ЧТО И КАК СПРАШИВАЮТ В СОЦИАЛЬНЫХ ВОПРОСНО-ОТВЕТНЫХ СЕРВИСАХ ПО-РУССКИ?

Мухин М. (mfly@sky.ru),
Уральский федеральный университет, Екатеринбург, Россия

Браславский П. (pbras@yandex.ru),
Kontur Labs/Уральский федеральный университет,
Екатеринбург, Россия

Ключевые слова: социальные вопросно-ответные сервисы, вопросительное предложение, классификация вопросов, Ответы@Mail.Ru

WHAT DO PEOPLE ASK THE COMMUNITY QUESTION ANSWERING SERVICES AND HOW DO THEY DO IT IN RUSSIAN?

Mukhin M. (mfly@sky.ru)
Ural Federal University, Yekaterinburg, Russia

Braslavski P. (pbras@yandex.ru)
Kontur Labs/Ural Federal University, Yekaterinburg, Russia

In our study we surveyed different approaches to the study of questions in traditional linguistics, question answering (QA), and, recently, in community question answering (CQA). We adapted a functional-semantic classification scheme for CQA data and manually labeled 2,000 questions in Russian originating from Otvety@Mail.Ru CQA service. About half of them are purely conversational and do not aim at obtaining actual information. In the subset of meaningful questions the major classes are requests for recommendations, or how-questions, and fact-seeking questions. The data demonstrate a variety of interrogative sentences as well as a host of formally non-interrogative expressions with the meaning of questions and requests. The observations can be of interest both for linguistics and for practical applications.

Keywords: community question answering, interrogative sentence, question types, Otvety@Mail.Ru

Introduction

Community question answering (CQA) is a popular on-line social activity. CQA sites allow users to pose questions to other community members, to answer questions, rate questions and answers, get scores, etc. Yahoo! Answers[1], Answers.Com[2] and Otvety@Mail.Ru[3] are examples of popular general-purpose CQA services. Stackoverflow[4] is an example of a domain-specific CQA service which specializes in software programming. Quora[5] represents a newer type of such service where questions and answers can be updated, followed, interlinked, etc., thus generating potentially higher-quality content and making it more reusable.

CQA became a good complement to Web search engines: they satisfy users' complex information needs, find answers to opinionated questions and questions that imply practical experience and accounting for context. To date CQA services have collected a vast amount of data: for example, Yahoo! Answers claimed reaching one billion questions & answers in October 2009.[6] On the one hand, CQA data (not only textual, but also user activity and interaction data) help improve existing services, rethink question answering (QA) and build value-added services on top of the collected data. On the other hand, the data are a valuable linguistic resource where researchers get access to a large amount of living language material from millions of informants that is partially structured (question — list of answers) and categorized by topics. Although many CQA services look alike and some services are operated globally, the usage patterns can vary in different countries, influenced by local traditions and culture as [24] showed. Thus, analysis of language- and country-specific data is important.

In our study we develop a framework for classification of Russian-language questions originating from a popular CQA service and manually classify 2,000 questions. This task is of interest both for linguistics and for practical applications. To the best of our knowledge, this is the first study of language material of the sort.

The rest of the paper is organized as follows. The following section surveys the literature on linguistic approaches to questions and the work on classification of questions in context of QA and CQA. Section 3 briefly describes Otvety@Mail.Ru service and the data used in our study. In Section 4 we introduce a framework for classification of CQA questions; Section 5 summarizes the results of manual classification of 2,000 questions. The sixth section contains conclusions and outlines directions for future research.

[1] http://answers.yahoo.com/

[2] http://www.answers.com/

[3] http://otvet.mail.ru/

[4] http://stackoverflow.com/

[5] http://www.quora.com/

[6] http://yanswersblog.com/index.php/archives/2009/10/05/did-you-know/

Related Work

In this section we survey the three groups of related work: 1) classification approaches to interrogative sentences and questions in linguistics, 2) question classification approaches in question answering (QA), and 3) question typologies introduced within CQA research.

Question as a semantic category and interrogative sentence are asymmetric phenomena; in the semantic description of the two it is crucial to set "a clear distinction between interrogative sentence as a syntactic notion and question as a semantic category required by the information structure" [17: 233]. Most linguistic studies focus on interrogative sentences rather than questions as such. However, formal and semantic aspects are often inseparable in traditional classifications. Thus, classical studies consider interrogative sentence among the other types of sentences of different purposes of communication (such as declarative, imperative, and, in some classifications, optative sentences), which "serves to express the question posed to the other party. With the help of the question the speaker seeks to obtain new information about something..." [11: 302]. Similar definitions can be found in [7, 19], and others.

Question as a type of statement with a particular communicative task — that of inducement to obtain information [5: 707] — can be structured both in the form of an interrogative and a non-interrogative sentence. By its nature, it can be a request, a demand, etc. On the other hand, interrogative sentence in its primary function may or may not express the speaker's desire to obtain new information, i. e., to be "properly interrogative" or "improperly interrogative" [5: 708], and to have "standard" or "non-standard" interrogative semantics [17: 233–234][7]. Based on this, Bulygina and Shmelev proposed the two issues for linguistic consideration: "1) how questions are expressed (besides interrogative sentences), and 2) what function interrogative sentences fulfill (other than their primary function of expressing questions)" [5: 111]. The second problem is reflected in a large number of studies. As for the former issue, we think that social services on the web may help to make a decision on it. They are characterized by large amounts of data, different ways of information search, informal register of communication, along with the necessary restrictions on the dialogue, which are missing in regular online forums. Among these relatively new linguistic data we are primarily interested in the functional-semantic question types (not interrogative sentences as such), as well as in characteristics of potential answers embedded in the question.

Automatic classification of questions is an important problem in the area of question answering (QA). QA is a subfield of information retrieval (IR), where user information need is formulated as a natural language question (rather than a list of keywords), resulting in an exact answer or a concise document fragment containing the answer — in contrast to a ranked list of documents in a classical IR scenario. This direction of research has largely shaped and demonstrated progress thanks to the TREC

[7] From this point of view classifications of interrogative sentences (categorized as direct and indirect speech acts) are presented, besides those already mentioned works, in [3, 5, 6, 13, 20, 25], and others.

QA track (see http://trec.nist.gov/data/qa.html, [22]). The main type of questions used for QA evaluation within TREC was open-domain factual questions, or factoids, e. g. *What was the monetary value of the Nobel Peace Prize in 1989?* (Later, questions seeking for definitions and relational information were added.) The overall performance of a QA system is heavily influenced by the ability of the system to predict the type of the expected answer based on the question. TREC participants used a wide range of question typologies and classification approaches that were tightly connected both with the TREC data and the named entity recognition (NER) output. The collection of about 5,500 labeled questions known as the UIUC dataset became a de-facto standard in the field ([12], http://cogcomp.cs.illinois.edu/Data/QA/QC/). The proposed hierarchy contains six coarse classes and 50 fine classes, see Fig. 1.

Coarse	**Fine**
ABBREVIATION	abbreviation, expansion
ENTITY	definition, description, manner, reason
DESCRIPTION	animal, body, color, creation, currency, disease/medical, event, food, instrument, language, letter, other, plant, product, religion, sport, substance, symbol, technique, term, vehicle, word
HUMAN	description, group, individual, title
LOCATION	city, country, mountain, other, state
NUMERIC VALUE	code, count, date, distance, money, order, other, percent, period, speed, temperature, size, weight

Figure 1. UIUC dataset question typology [13]

Several attempts have been made to enrich the UIUC typology or to tailor it to a particular task. For example, a recent study [16] analyzes questions in Korean, including those from a search engine query log and proposes a classification scheme based on the three facets:

- Answer format (AF): *single (factoid)*, *multiple (list)*, *descriptive (definition)*, and *yes/no*.
- Answer theme (AT, similar to UIUC types) is the class of the object sought by the question, such as *person*, *location*, or *date*. A total of 147 themes are organized in a hierarchy and are derived from a NER task.
- Question qualifier (QQ) reflects a question's semantics or pragmatics; the possible values are *specification*, *superlative*, *ordering*, *definition*, etc.

CQA has recently attracted attention of researchers from different fields — information retrieval, linguistics, as well as sociology and related disciplines. [1] gives a good insight into the nature of CQA services and analyzes various aspects and characteristics of Yahoo! Answers: the differences and similarities among categories, the activity of users, their interactions, etc. CQA sites contain a vast amount of user generated content (UGC) that significantly varies in quality. [2] addresses the problem of automatic identification of high-quality questions and answers in a dataset obtained from Yahoo! Answers. Using a wide range of features — content features, usage statistics, and user relationships — the authors were able to separate high-quality items from the rest with high accuracy.

Several studies have been done on classification of questions asked on CQA services. [8] introduced the question dichotomy *conversational* vs. *informational*, where the former questions are asked purely to start discussion and the latter are aimed at satisfying an actual information need. About 500 questions from different CQA services were annotated manually according to the scheme. Then, the authors implemented a classifier based on category, question text and asker's social network characteristics. [14] investigated a similar facet of Q&A threads, namely *social* vs. *non-social* intent of the users: all questions intended for purely social engagement are considered social, while those that seek information or advice are considered non-social but instigating a knowledge sharing engagement. 4,000 questions from two different CQA services were labeled.

Harper et al. later proposed a rhetorical question typology consisting of the three type pairs [9]: Advice and Identification, (Dis)Approval and Quality, and Prescriptive and Factual. Each pair is considered to be a subspecies of Aristotelian rhetorical genres: deliberative, epideictic, and forensic, respectively. 300 Yahoo! Answers questions were labeled manually according to this typology.

[10] adopted a psycholinguistic typology for labeling about 800 questions from Yahoo! Answers focused on the following topics: data mining, natural language processing (NLP), and eLearning. The annotation scheme consisted of the nine types: Concept Completion, Definition, Procedural, Comparison, Disjunctive, Verification, Quantification, Causal, and General Information Need.

[15] investigated how people ask and answer questions in online social networks (primarily on Facebook and Twitter). The authors classified 249 questions provided by survey participants into the following categories: Recommendation, Opinion, Factual knowledge, Rhetorical, Invitation, Favor, Social connection, and Offer.

Data

Otvety@Mail.Ru (*otvety* means *answers* in Russian) is a service of Mail.Ru, one of the leading Russian web portals. Otvety@Mail.Ru is a Russian counterpart of Yahoo! Answers, with similar rules and incentives (see Fig. 2). It was launched in August 2006, and after five years has reached 50M+ users, 60M+ questions, and 335M+ answers. The service claims to have 58K new users, 52K questions, and 235K answers daily.[8]

[8] http://otvet.mail.ru/news/#hbd2011

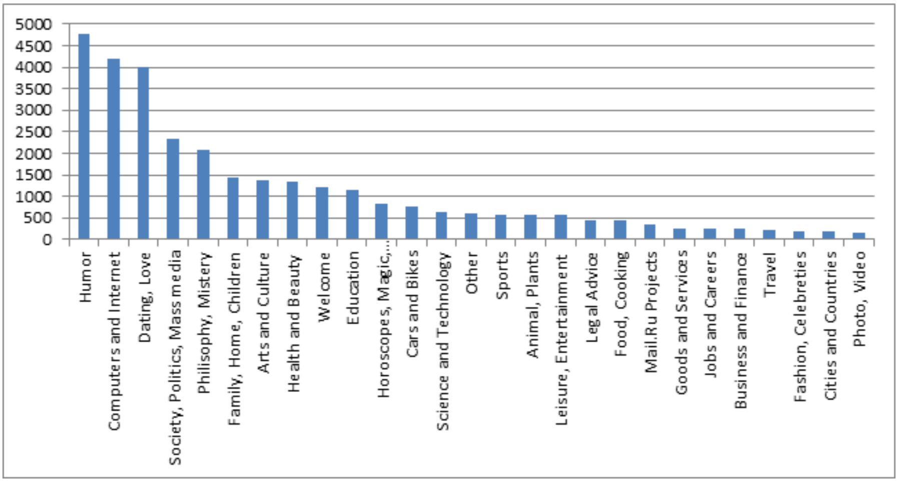
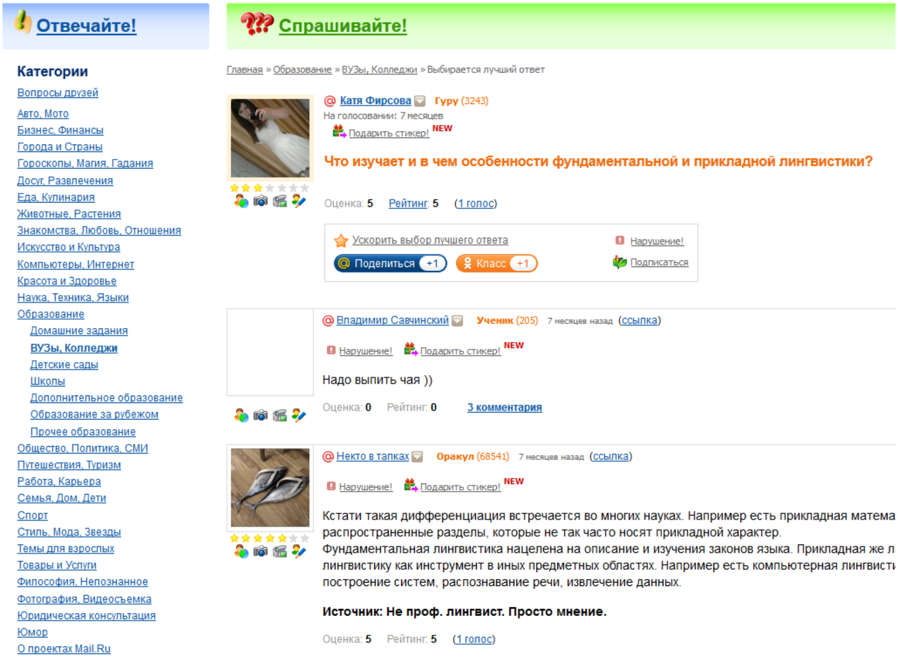


Figure 2. Otvety@Mail.Ru user interface

An Otvety@Mail.Ru question consists of a title (often it is the question itself, up to 120 characters), a detailed question description (optional) that may contain links, images, and video along with the text; tags (optional), category and subcategory (mandatory, are chosen manually by the asker from the drop-down lists).

For our initial experiments we downloaded every 1,000th question and its answers for the period from September 2009 to November 2010. It resulted in 31,223 non-empty pages. Fig. 3 shows the distribution of pages across the top-level categories. The average question length (concatenation of question title and optional question body) in our dataset is 22.5 words, while the average answer length is 19.7 words; a question receives 5.3 answers in average.



Отвечайте!

Категории

- Вопросы друзей
- Авто, Мото
- Бизнес, Финансы
- Города и Страны
- Гороскопы, Магия, Гадания
- Досуг, Развлечения
- Еда, Кулинария
- Животные, Растения
- Знакомства, Любовь, Отношения
- Искусство и Культура
- Компьютеры, Интернет
- Красота и Здоровье
- Наука, Техника, Языки
- Образование
  - Домашние задания
  - **ВУЗы, Колледжи**
  - Детские сады
  - Школы
  - Дополнительное образование
  - Образование за рубежом
  - Прочее образование
- Общество, Политика, СМИ
- Путешествия, Туризм
- Работа, Карьера
- Семья, Дом, Дети
- Спорт
- Стиль, Мода, Звезды
- Темы для взрослых
- Товары и Услуги
- Философия, Непознанное
- Фотография, Видеосъемка
- Юридическая консультация
- Юмор
- О проектах Mail.Ru

Спрашивайте!

Главная » Образование » ВУЗы, Колледжи » Выбирается лучший ответ

@ Катя Фирсова **Гуру** (3243)
На голосовании: 7 месяцев
Подарить стикер! NEW

Что изучает и в чем особенности фундаментальной и прикладной лингвистики?

Оценка: **5** Рейтинг: **5** (1 голос)

Ускорить выбор лучшего ответа Нарушение!
Поделиться +1 Класс +1 Подписаться

@ Владимир Савчинский **Ученик** (205) 7 месяцев назад (ссылка)

Нарушение! Подарить стикер! NEW

Надо выпить чая))

Оценка: **0** Рейтинг: **0** **3 комментария**

@ Некто в тапках **Оракул** (68541) 7 месяцев назад (ссылка)

Нарушение! Подарить стикер! NEW

Кстати такая дифференциация встречается во многих науках. Например есть прикладная матема
распространенные разделы, которые не так часто носят прикладной характер.
Фундаментальная лингвистика нацелена на описание и изучения законов языка. Прикладная же л
лингвистику как инструмент в иных предметных областях. Например есть компьютерная лингвисти
построение систем, распознавание речи, извлечение данных.

Источник: Не проф. лингвист. Просто мнение.

Оценка: **5** Рейтинг: **5** (1 голос)

Figure 3. Distribution of Q&A pages by category

An initial examination of the data allowed us to make some observations that imply a broad understanding of the question category in CQA.

1. A large number of "questions" with non-standard semantics do not relate to information search *per se*, but are rather invitations to conversation or an opportunity for the asker to express herself (to make a joke, to shock others, etc.), i.e. they carry only a secondary function. Examples: *Подойдет ли монтажная пена для макияжа!?))) [Is foam sealant suitable for makeup!?)))]; Как вы думаете будет ли такое время что все люди будут жить в единстве в мире и любви? [Do you think there will be a time when all people will live in unity, peace and love?]*

2. Many formally non-interrogative structures can be paraphrased and represented as traditional questions, for instance: *антивирус не обновляется из-за ошибки компилятора*; *Выбираю авто [the antivirus does not update due to compiler's error; Choosing auto mode]; Opel Astra, многие её хвалят, но хотелось бы узнать ваше мнение.[Opel Astra, many praise it, would like to know your opinion]*

3. Several questions that can imply different types of responses are combined together, e.g.: *Является ли философия наукой? И если да, то почему в ней так*

слабо развит математический аппарат?[Is philosophy a science? If yes, why is its mathematical apparatus so underdeveloped?]

4. Different uses of question title and body fields:

- Question topic in the title, detailed question in the body (well-formed structure): *Проблемы с Windows 7 || Как запускаю игру, мой ноут просто выключается. В чем проблема? [Problems with Windows 7|| When starting the game my laptop turns off. What is the problem?]*; *Капитанская дочка || Почему Маша, любя Гринёва, отказывается выйти за него замуж?! [The Captain's Daughter[9]||Why does Masha, being in love with Grinev, refuse to marry him?]*;
- The title contains the question, while the body is either empty, or contains a clarification or a request for help or answer: *Жёсткий диск какой фирмы посоветуете приобрести? || Баракуду не предлагать [What manufacturer's hard drive would you recommend?||Do not suggest Barracuda]; А кто солил арбузы? || Киньте рецептик, коль не жалко <…> [Who had made pickled watermelons? || Drop in a receipt, please…]*;
- The title contains an appeal for help, an address, or the beginning of an answer: *ответьте на вопросы плиззз))) || Для чего Князь Андрей отправляется на войну [Answer the question pleazzze))) || What for does Prince Andrew go to war[10]]; Дорогие хозяйки своего очага, подскажите, пожалуйста, || как и чем вывести жирное пятно (от крема) с дивана (шинил)? [Dear housewives, tell me, please, || how to remove a greasy spot (cream) from a sofa (chenille)?]*;
- The title and the body contain different, even if related questions: *Какова вероятность забеременеть сразу, если был незаметный выкидыш? || И что назначают врачи после выкидыша? [What are the odds of becoming pregnant shortly after a minor miscarriage? || And what do doctors prescribe after a miscarriage?]*;
- The question body merely repeats the title: *Что подарить девушке на 17 лет?[What is a good present for a girl's 17th birthday?].*

5. The wide topical range of questions spans from requests to help in solving a math problem or a crossword puzzle to requests for legal advice or a desired link, etc.

Question typology

When analyzing Otvety@Mail.Ru data we relied both on traditional approaches to question classification, and on recent work on questions in online social networks and on search queries in the form of questions (see Section 2). The proposed question

[9] A novel by Alexander Pushkin, studied in middle school.

[10] The question relates to the novel "War and Peace" by Leo Tolstoy, also studied in the school.

classification is functional-semantic by its nature, generalizing the substantial characteristics of questions rather than those of interrogative sentences.

The nature of the data implies iterative refinement of the classification (union or, in contrast, subdivision of classes, similarly to [14]) and suggests the following types of questions.

According to the main function we can distinguish the following classes:

1) Actual questions seeking for information, in broad terms: it can be traditional questions, requests for help or advice, as well as invitations to join a community, to make use of something (in this case we are talking about real, concrete facts, events, or matters);

2) *Rhetorical* questions and remarks that do not ask for information, 'chat' — the same as *conversational* in [8] and *rhetorical* in [15]: it can be an invitation to a conversation (even on a serious topic), a joke or an emotional expression:

Вы один из тех, каких много или считаете себя особенным?))) [Are you one of those who consider themselves extraordinary?];

Есть ведро солёных огурцов. Сколько надо вёдер водки?[There is a bucket of pickles. How many buckets of vodka are needed?]

Explicit questions can be further characterized by their particular functions, defined through the expected answer type or the action of the potential interlocutor. For our study we adopted the classification scheme from [15] as the top-level categories and elaborated a finer-grained layer for the two major classes.

1. *Factual knowledge* — the search for factual information. This category is further divided into the following subclasses:

- Object (*Как называется…?*[11], *Кто/Что это? [What/Who is…?]*);
- Object property (*Чем отличается…?*; *Как выглядит…?[How does X look…?]*; *Как действует…?[How does X work…?]*);
- Possibility (*Можно ли…?[Is it possible…?], Могу ли я…? [Can I…?]*, i. e. a question that asks about possibility/impossibility of doing something);
- Reason (*Почему это так…? [Why…?], О чем говорит…?[What does it mean…?]* — about the objective reason of a property or event);
- Aim (*Зачем нужен…?, Для чего…? [What for…?]*);
- Time (*Во сколько…?, Когда произойдет…? [When…? At what time will X happen?]*).

2. *Recommendation* — a question or request of the type «*please tell me, how to…*». A further specification is the following:

- Method (literally *Как сделать…? [How to make…?]*);
- Information search (predominantly on-line — *Где найти…?[Where can I find…?], Как узнать?[How can I know…?]* etc.);

[11] Here we consider both formal interrogative constructions and statements that can be reformulated as a question. For example, a question of type "Object" can be simply a link to an online image.

- Location, directions (*Куда поехать?*, *Где находится…? [Where is X located…? How can I get to…?]*— in the geographic sense).

3. *Opinion — Как вы относитесь к…? [What do you think about…?], Что вы предпочитаете?[What do you prefer…?].*

4. *Favor* — asking for help: *Пришлите ссылку… [Please send a link…], Решите задачу… [Please solve a problem…],* etc.

5. *Offer* (*Кому нужен… [Who needs…?], Отдам/Продам…[…to sell/…to give away]*).

6. *Social connections* — the search for people, business companions, friendship, love and sexual relations.

We also complemented the hierarchy above with the expected answer type following [16] (see Section 2).

Results

2,000 randomly sampled questions from the dataset were labeled manually by one of the authors according to the proposed scheme (Section 4). What follows is a summary of the main results.

993 cases (49.7 %) were assigned to the *rhetorical* class ('chat'); 1007 (50.3 %) were actual questions seeking for meaningful information. A further division of the question class is presented in Table 1. As one can see, the majority of the questions are seeking for procedural knowledge (*Recommendation*) or facts (*Factual Knowledge*); are asking for a favor (*Favor*) or inquire *Opinions* of others. Many questions of the *Favor* type are connected with acquisition of procedural knowledge (e. g. asking for help with a math assignment), thus we can conclude that the questions that can be reformulated as *How to make something?* prevail at Otvety@Mail.Ru. Questions that imply social interaction (*Offer* and *Social Connection*) are presented marginally.

Table 1. Subclasses of actual question class (total 1007)

Question type	Count	%
1. Factual knowledge	343	34.1
2. Recommendation	391	38.8
3. Opinion	123	12.2
4. Favor	135	13.4
5. Offer	4	0.4
6. Social connection	11	1.1

Table 2. Subdivision of two major question classes

	Count	%
1. Factual knowledge	**343**	**100**
1.1. Object	119	34.7
1.2. Object property	135	39.4
1.3. Possibility	41	12.0
1.4. Reason	35	10.2
1.5. Aim	7	2.0
1.6. Time	6	1.7
2. Recommendation	**391**	**100**
2.1. Method	251	64.2
2.2. Information search	120	30.7
2.3. Location, directions	20	5.1

The data in Table 2 are fairly predictable, as factual questions deal mostly with objects and their properties. Note the presence of the subgroup *Possibility* with such questions as whether it is possible to file a court petition or, for example, to buy a new SIM card.

Table 3. Breakdown of the sample by expected answer type

Answer type	Count	%
Yes/no	66	6.6
Single	241	23.9
Multiple	166	16.5
Descriptive	534	53.0

Table 3 shows that the majority of questions imply a detailed answer (description), i. e. a simple statement of a fact would not be sufficient. In contrast, there are very few yes/no questions (*Is the temperature of 46 degrees OK for an Intel Core 2 processor? Can I have a badger at home?*). As Table 4 reveals, the majority of the *Recommendation* questions imply a detailed response. A portion of the fact-seeking questions expects yes/no; however, short and descriptive answers prevail.

Table 4. Distribution of questions by expected answer type in two major classes

Answer type	Factual knowledge (total 343)	%	Recommendation (total 391)	%
Yes/no	44	12.8	1	0.3
Single	118	34.4	61	15.6
Multiple	40	11.7	88	22.5
Descriptive	141	41.1	241	61.6

Conclusions

In our study we refined and combined the question classification schemes reported in the literature. 2,000 questions in Russian originating from Otvety@Mail.Ru CQA service were tagged manually. The dataset is one of the biggest manually tagged collections of CQA questions reported in the literature (only the dataset reported in [14] excels in size) and the first Russian-language collection of this kind, to the best of our knowledge. About half of the questions is aimed rather at self-expression, joking and chatting than seeking for information and knowledge sharing. The portion of the "entertaining" questions seems to be higher in Otvety@Mail.Ru than in other similar CQA services according to [8, 14] (however, all the studies including ours use different classification schemes, so any comparisons should be done carefully). Obviously, this proportion of meaningful and conversational questions is determined by many features and characteristics of a CQA service: the broad audience, its social and demographic characteristics, absence of topical focus, the system of incentives, and the moderation policy.

Questions on CQA services demonstrate a different behavior compared to question-like status messages on Facebook and Twitter [15] — for instance, there are far fewer rhetorical and fact-seeking questions. The reason could be that in online social networks users can satisfy their needs in socializing and conversations without resorting to question-like messages; users seek chiefly for recommendations and opinions from their immediate social network.

It is interesting to note that the prevalence of *recommendations*, or *how* questions in our data mirrors the trend in search engine query logs: the *how* question-like queries surpass other pragmatics [18, 23]. The situation has changed a lot during the last decade: in the late 90s most search engine queries in question form sought for factual information [21].

The processed data demonstrate a variety of interrogative sentences as well as of formally non-interrogative expressions with meaning of questions and requests. The obtained data can be of interest for linguists, sociologists, and communication researchers. The data can also be used for improving existing CQA services and can contribute to question answering research. The manually tagged sub-corpus of questions is available for research purposes at http://kansas.ru/cqa/data/.

In our future research we are going to address the problem of automatic classification of questions, refine our classification scheme, as well as compare CQA questions with questions in search engine query logs.

We would like to thank Mail.Ru for granting us access to the data and Tanya Kondakova for initial data preparation.

References

1. *Adamic L. A., Zhang J., Bakshy E., and Ackerman M. S.* Knowledge Sharing and Yahoo! Answers: Everyone Knows Something. Proceedings of the 17th International Conference on World Wide Web (WWW '08). ACM, New York, NY, USA, 2008, pp. 665–674.
2. *Agichtein E., Castillo C., Donato D., Gionis A., Mishne G.* Finding High Quality Content in Social Media. Proceedings of the International Conference on Web Search and Web Data Mining (WSDM '08). ACM, New York, NY, USA, 2008, pp. 183–194.
3. *Bally Ch.* Linguistique générale et linguistique française [Obshchaja lingvistika i voprosy francuzskogo jazyka]. Moscow, Inostrannaya Literatura, 1955, p. 416.
4. *Beloshapkova V. A.* Types of Sentences for Aim [Tipy predlozhenij po celeustanovke]. Sovremennyj russkij jazyk: Uchebnik dlja vuzov [The Modern Russian Language: Textbook for Universities]. Edited by V. A. Beloshapkova, 3rd ed. Moscow, Azbukovnik, 1999, pp. 705–708.
5. *Bulygina T. V., Shmelev A. D.* Interrogative sentences and Questions [Voprositel'nye predlozhenija i voprosy]. Chelovecheskij faktor v jazyke: Kommunikacija, modal'nost', dejksis [The Human Factor in Language: Communication, Modality, Deixis]. Moscow, Nauka, 1992, pp. 111–116.
6. *Conrad R.* Interrogative Sentences as Indirect Speech Acts [Voprositel'nye predlozhenija kak kosvennye rechevye akty]. Novoe v zarubezhnoj lingvistike. Vyp. 16: Lingvisticheskaja pragmatika [Novelty in Foreign Linguistics: Linguistic Pragmatics]. Vol. 16. Moscow, Progress, 1985, pp. 349–383.
7. *Gvozdev A. N.* The Modern Russian Literary Language: Syntax: Part 2 [Sovremennyj russkij literaturnyj jazyk: Sintaksis: Ch. 2]. 5th ed. Moscow, Librokom, 2009. 352 p.
8. *Harper F. M., Moy D., and Konstan J. A.* Facts or Friends? Distinguishing Informational and Conversational Questions in Social Q&A Sites. Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09). ACM, New York, NY, USA, 2009, pp. 759–768.
9. *Harper F. M., Weinberg J., Logie J., and Konstan J.* (2010), “Question types in social Q&A sites”. First Monday [Online], Volume 15 Number 7 (4 July 2010), available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2913/2571
10. *Ignatova K., Toprak C., Bernhard D., and Gurevych I.* Annotating question types in social Q&A sites. Tagungsband des GSCL Symposiums “Sprachtechnologie und eHumanities” Abteilung für Informatik und Angewandte Kognitionswissenschaft Fakultät für Ingenieurwissenschaften Universität Duisburg-Essen, 2009, pp. 44–49.
11. *Lekant P. A.* Interrogative Sentences [Voprositel'nye predlozhenija]. Sovremennyj russkij literaturnyj jazyk: Uchebnik dlja vuzov [The Modern Russian Literary Language: Textbook for Universities]. Edited by P. A. Lekant, 4th ed. Moscow, Vysshaja Shkola, 1999, pp. 302–305.
12. *Li X., Roth D.* Learning Question Classifiers. Proceedings of the 19th International Conference on Computational Linguistics — Volume 1 (COLING '02),

Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 2002, pp. 1–7.

13. *Lindstrem E. N.* Classification of Russian Interrogative in Form Statements on the Basis of a Pragmatic-Based Universal Model [Klassifikacija russkih voprositel'nyh po forme vyskazyvanij na baze pragmaticheski obosnovannoj universal'noj modeli]. PhD diss. Petrozavodsk, 2003.
14. *Mendes-Rodrigues E. and Milic-Frayling N.* Socializing or Knowledge Sharing? Characterizing Social Intent in Community Question Answering. Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09). ACM, New York, NY, USA, 2009, pp. 1127–1136.
15. *Morris M. R., Teevan J., and Panovich K.* What Do People Ask Their Social Networks, And Why?: A Survey Study Of Status Message Q&A Behavior. Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI '10). ACM, New York, NY, USA, 2010, pp. 1739–1748.
16. *Oh H.-J., Sung K.-Y., Jang M.-G., Myaeng S. H.* (2011) Compositional Question Answering: A Divide and Conquer Approach. Information Processing & Management, Vol. 47, Issue 6, pp. 808–824.
17. *Paducheva E. V.* The Statement and Its Correlation with Reality [Vyskazyvanie i ego sootnesennost' s dejstvitel'nost'ju]. Moscow, Nauka, 1985. 271 p.
18. *Pang B., Kumar K.* Search in the Lost Sense of “Query”: Question Formulation in Web Search Queries and its Temporal Changes. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers — Volume 2 (HLT '11), Vol. 2. Association for Computational Linguistics, Stroudsburg, PA, USA, 2011, pp. 135–140.
19. *Rozental' D. E., Telenkova M. A.* Dictionary of Linguistic Terms [Slovar'-spravochnik lingvisticheskih terminov]. Moscow, Prosveshchenie, 1976.
20. *Russian* Grammar [Russkaja Grammatika]. Moscow, Nauka, 1980. Vol. 2.
21. *Spink A., Ozmultu H. C.* (2002) Characteristics of Question Format Web Queries: An Exploratory Study. Information Processing & Management, Vol. 38, Issue 4, pp. 453–471.
22. *Voorhees E. M., Dang H. T.* Overview of the TREC 2005 Question Answering Track. Proceedings of TREC-14, Gaithersburg, MD, USA, 2005.
23. *Yandex* (2010), “Questions to Yandex” (in Russian), available at: http://company.yandex.ru/researches/figures/ya_questions_2010.xml
24. *Yang J., Morris M. R., Teevan J., Adamic L. A., and Ackerman M. S.* Culture Matters: A Survey Study of Social Q&A Behavior. International AAAI Conference on Weblogs and Social Media, North America, 2011. Available at: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2755.
25. *Zhinkin N. I.* Question and Interrogative Sentence [Vopros i voprositel'noe predlozhenie]. Jazyk — Rech' — Tvorchestvo. Issledovanija po semiotike, psiholingvistike, pojetike [Language — Speech — Creativity. Research on Semiotics, Psycholinguistics, Poetics]. Moscow, Labirint, 1998, pp. 87–103.