (2) common one-word queries as reported by Google[4] and Metacrawler[5]. These words were all verified to occur within the trec5-fbis corpus.

- Since our database is a trec-5 collection, we analyzed the TREC-5 topics 251-300 that could be narrowed down to one word. We used 25 such queries. For example TREC topic 269 is "Foreign Trade," which we generalized to "trade".

We report below analysis of the clarity impact in the aggregate and then present detailed analysis of some of the query words.

6.1 Overall trends

Overall, the query clarity improves an average of 41% for TREC queries and 25% for Web queries (per-query average). The top of Figure 4 shows a breakdown of the improvement on a per-query basis for TREC queries and the bottom half shows the same for Web queries. Almost all queries have average improvements in clarity, and even when the average is negative, most of the clarification questions reduce ambiguity.
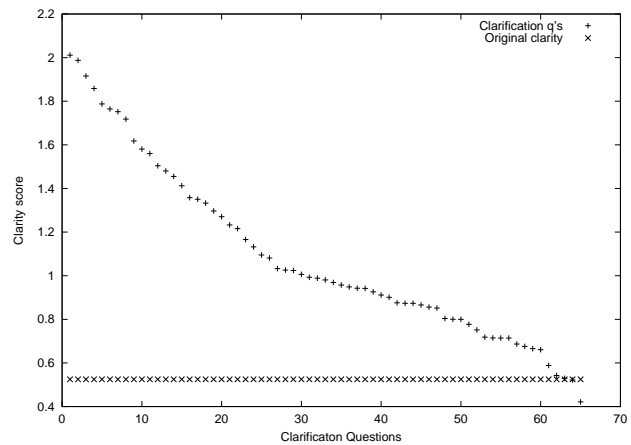

Figure 5: Graph shows the clarity values for the clarifying questions related to *tour*, the values shown in Figure 6. The values were sorted in order of clarity. The original clarity of *tour* is the solid horizontal line at 0.525

For example, for the original query word *defense* there were 85 patterns that occurred often enough to be recorded. The clarity of *defense* was 0.511 (not shown in the table). Of the 85 patterns, 71 of them (83.52%) showed an improvement in clarity, averaging 0.171 or 33% (the values ranged from -25% to 200%).

About 95% of the questions for the trec5 questions and 62% for the general web queries improved query clarity, strongly suggesting their value in reducing ambiguity. The reasonable similarity between the pooled and query-weighted averages suggests that the results are comparable across queries.

The percent improvement in clarity ranged from 347% to improve *mines* to *JJ mines:NNS NP* down to -51% trying to improve *heart* to *CD heart:NN CC* in the set of TREC questions. In the set of general Web queries this range is from 412% improving *apartment* to *AT apartment:NN VBN* to -62% for trying to improve *computer* to *NN computer:NN NNS*. It appears that the queries where ambiguity could be reduced the most were cases where the word could be either a noun or a verb (e.g., *tour, defense*) or where the word often appeared as a proper noun (e.g., *airline*).

If clarity numbers are calculated or approximated in advance, it would be possible to remove clarification questions that actually reduce clarity (i.e., increase ambiguity), making this technique even more useful.

6.2 Detailed analysis

The query word *tour* is not very common in the trec5-fbis corpus that we used: it occurs in 1668 out of the 178,000 stories. The word *tour* has a clarity value of 0.525. We show a partial breakdown of the results for *tour* clarification in Figure 6. The results are ordered by percent improvement in the original clarity value, ranging from a 283% improvement to a drop of almost 20%. Figure 5 shows a graph of the clarity values, with a straight line representing the original clarity of the query. It is clear that improvement is usually seen, and often quite large.

The pattern *JJ tourist:NN VBG* corresponds to phrases

TREC topic	Query word	Num. Quests	Number improved	Percent improved	Change in clarity	Percent improved
257	cigarette	15	5	33.33	-0.15	-9.69
282	crime	93	47	50.53	0.08	10.41
268	defense	85	71	83.52	0.17	33.47
288	diet	16	8	50.00	0.04	4.16
282,284	drug	95	78	82.10	0.18	23.05
255	environment	97	72	74.22	0.13	23.65
298	gun	80	32	40.00	-0.07	-6.13
275	health	82	46	56.09	0.07	7.73
254	heart	37	15	40.54	-0.03	-2.85
289	hospital	70	59	84.28	0.31	27.67
294	husbandry	16	15	93.75	0.51	86.22
279	magnetic	29	25	86.20	0.52	55.14
293	military	91	58	63.73	0.09	18.63
277	mines	51	50	98.03	0.67	129.65
258	security	88	67	76.13	0.13	26.91
252	smuggling	64	50	78.12	0.40	54.96
271	solar	29	25	86.20	0.31	35.51
285	submarine	64	63	98.43	0.57	90.30
287	surveillance	30	28	93.33	0.70	102.57
291	tax	97	42	43.29	-0.01	-1.48
269,283	trade	94	64	68.08	0.07	14.81
300	traffic	87	69	79.31	0.22	26.89
276	uniform	44	40	90.90	0.53	64.63
265	violence	73	53	72.60	0.20	22.60
292	welfare	46	33	71.73	0.32	51.80
Per-question averages		1165	1118	95.96	0.24	35.68
Per-query averages		46	45	97.8	0.27	40.96
	airlines	36	19	52.77	0.05	4.15
	apartment	45	30	66.66	0.18	17.27
	attorney	29	20	68.96	0.24	21.80
	author	92	32	34.78	-0.04	-5.14
	biotech	24	24	100.00	0.55	76.012
	birthday	16	12	75.00	0.26	22.95
	cheap	44	40	90.90	0.48	64.65
	computer	75	11	14.66	-0.34	-24.50
	divorce	13	3	23.08	-0.22	-14.73
	employment	92	76	82.60	0.23	44.91
	fish	72	58	80.55	0.34	43.16
	free	96	66	68.75	0.17	26.56
	homes	23	2	8.69	-0.70	-34.10
	house	83	65	78.31	0.27	37.66
	international	94	82.00	87.23	0.29	74.17
	investing	29	22	75.86	0.29	49.72
	job	78	34	43.58	0.04	5.064
	loan	82	36	43.90	0.02	2.59
	marijuana	37	37	100.00	0.53	54.64
	master	38	21	55.26	0.13	13.60
	software	50	16	32.00	-0.18	-15.19
	tour	65	63	96.92	0.59	113.60
	travel	67	31	46.26	0.02	1.63
	vacation	23	19	82.60	0.22	23.02
	vietnam	45	20	44.44	-0.09	-9.12
Per-question averages		1351	842	62.32	0.14	23.76
Per-query averages		54	33	61.10	0.14	25.43

Figure 4: Summary evaluation of change in clarity broken down by query and overall for TREC queries (top) and general Web queries (bottom).

Pattern	Count	Clarity	Percent Improvement
np tourists:nns cc	12	2.01	283.1
nn toured:vbd at	20	1.98	278.5
jj tour:nn np	34	1.91	264.8
np tourist:np np	23	1.85	253.9
np tourist:nn nn	24	1.78	240.4
cc tourism:nn np	27	1.76	236.0
jj tourists:nns cc	22	1.75	233.6
cc tourism:nn nn	39	1.71	227.2
jj tour:nn wdt	24	1.61	208.1
jj tour:nn in	188	1.56	197.1
nn tourism:nn nns	15	1.50	186.4
cd tourist:nn nns	14	1.48	181.9
at tour:nn nn	35	1.45	177.1
nn tourism:in nn	14	1.41	169.0
in tour:at nn	33	1.35	158.6
jj tours:nns in	13	1.35	157.2
cc tourism:nn at	21	1.33	153.8
in tourism:nn at	15	1.29	147.0
...			
...			
cc tourist:nn nns	40	0.52	0.8
jj tourists:nns vbg	19	0.52	-0.6
np tour:nn in	54	0.42	-19.7

Figure 6: Breakdown of changes in clarity for the patterns recorded for the query _tour_. The original clarity of that query was 0.525

Pattern	Count	Clarity	Percent Improvement
in defense:np bez	27	2.01	302.3
np defense:np in	56	1.98	297.4
punc defense:np np	147	1.91	283.1
at defense:nn cc	84	1.85	271.6
at defense:np in	109	1.78	257.5
in defense:np rb	32	0.99	98.5
in defense:nn at	49	0.98	96.1
nns defense:nn np	39	0.96	93.8
in defense:np cc	192	0.95	91.4
np defense:np np	3978	0.94	89.7
in defense:nn nns	272	0.94	88.5
nn defense:np np	137	0.94	88.3
nn defense:cc nn	59	0.90	80.2
at defense:nn nn	1470	0.87	75.2
in defense:np nn	47	0.87	74.7
np defense:np nn	46	0.86	73.2
cc defense:np np	265	0.85	71.2
nn defense:np	31	0.85	70.3
nn defense:in nn	84	0.80	60.8
...			
...			
toin defense:np np	55	0.52	4.4
in defense:nn in	373	0.42	-15.7

Figure 7: Breakdown of changes in clarity for the patterns recorded for the query _defense_. The original clarity of that query was 0.511

such as "typical tourist visiting" and does not provide much focusing of the query. On the other hand, the pattern _NP tourist:NP NP_ represents phrases such as "National Tourist Board" and the documents in which such phrases are present focus more on information, announcements, etc., for tourists, providing a fairly tight set of retrieved documents.

Figure 7 shows a similar effect for the query word _defense_, the word with a high improvement of clarity in the TREC query set. Note that _defense_ also occurs about as often as _tour_: 1199 times in trec5-fbis. The clearest clarification in this case is "IN defense:NP BEZ " which corresponds to "[Department/Ministry] of Defense is" and occurs 27 times. The clarification "CD defense:NN CC" corresponding to phrases such as "[team will use] a 52 defense and" which although it occurs an equal number of times, does not show as great an improvement in clarity. In the first case almost all documents are military related, whereas _defense_ as a common noun could appear in any context—viz. defense, sports, law, etc. There is no apparent relation between the number of occurrences of a pattern and clarity.

7. CONCLUSION AND FUTURE WORK

We have shown a method for extracting patterns of word usage from a corpus and for using those patterns to help the user clarity an ambiguous query. The patterns represent sequences of parts of speech that occur around the query word often enough in the corpus that they are likely to be meaningful. Frequent patterns are mapped to human-generated clarification questions that the user may choose from. Note that we envision this as an optional side-bar to the return of query results—there seems little value in requiring the searcher to clarify the question if his or her desired document is already in the top ranks.

Once the user has selected a clarification question, we can re-issue the query based on the part of speech patterns, though slightly relaxed to increase recall. A list of sample patterns is generally helpful for the user interpreting the question.

We used a statistical measure called query clarity [8] to demonstrate that the clarification questions generally provide a much more focused set of documents.

We are working on extending these ideas to build a better set of clarification questions for the patterns and to determine ways for automatically constructing such questions (because it is tedious to do so by hand). In Figure 3 it is apparent that certain clarification questions have multiple patterns associated with them. We are studying methods by which these patterns for a given question can be learned using training data. We are also building a system that implements these ideas to field test an interface to find out what works best in live settings.

Acknowledgments

8. REFERENCES

[1] A.Leuski. Evaluating document clustering for interactive information retrieval. In *Proceedings of the ACM CIKM 2001 Tenth International Conference on Information and Knowledge Management*, pages 33–40, 2001.

[2] A.Leuski. *Interactive Information Organization: Techniques and Evaluation*. PhD thesis, University of Massachusetts, Amherst, 2001.

[3] Peter G. Anick and Suresh Tipirneni. The paraphrase search assistant: Terminological feedback for iterative information seeking. In *Proceedings of the ACM Conference on Research in Information Retrieval (SIGIR)*, pages 153–159, 1999.

[4] P.D. Bruza and S. Dennis. Query-reformulation on the internet: empirical data and the hyperindex search engine. In *Proceedings of the RIAO Conference: Intelligent Text and Image Handling*, pages 488–499, Montreal, Canada, 1997.

[5] Peter Bruza, Robert McArthur, and Simon Dennis. Interactive internet search: Keyword, directory and query reformulation mechanisms compared. In *Proceedings of the ACM Conference on Research in Information Retrieval (SIGIR)*, pages 280–287, 2000.

[6] James P. Callan, W. Bruce Croft, and Stephen M. Harding. The INQUERY retrieval system. In *Proceedings of DEXA-92, 3rd International Conference on Database and Expert Systems Applications*, pages 78–83, 1992.

[7] K.W. Church, W.Gale, P.Hanks, and D.Hindle. Using statistics in lexical analysis. In *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pages 115–164, 1991.

[8] W.B. Croft, S. Cronen Townsend, and V.Lavrenko. Relevance feedback and personalization: A language modeling perspective. In *Proceedings of the DELOS-NSF Worshop on Personalization and Recommender Systems in Digital Libraries*, pages 49–54, 2001.

[9] Steve Cronen-Townsend and W. Bruce Croft. Quantifying query ambiguity. In *Proceedings of the Conference on Human Language Technology (HLT 2002)*, pages 94–98, 2002.

[10] Steve Cronen-Townsend, Yun Zhou, and W. Bruce Croft. Predicting query performance. In *Proceedings of the ACM Conference on Research in Information Retrieval (SIGIR)*, 2002.

[11] D.Lawrie, W.B. Croft, and A.Rosenberg. Finding topic words for hierarchical summarization. In *Proceedings of the ACM Conference on Research in Information Retrieval (SIGIR)*, pages 349–357, 2001.

[12] G.Grefenstette. SQLET: Short query linguistic expansion techniques: Palliating one or two-word queries by providing intermediate structure to WWW pages. In *Proceedings of RIAO '97. (1997)*, 1997.

[13] Marti Hearst and Jan Pedersen. Reexamining the cluster hypothesis: Scatter/gather on retrieval results. In *Proceedings of the ACM Conference on Research in Information Retrieval (SIGIR)*, pages 76–84, 1996.

[14] J.Allan, A.Leuski, R. Swan, and D.Byrd. Evaluating combinations of ranked lists and visualizations of inter-document similarity. In *Information Processing and Management (IPM), vol. 37*, pages 435–458, 2000.

[15] K.Franzen and J.Karlgren. Verbosity and interface design. Technical Report T2000:04, Swedish Institute of Computer Science (SICS), 2000. Also available at http://citeseer.nj.nec.com/313985.html.

[16] R. Krovetz. *Word Sense Disambiguation for Large Text Databases*. PhD thesis, University of Massachusetts, Amherst, 1995.

[17] Dawn Lawrie and W.B. Croft. Discovering and comparing topic hierarchies. In *Proceedings of RIAO 2000 Conference*, pages 314–330, 2000.

[18] Anton Leuski and James Allan. Lighthouse: Showing the way to relevant information. In *Proceedings of the IEEE Symposium on Information Visualization 2000 (InfoVis 2000)*, pages 125–130, 2000.

[19] M.Sanderson and D.Lawrie. Building, testing, and applying concept hierarchies. In W. Bruce Croft, editor, *Advances in Information Retrieval: Recent Research from the CIIR, W. Bruce Croft, ed., Kluwer Academic Publishers*, chapter 9, pages 235–266. Kluwer Academic Press, 2000.

[20] Mark Sanderson. Word sense disambiguation and information retrieval. In *Proceedings of the ACM Conference on Research in Information Retrieval (SIGIR)*, pages 142–151, 1994.

[21] Mark Sanderson and W. Bruce Croft. Deriving concept hierarchies from text. In *Proceedings of the ACM Conference on Research in Information Retrieval (SIGIR)*, pages 206–213, 1999.

[22] Richard Seltzer. Taking "snapshots" with livetopics: Watch out for mirages, 1997. http://www.samizdat.com/cobb/oct97.html.

[23] Russell C. Swan and James Allan. Aspect windows, 3-d visualizations, and indirect comparisons of information retrieval systems. In *Proceedings of the ACM Conference on Research in Information Retrieval (SIGIR)*, pages 173–181, 1998.

[24] J. Xu, J. Broglio, and W.B. Croft. The design and implementation of a part of speech tagger for english. Technical Report IR-52, University of Massachusetts, Amherst, Center for Intelligent Information Retrieval, 1994.