# Towards Automatic Evaluation
# of Health-Related CQA Data

Alexander Beloborodov[1], Pavel Braslavski[1,2], and Marina Driker[3]

[1] Ural Federal University `xander-beloborodov@yandex.ru`
[2] Kontur Labs `pbras@yandex.ru`
[3] Ural State Medical University `mdriker@yandex.ru`

**Abstract.** The paper reports on evaluation of Russian community question answering (CQA) data in health domain. About 1,500 question–answer pairs were manually evaluated by medical professionals, in addition automatic evaluation based on reference disease–medicine pairs was performed. Although the results of the manual and automatic evaluation do not fully match, we find the method still promising and propose several improvements. Automatic processing can be used to dynamically monitor the quality of the CQA content and to compare different data sources. Moreover, the approach can be useful for symptomatic surveillance and health education campaigns.

## 1   Introduction

The web has become an important source of health information for lay-people. In 2012, 59% of the US adults looked online for health information; 45% of them searched for specific disease or medical problem [9]. These figures are lower in Russia, but still substantial and growing: in 2013, 13% of Russian population searched for health-related information online, with the proportion much higher in big cities [1]. Although general search engines remain the primary tool for searching medical information online, there are also other options. One of the popular destinations is community question answering (CQA) sites that allow users to post questions on virtually any subject to other community members, answer questions, rate and comment answers, and gain points and badges. Yahoo!Answers[4] and WikiAnswers[5] are examples of popular CQA platforms. CQA is a good complement to web search that allows for a more detailed description of information need, delivers more social and personalized search experience, suits users with low search engine proficiency, etc. CQA data are large, diverse, and dynamic, but content quality can be the major issue, which is critical in case of medical information.

There are guidelines for medical websites with editorial content that enable veracity of the information provided online. For example, the Health On the

---

[4] https://answers.yahoo.com/
[5] http://wiki.answers.com/

Net Foundation (HON)[6] elaborated HONcode and reviews websites to comply with it. More than 7,300 sites are certified by the HON to date. Obviously, this approach cannot be applied to sites that are fueled with extensive and loosely controlled users' input.

Several approaches aimed at automatic detection of high-quality content in CQA were proposed in the literature (see next section). In contrast to these studies we experiment with an automatic method for quality assessment focused on *health-related* CQA. The idea of the approach is to perform evaluation on a narrowed subset – questions asking for medication for a specific symptom or disease, for example (all examples are originally in Russian):

> **Q:** *Please suggest good [runny nose] drops. need to hold out an important meeting, at least 2 hours without snivel...*
> **A:** *Take [Sanorin] or [Nazol Advance]*

According to different studies [1, 9, 7, 21] this type of information need is among the most common ones. We hypothesize that this type of questions is exemplary enough to reflect the overall quality of CQA health-related content. The approach can be used for comparison of different CQA services in the health domain or longitudinal observation of a CQA subcategory, rather than as a technique to evaluate individual items.

In our experiment we used health-related questions and answers from the Russian CQA platform Otvety@Mail.Ru[7] (*otvety* means *answers* in Russian). First, we compiled a list of unambiguous medicine designations from a comprehensive registry of drugs. Second, we composed a list of 13 diseases and symptoms, mined corresponding reference drugs from an online resource and performed manual post-processing of the obtained table. Automatic evaluation is straightforward: we count correct and incorrect 'disease-in-question – drug-in-answer' occurrences. This approach is motivated by analysis of disease–drug pair frequencies in our previous study [5]. In parallel we carried out manual evaluation of about 1,500 question–answer pairs. Juxtaposition of both manual and automatic evaluation showed low consistency rate. This can be partly due to discrepant manual evaluation that was quasi crowd-sourced. Another possible reason is that the automatic approach is too shallow and simplistic. Nevertheless we find the method still promising and propose several improvements.

The paper is organized as follows. Next section surveys literature on detection of high-quality content in CQA and analysis of health-related content in social media. Section 3 describes data used in the study: a general list of drugs, a list of 13 diseases with their reference drugs, and Otvety@Mail.Ru dataset. Section 4 describes manual and automatic evaluation results and comparison of both. Section 5 discusses the obtained results. Section 6 defines directions for future research and concludes.

---

[6] https://www.hon.ch/
[7] http://otvety.mail.ru/

## 2   Related Work

Our work is related to prior research in detection of high-quality content in CQA and analysis of social media in health domain.

*Content Quality in CQA.* There are several methods aimed at automatic evaluation of CQA data described in the literature. There are slight variations in the notion of a good question or answer. E.g. studies distinguish between asker's [4, 18] and 'external' perception of answer quality [3]. Some of the studies consider answers quality only [2, 18], some look into questions[10, 15, 14], the other examine questions and answers simultaneously [3]. Authors classify questions into 'information-seeking' vs. conversational, or entertaining ones [10, 15]. Some approaches seek to evaluate content quality on-the-fly, right upon question or answer arrival; the other work with archival data, i.e. containing users' ratings and comments, usage statistics, etc. The methods described in the literature employ machine learning techniques and a wide range of features. The features may include: 1) text features grasping text grammaticality, spelling, visual neatness, readability, etc., 2) user features such as user rating, activity, badges, expertise in the topic, interactions with other users, and 3) content popularity reflected in click statistics. None of the features are domain-specific, and methods can be applied presumably to any subject area. However, authors point out that different CQA categories vary in user interaction patterns, vocabulary, etc. Studies devoted to analysis of questions posted on Stackoverflow[8] (a CQA platform on software programming) take into account domain specifics and make use of dedicated dictionaries [13, 8].

*Health Information in Social Media.* It has been shown that Twitter data have a great potential for public health, e.g. for symptomatic surveillance, analyzing symptoms and medication usage, behavioral risk factors, geographic localization of disease outbreaks, etc.[12, 17]. In our previous study [5] we applied a similar set of tools to CQA content in health domain and hypothesized that the approach can be used for content quality evaluation. Bhattacharya et al.[6] analyzed health beliefs of the type *smoking causes death* or *tea tree oil treats infection* expressed on Twitter. The way the data were gathered and processed is similar to our approach. Wong et al.[20] proposed an experimental dialog system that uses Yahoo!Answers data for guided conversations on health-related subjects. However, quality issue of CQA data is not addressed in the study.

   Dedicated studies on CQA content quality in health domain are relatively few and rely mostly on manual processing. Zhang's study [21] described linguistic features, users' motivations and question types, as well as temporal, emotional and cognitive aspects of a sample of about 270 questions in *Health* category of Yahoo!Answers. Oh et al.[16] outlined a draft of a quality evaluation experiment of health-related content of Yahoo!Answers. The plan features multi-faceted answers judgment by representatives of three groups – questioners, health reference librarians, and nurses; preliminary results for 10 questions are quoted. Kim

---
[8] http://stackoverflow.com/

et al.[11] semi-automatically assessed around 5,400 questions and answers on H1N1 influenza strain posted on Yahoo!Answers. The authors identified major subtopics in H1N1 questions, types of resources askers and answerers referred to, and medical concepts mentioned in the data.

## 3 Resources & Data

### 3.1 Disease and Medicine Dictionaries

The starting point for the formation of the 'disease – medicine' table was data gathered from the Registry of Medicine. The reference book (and its online counterpart[9]) combines comprehensive information on drugs and International Classification of Diseases (ICD)[10] codes.

For the experiment, we selected 13 frequent symptoms/diseases: *allergy/ urticaria, rhinitis, tonsillitis, gastritis, diarrhea, influenza-like illness (ILI), candidiasis, herpes, heartburn, stomatitis, hemorrhoids, dysbiosis,* and *otitis.* In the context of our study it is particularly important that these diseases are often self-treated without consulting a doctor; complementary medicine is often used to treat these symptoms.

ICD is too detailed for our purposes, so we merged groups of diseases with similar symptoms (and corresponding lists of drugs). For example, influenza-like illness (ILI) group is combined of three ICD codes:

> *J06. Acute upper respiratory infections of multiple and unspecified sites;*
> *J10. Influenza due to other identified influenza virus;*
> *J11. Influenza, virus not identified.*

We also provided each disease with its synonyms, including vernacular names – e.g. *runny nose* has nine name variants.

General medicine names were taken from the State Register of Approved Drugs[11] and converted to a list of unambiguous one-word names, toward this end were removed 1) pharmaceutical form designations such as *drops, cream, solution, tablets,* etc.; 2) modifiers of medicine names (e.g. *Aspirin Cardio, 1000 Aspirin, Aspirin York, Aspirin Express → Aspirin*); 3) names consisting of common words. As a result, of 11,926 unique entries in the initial list we came up to the list of 4,120 drugs.

In addition, we manually processed the list of drugs corresponding to the 13 selected diseases. We excluded 1) dietary supplements, multivitamin and mineral complexes with a wide range of indications – selected symptoms are not specific for them; 2) baby nutrition products; 3) psychotropic drugs; 4) drugs with expired registration, as well as drugs that are not currently used in clinical practice because of their low efficiency or high toxicity; 5) drugs with missing formula (only trade names are presented in the registry). Table 1 shows statistics of 'disease – drug' table (a drug may correspond to several diseases).

---

[9] http://rlsnet.ru
[10] http://www.who.int/classifications/icd/en/
[11] http://grls.rosminzdrav.ru/

**Table 1.** Number of unique drug names for each of 13 diseases

| Disease or symptom | # of drugs |
|---|---|
| ILI | 294 |
| rhinitis | 260 |
| tonsillitis | 167 |
| diarrhea | 155 |
| otitis | 149 |
| allergy/urticaria | 118 |
| gastritis | 106 |
| stomatitis | 90 |
| candidiasis | 84 |
| herpes | 64 |
| hemorrhoids | 57 |
| dysbiosis | 53 |
| heartburn | 46 |

### 3.2 Otvety@Mail.Ru

Otvety@Mail.Ru is a Russian counterpart of Yahoo!Answers with similar rules and incentives. The site was launched in 2006 and has accumulated almost 80 million questions and more than 400 million answers by August 2012[12]. Otvety@Mail.Ru has two-level directory with about 30 top-level categories, including *Health and Beauty*, and about 200 subcategories. The users have to assign their questions to a second-level category using drop-down lists.

The initial data set contained 128, 370 questions and corresponding answers from four second-level categories: *Diseases and Medicines; Doctors, Clinics, and Insurance; Doctors' answers; and Kids' Health* in the timespan from 1 April 2011 to 31 March 2012. Tables 2 and 3 summarize main characteristics of the corpus used in the experiment. Additional details about Q&A topics and users' demographics can be found in our earlier paper [5].

**Table 2.** Otvety@Mail.Ru Health corpus statistics

| | |
|---|---|
| Number of questions | 128,370 |
| Average number of answers per question | 5 |
| Average question length (words) | 10.1 |
| Average answer length (words) | 21.6 |
| Questions mentioning one of the 13 diseases | 7,147 |
| ...of which with answers mentioning a drug | 4,054 |

---

**Table 3.** Most frequent diseases and drugs

| Top-5 diseases | |
|---|---:|
| rhinitis | 1,606 |
| allergy/urticaria | 1,182 |
| tonsillitis | 802 |
| ILI | 730 |
| candidiasis | 494 |
| **Top-5 medicines** | |
| iodine | 3,291 |
| activated carbon | 2,526 |
| hydrogen peroxide | 2,057 |
| aspirin | 1,873 |
| analgin | 1,531 |
| **Top-5 disease-medicine pairs** | |
| herpes–aciclovir | 274 |
| allergy/urticaria–suprastin | 196 |
| candidiasis–flucostat | 157 |
| herpes–zovirax | 132 |
| diarrhea–activated carbon | 131 |

## 4 Experiment

### 4.1 Data Preparation

As mentioned above, our goal was to automatically evaluate question–answer pairs, where the question is of type "What should one take in case of $X$?" or "How to treat $X$?" ($X$ is a disease or symptom from a predefined list). We indexed Otvety@Mail.Ru data by 13 diseases and their variants (see section 3.1), as well as by all medicines. Medicine and disease names are often misspelled; to improve the retrieval quality we implemented a fuzzy matching algorithm with learned edit distance that captures misspelled mentions of both diseases and medicines.

We sampled randomly about 1,000 questions with their corresponding answers according to a straightforward template *'disease in question – medicine in answer'*. Manual investigation showed that for questions with at least one answer containing medicine mention, only 53% belong to the sought type. When we additionally require that at least one answer contains *two or more medicine mentions*, this rate increases to 79%. We opted for this simple criterion with satisfactory precision for data selection, although it can deliver false positives of the kind:

**Q:** *I have got a [coldsore] on my lip. Should I go to school?*
**A:** *Haha... 90% of the population have herpes and go to school and work. You should buy [Kagocel] and [Aciclovir] at the pharmacy. Get well )*

For manual evaluation we selected 255 questions with at least one answer containing two or more medicine mentions (from 1,000 sampled previously), which resulted in 977 question-answer pairs. Additionally, we randomly sampled 500 question-answer pairs as a complement to the questions of the particular type.

## 4.2 Manual Evaluation

We sent an invitation to participate in the evaluation of health-related community questions and answers to selected staff members of the Ural State Medical University. Seven assessors took part in the evaluation (including one of the authors – MD), all holding a university degree in medicine. The evaluation was conducted using an online tool, no personal meetings and instruction took place. Before the start of the evaluation assessors were exposed an instruction and several examples. Assessors could pause evaluation any time and return to it at a convenient time; there was also no 'minimum assignment' – assessors could determine the comfortable volume of labeling by themselves, which resulted in an uneven distribution of answers labelled by individual assessors (406 : 267 : 197 : 102 : 58 : 50 : 11). Evaluation queue was generated randomly, so that the answers to the same question did not appear straightly one after another. Assessors were presented a question-answer pair that they could evaluate on a three-grade scale: 0 – low quality (potentially harmful), 1 – potentially useful answer, 2 – high-quality answer. The left-hand side of table 4 summarizes the results of manual evaluation, including randomly sampled question–answer pairs (i.e. not of the type 'how to treat a particular disease or symptom').

Inter-assessor agreement calculated on 100 double-judged items is quite low: raters' labels coincide only in 21% of cases, Cohen's $\kappa$ is 0.51 (calculated taking into account distances between labels).

## 4.3 Automatic Matching

Automatic processing of question–answer pairs is straightforward: we count 'correct' and 'incorrect' disease-in-question—drug-in-answer pairs; when multiple drug mentions occur in answer, the final score (0 or 1) is calculated based on majority voting (e.g. two 'correct' mentions out of three result in 1). Right-hand side of table 4 shows automatically obtained scores for manually labeled data; the rightmost column reports the proportion of matching scores (manual scores are binarized as follows: $0 \rightarrow 0$; $1, 2 \rightarrow 1$). Table 5 shows automatic scores for the whole subset of question–answer pairs matching the pattern 'questions with at least one answer with 1+ drug mentions'; figure 1 depicts the distribution of the values over the year.

Table 4. Manual and automatic evaluation results

| Disease | # | Manual labels (M) | | | | | | Automatic scores (A) | | | | M&A | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | | 1 | | 2 | | 0 | | 1 | | matches | |
| rhinitis | 182 | 68 | 0.37 | 70 | 0.38 | 44 | 0.24 | 60 | 0.33 | 122 | 0.67 | 112 | 0.62 |
| allergy/urticaria | 149 | 81 | 0.54 | 39 | 0.26 | 29 | 0.19 | 24 | 0.16 | 125 | 0.84 | 56 | 0.38 |
| candidiasis | 148 | 48 | 0.32 | 66 | 0.45 | 34 | 0.23 | 18 | 0.12 | 130 | 0.88 | 96 | 0.65 |
| herpes | 136 | 49 | 0.36 | 57 | 0.42 | 30 | 0.22 | 15 | 0.11 | 121 | 0.89 | 88 | 0.65 |
| tonsillitis | 107 | 23 | 0.21 | 50 | 0.47 | 34 | 0.32 | 15 | 0.14 | 92 | 0.86 | 79 | 0.74 |
| diarrhea | 70 | 27 | 0.39 | 28 | 0.40 | 15 | 0.21 | 7 | 0.10 | 63 | 0.90 | 44 | 0.63 |
| heartburn | 53 | 17 | 0.32 | 20 | 0.38 | 16 | 0.30 | 5 | 0.09 | 48 | 0.91 | 33 | 0.62 |
| stomatitis | 48 | 18 | 0.38 | 24 | 0.50 | 6 | 0.13 | 7 | 0.15 | 41 | 0.85 | 31 | 0.65 |
| dysbiosis | 28 | 16 | 0.57 | 6 | 0.21 | 6 | 0.21 | 2 | 0.07 | 26 | 0.93 | 14 | 0.50 |
| ILI | 24 | 7 | 0.29 | 11 | 0.46 | 6 | 0.25 | 2 | 0.08 | 22 | 0.92 | 17 | 0.71 |
| otitis | 13 | 3 | 0.23 | 4 | 0.31 | 6 | 0.46 | 5 | 0.38 | 8 | 0.62 | 9 | 0.69 |
| gastritis | 12 | 5 | 0.42 | 4 | 0.33 | 3 | 0.25 | 3 | 0.25 | 9 | 0.75 | 6 | 0.50 |
| hemorrhoids | 7 | 2 | 0.29 | 3 | 0.43 | 2 | 0.29 | 3 | 0.43 | 4 | 0.57 | 4 | 0.57 |
| **Total** | **977** | **364** | **0.37** | **382** | **0.39** | **231** | **0.24** | **166** | **0.17** | **811** | **0.83** | **589** | **0.60** |
| Random sample | 500 | 207 | 0.41 | 199 | 0.4 | 94 | 0.19 | | | | | | |

## 5 Discussion

### 5.1 Quality of Manual Assessment

After completion of the manual assessment, we reviewed the results and made certain that many labels appear inconsistent. We had a post-assessment phone interviews with most active assessors. It turned out that doctors were skeptical about the web as a source of reliable and trustworthy medical information in general. In addition, doctors perceive online media as a competitor in some sense – many patients come to doctor's appointment with a diagnosis self-established upon consulting the web. Additionally, doctors tend to rate answers containing *'consult your doctor'* advice higher.

Doctors can hardly disengage from a concrete clinical case, they assume more than is given in the question and may have an individual opinion on the effectiveness and applicability of a specific drug. In general, they tend to underrate answers following the principle *primum non nocere*. At the same time, contrary cases may occur: doctors put a non-zero score to an answer that is apparently not correct but *'will not hurt'*.

Thus, our hypothesis that we can perform a online quasi crowd-sourced evaluation with minimal investment into instructions and training of assessors is not justified. Medicine and Health is a too specific area and medical professionals are too unique to be 'averaged'. A solid preparatory work is needed even in case of a simple task like ours.
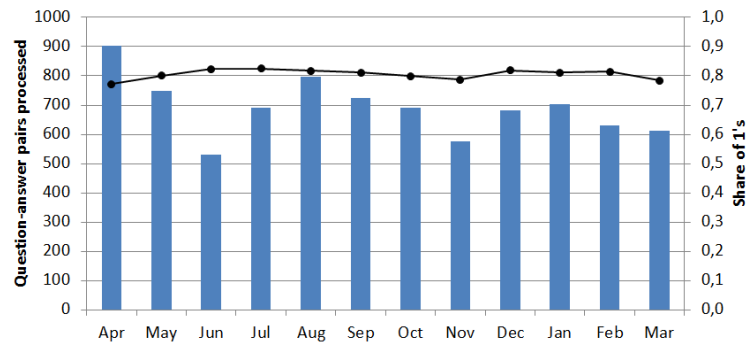
**Fig. 1.** The total number of processed question–answer pairs (bars) and the ratio of 1's (line) by month.

## 5.2 Inconsistency of Automatic vs. Manual Labels

We were able to identify several typical cases when automatic assessment did not match manual labeling.

*Diagnosis shift.* The questioner describes their conditions and assumes a diagnosis; the answerer disputes it and suggests treating another illness, for example:

> **Q:** *I have a strange allergy!!! small blisters all over my body, very itchy! help!*
> **A:** *<..>If the spots break out in the area where you scratched, then you have a skin infection. <..>*

*Sentiment of drug mentions.* A medicine's mention does not necessarily mean recommendation. Spam and advertising of drugs (often belonging to alternative medicine) presented in the data are often accompanied by criticism of traditionally used drugs. Users can doubt usefulness of a drug or even warn against taking the drug, for example:

> **Q:** *How can I get rid of herpes?*
> **A:** *Aciclovir does not help me at all.*

> **Q:** *What would help my kid at [diarrhea]? (2 and a half years) A week ago took a course of antibiotics treatment*
> **A:** *<..> don't you get enough chemistry during your disease? Whatever "natural" all [Linex], [Hilak] forte, etc. are – they are produced chemically from inorganic products.*

*Multiple drugs in answer.* Binary score calculated using simple majority voting does not reflect answer quality well. Users often suggest several drugs that they believe will solve the problem comprehensively. For example, in case of herpes – antiherpes cream and immunomodulators; in case of otitis – antibiotics along probiotics to fight dysbiosis as a possible side-effect of antibiotic treatment.

**Table 5.** Automatic evaluation results

| Disease | # | 0 | | 1 | |
|---|---|---|---|---|---|
| rhinitis | 1,653 | 509 | 0.31 | 1,144 | 0.69 |
| allergy/urticaria | 926 | 211 | 0.23 | 715 | 0.77 |
| candidiasis | 771 | 95 | 0.12 | 676 | 0.88 |
| herpes | 920 | 94 | 0.10 | 826 | 0.90 |
| tonsillitis | 889 | 143 | 0.16 | 746 | 0.84 |
| diarrhea | 841 | 74 | 0.09 | 767 | 0.91 |
| heartburn | 440 | 130 | 0.30 | 310 | 0.70 |
| stomatitis | 359 | 68 | 0.19 | 291 | 0.81 |
| dysbiosis | 420 | 63 | 0.15 | 357 | 0.85 |
| ILI | 513 | 88 | 0.17 | 425 | 0.83 |
| otitis | 147 | 45 | 0.31 | 102 | 0.69 |
| gastritis | 269 | 76 | 0.28 | 193 | 0.72 |
| hemorrhoids | 137 | 22 | 0.16 | 115 | 0.84 |
| **Total:** | **8,285** | **1,618** | **0.2** | **6,667** | **0.8** |

### 5.3 Analysis of User Opinions

Analysis of the answers shows that users tend to advise 'traditional remedies', criticizing the unnaturalness, high prices, and side effects of modern drugs. 'Old good' activated carbon and medical herbs (salvia, calendula, chamomile, etc.) are recommended as a remedy against almost every disease. Widely advertised immunomodulators developed and produced in Russia such as *kagocel* or *arbidol* are very popular in answers; although many medical professionals are skeptical about them, since there are no studies proving their effectiveness in terms of evidence-based medicine. There are also strong misconceptions about the applicability of certain drugs. For example, the advice to apply *corvalol* (barbiturate-based sedative) to herpetic blisters is quite frequent.

## 6 Conclusion

In the course of experiment, about 1,500 question–answer pairs were manually evaluated by medical professionals, which makes this study one of the largest on the subject. Evaluation data are freely available for research purposes[13]. Distribution of assessors' labels on the selected type of questions is very similar to the distribution on the randomly sampled question–answer pairs. Thus, we can assume that the selected type of questions is exemplary for the whole dataset in terms of content quality. However, manual assessment results have low consistency, and we will address the issue in our future work.

In addition, automatic evaluation based on reference disease–medicine correspondence was performed. Automatic scores for the selected set of diseases are

---

[13] http://kansas.ru/cqa/data2/

quite stable over time and do not depend on the volume of the processed data. To the best of our knowledge, the described experiment is the first attempt to automate the evaluation of community questions and answers in health-related domain. Although the results of the manual and automatic evaluations do not fully match, we find the method still promising. We see several directions that can potentially improve the automatic method:

– combination of the proposed domain-specific approach with the general approaches to CQA content evaluation described in the literature;
– application of sentiment analysis methods to detect polarity of drug mentions in answers (i.e. positive/neutral/negative);
– finer-grained selection of content items to be automatically assessed (e.g. classification of sentences into describing diseases vs. describing drugs similarly to [19]).

We will investigate the potential of the method to compare the quality of several datasets of questions and answers in health domain.

Besides the immediate task of automatic estimation of CQA content quality the proposed approach and its modifications can be applied to symptomatic surveillance, survey of drug usage, identifying common misconceptions in lay people thus guiding health education campaigns.

# References

1. Internet kak istochnik informazii o zdorov'e, medizine i lekarstvennykh preparatakh (Internet as a source of information about health, medicine, and medication). Synovate Comcon Healthcare (April 2014), http://www.comcon-2.ru/download.asp?544
2. Adamic, L.A., Zhang, J., Bakshy, E., Ackerman, M.S.: Knowledge sharing and yahoo answers: Everyone knows something. In: Proceedings of WWW'08. pp. 665–674 (2008)
3. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. In: Proceedings of WSDM'08. pp. 183–194 (2008)
4. Agichtein, E., Liu, Y., Bian, J.: Modeling information-seeker satisfaction in community question answering. ACM Trans. Knowl. Discov. Data 3(2), 10:1–10:27 (2009)
5. Beloborodov, A., Kuznetsov, A., Braslavski, P.: Characterizing health-related community question answering. In: Advances in Information Retrieval, LNCS, vol. 7814, pp. 680–683. Springer Berlin Heidelberg (2013)
6. Bhattacharya, S., Tran, H., Srinivasan, P.: Discovering health beliefs in twitter. In: AAAI Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text (2012)

7. Cartright, M.A., White, R.W., Horvitz, E.: Intentions and attention in exploratory health search. In: Proceedings of SIGIR'11. pp. 65–74 (2011)
8. Correa, D., Sureka, A.: Fit or unfit: Analysis and prediction of 'closed questions' on stack overflow. In: Proceedings of COSN'13. pp. 201–212 (2013)
9. Fox, S., Duggan, M.: Health online 2013 (January 2013), http://www.pewinternet.org/2013/01/15/health-online-2013/
10. Harper, F.M., Moy, D., Konstan, J.A.: Facts or friends?: Distinguishing informational and conversational questions in social q&a sites. In: Proceedings of CHI'09. pp. 759–768 (2009)
11. Kim, S., Pinkerton, T., Ganesh, N.: Assessment of h1n1 questions and answers posted on the web. American Journal of Infection Control 40(3), 211–217 (April 2012)
12. Lampos, V., De Bie, T., Cristianini, N.: Flu Detector - Tracking Epidemics on Twitter, Lecture Notes in Computer Science, vol. 6323, pp. 599–602. Springer Berlin Heidelberg (2010)
13. Lezina, G., Kuznezov, A., Braslavski, P.: Learning to predict closed questions on stack overflow. Proceedings of Kazan Federal University *(in press)* (2014)
14. Li, B., Jin, T., Lyu, M.R., King, I., Mak, B.: Analyzing and predicting question quality in community question answering services. In: Proceedings of WWW'12 (Companion volume). pp. 775–782 (2012)
15. Mendes Rodrigues, E., Milic-Frayling, N.: Socializing or knowledge sharing?: Characterizing social intent in community question answering. In: Proceedings of CIKM'09. pp. 1127–1136 (2009)
16. Oh, S., Worrall, A., Yi, Y.J.: Quality evaluation of health answers in yahoo! answers: A comparison between experts and users. In: Proceedings of the American Society for Information Science and Technology. vol. 48, pp. 1–3 (2011)
17. Paul, M., Dredze, M.: You are what you tweet: Analyzing twitter for public health. In: Proceedings of ICWSM'11 (2011)
18. Shah, C., Pomerantz, J.: Evaluating and predicting answer quality in community qa. In: Proceedings of SIGIR'2010. pp. 411–418 (2010)
19. Sondhi, P., Gupta, M., Zhai, C., Hockenmaier, J.: Shallow information extraction from medical forum data. In: Proceedinhs of COLING'2010 (Posters). pp. 1158–1166 (August 2010)
20. Wong, W., Thangarajah, J., Padgham, L.: Health conversational system based on contextual matching of community-driven question-answer pairs. In: Proceedings of CIKM'11. pp. 2577–2580 (2011)
21. Zhang, Y.: Contextualizing consumer health information searching: An analysis of questions in a social q&a community. In: Proceedings of IHI'10. pp. 210–219 (2010)