DEIM Forum 2012 D2-4

質問応答の連結によるウェブクエリの意図表現

尹 聖雄† アダム ヤトフト† 田中 克己†

† 京都大学 大学院 情報学研究科 〒 606-8501 京都市左京区吉田本町
E-mail: †{yoon,adam,tanaka}@dl.kuis.kyoto-u.ac.jp

Connecting QAs to Show Intent underlying Web Query

Soungwoong YOON†, Adam JATOWT†, and Katsumi TANAKA†

† School of Informatics, Kyoto University Yoshida honmachi, Sakyo, Kyoto 606-8501 Japan
E-mail: †{yoon,adam,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract People want to *ask* something to find relevant *answer* following their information needs. Discovering intents behind Web query is helpful for users to find useful Web search results or recognize search objectives. In this paper, we try to discover structured intents of Web query using Community Question-Answer (CQA) information following the assumption that QAs concerned with query keywords represent practical intents of the query. Through syntactic and structural analysis of CQA contents, we extract candidate features of intent which form intent expressions and measure relations of answers toward questions by using QA similarities and differences. Finally we connect QAs using intent expressions and answer - question relations. Experimental results show that QA connections are useful as fragments of knowledge to understand Web query intents and their changes.

Key words Intent, Query Intent, Community Question-Answer

1. Introduction

Web is the biggest repository of information but still not well-arranged knowledge source. Users' information needs are important to find relevant Web information. We simply define these information needs in Web search environment as *intent*, which means that discovering intents will be a premise to satisfy users' needs. However it is difficult to designate intent precisely for information providers as well as the users themselves, because they may not possess sufficient knowledge to imagine and conceptualize search objective. Moreover, the intent may change following search environment [13], [18].

Finding intent in Web query has been an active research topic for the last years with two main dimensions: query topic classification and query intent classification [3]. Query topic classification consists of identifying a query as belonging to one or more categories from a *predefined* taxonomy which contains from 12 [16] to 6,000 [5] entries, and query intent classification consists of identifying the underlying goal of the user when submitting one particular query by studying *query log* which is manually labeled as training data or ground-truth. Researchers characterized intent as specified categories or clusters which were found by analyzing user search history [4], [9], [10]. But these aggregation or extraction methods can provide ways to find general intent-driven clusters, rather than detect the actual user intents.(注1)

When a user starts searching by the query, her objective usually cannot be fulfilled at a time. Figure 1 shows an example of complex search activity representing search intent and its changes constructed based on question-answering(注2). Question 1 (Q1) covers the overall possibilities of intents, and its answer (A1) shows a possible search direction. Then A1 is connected naturally to other question Q2 to find the precise information. With the same manner, there are possible connections such as A2 → Q3 and A3 → Q4, though there may be no term overlaps such as A2 and Q3. (注3)

(注1)：There are 46.5% non-clicks in AOL 500K User Session Collection, which means that one should analyze users' clicks as well as non-clicks for estimating user satisfaction in search activity.
(注2)：Note that this is an output of our experiment.
(注3)：We assume that the question is semantically connected with its

Q1. I have tickets and a plan to visit Tokyo, Osaka and Kyoto in June. Is it safe and OK to travel?

A1 : Sure. You can travel that area with JR rail pass.

Q2. Japan rail pass and Kyoto travel... Please help?

A2 : You can access to JR Kyoto station using one train.

Q3 : Where should I stay if I am traveling Tokyo, Osaka and Kyoto?

A3 : Check this hotel site. Also they guide must-see things.

Q4 : What are some must-see things in Tokyo & Kyoto?

A4 : Check this blog. You can see Nijo Castle in Kyoto.

図 1 Example of search intent changes by the query 'kyoto travel'

These connections will be useful for a user who wants to visit Kyoto, and also others who already knew something but want to verify their plans. Each phase represents an intent, but the whole sequence traces human's search activity and implies the intent and its changes about 'travel' concerned with 'Kyoto.'

In this paper, we attempt to discover search sequences following intent through finding their features using Community Question-Answer (CQA) corpus. We assume that a question-answer pair (QA) in CQA represents a practical intent, and then extract candidate features of intent in QAs through syntactic and structural analysis. These features are: a) combined to find expressions of intent for setting the starting point of QA connections, and b) used to assess the relation between answers and questions. Finally we form chain-shaped QA connections by using intent expressions and answer - question relations. Experimental results show that our method can reveal possible search activities to designate their intents with around 50% accuracy.

A novel aspect of our work is that we propose the methodology of using social knowledge towards discovering search intents by the query. To certain extent, our task is to predict the most probable or typical phases of intent that a user may follow. The contributions are: 1) We try to supplement human judgments by using CQA knowledge, 2) Our method is lightweight and covers social information such as the one found in CQA to find practical intents, and 3) We show experimental results that prove the effectiveness of our method.

The remainder of the paper is organized as follows. Section 2 discusses the related research. We describe the methodology to discover QA connections based on query intent in Section 3. Experimental results and discussions are shown in Section 4. Finally, we conclude the paper in the last section and discuss our future work.

answer(s).

2. Related Research

2.1 Intent Discovery

Research on search intent discovery originated from the analysis of click-through data and query-intent categorization. Following the well-known query classification first proposed by Broder [4], Jansen et al. [9] stated that user intent can be categorized into three general intent classes: navigational, transactional and informational. The characteristics of user intent have been conventionally defined by analyzing click-through data [4], [9], [10]. Using large amounts of data containing evidence of user search-related activities made it possible to not only understand user needs, but also to depict user behavior in browsing Web search results [6] or support non-informational search intent on the Web [11].

However, the usefulness of these categorizations is limited by the data sets used and the efficiency of post-processing. There is a risk of the over-generalization being reflected in mismatches in classification between automatic and manual categorization. This is because the above researches were based on their own rigid classification schemes, and biased by the data sets they used. Furthermore, the previously proposed methods often have failed to represent the actual user intent as the scope of possible intents may simply be too large and too heterogeneous to be accurately reflected in any fixed taxonomy. Moreover, one should realize that the information needs of Web users are constantly changing and so does the Web itself.

In this work, we try to find the intent itself to overcome these broad-categorization problems. Both user-specific and statistically dominant viewpoints as well as direct intent expression can help users to choose their intent within possible suggestions, which can cover up their laziness when finding information on the Web.

2.2 CQA Analysis

The history of Social Network such as blogs, social annotations and social tagging, and CQA is just several years, but enormous development of Web technologies and needs to communicate together conjoins to the importance of one-person media and knowledge sharing community such as social search. Besides of this simple structure of information sharing community, there are semi-structured social knowledge such as CQA and ODP[注4], Wikipedia[注5], which is flex-

(注4) : http://www.dmoz.org
(注5) : http://www.wikipedia.org

ible compared to structured knowledge such as WordNet(注6).

CQA is one of significant knowledge bases created by people. People can post their questions and answers, and, moreover, analyze these interactions and vote its quality. Once these interactions are set, SE also uses CQA information to enhance user convenience.

There were studies concerned with CQA [1], [7], [8], [12], [15], [17] mainly about the structure of CQA or types of user interactions. In this paper, we point out the semantic structure of questions and answers in CQA to extract useful information to express human thinking.

3. Connecting QAs by Intent of Web Query

3.1 Assumptions

Our basic motivation is that intent behind a simple query can be shown by certain words and phrases. Our work is trying to back-track the procedure to find intent expressions from the query.

Assumption 1. Questions concerned with query keywords contain possible expressions of query intent.

Assumption 2. A question is semantically connected with its answer(s).

Following Assumption 1, list of questions and its answers, $\mathbf{QA} = \{\{Q_i, \{A_i^j\}\}\}$, is defined as CQA contents matched with a given query. For simplicity, we set $\mathbf{QA}$, $|\mathbf{QA}| = n$, as top n QAs retrived for the query, $n > 0$. Note that the number of questions in $\mathbf{QA}$ is n, but number of answers are at least n because there may be more than one answer in a question(注7). We call the answers which are posted for the question, $\{A_i^j\}$ as the answer set of Q_i.

3.2 Extracting Intent Features

Intent features are defined in this paper as words that directly lead to understand the intent behind the query, variations or conceptualizations of possible intents.

Figure 2 demonstrates the concept of finding intent features. The query 'kyoto travel' has possible intent expressions such as 'what is the cheapest way to go to Kyoto?', 'good hotel in Kyoto' and many others. Using these expressions, we can detect explicit features which should be useful to represent possible intents such as 'cheap(est) way,' 'Kyoto' or 'good hotel', and implicit features which support intent such as 'travel.' Our objective is finding these features and

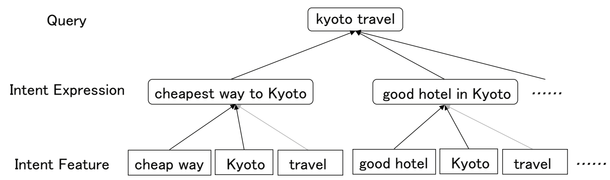

図 2 Concept of intent features

showing their usefulness in relation to concrete intent.

So we can say that intent features cover not only related concepts and co-occurred terms, but also terms which have semantic concern. Although recognizing these features is comparably easy for human, it is difficult for machines because the search engine cannot measure the degree of intent. In example of Figure 2, the given two expressions clearly support the intent '(how to) travel to Kyoto,' but simultaneously they can be also divided into two sub-intents: 'flight ticket to Kyoto' and 'hotel in Kyoto', even though the query 'kyoto travel' is the same. Our objective is to find ways to discover these useful intents through discovering their features by harnessing knowledge accumulated on the Web without human supervision.

3.2.1 Extracting Intent Features in questions

First, we need to detect important words of questions. Yoon et al. [18], [19] showed several methods to detect intent features in questions.

There are useful clues to find intent under the questions using data as follows.

- Query words: The explicit hint to motivate intent.
- Words which are indexed by SE and shown in Web search results of the query: These words are shown as bold character within search results. Mostly same with query terms, but different words such as abbreviations, synonyms and spell corrections are also shown which can supplement the query semantics. We call these terms as *coordinated terms*.
- Headwords [14], [18] in $\mathbf{QA}$: Each headword is regarded as the semantic core of a question phrase.

For finding intent feature sets of questions, $\mathbf{FQ} = \{FQ_i\}$, we first calculate TFIDF of $\mathbf{QA}$ and choose top k terms as candidate intent features of questions. Then we check each Q_i to find candidate features. Terms found according to this procedure are contained with i^{th} question intent feature set FQ_i.

Next, we input headword of question phrase and collect coordinated terms into FQ_i.

3.2.2 Extracting Intent Features in answers

Methods for collecting $\mathbf{FQ}$ however are not useful for extracting intent features in answers, $\mathbf{FA} = \{FA_i\}$, because 1)

(注6) : WordNet, a lexical database for English language. Princeton University, http://wordnet.princeton.edu

(注7) : Though there are not only resolved questions but also open / undecided questions in CQA, we collect QAs which suffice this condition.

Answer phrases are usually larger than a sentence, comparably longer than question sentence, and have no structural significance, 2) Themes in answers may be diverse following question intents and noisy terms may be included. For collecting **FA**, we have two scenarios of answering to a question in CQA following the observation of answer sentences.

- Similar answers: All answers of the question have similar contents, especially with the best answer. This condition is from the best answer contains the most important information, and other answers supplement and/or comment it.
- Different answers: Answers are different, mainly with the best answer. This case shows that the best answer was chosen as the alternative by the questioner, and other answers reflect different options or opinions.

In both cases, impact of features within an answer is judged by frequent terms within the answer set, and also unique terms in each answer. We assume that frequent terms point out the *objects* which are commonly described in answer set, and unique terms are *descriptions* of objects. Following this observation, we calculate document frequency (DF) of terms within answer set, and then choose top k DF terms and k least frequent DF terms which are shown in j^{th} answer of i^{th} answer set as FA_i^j, $j \leqq 0$.

3.3 Finding Related Question of Answer

Now QA_i have features FQ_i and FA_i, and an answer is needed to be related with other questions to show connections of QAs. The simplest way of these connections are using textual similarities, but we need to regard the difference between QAs too. There are two factors when measuring the relation between an answer as the source and a question as the target.

- Overlaps of intent features: This factor directly shows the semantic overlap between two contents.
- Overall similarity: This factor indicates the syntactic similarity between two contents.

Following assumption 2, a question is already connected with its answer(s). The problem is finding other question related with an answer.(注8)

We use both similarity and difference between QAs to measure answer-question relations, shown in Figure 3. First, we check the overlaps of intent features between the answer and all other questions in **QA** (line 01-04). If there are overlaps, we calculate the cosine *similarity* between two contents and sum two scores as relation score of i^{th} question toward j^{th} answer (line 05-08). Next, if relation score of m^{th} question is higher than threshold, we calculate the dissimilarity between m^{th} question and the question of j^{th} answer, and choose the most dissimilar question as the related question of j^{th} answer (line 09-12).

Input: QA list **QA**, Source answer $A[j]$
Output: Target question $Q[m]$

```
01 Foreach (QA[i] in QA)
02  If (i != j)
03   ScoreFeature[i]
       = IntentFeature Overlap ratio (Q[i], A[j])
04   If (ScoreFeature[i] > 0)
05    ScoreASim[i]
        = ScoreFeature[i] + CosSim(Q[i], A[j])
06   EndIf
07  EndIf
08 EndForeach
09 Foreach (QA[m] When ScoreASim[m] > t)
10  ScoreQSim[m] = 1 - CosSim(Q[m], Q[j])
11 EndForeach
12 Q[m] = Q[m] which has Max(ScoreQSim)
```

図 3 Pseudo code for calculating answer-question relation

This procedure is done for every answers in **QA**. Note that sometimes there are answers which have no related questions.

After this procedure, bipartite connection of answers and questions in **QA** is formed.

3.4 Connecting QAs

3.4.1 Intent Expressions by combining Features

For connecting QAs, we need to estimate the starting point of intent. Terms in the query and features extracted from **QA** are possible clues, but combining their importance is a difficult problem. As a solution, we arbitrarily combine features and select the existing patterns in **QA**, shown in Figure 4.

First, we combine all possible two-word pairs of intent features (line 01). This idea was originally taken from [2], but we simply use all possible two-word pairs. Then we check these pairs' inclusions in each QA sentences. If number of terms between two terms of a feature pair is less than l excluding stopwords(注9), we list up this feature pair as a possible intent expression (line 02-08). After collecting possible expressions, finally we merge similar phrases to generate set of intent expressions (line 09-17).

For example, in a question phrase 'Where is good hotel in Kyoto?' obtained by the query 'kyoto travel', intent features are 'kyoto', 'travel' and 'good hotel'. Then two-word combinations of features are 'kyoto, travel', 'kyoto, good hotel' and

(注8): Note that the objective for extracting this answer-question relation is for detecting the 'next' phase which contains similar intent.

(注9): http://armandbrahaj.blog.al/2009/04/14/list-of-english-stop-words/

Input: Intent features $\mathbf{F} = \{\mathbf{FQ}, \mathbf{FA}\}$, QA list $\mathbf{QA}$

Output: Intent expression set $\mathbf{E}$

```
2WordCombinations[] = Combination(F, 2)
Foreach ([w1, w2] in 2WordCombinations[])
 Foreach (question & answer sentence in QA)
  If (Words between w1 and w2 are less than l)
   Add Substring(sentence from w1 and w2) to E
  EndIf
 EndForeach
EndForeach
For (e[i] in E)
 For (e[j] in E)
  If (Overlap(e[i],e[j]) > 0 AND i != j)
   Add Merge(e[i],e[j]) to E
   Erase e[i], e[j] from E
  EndIf
 EndFor
EndFor
Return E
```

図 4 Pseudo code for generating intent expressions

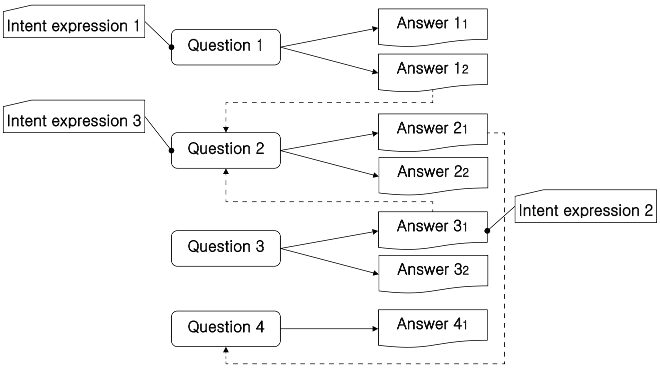

図 5 Example of intent expressions and related questions

'travel, good hotel', and the intent expression in the given question is 'good hotel in Kyoto'.

3.4.2 Connecting QAs by Intent Expressions and Related Question of Answer

Using intent expressions, question(s) which meet with an intent expression are initiated as starting point of QA connection. We connect this question with its all answers following assumption 2. If an answer phrase has intent expressions, we simply one-step go backward to its question as an initial point of QA connection. Then we have the answer-question relations measured in Section 3.3.

Figure 5 shows the example of all connections in **QA**, where solid lines indicate explicit connections from questions to its answers and dashed lines indicate the answer-question relations.

From the initial points by intent expressions, we connect a QA to other QA using answer-question relation. As a result, the tree-shape QA connections are extracted. Next, we join subtrees to organize the chain-shape QA connections from trees by choosing the longest possible connection to generate chains. In former example, a QA connection is $Q1 \rightarrow A1_2 \rightarrow Q2 \rightarrow A2_1 \rightarrow Q4 \rightarrow A4_1$, and other is $Q3 \rightarrow A3_1 \rightarrow Q2 \rightarrow A2_1 \rightarrow Q4 \rightarrow A4_1$.[注10]

表 1 Test queries

General query	Practical query
google	google maps, google maps navigation
yahoo	yahoo mail, yahoo answers API
ebay	ebay business, ebay shipping fee
internet	internet provider, craigslist
microsoft	msn, windows live hotmail
myspace	myspace layouts
dictionary	english dictionary, electronic dictionary
travel	southwest airlines, kyoto travel
paul krugman	paul krugman newspaper, krugman academic speciality
japan	japan earthquake

4. Experiments

4.1 Test Queries & Evaluation Criteria

To verify the QA connections following intent of the query, we need test queries for collecting CQA information. Intuitively, queries can be characterized by the concreteness of their intent. For example, the query 'travel' may contain more general intent than 'kyoto travel'.

For covering this difference, we first choose 10 queries which have general intents from two different sources[注11], and 18 practical queries are manually generated from them. Table 1 shows the test queries.

For each query, we collect 50 QAs using Yahoo! Answers API[注12] in April 2011. The constraints for the experiments are set as: k (usage of DF terms) is 3, l (number of terms between combined two features) is 5, and answer-question similarity threshold is 0.5. We vary the number of **QA**, n from 10 to 50 and evaluate the results because n is important factor to judge the number and quality of intent evolution patterns. Intuitively, the more QAs are used, the more information will be grabbed but the more noisy terms are also extracted.

(注10)：Though there is overlap between two patterns, we regard those patterns as different ones. We will discuss the importance of patterns following its support ratio by seed features in further research.

(注11)：AOL 500K User Session Collection (http://www. gregsadetsky .com/aol-data/) and NIST TREC QA track 2007 (http://trec.nist.gov /data/qa/2007_qadata.html)

(注12)：http://developer.yahoo.com/answers

表 2 Precision and portions of QA connections following three intent types

# QAs	Precision	S (%)	G (%)	P (%)
10	0.377	0.075 (20)	0.018 (5)	0.283 (75)
20	0.492	0.100 (20)	0.031 (6)	0.361 (74)
30	0.491	0.110 (22)	0.031 (6)	0.350 (72)
40	0.443	0.053 (12)	0.042 (9)	0.348 (79)
50	0.442	0.070 (16)	0.034 (8)	0.338 (76)

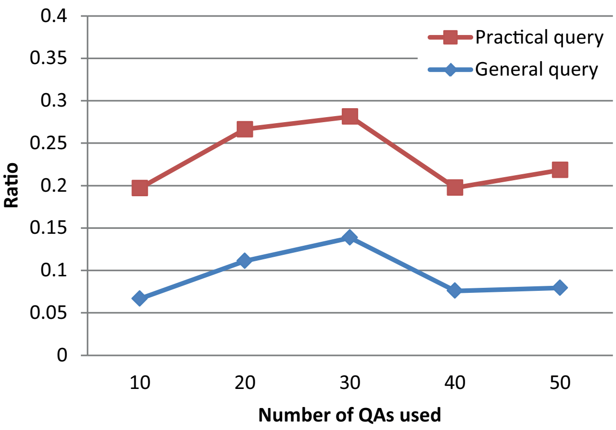

図 6 Average ratio of Specification / Generalization patterns

For assessing the quality of QA connections, we set three types of possible QA connection patterns.

- Specification (S): Precise description or solution of given problem following general expression, which is matched with the motivation.

- Generalization (G): If there is no answer with practical problem, generalizing idea to concerned topics will be useful to set the search goal. This pattern is also matched with our motivation because the reverse order of this pattern will be the same with S pattern, though separating G and S pattern is still problematic. In many cases, S and G patterns are shown in a QA connection.

- Parallel shifting (P): Enumerating similar idea will be shifting topics, which are also frequent cases. Thinking about another problem or shifting topics of given problem will be possible way to find intent.

Our assumption is that the general queries contain various directions of intent compared to practical queries, thus the precision of intended QA connections and number of QAs within a QA connection in practical query cases are larger than those of general query cases.

4. 2 Evaluation

Table 2 shows the precision scores of found QA connections and proportions of QA connections following intent types. The precisions of intended QA connections are not significantly changed following the increment of QAs used (n), and maximum 0.492 when n=20. Table 3 shows examples of QA connections by intent type.

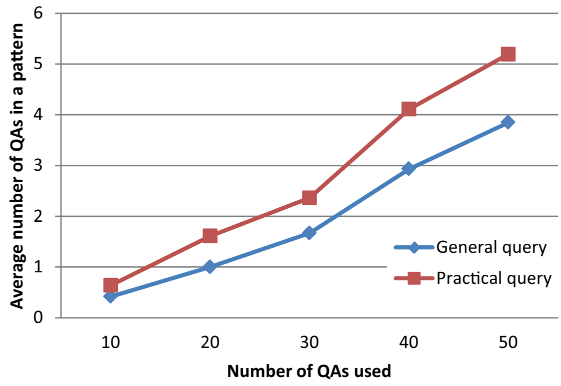

図 7 Average number of features in a pattern

Next, we evaluate the effect of query characteristics. Figure 6 shows the portion of S, G patterns. In all cases, more S and G patterned QA connections are found when using practical queries than cases of general queries, which supports our hypothesis.

Figure 7 shows the number of QAs in intended QA connections. Compared to the case of general queries, QA connections of practical queries contain more QAs within a QA connection, which implies that there are more stages of thinking in practical query.

4. 3 Discussion

In this section, we list some discussion points.

（1） Truly *asking* queries: The quantity and characteristics of QAs mainly depend on the query, but it is different in case of QA connections. If the query is semantically concrete, detecting dominant intent of query will be comparably easy because CQA users may have similar interests of that query. Although showing staged patterns can be useful to enumerate user's choices, there are queries which are fit to our analysis like the query categorization approach proposed by [9].

（2） Quality of CQA contents: Our assumption is that the query contains intent(s), but the quality of CQA contents by the query will be different problem. Our filtering process such as choosing top n QAs, finding intent expressions by using random combination of intent features can decrease the noise and partially complement the problem.

（3） Finding highly probable QA connections: We enumerate the QA connections and check each connection's usefulness, but their probability toward user intent will be different matter. Unfortunately, frequent QA connections look not always important.

（4） Ontological background knowledge: Typical se-

表 3 Example of intent evolution patterns by type

Pattern type	Query	Example
Specification	google maps navigation	Q33. Navigation for my MotoCliq..? A33. Make sure you have the latest update of android, google maps should... ↪Q14. Navigation using offline google maps and usb gps? A14. Google maps is not “downloadable” that’s your biggest problem. ↪Q2. How can i get google maps navigation on the iphone even though it is only released on android? A2-2. You don’t. However, supposedly Mapquest is working on an app for the iPhone that offers voice nav like the Google Maps for Android... ↪Q20. Does the Google Maps app on iPhone have voice navigation? A20. No, it doesnt. Mapquest does though.
	ebay business	Q31. How do I begin registering my eBay business with the state. government, etc? A31. Here is a state by state guide to opening a business ↪Q47. Will I be charged federal and/or state tax on my Ebay business inventory? A47. There is no federal tax on inventory. If you ... ↪Q26. How much would it cost for a tax preparer with the new 1099K form for my ebay business? A26. ...You are required to file a tax return if your net icnome is more than $400. ...
Generalization	google maps navigation	Q27. Does the palm pre plus support google maps? A27. Yes. the mobile version of google maps is supported on the Palm OS. You can download it... ↪Q32. Google Map free turn by turn GPS navigation Help? A32. Same problem. ↪Q30. What are the current digital map navigation problems? A30. An application that would be nice is a nautical route plotter...
	craigslist	Q20. How can I find out if a Craigslist apartment rental is for real? A20-2. ...No ”real” landlord will accept cash on the first meeting... ↪Q28. What kind of things would you buy from craigslist? A28. ...You can’t pay full price for but are still good used items. Then try Ebay.
Parallel shifting	google maps navigation	Q42. What are the best offline navigation apps for android (HTC Legend)? A42-2. I made a video with my opinion on the best offline Nav apps for android ... ↪Q18. Why won’t Google Navigation work on my Verizon Samsung Fascinate android phone? A18. Verizon removed the google navigation, ... ↪Q20. Does the Google Maps app on iPhone have voice navigation? A20. No, it doesnt. Mapquest does though.
	msn	Q21. How do I retrieve a msn messenger contact i accidentally deleted? A21. ...The option ‘remove’ on ‘blocked list’ or ‘allowed list’ will be available... ↪Q26. How do you export your msn contacts list to a different email account? A26-2. Save your contact list then import on your new account... ↪Q30. What is a good alternative for msn messenger? A30. ... Our hotmail / live account > login > up on the web > click on MESSENGER > sign in, or try one of these messenger programs e.g. Yahoo messenger, skype ...

quences of activities or concepts such as 'hotel' - 'restaurant' - 'travel spot' in 'travel' intent could improve our method's performance. Useful knowledge source could be acquired from external corpus and/or user activity logs.

5. Conclusion & Future Directions

In this paper, we propose the method of connecting QAs to show intent underlying Web query. Based on the intent features extracted from CQA contents, we generate intent expressions and find related questions of given answers for organizing QA connections. Experimental results show that 49.2% of QA connections are valuable as the knowledge fragment of intent and observed characteristics are matched with our hypothesis, which means that we can extract clues of intent by using query and social knowledge e.g. CQA corpus.

Obviously, there are other factors which affect the intent behind the query. We need to validate them to reveal the true query intent. We showed the QA connections to express intent, but we need to show 'answers' of that intended information also in further research.

6. Acknowledgements

This research was supported in part by the Kyoto University Global COE Program: Informatics Education and Research Center for Knowledge-Circulating Society.

文 献

[1] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge sharing and yahoo answers: everyone knows something. In *Proceeding of the 17th international conference on World Wide Web*, WWW '08, pages 665–674, New York, NY, USA, 2008. ACM.

[2] D. Bollegala, Y. Matsuo, and M. Ishizuka. Measuring the similarity between implicit semantic relations using web search engines. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, pages 104–113, New York, NY, USA, 2009. ACM.

[3] D. J. Brenes, D. Gayo-Avello, and K. Pérez-González. Survey and evaluation of query intent detection methods. In *Proceedings of the 2009 workshop on Web Search Click Data*, WSCD '09, pages 1–7, New York, NY, USA, 2009. ACM.

[4] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36:3–10, September 2002.

[5] A. Z. Broder, M. Fontoura, E. Gabrilovich, A. Joshi, V. Josifovski, and T. Zhang. Robust classification of rare queries using web knowledge. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 231–238, New York, NY, USA, 2007. ACM.

[6] Q. Guo and E. Agichtein. Exploring mouse movements for inferring query intent. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 707–708, New York, NY, USA, 2008. ACM.

[7] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina. Questioning yahoo! answers. Technical Report 2007-35, Stanford InfoLab, 2007.

[8] U. Hermjakob. Parsing and question classification for question answering. In *Proceedings of the workshop on Open-domain question answering - Volume 12*, ODQA '01, pages 1–6, Stroudsburg, PA, USA, 2001. Association for Computational Linguistics.

[9] B. J. Jansen, D. L. Booth, and A. Spink. Determining the informational, navigational, and transactional intent of web queries. *Inf. Process. Manage.*, 44:1251–1266, May 2008.

[10] U. Lee, Z. Liu, and J. Cho. Automatic identification of user goals in web search. In *Proceedings of the 14th international conference on World Wide Web*, WWW '05, pages 391–400, New York, NY, USA, 2005. ACM.

[11] Y. Li, R. Krishnamurthy, S. Vaithyanathan, and H. V. Jagadish. Getting work done on the web: supporting transactional queries. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '06, pages 557–564, New York, NY, USA, 2006. ACM.

[12] Y. Liu, J. Bian, and E. Agichtein. Predicting information seeker satisfaction in community question answering. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 483–490, New York, NY, USA, 2008. ACM.

[13] G. Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49:41–46, April 2006.

[14] D. Metzler and W. B. Croft. Analysis of statistical question classification for fact-based questions. *Inf. Retr.*, 8:481–504, May 2005.

[15] Y. Miao, L. Zhao, C. Li, and J. Tang. Automatically grouping questions in yahoo! answers. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '10, pages 350–357, Washington, DC, USA, 2010. IEEE Computer Society.

[16] A. Spink, D. Wolfram, M. B. J. Jansen, and T. Saracevic. Searching the web: the public and their queries. *J. Am. Soc. Inf. Sci. Technol.*, 52:226–234, February 2001.

[17] X.-J. Wang, X. Tu, D. Feng, and L. Zhang. Ranking community answers by modeling question-answer relationships via analogical reasoning. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 179–186, New York, NY, USA, 2009. ACM.

[18] S. Yoon, A. Jatowt, and K. Tanaka. Intent-based categorization of search results using questions from web q&a corpus. In *Proceedings of the 10th International Conference on Web Information Systems Engineering*, WISE '09, pages 145–158, Berlin, Heidelberg, 2009. Springer-Verlag.

[19] S. Yoon, A. Jatowt, and K. Tanaka. Detecting intent of web queries using questions and answers in cqa corpus. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '11, pages 352–355, Washington, DC, USA, 2011. IEEE Computer Society.