Facebook Application for Asking, Answering and Searching Social Questions

Ram Kandasamy
Caltech
1200 East California Boulevard
Pasadena, California 91125

ramk@caltech.edu

Victor Li
Caltech
1200 East California Boulevard
Pasadena, California 91125

vli@caltech.edu

Esther Wang
Caltech
1200 East California Boulevard
Pasadena, California 91125

eswang@caltech.edu

ABSTRACT
This paper describes the motivation and features of Spoke, a Facebook application we had created for harboring a natural, friendly environment where Facebook users can discuss and browse social questions with their friends.

1. INTRODUCTION
We can rely on most search engines to answer questions with "factual" answers, such as the location of Corner Bakery in Pasadena, the name of The Beatles' first album, etc. However, for queries without definitive, objective answers we often resort to special-interest forums such as Yelp, College Confidential, and Math Overflow in hopes of finding reliable answers and opinions. As for the esoteric questions that do not belong to any specific area of interest, we resort to forums such as Yahoo Answers. Unfortunately, the usefulness and trustworthiness of answers decrease significantly in these forums, since anyone with access to the internet can provide answers.

The most natural way for us to find answers to "social questions" without definitive answers would be ask our network of friends whom we can trust and are accountable for their answers. Asking a network of friends and friends of friends (we will denote this group as 2-friends, because there is two degrees of separation) would be the more efficient and reliable method of getting our queries answered. Examples of "social questions" we may want our friends' input on are: what should I get my boyfriend for graduation? Romantic date ideas around Los Angeles? What is the best bike repair shop within walking distance from Caltech? What is the best way to avoid getting scammed on Craigslist? What was the most difficult but interesting class you have taken at Caltech?

Many people use their Facebook statuses, groups or tagged notes to receive feedback on such "social questions". However, these media have limited capabilities, since they were not designed to serve the purpose of pooling toegher opinions or finding answers to questions. Users are also limited to their immediate network of friends (no 2-friends). There is also no organized way of archiving answers and questions that originated from users' Facebook statuses. Consequently, there is no easy way for users to browse through social questions and answers their friends' have asked without combing through pages and pages of Facebook status updates, most of which are not even questions. Moreover, at the time of Spoke's conception (January 2010), Facebook's search engine did not encompass searching statuses, so there was no easy way of looking up a useful question thread another user may have started on his Facebook profile. Even though search for statuses have now been enabled, the search capability is still imperfect.

The goal of our Q&A service Facebook Application, Spoke, is to create a natural environment where Facebook users can discuss and browse social questions with their friends and the Spoke network. To this end, we created an easy to navigate user interface, where questions tagged by clickable keywords that sorted into categories for the users to browse. Users are asked to declare their "areas of expertise", i.e. topics they would like to answer questions about, which share the same database of keywords as the questions. Users are also encouraged to provide feedback on the quality of questions and answers by voting up via voting up questions and answers they liked.

2. REVIEW OF RELATED WORKS
Reliable question and answer services for social questions have garnered a lot of attention recently. The top three products in the "social Q&A" space are Aardvark, Quora and Facebook Questions.

2.1 Aardvark
Aardvark, founded in 2007 and acquired by Google for 50 million USD in February 2010, is conceived as a "Social Search Engine: a way to find people, not webpages, that have specific information." Aardvark users ask a question, either by instant message, email, web input, text message, or voice [3]. With an algorithm detailed in [1], Aardvark routes the question to the person in the user's extended social network, which includes friends and friends of friends on Facebook or Google contacts, that is most likely to be able to answer that question. Aardvark is composed of four main components: (1) Crawler and Indexer that crawls through each user's Facebook profile and blog to collect "labeled" resources, (2) Query Analyzer that analyzes the users' questions, (3) Ranking function that returns the best users to provide the information, and (4) UI that presents the information to the user (see Figure 1).

While the Aardvark network is extensive—includes Friends, Friends of Friends and the entire Aardvark network, it does not provide direct user to user interaction. Users are limited to answering questions routed to them by Aardvark's algorithm and questions are directed to other Aardvark users via Aardvark. The Aardvark homepage displays conversation threads people are having real time, but there is not a way for the user to browse these conversations freely.

Figure 1 Aardvark main page

2.2 Quora

Quora, a website founded by former Facebook CTO Adam D'Angelo in June 2008, first launched in private beta in January 2010. It was recently evaluated by venture capitalist to be worth 86 million USD. We did not get access to Quora, which available by invitation only, until more than halfway through the completion of our project. The goal of Quora is to create a "continually improving collection of questions and answers created, edited, and organized by everyone who uses it. The most important thing is to have each question page become the best possible resource for someone who wants to know about the question [3]".

Quora's user interface has several important features, which we had also independently designed for Spoke to have. Users can add "Topics" to follow and receive endorsements (awarded by other users). Questions and responses can be voted up and down by users (see Figure 2).

One thing that sets Quora apart from other Q&A services is that Quora has a very academic, Wikipedia-y vibe. For example one of our answers in a question thread was flagged for update because one of the sentences was missing a capitalization. We envisioned Spoke to be more lighthearted, and focused on fun Q&A based user interactions, rather than serious "information aggregation".
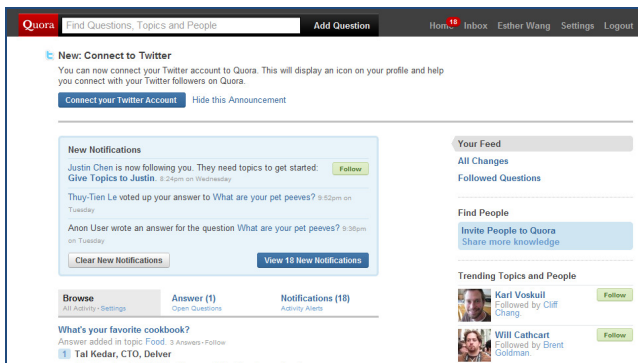

Figure 2 Quora main page

2.3 Facebook Questions

On May 23rd, 2010, Techcrunch published an article featuring inslide look Facebook Questions, a Q&A service Facebook has been working on [4]. The article included three screen shots—the main page of the application, which includes a full page of questions and answers (see Figure 3), a particular question thread (see Figure 4), and answer response box (see Figure 5). Since none of us got approved to test for Facebook Questions, we will quote the Techcrunch article: "Facebook Questions feels different from Quora because it seems to be more intimate/fun/terse than intellectual/useful/detailed." This is the vibe we want Spoke to have.

Unfortunately, the article does not go into detail regarding how the questions are organized and how users can browse the questions.
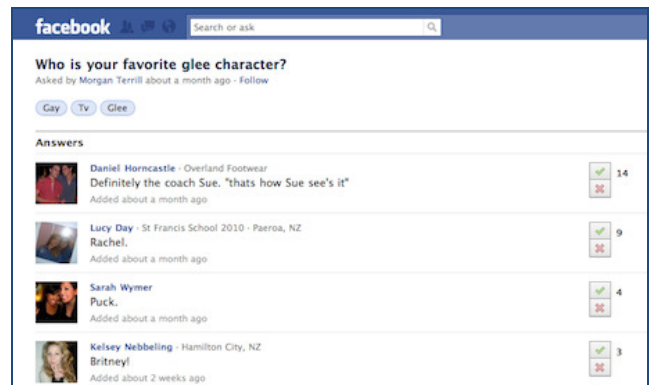

Figure 3 Facebook Questons main page

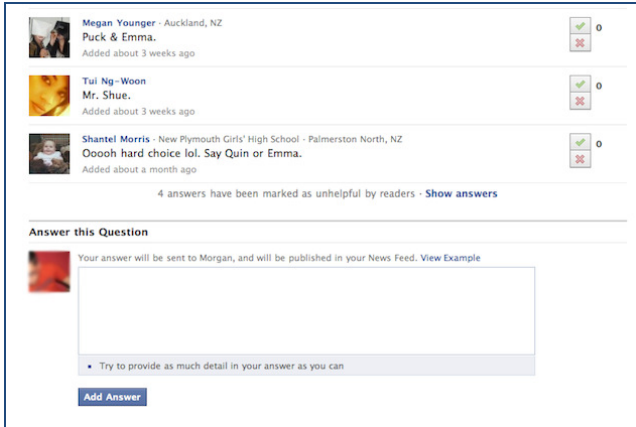**Figure 4 Facebook Questions example question thread**

Figure 5 Facebook Questons example answer box

3. SPOKE USER INTERFACE

Facebook users can sign up for Spoke at the following link: http://apps.facebook.com/spokeit/. The following sections detail the purposes of each page/tab of the application.

3.1 Areas of Expertise

When a user signs up for Spoke, he is prompted to enter three "areas of expertise", i.e. three topics he would like to answer questions about (see Figure 6). We currently limit the topics a user can enter to keywords in our database of tags via auto-complete and drop down menu (see Figure 8). The user can edit his areas of expertise on his profile page; he can add up to 10 topics but must always have at least three (see Figure 8). On the profile page, the user can also click on each of his areas of expertise to view questions tagged with his area of expertise (see Figure 7).

Figure 6 Spoke sign up page
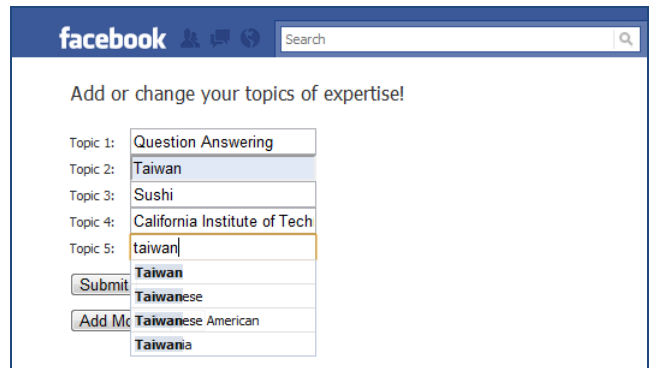

Figure 7 Spoke sample user profile page

Figure 8 Spoke sample user expertise page

3.2 Questions and Answers

Our goal is to provide users with an easy way of asking, answering, searching and browsing questions.

3.2.1 Question Asking

The user can ask a question on the right hand side of each page (see Figure 7). The user is required to tag his question with one of the tags from our database. The question can be directed to every user on Spoke or only the user's friend. This allows for the user to differentiate general questions from perhaps more personal questions he does not wish to share with the Spoke network.

3.2.2 Question Answering

The user can answer any viewable question (that is, questions that are not filtered by a user who is not his friend) by clicking on the question. This leads him to the question's page, where he can add his answer to the thread (see Figure 9).

The user can also view the list of answers he has provided on the My Answers page (see Figure 10).
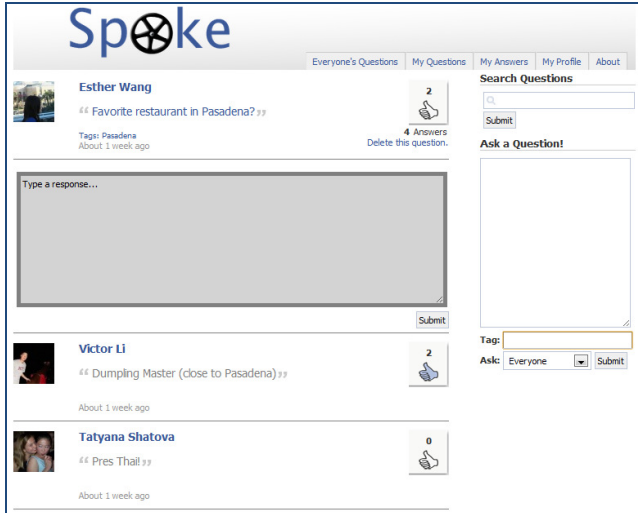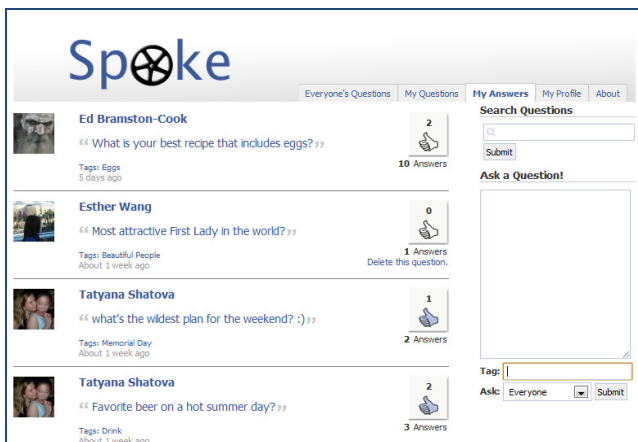
Figure 9 Spoke sample question & answer thread

Figure 10 Spoke sample My Answers page

3.2.3 Question Searching

The user can search questions by entering a string of text on the right hand side of each page (see Figure 11).

The search algorithm is explained in more detail in Section 4.3.

Figure 9 Spoke sample search result for query "song"

3.2.4 Question Browsing

Users can browse all questions he has asked by going to My Questions (see Figure 12).
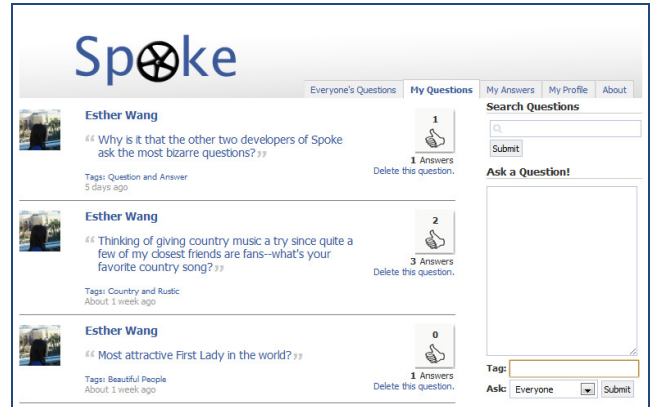

Figure 10 Spoke sample My Questions page

Users can view questions other users have asked by going to Everyone's Questions (see Figure 13). We have implemented three filters for filtering the displayed questions: (1) My Expertise, (2) Asked By, and (3) Organize By. With the My Expertise filter, the user can choose to display questions pertaining to his areas of expertise. The motivation for implementing this filter is that the user is probably more interested in viewing and answering questions in his areas of interest/expertise. We also want users to answer questions within his areas of expertise to improve the quality of answers provided for questions under that topic, assuming the user is actually an "authority" in his declared areas of expertise. With the Asked By filter the user can choose to display questions either asked by everyone or by his friends only. With the Organized By tab, the user can choose to display questions from newest to oldest, most votes to fewest votes or least answered to most answered. By default, all questions are displayed from newest to oldest.

Figure 11 Spoke Everyone's Questions page

The user can also browse questions by categories the questions' respective tags fall under (see Figure 14). Our method of categorizing tags is detailed in Section 4.2.
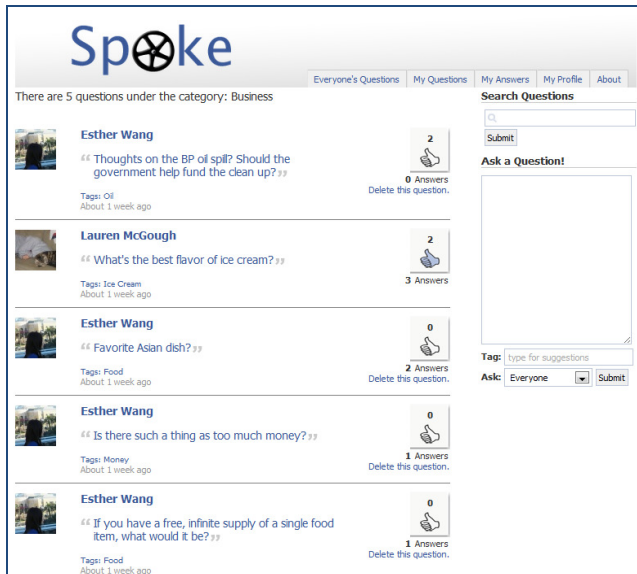
Figure 12 Spoke sample result from clicking a category, "Business"

Lastly, the user can also click on the question tags (which share the same keywords as user's Areas of Expertise) to browse questions (Figure 15).
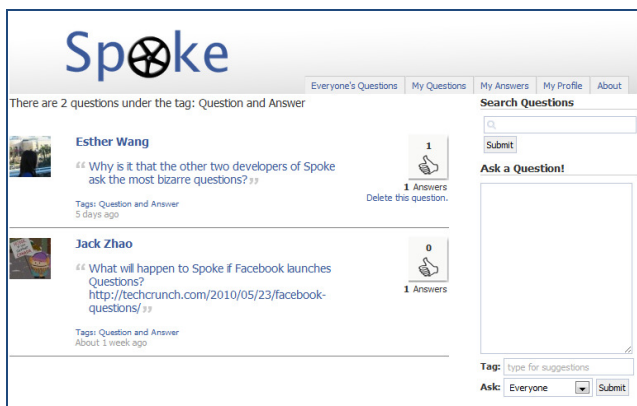

Figure 13 Spoke sample result from clicking a tag, "Question and Answer"

3.3 User interaction with other users

Users can directly interact with other users on Spoke by asking questions, answering questions and voting up questions or answers other users have supplied. The user can also view every other Spoke profile page by clicking their names next to their questions or answers. Again, if the two users are not Facebook friends, then questions addressed only to friends cannot be viewed. Users can answer all viewable questions. Finally, users can vote up questions and answers they like and subsequently boost the reputation points (explained in Section 3.4) of the users who supplied those questions and answers.

3.4 Reputation and Awards System

To ensure the quality of questions and answers provided and to reward users for positively contributing to Spoke, we have created reputation and awards systems. Reputation points are awarded as

follows: users get 15 reputations for every answer provided that is voted up by the asker; 10 reputations for every answer provided that is voted up; 5 reputations for every question asked that is voted up. We currently give out two types of awards for quality questions and quality answers—the user is awarded gold, silver or bronze rims for answers/questions that receive 50, 30 and 15 votes respectively. The reputation points and medals are displayed on the user's profile page (Figure 7).

4. SPOKE BACKEND
4.1 Data Storage

The backend database of Spoke is comprised of five major sections: (1) activity database, (2) questions inverted index, (3) tags inverted index, (4) category inverted index, and (5) open directory project.

The activity database is organized by Facebook user id. Each user has three tables to store his questions, answers, all activities (related to question asking, answering and voting, for example). Since everything is ordered by user id, the database enables fast look up of questions, user's voting, and a question's answers.

The inverted index databases contains questions that have been stop-word and punctuation removed and stemmed. The question inverted index stores each word of every question asked, and points to all questions that contain this word. This allows for the search function to grab a list of relevant questions based on a query and order them based on our search algorithm.

There are two tag inverted indexes. One is based on individual words of the tags, and the other is based on the full tag. The former index is used in the search algorithm while the latter is used for easy lookup of questions under a certain tag. The category inverted index simply stores a list of questions under the various categories for easy lookup.

The open directory project database stores a list of keywords from the Open Directory Project (http://www.dmoz.org). This database contains a user-generated list of keywords and the categories for which the keywords fall under. Unfortunately, the project still needs a lot of work—the categorizations are strange and the database does not contain all "relevant" keywords. The current categories are also not the most useful or obvious—for example we have a category for "Reference" and "Kids and Teens". For now, we are using a modified version of the keywords and categories to organize our questions tags.

4.2 Tags Cleanup

In order to use the Open Directory Project keywords for our purposes, we needed cull and reorganize the list of keywords and categories. We were able to cull the list of keywords by removing roughly 200,000 keywords relating to foreign subjects, and use the remaining keywords as our question tags. The Open Directory Project seems to frequently arbitrarily map a tag to multiple categories, which leaves us the task of determining the single category for the tag. For example, according to the Open Directory Project, the keyword "Washington D.C." falls under the category of "Los Angeles," among many others. Since the Open Directory Project is set up as a tree of categories, our strategy of categorizing the tags (given the existing paths of categorization), is to pick the category that is closest to the root of the tree (the question tag). For the most part, this strategy is adequate but far from perfect. For example, under the category of "Business" we

have the following tags: "Oil", "Ice Cream", "Food" and "Money" (see Figure).

4.3 Search Algorithm
The search algorithm uses the question inverted index and tag inverted index to get a list of relevant questions. From this list, we rank the results by weighting a match to a tag highest, a match to a word in the question second highest, and the time of the question last. We weight tags highest because the tag captures the essence of the question in a very few number of words, whereas it is harder to determine how important a word is in a question. We rank newer questions higher than older questions because they're more likely to have fresher content.

5. USER FEEDBACK
We notified our Facebook friends that the application was ready for testing on May 27th, 2010. As of June 4th, 2010, Spoke has 42 monthly active users, according to the statistics on Spoke's Facebook application page:

http://www.facebook.com/?ref=logo#!/apps/application.php?id=108357272521370

We also conducted an anonymous survey, which can be found at:

http://www.surveymonkey.com/s/L6S3BXT

Our friends suggested very helpful user interface improvements and caught many bugs that escaped us, which have now been fixed. For example, if a user clicks inside the response box after having entered text in the box, the text inside the box would be erased.

For the most part Spoke was fairly well-received. Most of the users found the features of Spoke to be useful and easy to understand.

In terms of UI suggestions, we fulfilled the simple aesthetics requests such as alignment but the more fundamental changes we leave to future work. The main concern here is that users would prefer user generated tags and more intuitive categories as opposed to being restricted to our choice of tags and the categories culled from the Open Directory Project. Some users have also suggested allowing for multiple tags per questions, as opposed to single tag per question. An interesting suggestion is to order the questions displayed on the user profile page by the last time the question received an answer, which would help generate discussions on good questions. Currently, the user sees the latest question posted by each of his friends on his profile page.

6. FUTURE WORK
There are many ways in which we can expand and improve Spoke's current capabilities.

6.1 Spoke Network
Spoke's network currently only gives users the option to make their questions viewable to either friends or the entire Spoke network. It would be good to have an intermediary level of 2-friends, like Aardvark.

6.2 Tags and Categories
Ideally, we would allow users to enter any area of expertise and assign any tag to their questions (with minimal spell checking and check for spam), instead of limiting them to keywords from the Open Directory Project. However, because we currently rely on the Open Directory Project tags-categories tree for assigning tags to their proper categories, we must impose such limitations.

To create the ideal categorization system, we would first determine a set of say, 20 useful categories. Then, to create our training set, we would manually label tags/questions as belonging to one of the 20 categories. We did not have enough training data (i.e. questions and tags) or time to create such a learning set. Once we have such a training set, we can then implement and test different machine learning algorithms for automatic classification of tags.

As an interesting side note, Quora does not currently have the capability of automatic question categorization due to the same reasons [5].

6.3 Awards System
We could have implemented more complicated awards system that makes use of user expertise information. For example, we can incentivize users to answer questions in their "declared" areas of expertise by rewarding them with more reputation points when they answer such questions. The awards system can easily be expanded to encompass more specific awards, such as most avid Spoke user.

6.4 Miscellaneous
Perhaps we should have required users to enter a question upon signing up for Spoke, in addition to entering three areas of expertise. For example, the application for testing Facebook Questions requires the applicant to enter three questions and corresponding answers for review.

Aesthetics wise, we could try to make the transition between pages "smoother". Moreover, we could make the logo and sign up page prettier, had we had more time to try different designs.

7. ACKNOWLEDGMENTS
Our thanks to Professor Adam Wierman, Minghong Lin and our friends for their guidance and useful feedback throughout the course of the project.

8. REFERENCES
[1] Horowitz, D. and Kamvar, S. 2010. *The Anatomy of a Large-Social Search Engine.*

[2] http://vark.com/about

[3] http://www.quora.com/about

[4] http://techcrunch.com/2010/05/23/facebook-questions/

[5] http://www.quora.com/Can-Quora-automatically-categorize-my-question?q=automatic+cate