Russian-Language Question Classification: A New Typology and First Results

Kirill Nikolaev 🆔 and Alexey Malafeev(✉) 🆔

National Research University Higher School of Economics,
Nizhny Novgorod, Russia
`kinikolaev@edu.hse.ru`, `aumalafeev@hse.ru`

Abstract. This paper deals with automatic classification of questions in the Russian language, a natural early step in building a question answering system. We developed a typology of Russian questions using interrogative particles, pronouns and word order as the main features. A corpus of 2008 questions was manually compiled and annotated according to our typology. We used a fine-grained class set and a coarse-grained one (23 and 14 classes, respectively). The training data, represented as character bi-/trigrams and word uni-/bi-/trigrams, was used to approach the task of question classification. We tested several widely used machine-learning methods (logistic regression, support vector machines, naïve Bayes) against a regular expression baseline on a held-out test corpus annotated by an external expert. The best results were achieved by a SVM classifier (linear kernel) that achieved the accuracy of 65.3% (fine-grained) and 68.7% (coarse-grained), while the baseline regular expression model showed 52.7% accuracy.

Keywords: Question answering · Question answering systems · QA systems
Russian-language questions · Question classification · Question tagging
Russian question typology

1 Introduction

Interrogative sentences have a very important function in communication. Questions have a cognitive meaning without a common attributive proposition, because they do not usually include affirmation or negation. They are mostly used for acquiring knowledge and are key tools of cognition (Nevolnikova 2004).

From the natural language processing perspective, interrogative sentences are often key language material when one works on such tasks as question answering, building dialogue agents, and discourse modeling. In this paper, we focus on automatic question typology for the domain of question answering. According to Burger et al. (2001), identifying question classes is the first step in building a QA system.

In contrast to information retrieval (IR) systems, which usually return a set of documents relevant to some keywords, question-answering (QA) systems aim at yielding an exact answer to the question (Monz 2003a, b). Question answering is considered a more difficult task due to the constraints on input (natural language questions vs keywords) and output (focused answers vs entire documents) representation (Bunescu and Huang 2010).

Yet, the benefit of QA systems is that they do not overwhelm the user with excess information (Galea 2003).

Various systems have been developed for answering questions in English (e.g. TEQUESTA (Monz 2003a, b), START (Katz et al. 2006), OpenEphyra (van Zaanen 2008), IBM Watson (Ferucci et al. 2010), EAGLi (Gobeill et al. 2012), etc.). An excellent survey of state-of-the-art learning-based methods for English-language question classification was conducted by Loni (2011).

As far as Russian-language QA systems are concerned, they have had very limited coverage in literature. Although a monograph by Sosnin (2007) and a few research papers, such as (Suleymanov 2001; Tikhomirov 2006; Mozgovoy 2006; Solov'ev and Peskova 2010), have been published, they mostly contribute to the theory on the problem, rather than propose and/or evaluate efficient practical solutions.

Our work differs from the papers previously published by other researchers in that we attempt to solve the practical task of automatic typologization of questions in the Russian language using complex class sets (23 question types in the fine-grained and 14 in the coarse-grained set), while in other papers other approaches have been used. For example, in (Sosnin 2007), only three general question classes are distinguished: (1) "problem", (2) "task", and (3) "inquiry" (p. 118). Alternatively, in (Solov'ev and Peskova 2010), a paper devoted to question analysis for a Russian-language question answering system, a detailed question taxonomy was used. It was borrowed from (Ittycheriah 2008) with some modification, but we consider it too fine-grained to be practical. It has 39 classes, some of which are very specific and rare, e.g. *Organ*, *Salutation*, *Plant*, etc. Also, the differences between some of the classes do not seem very well-defined, e.g. *Areas* vs *Geological objects* vs *Location* vs *Country*, or *Company-roles* vs *Occupation*, etc. Furthermore, the question analysis technique used was trivial: a limited number of key words were simply searched for in the question. Expectedly, the performance of the module was quite low – 67% error (Solov'ev and Peskova 2010, p. 48). In contrast, we report question tagging accuracy of up to 68.7%. That said, the results are obviously not directly comparable because different classifications are used.

In the next section, we describe the question typology that the developed classifier is based on. Sections 3 and 4 are devoted to the regular expression baseline and the machine learning classification methods used, respectively. Section 5 discusses the results attained. Finally, in Sect. 6 we draw conclusions and outline some directions for future work.

2 Interrogative Sentences in the Russian Language

In our research, we focus on the functional aspect of interrogative sentences. According to Shvedova, information gathered via interrogative sentences can be of various nature: about the subject of some action (*Кто это сделал?/Who did it?*), the object (*Что было сделано?/What was done?*), the goal (*Для чего ей это?/Why does she need this?*), etc. (Shvedova 1980). The main question formation means are intonation, interrogative particles (*ли, так, верно, как, что ли, etc.*), interrogative pronominal

words (*где, куда, когда, откуда, почему, зачем, как, etc.*) and word order. It is possible to use some of these means as features in classification.

The first question typology that we considered was developed by Shvedova (1980). She divides interrogative sentences as follows:

1. According to the type and volume of required information:
    a. General: acquiring the information about the situation in whole. *Что происходит в Китае?/What is happening in China?*
    b. Special: acquiring the information about an aspect of the matter. *Какие песни сделали её известной?/Which songs made her famous?*
2. According to what answer is expected:
    a. Requiring a confirmation: yes/no or true/false: *Лондон – столица Англии?/Is London the capital of England?*
    b. Requiring information: *Зачем они пьют чай?/What do they drink tea for?*

It can be seen that this typology is too general for the purposes of automatic question answering. Another functional classification we used for building our own typology is Graesser's Taxonomy of Inquiries (Lauer et al. 2013). The taxonomy is presented in Table 1.

Table 1. Arthur Graesser's Taxonomy of Inquiries

Question	Abstract specification
Verification	Is a fact true?
	Did an event occur?
Comparison	How is X similar to /different from Y?
Disjunctive	Is X or Y the case?
Concept completion	Who? What? When? Where?
Definition	What does X mean?
Example	What is an example of X?
Interpretation	How is the particular event interpreted or summarized?
Feature specification	What qualitative attributes does X have?
Quantification	What is the value of a quantitative variable?
Casual antecedent	What caused some event to occur?
Casual consequence	What are the consequences of an event/state?
Goal orientation	What are the motives behind an agent's actions?
Enablement	What object or resources enable an agent to perform an action?
Instrumental	How does an agent accomplish the goal?
Expectational	Why did some expected event not occur?
Judgmental	The questioner wants the answerer to judge an idea or give an advice what to do
Assertion	The speaker expresses that he or she is missing some information
Request/Directive	The speaker directly requests the information

Based on these classifications and drawing upon the functional aspect of interrogative sentences in Russian, we created a new question typology for the purposes of our research (Table 2).

Table 2. Russian question typology

Tag	Numeric tag	Wording examples
General	1	*Что происходит в …?/What is happening in …?*
Verification	2	*Правда ли, что …?/Is it true that …?*
Definition	3	*Что означает/такое?/What is …? What does … mean?*
Example	4	*Приведи пример…?/Give an example of …?*
Comparison	5	*Чем похожи/отличаются…?/What are the similarities/differences between …?*
Choice	6	*X или Y?/X or Y?*
Concept Completion	7	[Further subdivided into:]
(a) Agent	71	*Кто?/Who?*
(b) Object	72	*Что?/What?*
(c) Location	73	*Где? Куда? Откуда?/Where? From where?*
(d) Date:	74	*Когда? Какого числа?/When? What date?*
• Date-birth	741	*Когда родился?/When was … born?*
• Date-death	742	*Когда умер?/When did … die?*
(e) Time	75	*Во сколько?/What time?*
Quality	8	*Какой?/What kind of?*
Quantity	9	*Сколько? Как много?/How many/much?*
(a) Quantity-age	91	*Сколько лет? В каком возрасте?/How old? At what age?*
(b) Quantity-time	92	*Сколько времени?/What time?*
(c) Price	93	*Сколько стоит?/How much is …?*
Action	10	*Что делать, чтобы…?/What do I do to/if …?*
Instrument	11	*С помощью чего? Каким методом?/With what? How?*
Goal	12	*К чему? Зачем?/For what?*
Reason	13	*Почему?/Why?*
Consequence	14	*Каковы последствия?/What are the consequences of …?*

As can be seen from the Table, there are 23 distinct question classes in our typology. Since some classes are further subclassed, this is also a taxonomy of question types. If we flatten the taxonomy, we get 14 general classes (the coarse-grained class set). We use both fine-grained and coarse-grained class sets in our experiments (Sect. 4).

Some comments need to be made on our classification. Firstly, the proposed classification is by no means a complete listing of all possible question types in Russian. When developing the classification, we aimed at the balance between completeness and practicality. Thus, certain question types not so commonly used in

practice were omitted: e.g. *как часто? (how often?)* or *до какого времени? (up to what time?).*

Another important comment is that not all wordings for each question type are listed in the table; for the sake of brevity, we list only the most common ones, although other wordings are possible. For example, apart from *почему? (why?)*, a question of the "Reason" type may be worded as: по какой причине? (*for what reason?*), *отчего? (why?)*, etc.

The "General" question type might need explanation. In this type of questions, a general summary of a situation is requested, e.g. *Что происходит в Китае?/What is happening in China?* In contrast, "Verification" questions require a short yes/no answer, e.g. *Верно ли, что кошки не летают?/Is it true that cats don't fly?*

In the following two sections, we will describe our attempt at solving the Russian question classification task using the developed typology of questions.

3   Baseline Question Tagging Method: Regular Expressions

Most QA systems first classify questions based on the type (related to such question words as "What", "Why", "Who", "How", "Where"), which is followed by the identification of the answer type (Damljanovic et al. 2010). Manually designed regular expression are the most obvious tool for identifying the question type, and they are successfully used in many QA systems, e.g. TEQUESTA (Monz 2003a, b).

We implemented our own regular expressions-based question classifier in Python 3. With a more or less complex pattern for each question type, this classifier was used as the baseline method in our work. Some examples of the patterns are given in Table 3.

Table 3. Sample regular expression patterns for Russian interrogative sentences

Tag	Numeral tag	Example patterns
Example	4	/.*(([пП]риведи|[кК]акой пример|образец)|(([чЧ]то|[кК]то) ((может ((служить)|(выступ(ать|ить))))|((по?)служит|выступ(ает|ит))) (как)? ((пример(ом?))|(образ(ец|цом))))).*[\?\.]/
Quality	8	/.*(([кК]ак(ой|ая|ое|ом|ие))|([кК]ак(им|ими|ие) ((свойств(ом|ами|а))|(качеств(ом|ами|а))) (наделен|обладает|характеризуется| отличается|имеет))).*[\?\.]/
Goal	12	/.*((([кК] чему)|([зЗ]ачем)|([сС] как(ой|ими) (цел(ью|ями)|(задач(ей| ами)))|([дД]ля чего)|(([вВ]о имя)|([дД]ля как(ой|их)) (цел(и|ей))| (задач(и?)))|([кК]ак(ую|ие) (цел(ь|и))|(задач(у|и)))).*[\?\.]/

In our implementation, identifying the type of the question with regular expressions is incremental. A string variable is gradually matched, via re.match(pattern, string) method, against all example patterns, starting with the simplest (1, 2, 71, 72, 73), and ending with the most complex (8, 10, 11). The latest matching pattern is chosen as the hypothetic answer.

We evaluated the baseline classifier (as well as other classifiers described in the next section) on a held-out test set of 150 questions manually annotated by an external expert in accordance with our classification. The regular expression classifier correctly matched 79 questions (52.7% accuracy).

4 Applying Machine Learning to Question Tagging in Russian

Many early approaches to determining question types employ manually designed rules or regular expressions and are non-probabilistic in that a pattern/set of conditions is either matched or not. Pinchak and Lin point out in (2006), however, that such an approach has two major drawbacks:

1. There will always be questions whose types do not match the patterns;
2. The predetermined granularity of the categories leads to a trade-off between how well they match the actual question types and how easy it is to build taggers and classifiers for them.

Thus, a probabilistic answer type model that directly computes the degree of match between a potential answer and the question context is much more effective. Such algorithms are used in some works on QA for English: (Li and Rot 2002; Zhang and Le 2003; Pereira et al. 2009), etc.

To train our question classifiers, we needed a set of questions in Russian, annotated for question types. Due to the unavailability of a suitable collection of questions, we semi-automatically extracted a set of 2008 questions from the Russian Internet Corpus (Sharoff et al. 2006), and manually tagged each question in accordance with our typology. We used two different annotation sets: with the simplified "Concept Completion" and "Quantity" class groups (14 classes total) and the original detailed classification (23 classes).

The questions were converted to five different bag-of-word representations each (character bigrams, character trigrams, word unigrams, word bigrams, word trigrams) and then used for the learning of some traditional machine learning algorithms. This was done in RapidMiner, a cross-platform software framework developed on an open-core model and providing multiple solutions for machine learning, data and text mining, predictive analytics, etc. (Klinkenberg 2013).

For automatic classification of questions, we trained (using the '1 vs All' classification strategy) three different machine learning algorithms - naïve Bayes, support vector machine and logistic regression. Ten different datasets, depending on the set of classes – fine-grained or coarse grained, – and type of representation (word unigrams/bigrams/trigrams, and character bigrams/trigrams) were used with each algorithm. During evaluation, we ensured that the system relied only on the n-grams observed in training. All n-grams not seen in the training data were ignored, i.e. no strategy for dealing with unknown (not previously seen) n-grams was implemented.

The trained models were tested on a held-out test set of 150 questions manually annotated by an external expert in accordance with our classification. The training and test sets of questions, as well as the models used in this study, are made freely available

to the community[1]. The distribution of coarse-grained question types in the training and test sets is illustrated in Table 4. Table 5 shows the evaluation results for both the fine-grained and coarse-grained class sets.

Table 4. Question type distribution in the training and test sets.

Set	1	2	3	4	5	6	7	Total
Training	153	424	64	10	15	76	579	
Test	15	34	6	3	2	10	25	
	8	9	10	11	12	13	14	
Training	186	49	59	9	116	201	67	2008
Test	17	10	1	2	5	16	4	150

Table 5. Prediction accuracy for question tagging.

Algorithm and model	Correct matches (fine-grained)	Accuracy (fine-grained)	Correct matches (coarse-grained)	Accuracy (coarse-grained)
NB, c-bi	23	15.3%	19	12.7%
NB, c-tri	39	26%	39	26%
NB, w-uni	38	25.3%	39	26%
NB, w-bi	61	40.7%	61	40.7%
NB, w-tri	61	40.7%	62	41.3%
SVM, c-bi	65	43.3%	76	50.7%
SVM, c-tri	69	46%	73	48.7%
SVM, w-uni	67	44.7%	74	49.3%
SVM, w-bi	73	48.7%	82	54.7%
SVM, w-tri	68	45.3%	81	54%
LR, c-bi	55	36.7%	63	42%
LR, c-tri	20	13.3%	40	26.7%
LR, w-uni	58	38.7%	60	40%
LR, w-bi	90	**59.3%**	92	61.3%
LR, w-tri	88	56%	97	**64.7%**

We also studied how prediction quality changed with the normalization of the training data set and using different SVM kernels. This is illustrated in Table 6:

Thus, the best classification result (65.3% acc. fine-grained and 68.7% acc. coarse-grained) was achieved by a support vector machine with a linear kernel, using the word trigram text representation. This is consistent with the state-of-the-art results for English question classification reported in (Silva et al. 2011), also obtained with a linear SVM classifier.

---

[1] https://github.com/Pythonimous/Q-A-System.

Table 6. Normalization, linear SVM kernel and system performance

Algorithm, setup	Correct matches (fine-grained)	Accuracy	Correct matches (coarse-grained)	Accuracy
SVM (dot kernel), w-tri, normalized proportion	68	45.3%	81	54%
SVM (linear kernel), w-tri, normalized proportion	98	**65.3%**	103	**68.7%**

5 Discussion

After evaluating all classification algorithms trained on the different datasets, we can observe the following. Firstly, as expected, using the coarse-grained class set was an easier task that resulted in higher accuracy than for the fine-grained set. The accuracy difference ranged from 0.6% for naïve Bayes to 8.7% for logistic regression (considering the best text representation models for each ML algorithm). Using normalization and the linear kernel for the SVM allowed us to increase classification accuracy to its highest of 65.3% and 68.7% for the fine-grained and the coarse-grained class sets, respectively. To compare, the regular expression baseline showed a stable 52.7% accuracy, which proved considerably better than naïve Bayes, but less effective than the two other algorithms.

The accuracy is quite low in comparison with the results obtained for English-language datasets. Indeed, the state-of-the-art result reported by Silva et al. (2011) is 95% and 90.8% for 6 coarse-grained and 50 fine-grained classes, respectively. However, this is to be expected for a number of reasons. Firstly, Silva et al. used a much larger dataset, published by Li and Roth (2002), consisting of 5500 training and 500 test questions. Secondly, there are many freely available NLP tools for English, which allows for extracting various features useful for question classification. In particular, apart from word unigrams, the classifier by Silva et al. used such features as headwords, hypernyms and indirect hypernyms. Lastly, a lot of research was published on question classification for English, as shown in (Loni 2011), which cannot be said about Russian-language question tagging. Since the Russian language is very different from English in both morphology and syntax, the methods used for the classification of questions in English are not always equally effective or even applicable for Russian.

The confusion matrix (not presented here due to size constraints, but available upon request) for the top-accuracy algorithm was also analyzed. It was found that the "Instrument", "Example", "Action", and "Definition" questions were predicted with less than 50% accuracy. These categories were present in training set with a low frequency: 0.004, 0.004, 0.029, and 0.031 per 1000 questions, respectively. The most accurately predicted question types were "Consequence", Reason" and "Goal".

6 Conclusion and Future Work

The main theoretic contribution of our paper is a classification of Russian questions, consisting of 14 coarse-grained and 23 fine-grained classes. Using this typology, we have applied machine learning methods to the task of automatic classification of

Russian questions. We tested a regular expression baseline and three different classifiers using five different sets of features (character bi-/trigrams, word uni-/bi-/trigrams). The best classifier, SVM (linear kernel), trained on pre-normalized (via proportion transformation) word trigrams), achieved the classification accuracy of 65.3% and 68.7% for the fine-grained and the coarse-grained class sets, respectively, in contrast to the 52.7% baseline result (regular expression model).

Presented in this paper is one of the very few attempts to solve the Russian question tagging task using large class sets. The methods employed in our work showed relatively good results (given the class set sizes) in comparison with what is reported for Russian-language question classification in the literature, although there is still a lot of room for improvement. Indeed, much work needs to be done to approach the above 90% accuracy achieved for English by the research community.

Our work can be used for building a complete Russian QA pipeline or as a standalone question tagging solution. We believe that it should be possible to improve our results by further expanding the training data set, as well as using more complex machine learning algorithms (neural networks) and features (word2vec).

References

Bunescu, R., Huang, Y.: Towards a general model of answer typing: question focus identification. In: Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2010), RCS Volume, pp. 231–242 (2010)

Burger, J., Cardie C., Chaudhri V., Gaizauskas R., Harabagiu S., Israel D., Jacquemin, C., Lin, C. Y., Maiorano, S., Miller, G., Moldovan, D.: Issues, tasks and program structures to roadmap research in question & answering (Q&A). In: Document Understanding Conferences Roadmapping Documents, pp. 1–35 (2001)

Damljanovic, D., Agatonovic, M., Cunningham, H.: Identification of the Question focus: combining syntactic analysis and ontology-based lookup through the user interaction. In: LREC (2010)

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A.A., Lally, A., Murdock, J.W., Nyberg, E., Prager, J., Schlaefer, N.: Building Watson: an overview of the DeepQA project. AI Mag. **31**(3), 59–79 (2010)

Galea, A.: Open-domain surface-based question answering system. In: Proceedings of CSAW, vol. 3 (2003)

Gobeill, J., Pasche, E., Teodoro, D., Veuthey, A.L., Ruch, P.: Answering gene ontology terms to proteomics questions by supervised macro reading in Medline. EMBnet Journal **18**(B), 29–31 (2012)

Ittycheriah, A.: A statistical approach for open domain question answering. In: Strzalkowski, T., Harabagiu, S.M. (eds.) Advances in Open Domain Question Answering, vol. 32, pp. 35–69. Springer, Dordrecht (2008). https://doi.org/10.1007/978-1-4020-4746-6_2

Katz, B., Borchardt, G.C., Felshin, S.: Natural language annotations for question answering. In: FLAIRS Conference, pp. 303–306 (2006)

Klinkenberg, R. (ed.): RapidMiner: Data Mining Use Cases and Business Analytics Applications. Chapman and Hall/CRC, Boca Raton (2013)

Lauer, T.W., Peacock, E., Graesser, A.C.: Questions and Information Systems. Psychology Press, Routledge (2013)

Li, X., Roth, D.: Learning question classifiers. In: Proceedings of the 19th International Conference on Computational Linguistics, Association for Computational Linguistics, vol. 1, pp. 1–7 (2002)

Loni, B.: A survey of state-of-the-art methods on question classification. Literature survey, Published on TU Delft Repository (2011)

Monz, C.: Document retrieval in the context of question answering. In: Sebastiani, F. (ed.) Advances in Information Retrieval, ECIR 2003. LNCS, vol. 2633, pp. 571–579. Springer, Heidelberg (2003a). https://doi.org/10.1007/3-540-36618-0_44

Monz, C.: From Document Retrieval to Question Answering. Institute for Logic, Language and Computation (2003b)

Mozgovoy, M.V.: A simple question-answering system based on a semantic analyzer for the Russian language [Prostaya voprosno-otvetnaya sistema na osnove semanticheskogo analizatora russkogo yazyka], Vestnik of the St. Petersburg University. Series 10. Applied mathematics. Informatics. Management processes [Vestnik SPbGU. Seriya 10. Prikladnaya matematika. Informatika. Protsessy upravleniya], no. 1, pp. 116–122 (2006)

Nevolnikova, S.V.: Functional and semantic types of Russian interrogative sentences and their role in text formation [Funktsional'no-semanticheskie raznovidnosti russkikh voprositel'nykh predlozheniy i ikh rol' v tekstoobrazovanii]. Rostov-on-Don (2004)

Pereira, F., Mitchell, T., Botvinick, M.: Machine learning classifiers and fMRI: a tutorial overview. Neuroimage **45**(1), S199–S209 (2009)

Pinchak, C., Lin, D.A.: Probabilistic Answer Type Model. In: EACL (2006)

Sharoff, S.: Creating general-purpose corpora using automated search engine queries. In: WaCky, pp. 63–98 (2006)

Shvedova, N.Y.: Russkaja Grammatika [Russian Grammar]. AN SSSR Publ, Moscow (1980)

Silva, J., Coheur, L., Mendes, A.C., Wichert, A.: From symbolic to sub-symbolic information in question classification. Artif. Intell. Rev. **35**(2), 137–154 (2011)

Solov'ev, A.A., Peskova, O.V.: Building a question-answering system for the Russian language: question analysis module [Postroenie voprosno-otvetnoy sistemy dlya russkogo yazyka: modul' analiza voprosov], New information technologies in automated systems [Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh], no. 13, pp. 41–49 (2010)

Sosnin, P.I.: Question-Answer Modeling in the Development of Automated Systems [Voprosno-otvetnoe modelirovanie v razrabotke avtomatizovannykh sistem]. Ul'yanovsk, USTU (2007)

Suleymanov, D.S.: A study of the basic principles of building a semantic interpreter for questions and answers in natural language in AOS [Issledovanie bazovykh printsipov postroeniya semanticheskogo interpretatora voprosno-otvetnykh tekstov na estestvennom yazyke v AOS]. Educational technologies and society [Obrazovatel'nye tekhnologii i obshchestvo], no. 3, pp. 178–192 (2001)

Tikhomirov, I.A.: Question-answering search in the intelligent search system Exactus [Voprosno-otvetnyy poisk v intellektual'noy poiskovoy sisteme Exactus]. In: Proceedings of the Fourth Russian Seminar on Evaluation of Information Retrieval Methods ROMIP [Trudy chetvertogo rossiyskogo seminara po otsenke metodov informatsionnogo poiska ROMIP], pp. 80–85 (2006)

van Zaanen, M.: Multi-lingual Question Answering using OpenEphyra. CLEF (Working Notes) (2008)

Zhang, D., Lee, W.S.: Question classification using support vector machines. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, pp. 26–32 (2003)