Assessing Content Quality in Community Question Answering: Work-in-Progress

Pavel Braslavski

Yandex

Vadim Kantorov

Ural Federal University / ENS de Cachan
Outline

• CQA & content quality issue
• Our approach & dataset
• Data labeling
• Classification results
• Future directions
COMMUNITY QUESTION ANSWERING
Community question answering

- Users ask questions
- Users answer questions
- Answers are rated by the users
- Users gain points
Community QA
Where is a city of Kutaisi?

5 years ago

baso52

Best Answer - Chosen by Voters

One of the ancient regions of Georgia, Imereti, which constituted from the late 15th century. At the site where the Rioni emerges from the mountains into the plain stands the most ancient city of the Kingdom of Colchis (6th-1st centuries B.C.), Kutaisi (its name is derived from "kuata" meaning "strong"). Probably, you know from Greek mythology the legend of Jason and the Argonauts and their travel to Colchis for Golden Fleece. There is a romantic love story of Medea, the daughter of Ayates, king of Colchis, and Jason. Medea's story is told by Euripides in tragedy "Medea".

From the 10th century to 1112, when Tbilisi was liberated from foreign conquerors, Kutaisi was the residence of Georgian kings, and in the 15th century it became the capital of

Kutaisi 2011
Other Answers (4)

It is in the European country of Georgia.
5 years ago
0% 0 Votes

Kutaisi (Georgian: ქუთაისი; ancient names: Areai/Aia, Kutatisi, Kutaiissi) is Georgia's second largest city in the western province of Imereti. It is 221 km to the west of Tbilisi, with a population of about 230,000 (1989).
Source(s): http://en.wikipedia.org/wiki/Kutaisi
5 years ago
0% 0 Votes

Kutaisi is Georgia's second largest city in the western province of Imereti. It is 221 km to the west of Tbilisi, with a population of about 230,000 (1989).
5 years ago
0% 0 Votes

Sounds Indian to me.
5 years ago
0% 0 Votes
Resolved Question

Do you think I am cute, pretty or beautiful?

I think someone like Sophia Loren is beautiful and I think Cameron Diaz is pretty. What category would you put me in? I think possibly cute, but some would say beautiful. Opinions please - only genuine answers thanks.

4 years ago

Best Answer - Chosen by Asker

I'd say you're vain

4 years ago

8 people rated this as good

Asker's Rating: *****
I totally agree - I am too vain!
User-Generated Content & Quality

OUR APPROACH & DATASET
Harper et al. 2009

Conversational vs. informational questions

Agichtein et al. 2008

Graded quality of individual items

Our approach
• SE setting
• informational vs. non-informational
• in-page features
• Otvety@Mail.Ru
Our Approach

- An HTML page as a whole (SE indexing unit)
- Informational pages vs. others
- A less sophisticated model, but easier to maintain (SE settings)
- Only in-page features (*so far...*)
- Untouched source of data
Otvety@Mail.Ru

- A Russian counterpart of Y!Answers
- Launched 21 Aug 2006
- 50M+ users
- 60M+ questions
- 335M+ answers
- 58K new users, 52K questions, 235K answers daily

http://otvet.mail.ru/news/#hbd2011
Лучший футболист Торпедо Кутаиси всех времён?

Лучший ответ

Шенгелия

Другие ответы

Джемал Херхадзе-гвардеец команды.
Dataset

- Otvety@Mail.Ru
- Every 1000th question from 27.09.2009 to 07.11.2010
- 14,065 questions / 75,717 answers
- 11,625 users
- + graph of all their interactions
  → 7,631,566 edges ‘X answered Y’s question’
Categories by Size
Thread Size by Category

Kutaisi 2011
Time to First Answer

Kutaisi 2011
User Activity

#Answers by a user

#Questions by a user

Kutaisi 2011
User Activity – 2

[Graph depicting the relationship between the number of questions and the number of answers, with axes labeled #Questions and #Answers, showing a trend indicating that as the number of questions increases, so does the number of answers.]
Similarity of Categories

Kutaisi 2011
DATA LABELING
Labeling

2 assessors

Task: to tag thread as *informational* or *non-informational*
Оценка

Вопрос #4 из 20

Знакомства, Любовь, Отношения / Отношения

by Анюта Черненко (Ученик)

Что делать, если не можешь забыть любимого человека и чувствуешь, что он твоя судьба? Прошло уже 2 года.

by Sivilla &amp; (Профи)

Если не можешь забыть, надо встретиться и попытаться продолжить отношения, или хотя бы поговорить. От этого многое может измениться. Скорее всего это от незаконченности отношений.

by Сергей Monterey (Мышлитель)

Переключиться на другого, и почувствовать, что он твоя судьба )))))))))

Старая оценка: Нонинформативное

После нажатия на кнопку "Оставить комментарий", появится это поле.

Оставить комментарий
Tagging Results

• 2,000 documents
• 1,070 document judged by both assessors
• Agreement on 916 docs
  \(inf: 216, non-inf: 700\)
• Disagreement on 154 docs
• Average time per document: 15.33 sec
inf vs. non-inf by Category

Kutaisi 2011
CLASSIFICATION & FUTURE DIRECTIONS
Features

• Complexity and size (C): # of words, # of answers, avg. answer length, avg. word & sentence length

• Question (Q): first-person pronouns, question words

• Text well-formedness (W): misspelled words, capitalized sentences, well-formed punctuation, punctuation density, ...

• Informality (I): CAPS words, )))], ??!! , smileys

• User activity (U): efficiency of asker, activity level of asker, average & max levels of answerers (from Mail.Ru)

• External content (E): hyperlinks, images, video

• Other (O): best answer presence, top-level category
Preliminary Classification Results

- Method: SVM with RBF kernel
- Accuracy: 0.81
- Precision: 0.67
- Recall: 0.45
- F1: 0.53
Future Directions

- Feature engineering
- Feature selection
- Classifier tweaking
- Active learning approach
- Analysis of crowdsourced vs. assessor judgments
(Informational and conversational) questions?

pb@yandex-team.ru
vadimkantorov@gmail.com