

Foundations and Trends® in Information Retrieval  
Vol. 12, No. 1 (2018) 1–163  
© 2018 D. Hoogeveen, L. Wang, T. Baldwin,  
K. M. Verspoor  
DOI: 10.1561/15000000062



## Web Forum Retrieval and Text Analytics: a Survey

Doris Hoogeveen  
University of Melbourne  
doris.hoogeveen@gmail.com

Li Wang  
Evernote, California  
li@liwang.info

Timothy Baldwin  
University of Melbourne  
tb@ldwin.net

Karin M. Verspoor  
University of Melbourne  
karin.verspoor@unimelb.edu.au

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Types of forums . . . . .	3
1.2	A short history of forums . . . . .	7
1.3	Scope and outline . . . . .	13
1.4	Glossary . . . . .	13
1.5	Existing data sets . . . . .	16
<b>2</b>	<b>Post classification</b>	<b>18</b>
2.1	Post type classification . . . . .	19
2.2	Question classification . . . . .	21
2.3	Post quality assessment . . . . .	22
2.3.1	Features for post quality classification . . . . .	24
2.3.2	Completeness and answerability . . . . .	27
2.4	Subjectivity and viewpoint classification . . . . .	30
2.4.1	Subjectivity and viewpoint classification in cQA archives . . . . .	31
2.4.2	Subjectivity and viewpoint classification in discus- sion forums . . . . .	33
2.5	Post classification summary . . . . .	34
<b>3</b>	<b>Post retrieval</b>	<b>35</b>

3.1	Discussion forum post retrieval . . . . .	36
3.2	CQA question retrieval . . . . .	38
3.2.1	Question retrieval: statistical translation models . . . . .	40
3.2.2	Question retrieval: topic models . . . . .	46
3.2.3	Deep learning approaches to question retrieval . . . . .	48
3.2.4	Question retrieval: using category information . . . . .	53
3.2.5	Other question retrieval methods . . . . .	56
3.3	CQA answer retrieval . . . . .	59
3.3.1	Answer retrieval: topic models . . . . .	60
3.3.2	Answer retrieval: incorporating answer quality . . . . .	61
3.3.3	Answer retrieval: adding user information . . . . .	63
3.3.4	Machine learning approaches to answer retrieval . . . . .	64
3.3.5	Other answer retrieval methods . . . . .	68
3.3.6	Shared tasks on answer retrieval . . . . .	69
3.4	Post retrieval evaluation . . . . .	71
3.5	Post retrieval summary . . . . .	72
<b>4</b>	<b>Thread level tasks</b>	<b>73</b>
4.1	Task orientation and solvedness . . . . .	73
4.2	Thread discourse structure . . . . .	75
4.2.1	Thread linking structure recovery . . . . .	75
4.2.2	Dialogue act tagging . . . . .	76
4.2.3	Thread partitioning . . . . .	80
4.3	Discussion forum thread retrieval . . . . .	81
4.4	QA-pair extraction . . . . .	85
4.5	Thread summarisation . . . . .	87
4.5.1	Summarising discussion forum threads . . . . .	87
4.5.2	Summarising cQA answers . . . . .	92
4.6	Thread level tasks summary . . . . .	96
<b>5</b>	<b>Social forum analysis</b>	<b>97</b>
5.1	User satisfaction . . . . .	97
5.2	User and community analysis . . . . .	100
5.3	Expert finding . . . . .	103
5.3.1	Question recommendation and question routing . . . . .	106
5.4	Social forum analysis summary . . . . .	110

<b>6 Conclusion</b>	<b>111</b>
6.1 Standardization and comparison of methods . . . . .	111
6.2 Challenges . . . . .	112
6.3 Open research questions . . . . .	113
<b>Acknowledgements</b>	<b>115</b>
<b>References</b>	<b>116</b>

Preprint

## Abstract

This survey presents an overview of information retrieval, natural language processing and machine learning research that makes use of forum data, including both discussion forums and community question-answering (cQA) archives. The focus is on automated analysis, with the goal of gaining a better understanding of the data and its users.

We discuss the different strategies used for both retrieval tasks (post retrieval, question retrieval, and answer retrieval) and classification tasks (post type classification, question classification, post quality assessment, subjectivity, and viewpoint classification) at the post level, as well as at the thread level (thread retrieval, solvedness and task orientation, discourse structure recovery and dialogue act tagging, QA-pair extraction, and thread summarisation). We also review work on forum users, including user satisfaction, expert finding, question recommendation and routing, and community analysis.

The survey includes a brief history of forums, an overview of the different kinds of forums, a summary of publicly available datasets for forum research, and a short discussion on the evaluation of retrieval tasks using forum data.

The aim is to give a broad overview of the different kinds of forum research, a summary of the methods that have been applied, some insights into successful strategies, and potential areas for future research.

# 1

---

## Introduction

---

In this survey we will give an overview of a broad range of forum-related research. Forum research can be divided into two streams: discussion forums and community question-answering (cQA) archives. Both of these are websites that promote interaction and information sharing by the community, but they differ in their purpose, and because of that they often differ in their specific setup as well.

Forum data has been used for a large range of tasks and subtasks in information retrieval and natural language processing. Most of the tasks have to do with improving access to the rich information in the data, like post, question, or answer retrieval, thread summarisation, and expert finding. Subtasks cover specific aspects of the data and can be used to improve the results of the main tasks. Examples include dialogue act tagging, question and post type classification, post quality assessment, subjectivity and viewpoint classification, solvedness detection, thread type identification, topic detection, and user analysis. Forum research can also be used to improve the organization of the data, for instance by identifying duplicate questions, or categorizing posts.

In the remaining sections, we will present an overview of the different types of forums (§1.1), briefly discuss their history (§1.2), outline

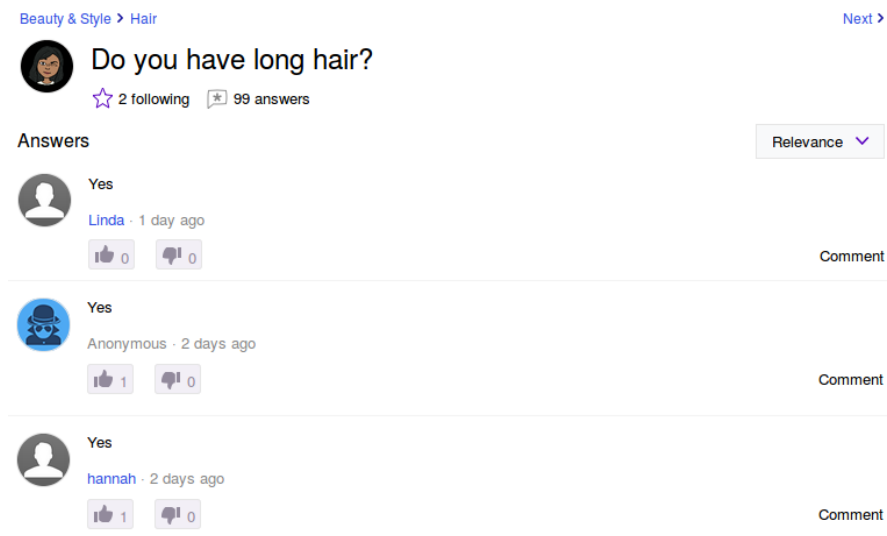
the scope of the survey (§1.3), present a glossary (§1.4), and present an overview of existing datasets used for forum research (§1.5).

## 1.1 Types of forums

In this section we will look at the differences between discussion forums and community question-answering archives. Both of these promote community interaction. Community question-answering archives are meant to help people to solve their problems and answer their questions. As soon as someone posts a good answer to a new question, the interaction is considered to be finished. Discussion forums on the other hand, are meant as a platform for people to discuss things.

This difference is not always strictly observed however. Some cQA archives contain questions like "Any1 from NY?", which do not express an information need, but rather a social need. Another example is requests for recommendations. Such questions do not have one correct answer and are therefore again more suited to discussion forums. Conversely, many factual questions and requests for help are posted on discussion forums, which might be more suitable for cQA archives.

Not much work has been published on the typology of forums. Choi et al. [2012] proposed a typology of online Q&A models consisting of four distinct types: community-based (e.g. Yahoo! Answers), collaborative (e.g. WikiAnswers), expert-based (e.g. the Internet Public Library (IPL) 'Ask a Librarian'-service), and social (e.g. Twitter, which we do not consider to be a forum). Shah et al. [2014] placed the four cQA forum types in a hierarchical structure of Q&A services, which also includes face-to-face Q&A, and automatic Q&A services. Discussion forums are not present in either of these taxonomies. Several dimensions along which we can classify internet communication tools (including forums) are presented in Long and Baecker [1997]. While slightly outdated, it includes aspects like *conversational style* and *audience membership*, which are still valid today. Similar relevant dimensions or aspects can be found in Davies et al. [2005] (e.g. degree of interaction, motivation/orientation, size, maintenance, etc.).



**Figure 1.1:** An example of a question on a cQA archive that may be intended to start a conversation. Source: Yahoo! Answers, <https://au.answers.yahoo.com/question/index?qid=20160921123000AA1wLIx>, accessed on 24th of September 2016.

In this survey we argue that forums exist on a spectrum with discussion threads on the one hand, where users have a high degree of freedom in what they post, and strict question-answering threads on the other, with heavy moderation to ensure only good answers are posted and threads are closed as soon as the question has been answered in a satisfactory way. In some cases the distinction is blurred. Linux Questions (<http://www.linuxquestions.org/>) for instance, looks like a forum, and has subforums dedicated to discussing Linux related topics, but also focuses on answering questions. Yahoo! Answers (<https://answers.yahoo.com/>), a cQA archive, contains questions that look like they are intended to spark a conversation. An example can be found in Figure 1.1. This also illustrates the lack of moderation on Yahoo! Answers.

On the far end of the cQA side of the spectrum there are cQA sites with a high degree of moderation supplied by the community itself. On



such websites there is often a reward system in place for users that ask good questions and provide good answers. StackExchange is a good example of this. Figure 1.2 shows an example of a thread from the StackExchange Cooking site.

As can be seen in the example, a distinction is made between answers and comments. Comments are used to ask for clarification, correct people, offer small suggestions, or make general remarks or even jokes. Answers are reserved for genuine answers. The number of reputation points and other rewards the users have obtained is shown next to their name. In this way, active contributors and experts can be distinguished from new users. This can be one way for users to consider which answer is the best one. Users can also look at the number of up votes and down votes an answer has received. These votes are cast by the community to indicate the quality of answers (and questions).

Another characteristic of most cQA archives, and something that discussion forums do not offer, is that question askers are encouraged to choose one of the answers as the best answer. That way other users know that the information need has been satisfied and they can focus their efforts on other questions. Repeated questions can be linked to archived ones, and an active effort is made by the community to keep the answers focused and not to stray away from the question. When it does happen, the question is usually closed. This is very different from discussion forums, where some threads can 'live' for very long and no one is bothered by it. A classic example of this is the famous "i am lonely will anyone speak to me" thread posted in the Moviecodec.com branch discussion forum, The Lounge Forums, in 2004.<sup>1</sup> It is still active today: more than twenty years since it was started.

Forums differ in how much access they offer to the outside world, but most of them make their content visible for everyone, while requiring people to sign up if they want to contribute. Some forums offer the option to sign up as an anonymous user. This makes the threshold to

---

<sup>1</sup><https://www.loungeforums.com/on-topic/i-am-lonely-will-anyone-speak-to-me-2420/>. It is more than 2000 pages long. Several magazines and newspapers have featured this thread. See for more information [https://en.wikipedia.org/wiki/I\\_am\\_lonely\\_will\\_anyone\\_speak\\_to\\_me](https://en.wikipedia.org/wiki/I_am_lonely_will_anyone_speak_to_me).

**Bread with no salt**

I've been doing a lot of yeast bread baking as of late and I got to wondering what the salt in the bread was for. Upon doing some research it turns out (aside from perhaps some flavor) the purpose is to "Control" the yeast during the rise. This got me to wondering if it was possible to make a yeast bread with no salt at all. My first attempt was met with defeat, and upon some more experimentation I was able to get the salt down 75% with success. Its the last 25% that alludes me.

Is it possible to make a yeast bread with no salt?

bread salt yeast

share improve this question

asked Jan 27 '14 at 14:02  
iamkrillin 133 • 1 • 5

edited Jan 27 '14 at 23:02  
SourDoh 8,547 • 14 • 31

2 What is the problem/symptoms if you leave out all the salt? I have done it (leaving all the salt out) frequently, and in general I find the result good enough. Also, how much yeast do you start with? If you are relying on salt to inhibit yeast growth, maybe you are just using too much yeast? – runtscho Jan 27 '14 at 14:06

What I usually run into is the top of the loaf collapsing in on itself. The recipe I am using calls for 2 teaspoons of yeast. – iamkrillin Jan 27 '14 at 14:06

1 "2 teaspoons" doesn't say much. What is important is the ratio of yeast to flour by weight, with usable amounts ranging from 0.5% to 10% of the flour weight in raw yeast (divide by 3 for active dry). Numbers at the edge of the interval are difficult to work with, 2% for lean and 3% for enriched breads are common, maybe 1% more for quick rises. (2% means 2 gram yeast to 100 g flour). – runtscho Jan 27 '14 at 14:14

@runtscho thanks! I'll have to do some more experimenting later today. – iamkrillin Jan 27 '14 at 14:17

3 Look up [Tuscan Bread](#) which is traditionally made with no salt. – derobert Jan 27 '14 at 16:54

show 1 more comment

**6 Answers** active oldest votes

5 If you make a bread without salt, you will have to make the dough dryer as well. Salt (for lack of a better word) competes with gluten and yeast for moisture. Without the salt, the yeast will work a bit faster (this effect isn't that pronounced) and the gluten will be very soft. The effect on the gluten usually causes loaves without salt to fall flat as the gluten is overly extensible but not very elastic. This could be part of why your loaves are collapsing, as without salt it is very hard to maintain the tension of the outer gluten sheath.

As SAJ14SAJ points out, the lack of salt will make the bread taste very "flat". While there are breads traditionally made without salt, they are usually served with very flavorful accompaniments like olives and sardines. Depending on what you're using the bread for, I'd suggest using an enriched dough as the added flavor of eggs, butter, and/or sugar will also help to cover the lack of salt.

Edited to add links to a couple of articles describing salt's effects on dough and one on salt taste in general.

share improve this answer

answered Jan 27 '14 at 15:17  
SourDoh 8,547 • 14 • 31

add a comment

2 It is certainly possible to make bread without salt. You would adjust the initial quantity of yeast and proofing times to get the desired outcome (it sounds like your loaves are over-proofing).

The thing is, it would taste terrible. Enhancing the flavor is the more important role of salt in bread, not just governing the growth rate of the yeast.

share improve this answer

answered Jan 27 '14 at 14:09  
SAJ14SAJ 59.8k • 9 • 99 • 169

2 I disagree with the "terrible" part. The enhancement in taste exists, but it is small, and while I can notice it, often I don't miss it when I decide to go without. – runtscho Jan 27 '14 at 14:11

@runtscho Okay, awful, flat, bland, boring, tasteless, dead, lifeless, icky... :-)- SAJ14SAJ Jan 27 '14 at 14:12

7 Certain Italian breads have no salt. The flavour comes from a long slow prove and really, really good olive oil for dipping :) – ElendiTheTall Jan 27 '14 at 14:16

1 I would think its very possible to make a flavorful bread with no salt. Perhaps by adding something else to make up for it not being there as @ElendiTheTall suggests. – iamkrillin Jan 27 '14 at 14:17

3 I conclude that while salt certainly changes the taste, it is a matter of personal preference if you find this change necessary or optional. I have never met somebody who has found it unwanted, so bakers are on the safe side when adding it. And sure, this answer is right that taste is the main reason for salt in bread. – runtscho Jan 27 '14 at 15:32

show 2 more comments

**Figure 1.2:** An example of a cQA thread. Source: StackExchange Cooking, <http://cooking.stackexchange.com/questions/41501/bread-with-no-salt>. Modified slightly by removing some answers, for presentational purposes. Accessed on 24th of September 2016.

contribute lower. In some forums that is seen as a good thing, because it lowers the bar of entry, but in forums that want to create a steady community of people that contribute regularly, these kinds of one-off contributions are discouraged. Having a system where people need to sign up before they can participate has the added benefit of making it difficult for bots to post spam, and it allows for personalisation of the forum. Some forums even offer member pages with all kinds of meta data such as when they became a member, how active they are, reputation points, question and answer history, and all the subforums they participate in, or topics they have expertise in. StackExchange<sup>2</sup> is once again a good example of this.

While many discussion forums explicitly show the discourse structure of the thread, i.e., which post is a reply to which earlier post, (see Figure 1.3 for an example), this is not always the case (see Figure 1.4). Quoted posts, allowed by some forums and illustrated in Figure 1.5, can be used to retrieve at least part of the discourse structure. We discuss this in §4.2.

CQA archives only have a simple two-part discourse structure, between a question and each of its answers. The original order of the answers is often not preserved. Instead, they are usually ordered based on the number of votes they have received from the community, with the answer that was accepted as the correct one by the question asker at the top.


## 1.2 A short history of forums

One of the earliest examples of a community question-answering service is The Straight Dope<sup>3</sup> founded in 1973. It started out as a column in several American newspapers, but these days it also has an online forum where people can ask questions and receive answers. The setup is closer to a discussion forum than a cQA archive however, with several subforums specifically created for discussion, such as the Elections subforum.

---

<sup>2</sup><http://stackexchange.com/>

<sup>3</sup><http://www.straightdope.com/>

5655  The reason why tomato soup and grilled cheese is such a good combo is because it's basically the same ingredients as pizza.   
 1 day ago by [JAMALDAVIS](#)   
 Edit: thanks for the love. I am also obligated to point out the fact that lasagna is also the same. Carry on   
 1001 comments share

**Top 200 Comments** show 500   
 sorted by: best

[\[-\] JekDarkness](#) 4096 points 1 day ago   
 Perhaps the reason pizza is so good is because it's the same ingredients as tomato soup and grilled cheese?   
 Why is tomato-based sauce, melted cheese, and crusty dough so good together in either arrangement?   
 permalink embed

[\[-\] FROM\\_THE\\_YEAR\\_2060](#) 3496 points 1 day ago   
 Tomatoes bring acidity, umami, and sweetness. Cheese brings umami and fat. Toasting bread caramelizes the sugars on the surface and gives it a crispy texture. All together, this creates a pleasurable blend of saltiness, sweetness, umami, fat, and texture. Each component releases different chemicals in the brain all associated with happiness, including caseomorphin from the cheese, which acts as an opiod when it hits the brain. The ratios of these ingredients have been perfected over thousands of years, and are found in similar proportions in the more modern food known as the dorito.   
 Edit: I was super baked when I wrote this post. Also, reducing your carbon footprint is the right thing to do. Be one with the earth.   
 Edit 2: I bet bots from whatever company makes Doritos upvoted the shit out of this post. Fuck Doritos and the company that makes them. Eat real food.   
 Edit 3: fuck lays   
 Edit 4: I have no knowledge of this manga or anime or whatever that you think I'm referencing. I just really like food.   
 permalink embed parent

[\[-\] flyingspex15](#) 2250 points 1 day ago   
 Each component releases different chemicals in the brain all associated with happiness, including caseomorphin from the cheese   
 More like quesomorphin   
 permalink embed parent

[\[-\] workdog](#) 560 points 1 day ago   
 Both queso and caseomorphin come from the Latin word for cheese "caseus". Morphine and similar terms come from Morpheus the Greek god of dreams. So, either way the term roughly means "cheese dreams".   
 permalink embed parent

[\[-\] Fidesphilo](#) 818 points 1 day ago   
 cheese dreams are made of Brie, who am I to disagree?   
 permalink embed parent

[\[-\] BigBooze77](#) 272 points 23 hours ago   
 Traveled the world all for 7 cheese... Everybody's lookin for Drommen   
 permalink embed parent

[\[-\] ButcherPiesMeats](#) 213 points 21 hours ago   
 Some of them want to fondue you.   
 Some of them want to be fondued by you.   
 permalink embed parent

[\[-\] Laffinkip](#) 101 points 20 hours ago   
 Some of them want to flambe, too   
 Every curd is churning for something   
 permalink embed parent

[\[-\] Bernstrademus](#) 45 points 18 hours ago   
 Sweet dreams are made of cheese. Who am I to dis a brie?   
 permalink embed parent

[\[-\] Gabocab](#) 17 points 17 hours ago   
 Tavoliere, with Serat sheep cheese   
 permalink embed parent   
 continue this thread

load more comments (8 replies)

[\[-\] stardly](#) 51 points 20 hours ago   
 Camembert wants to abuse you   
 Pepperjack wants to be abused   
 permalink embed parent

[\[-\] rykrtowe](#) 13 points 21 hours ago   
 /r/food is leaking   
 permalink embed parent

load more comments (2 replies)

[\[-\] PM\\_ME\\_YIFF\\_PICIS](#) 15 points 22 hours ago   
 We Shall Overcome   
 permalink embed parent

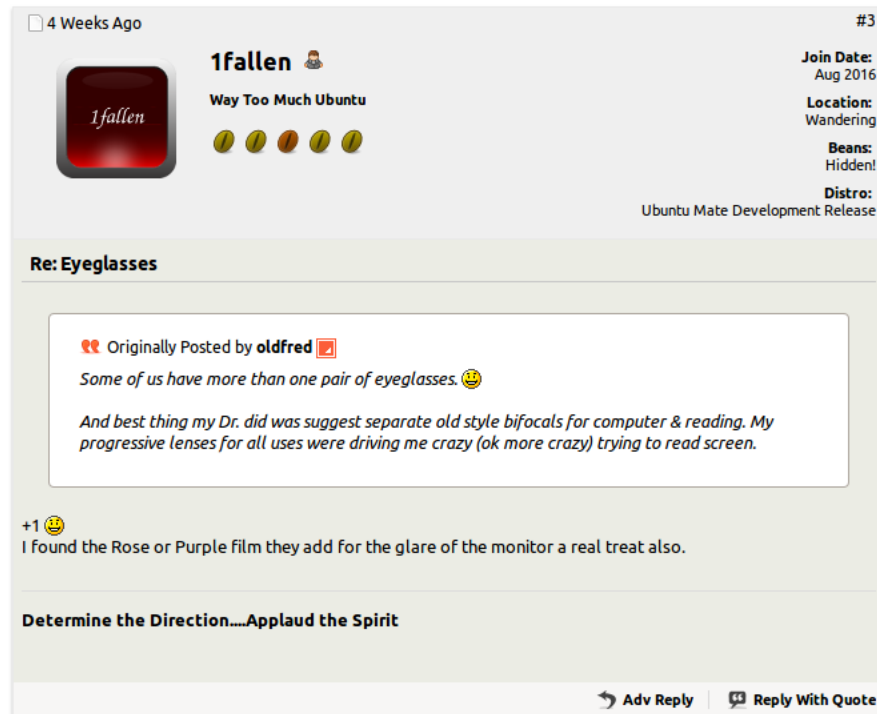
[\[-\] ratoaduckey](#) 28 points 22 hours ago   
 I love how this started with hardcore science talk and devolved into song lyrics   
 permalink embed parent   
 load more comments (12 replies)

**Figure 1.3:** An example of a discussion forum thread with explicit discourse structure. Source: Reddit, [https://www.reddit.com/r/Showerthoughts/comments/5403tk/the\\_reason\\_why\\_tomato\\_soup\\_and\\_grilled\\_cheese\\_is/](https://www.reddit.com/r/Showerthoughts/comments/5403tk/the_reason_why_tomato_soup_and_grilled_cheese_is/), accessed on 24th of September 2016.

Pages: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | >> (show all)

<p><b>fearme</b> addict</p> <p>Reged: 09/29/04 Posts: 465 Loc: CA</p>	<p><b>high stakes absolute players</b> #12075548 - 09/13/07 10:27 PM</p> <p>steamroller, greycat, doubledrag, win every day every session they play from talking to other high stakes players, has anyone ever booked a winning session against them cuz i havent anyway dont really play there these days because of what seems to be suspicious things on there</p> <p>i also saw steamroller chip dumping to someone named supercard i think, steam would cap flop, turn, and river then fold for 1 or something like that, and i remember greycat saying he lost 50+k to supercard</p> <p>ive never played with double drag but my friend told me ppl were very suspicious of his nl play and he ended up dumping 250k the next day to romaldo or something like that</p> <p><i>Edited by fearme (09/13/07 10:32 PM)</i></p> <p>Post Extras:   </p>
<p><b>ALL1N</b> lobster</p>  <p>Reged: 05/10/03 Posts: 2013 Loc: Melbourne 07</p>	<p><b>Re: high stakes absolute players [Re: fearme]</b> #12075633 - 09/13/07 10:34 PM</p> <p>Why does it worry you that some people would chipdump to others?</p> <p>Post Extras:   </p>
<p><b>fearme</b> addict</p> <p>Reged: 09/29/04 Posts: 465 Loc: CA</p>	<p><b>Re: high stakes absolute players [Re: ALL1N]</b> #12075650 - 09/13/07 10:35 PM</p> <p>im worried about more than chip dumping..</p> <p>Post Extras:   </p>
<p><b>Victor</b> Carpal \Tunnel</p> <p>Reged: 07/13/03 Posts: 11773</p>	<p><b>Re: high stakes absolute players [Re: fearme]</b> #12075707 - 09/13/07 10:38 PM</p> <p>so they win every session except the ones they chipdumb?</p> <p>i think your strategy should be to find the sessions where they are chipdumping and join.</p> <p>Post Extras:   </p>

**Figure 1.4:** An example of a discussion forum thread without explicit discourse structure. Source: The Two Plus Two Forum, <http://archives1.twoplustwo.com/showflat.php?Cat=0&Number=12075548>, accessed on 21st of October 2016.



**Figure 1.5:** An example of a post on a discussion forum, that quoted an earlier post, to make it clear what exactly is replied to, Source: Ubuntu Forums, <https://ubuntuforums.org/showthread.php?t=2337749>, accessed on 21st of October 2016.

Another early example is the Internet Oracle<sup>4</sup> founded in 1989. It specialises in humorous answers. Although it is a community question-answering service, questions and answers are submitted and distributed via e-mail.

Discussion forums also started to appear in the late 1980s. The Delphi Forums<sup>5</sup> was created in 1983 and is one of the earliest forums; it is still active today. Online discussion forums have their origins in bulletin boards and newsgroups such as Usenet, which has been around since 1980.

<sup>4</sup><http://internetoracle.org/>

<sup>5</sup><http://www.delphiforums.com/>

In the 1990s several cQA archives emerged. For instance:

- The Madsci Network:<sup>6</sup> It is heavily moderated and questions are all answered by scientists, rather than being open to anyone willing to contribute.
- Experts-Exchange<sup>7</sup>: This site is specifically for technology experts. It started out as purely community question-answering, but has expanded and now also offers help with code reviews, connecting freelancers to jobs, educating people, and live chat with an expert.
- 3form:<sup>8</sup> focuses on finding solutions to problems, rather than answers to questions. That is, questions are requests for information, either factual or not, while problems are questions for help in solving a particular issue.

Discussion forums also grew in popularity. In 1994 the W3C introduced WWW Interactive Talk (WIT),<sup>9</sup> a discussion forum that followed a set of design principles to display online discussions in such a way that it was easy to see which different topics were being discussed, and which points had been settled or not. Before WIT, many discussion forums suffered from the problem of people posting the same arguments over and over again, because there was no clear overview of a full thread. Although this was a step forward, and many alternatives sprang from this, to a certain extent we are still struggling with similar issues today.

In the first decade of the 2000s a large number of new cQA archives appeared, many of which are still extremely popular today: Baidu Knows,<sup>10</sup> WikiAnswers/Answers.com,<sup>11</sup> Quora,<sup>12</sup> Naver Knowl-

---

<sup>6</sup><http://www.madsci.org/> started in 1995 and still going.

<sup>7</sup><https://www.experts-exchange.com/> started in 1996 and still going.

<sup>8</sup><http://3form.org/> started in 1998 and still going.

<sup>9</sup>Official website: <https://www.w3.org/WIT/>, and more information can be found at [http://speed.eik.bme.hu/help/html/Special\\_Edition-Using\\_CGI/ch17.htm#WWWInteractiveTalk](http://speed.eik.bme.hu/help/html/Special_Edition-Using_CGI/ch17.htm#WWWInteractiveTalk).

<sup>10</sup><https://zhidao.baidu.com/> started in 2005.

<sup>11</sup><http://www.answers.com/> started on 2006 and its predecessor FAQForm in 2002.

<sup>12</sup><https://www.quora.com/> started in 2009.

edge Search,<sup>13</sup> Yahoo! Answers,<sup>14</sup> and the StackExchange<sup>15</sup> website, especially StackOverflow.<sup>16</sup> The only notable exception is Google Answers<sup>17</sup> which was started in 2002 but discontinued in 2006. Many of these large cQA archives are in English, but not all of them: Naver is Korean, and Baidu Knows is Chinese.

One specific example of a space where forums have been used and found to be helpful is education. There are several online cQA archives dedicated to questions about topics taught in schools. An example of this is Brainly,<sup>18</sup> which has the slogan “For students. By students.” The idea is that students help each other to learn. Other examples are Chegg,<sup>19</sup> and Piazza.<sup>20</sup> Lang-8<sup>21</sup> is a language learning platform that has many similarities to forums. Users write posts in a language they are learning. Native speakers of that language will then correct the post sentence by sentence and comment on it. The original poster can reply to the corrections, and other native speakers can join in the conversation too, to discuss linguistic constructs or explain semantic or syntactic points.

Many learning management systems include a forum to enable students to start discussions online, or ask questions. This is considered to be a vital ingredient of MOOCs for instance, where the number of students is so large that it is not possible for them to individually get in touch with the professor or tutors, and forums offer an alternative to ask for help or discuss the subject matter. In such a setting, the forums are used both as a cQA platform and as a discussion forum. One MOOC platform, EdX,<sup>22</sup> has recognised this dual nature of MOOC forums and allows people to choose what kind of post they make: a

<sup>13</sup><http://kin.naver.com/index.nhn> started in 2002.

<sup>14</sup><https://answers.yahoo.com/> started in 2005 and formally known as Yahoo! Q&A.

<sup>15</sup><http://stackexchange.com/> started in 2008.

<sup>16</sup><https://stackoverflow.com/>, the first cQA site of the StackExchange network.

<sup>17</sup><http://answers.google.com/answers/>. It grew out of Google Questions and Answers which was started in 2001.

<sup>18</sup><http://brainly.com/>

<sup>19</sup><https://www.chegg.com/>

<sup>20</sup><https://piazza.com/>

<sup>21</sup><http://lang-8.com/>

<sup>22</sup><https://www.edx.org/>



discussion sparking post, or a question-answer post. Threads are then labeled accordingly, so that other people know what kind of content a thread contains. The idea is that this labeling enhances information access.<sup>23</sup>

### 1.3 Scope and outline

In this survey we will describe research into automated analysis of forum data. That includes data from both discussion forums (also called web user forums; see, for instance, [Wang et al., 2013b]) and community question-answering (cQA) archives. These two forum types share a number of characteristics (as discussed in §1.1), which are not shared with other (semi) threaded discourses, like chat discussions, email threads, product reviews, or frequently asked question (FAQ) pages. These are therefore outside the scope of this survey.

At the start of this section we mentioned several tasks and subtasks. Each of these will be discussed in the following sections, divided into post classification (§2), post retrieval (§3), thread level tasks (§4), and social forum analysis or user studies (§5).

Previously published survey articles include Shah et al. [2009], who present an overview of early research in the cQA field, Gazan [2011], Li [2014], and Srba and Bielikova [2016], who all present an overview of cQA related research. Srba and Bielikova [2016] is the most recent and most comprehensive survey, discussing 265 research papers published before February 2015. They also show that the number of publications in this field has increased each year.

This survey covers 450 papers published until November 2016, and distinguishes itself from earlier survey papers by including discussion forums, instead of focusing on cQA archives only.

### 1.4 Glossary

The same or similar concepts sometimes appear in the literature under different names. We will try to use the same terminology for each con-

---

<sup>23</sup><http://blog.edx.org/navigating-two-kinds-online-discussion>

cept throughout this survey. This section summarises the important terminology we will use.

**Thread:** we use the term “thread” to refer to forum discussion threads, or a question on a cQA forum together with all of its answers (and comments). In discussion forums this is the full thread, which may span multiple pages (see PAGE below).

**Page:** in discussion forums, threads can sometimes become very large. If this happens, instead of displaying the full thread, only a certain number of posts are displayed at a time. So threads are divided into smaller units for easier display. Such chunks are called “pages”.

**Post/message:** the terms “post” and “message” are often used interchangeably in the research community to refer to each posting in a forum thread. In this survey we use “post” to denote forum thread post. The term “post” can also be used to refer to either the question post in a cQA archive, or an answer post. We use it as a general term when we want to refer to any text posted by a user, regardless of whether it is an initial post or question post, or an answer post. In situations where it matters we will distinguish clearly between the two, by calling them “initial post” (or “question post”) and “answer post”.

**Initial post:** this refers to the first post in a discussion forum thread, which starts a discussion. In the literature, it is sometimes also called the “root post/message” or “first post/message”.

**Question post:** this refers to the first post in a cQA thread, in which a question is asked. All other posts in a cQA thread are answers to this post.

**Answer post:** this refers to any post in a cQA thread that is not the question post, but rather a response to a question post.

**Word/term:** in this survey, “word” and “term” are used interchangeably to indicate a word unit in a post.

**Thread initiator:** the user who starts a new discussion thread (in discussion forums), or who posts a question (in cQA archives). This is the person that writes the INITIAL POST or QUESTION POST. In a cQA context we will sometimes refer to this person as the “question asker”.

**Quoted text:** in discussion forums a user may sometimes quote content from previous posts or email messages in his/her post. This quoted content is called “quoted text”. In cQA archives, quoted material often comes from other threads or from technical documentation. An example from a discussion forum can be found in Figure 1.5.

**Comment:** in some cQA archives, users can write comments to posts, in addition to answers. These two kinds of posts (comments and answers) serve a slightly different purpose. Answers are supposed to directly answer the question, while comments can be used to correct someone, ask for clarification on a certain point, make a small addition to a post, or provide similar short contributions that are not standalone answers.

**Thread structure:** The structure of a discussion forum thread can be viewed as a tree, with the initial post at the top, and reply posts branching out below it. Each post is placed below the post it responds to. This structure can be explicit, like in Figure 1.3, or not, like in Figure 1.4.

As background information we would like to very briefly introduce some IR evaluation metrics here, which will be mentioned in different places throughout this survey. Many different evaluation metrics are used for IR tasks using forum data, i.e. post retrieval, and IR in general. For instance, Mean Average Precision (MAP), Mean Reciprocal Rank (MRR) [Voorhees, 1999], Precision@ $n$ , nDCG [Järvelin and Kekäläinen, 2002], AUC (precision–recall or ROC), and Rank-Biased Precision [Moffat and Zobel, 2008]. Of these, MAP is the most widely used. It is the mean of the average precision at a given

cut-off point, calculated over all the queries in a set. The average precision is shown in Equation 1.1, in which  $N$  is the cut-off point,  $P$  is the precision, and  $R$  is an indicator of whether the document retrieved at  $i$  is relevant or not.

$$AP@N = \frac{\sum_{i=1}^N P(i) \cdot R(i)}{\# \text{ of relevant documents}} \quad (1.1)$$

## 1.5 Existing data sets

The field of forum related research has long suffered from a lack of publicly available datasets, but this is slowly changing. Over the years, many researchers have constructed their own sets using web forum crawling techniques, for instance using methods described in Wang et al. [2008] or Yang et al. [2009a]. Recently, some forums have started making (part of) their data available to the research community, and many top-tier conferences (e.g. the AAAI International Conference on Web and Social Media) encourage their authors to share their data and provide data sharing services specifically for this purpose. An overview of a large number of public and private datasets used in forum research can be found in Hoogeveen et al. [2015]. In this section we will present only the most important ones, which are openly available for research purposes. They are summarised in Table 1.1.

<b>The Yahoo! Webscope Dataset (L6)</b> Surdeanu et al. [2008] <a href="http://webscope.sandbox.yahoo.com/catalog.php?datatype=1">http://webscope.sandbox.yahoo.com/catalog.php?datatype=1</a>	4M question and answer pairs from a dump of Yahoo! Answers on 25/10/2007.
<b>The WikiAnswers Corpus</b> Fader et al. [2013] <a href="http://knowitall.cs.washington.edu/oqa/data/wikianswers/">http://knowitall.cs.washington.edu/oqa/data/wikianswers/</a>	30M clusters of questions from WikiAnswers, <sup>24</sup> tagged as paraphrases by users. Around 11% of them have an answer.
<b>TREC 2015 LiveQA data</b> Agichtein et al. [2015] <a href="http://trec.nist.gov/data/qa/2015_LiveQA.html">http://trec.nist.gov/data/qa/2015_LiveQA.html</a>	1000 Yahoo! Answers questions used as queries in the TREC 2015 LiveQA task, including answer strings from systems, with human judgements.
<b>The SemEval Task 3 cQA Dataset</b> Nakov et al. [2015] <a href="http://alt.qcri.org/semeval2015/task3/index.php?id=data-and-tools">http://alt.qcri.org/semeval2015/task3/index.php?id=data-and-tools</a>	2900 English questions and answers from the Qatar Living Forum, <sup>25</sup> and 1500 Arabic ones from the Fatwa forum on IslamWeb. <sup>26</sup>
<b>StackExchange dump</b> <a href="https://stackoverflow.blog/2009/06/stack-overflow-creative-commons-data-dump/">https://stackoverflow.blog/2009/06/stack-overflow-creative-commons-data-dump/</a>	A periodical dump of all the data on StackExchange, in XML format.
<b>CQADupStack</b> Hoogeveen et al. [2015] <a href="http://nlp.cis.unimelb.edu.au/resources/cqadupstack/">http://nlp.cis.unimelb.edu.au/resources/cqadupstack/</a>	All the data of twelve StackExchange forums, in JSON format.
<b>MSR Challenge Dataset</b> Bacchelli [2013] <a href="http://2013.msrfconf.org/challenge.php#challenge_data">http://2013.msrfconf.org/challenge.php#challenge_data</a>	Stripped version of a StackOverflow dump, in XML and postgresql formats.
<b>The NTCIR-8 cQA dataset</b> Ishikawa et al. [2010] <a href="http://research.nii.ac.jp/ntcir/permission/ntcir-8/perm-en-CQA.html">http://research.nii.ac.jp/ntcir/permission/ntcir-8/perm-en-CQA.html</a>	1500 questions and answers from Yahoo! Chiebukuro, the Japanese version of Yahoo! Answers, between April 2004 and October 2005.
<b>The Reddit Comment Corpus</b> <a href="https://www.reddit.com/r/datasets/comments/590re2/updated_reddit_comments_and_posts_updated_on/">https://www.reddit.com/r/datasets/comments/590re2/updated_reddit_comments_and_posts_updated_on/</a>	A periodical dump of all the comments. Some of it contains sentiment annotations.
<b>The Quora Dataset</b> <a href="https://data.quora.com/First-Quora-Dataset-Release-Question-Pairs">https://data.quora.com/First-Quora-Dataset-Release-Question-Pairs</a>	400.000 question pairs, annotated for duplicates. Released on 25/01/2017.

Table 1.1: An overview of publicly available forum data sets.

## References

---

- Ifeoma Adaji and Julita Vassileva. Towards Understanding User Participation in Stack Overflow Using Profile Data. In *Proceedings of the 8th International Conference on Social Informatics (SocInfo)*, volume Proceedings Part II, pages 3–13. Springer, 2016.
- Lada A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proceedings of the 17th International World Wide Web Conference*, pages 665–674. ACM, 2008.
- Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. Finding High-quality Content in Social Media. In *Proceedings of the 1st ACM International Conference on Web Search and Data Mining (WSDM)*, pages 183–194. ACM, 2008.
- Eugene Agichtein, Yandong Liu, and Jiang Bian. Modeling Information-Seeker Satisfaction in Community Question Answering. *TKDD*, 3(2):10:1–10:27, 2009.
- Eugene Agichtein, David Carmel, Donna Harman, Dan Pelleg, and Yuval Pinter. Overview of the TREC 2015 LiveQA Track. In *Proceedings of the 24th Text REtrieval Conference (TREC) (LiveQA Track)*, pages 1–9. NIST, 2015.
- Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. Diversifying Search Results. In *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining (WSDM)*, pages 5–14. ACM, 2009.

- Muhammad Ahasanuzzaman, Muhammad Asaduzzaman, Chanchal K. Roy, and Kevin A. Schneider. Mining Duplicate Questions in Stack Overflow. In *Proceedings of the 13th International Conference on Mining Software Repositories (MRS)*, pages 402–412. ACM, 2016.
- June Ahn, Brian S. Butler, Cindy Weng, and Sarah Webster. Learning to be a Better Q'er in Social Q&A Sites: Social Norms and Information Artifacts. *JASIST*, 50(1):1–10, 2013.
- Naoyoshi Aikawa, Tetsuya Sakai, and Hayato Yamana. Community QA Question Classification: Is the Asker Looking for Subjective Answers or Not? *IPSJ Online Transactions*, 4:160–168, 2011.
- Ameer Tawfik Albaham and Naomie Salim. Adapting Voting Techniques for Online Forum Thread Retrieval. In *Proceedings of the 1st International Conference on Advanced Machine Learning Technologies and Applications (AMLTA)*, pages 439–448. Springer, 2012.
- Ameer Tawfik Albaham and Naomie Salim. Quality Biased Thread Retrieval Using the Voting Model. In *Proceedings of the 18th Australasian Document Computing Symposium (ADCS)*, pages 97–100. ACM, 2013.
- Ameer Tawfik Albaham, Naomie Salim, and Obasa Isiaka Adekunle. Leveraging Post Level Quality Indicators in Online Forum Thread Retrieval. In *Proceedings of the 1st International Conference on Advanced Data and Information Engineering (DaEng)*, pages 417–425. Springer, 2014.
- James Allan, Jaime G. Carbonell, George Doddington, Jonathan Yamron, and Yiming Yang. Topic Detection and Tracking Pilot Study Final Report. In *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*. NIST, 1998.
- Hadi Amiri, Zheng-Jun Zha, and Tat-Seng Chua. A Pattern Matching Based Model for Implicit Opinion Question Identification. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pages 46–52. AAAI, 2013.
- Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow. In *Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 850–858. ACM, 2012.
- Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Steering User Behaviour with Badges. In *Proceedings of the 22nd International World Wide Web Conference*, pages 95–106. ACM, 2013.

- J. Anusha, V. Smrithi Rekha, and P. Bagavathi Sivakumar. A Machine Learning Approach to Cluster the Users of Stack Overflow Forum. In *Proceedings of the 2014 International Conference on Artificial Intelligence and Evolutionary Algorithms in Engineering Systems (ICAEES) (ICAEES)*, volume 2, pages 411–418. Springer, 2015.
- Jaime Arguello and Kyle Shaffer. Predicting Speech Acts in MOOC Forum Posts. In *Proceedings of the 9th AAAI International Conference on Weblogs and Social Media (ICWSM)*, pages 2–11. AAAI, 2015.
- Muhammad Asaduzzaman, Ahmed Shah Mashiyat, Chanchal K. Roy, and Kevin A. Schneider. Answering Questions about Unanswered Questions of Stack Overflow. In *Proceedings of the 10th Working Conference on Mining Software Repositories (MRS)*, pages 97–100. IEEE, 2013.
- J.L. Austin. *How to do Things with Words*. Oxford University Press, 1962.
- Alberto Bacchelli. Mining Challenge 2013: Stack Overflow. In *Proceedings of the 10th Working Conference on Mining Software Repositories (MRS)*, pages 53–56. IEEE, 2013.
- Timothy Baldwin, David Martinez, and Richard B. Penman. Automatic Thread Classification for Linux User Forum Information Access. In *Proceedings of the 12th Australasian Document Computing Symposium (ADCS)*, pages 72–79. ACM, 2007.
- Krisztian Balog, Yi Fang, Maarten de Rijke, Pavel Serdyukov, and Luo Si. Expertise Retrieval. *Foundations and Trends in Information Retrieval (FN-TIR)*, 6(2–3):127–256, 2012.
- Antoaneta Baltadzhieva and Grzegorz Chrupala. Question Quality in Community Question Answering Forums: a Survey. *ACM SIGKDD Explorations Newsletter*, 17(1):8–13, 2015.
- Xin-Qi Bao and Yun-Fang Wu. A Tensor Neural Network with Layerwise Pretraining: Towards Effective Answer Retrieval. *JCST*, 31(6):1151–1160, 2016.
- Alberto Barrón-Cedeno, Simone Filice, Giovanni Da San Martino, Shafiq Joty, Lluís Màrquez, Preslav Nakov, and Alessandro Moschitti. Thread-Level Information for Comment Classification in Community Question Answering. In *Proceedings of the Joint 53rd Annual Meeting of the Association for Computational Linguistics (ACL) and the 7th International Joint Conference on Natural Language Processing*, volume Volume 2: Short Papers, pages 687–693. ACL, 2015.



- Yonatan Belinkov, Mitra Mohtarami, Scott Cyphers, and James Glass. VectorSLU: A continuous word vector approach to answer selection in community question answering systems. In *Proceedings of the 9th Conference on Semantic Evaluation (SemEval)*, page 282. ACL, 2015.
- Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al. Greedy Layer-wise Training of Deep Networks. *NIPS*, 19:153, 2007.
- Adam Berger and John Lafferty. Information Retrieval as Statistical Translation. In *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 222–229. ACM, 1999.
- Adam Berger, Rich Caruana, David Cohn, Dayne Freitag, and Vibhu Mittal. Bridging the Lexical Chasm: Statistical Approaches to Answer-Finding. In *Proceedings of the 23rd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 192–199. ACM, 2000.
- Delphine Bernhard and Iryna Gurevych. Answering Learners’ Questions by Retrieving Question Paraphrases from Social Q&A Sites. In *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*, pages 44–52. ACL, 2008.
- Delphine Bernhard and Iryna Gurevych. Combining Lexical Semantic Resources with Question & Answer Archives for Translation-Based Answer Finding. In *Proceedings of the Joint 47th Annual Meeting of the Association for Computational Linguistics (ACL) and the 4th International Joint Conference on Natural Language Processing*, pages 728–736. ACL, 2009.
- Abraham Bernstein and Esther Kaufmann. GINO - A Guided Input Natural Language Ontology Editor. In *Proceedings of the 5th International Semantic Web Conference (ISWC)*, pages 144–157. Springer, 2006.
- Sumit Bhatia and Prasenjit Mitra. Adopting Inference Networks for Online Thread Retrieval. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 1300–1305. AAAI, 2010.
- Sumit Bhatia, Prakhar Biyani, and Prasenjit Mitra. Classifying User Messages for Managing Web Forum Data. In *Proceedings of the 15th International Workshop on the Web and Databases (WebDB)*, pages 13–18. ACM, 2012.
- Sumit Bhatia, Prakhar Biyani, and Prasenjit Mitra. Summarizing Online Forum Discussions - Can Dialog Acts of Individual Messages Help? In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2127–2131. ACL, 2014.

- Sumit Bhatia, Prakhar Biyani, and Prasenjit Mitra. Identifying the Role of Individual User Messages in an Online Discussion and its Use in Thread Retrieval. *JASIST*, 67(2):276–288, 2016.
- Jiang Bian, Yandong Liu, Eugene Agichtein, and Hongyuan Zha. Finding the Right Facts in the Crowd: Factoid Question Answering over Social Media. In *Proceedings of the 17th International World Wide Web Conference*, pages 467–476. ACM, 2008a.
- Jiang Bian, Yandong Liu, Eugene Agichtein, and Hongyuan Zha. A Few Bad Votes Too Many?: Towards Robust Ranking in Social Media. In *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*, pages 53–60. ACM, 2008b.
- Jiang Bian, Yandong Liu, Ding Zhou, Eugene Agichtein, and Hongyuan Zha. Learning to Recognize Reliable Users and Content in Social Media with Coupled Mutual Reinforcement. In *Proceedings of the 18th International World Wide Web Conference*, pages 51–60. ACM, 2009.
- Prakhar Biyani. *Analyzing Subjectivity and Sentiment of Online Forums*. PhD thesis, The Pennsylvania State University, 2014.
- Prakhar Biyani, Sumit Bhatia, Cornelia Caragea, and Prasenjit Mitra. Thread Specific Features are Helpful for Identifying Subjectivity Orientation of Online Forum Threads. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 295–310. ACL, 2012.
- Prakhar Biyani, Sumit Bhatia, Cornelia Caragea, and Prasenjit Mitra. Using Non-lexical Features for Identifying Factual and Opinionative Threads in Online Forums. *KBS*, 69:170–178, 2014.
- Prakhar Biyani, Sumit Bhatia, Cornelia Caragea, and Prasenjit Mitra. Using Subjectivity Analysis to Improve Thread Retrieval in Online Forums. In *Proceedings of the 37th Annual European Conference on Information Retrieval Research (ECIR): Advances in Information Retrieval*, pages 495–500. Springer, 2015.
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *JourMLR*, 3:993–1022, 2003.
- Mohan John Blooma, Alton Yeow-Kuan Chua, and Dion Hoe-Lian Goh. A Predictive Framework for Retrieving the Best Answer. In *Proceedings of the 23rd ACM Symposium on Applied Computing (SAC)*, pages 1107–1111. ACM, 2008.
- Mohan John Blooma, Alton Yeow-Kuan Chua, and Dion Hoe-Lian Goh. What Makes a High-Quality User-Generated Answer? *IEEE Internet Computing*, 15(1):66–71, 2011.

- Mohan John Blooma, Dion Hoe-Lian Goh, and Alton Yeow-Kuan Chua. Predictors of High-Quality Answers. *Online Information Review*, 36(3):383–400, 2012.
- Avrim Blum and Tom Mitchell. Combining Labeled and Unlabeled Data with Co-training. In *Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT)*, pages 92–100. ACM, 1998.
- Dasha Bogdanova and Jennifer Foster. This is how we do it: Answer Reranking for Open-Domain How Questions with Paragraph Vectors and Minimal Feature Engineering. In *Proceedings of the 2016 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 1290–1295. ACL, 2016.
- Ingwer Borg and Patrick J.F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005.
- Mohamed Bouguessa and Lotfi Ben Romdhane. Identifying Authorities in Online Communities. *TIST*, 6(3):30, 2015.
- Mohamed Bouguessa, Benoît Dumoulin, and Shengrui Wang. Identifying Authoritative Actors in Question-Answering Forums: the Case of Yahoo! Answers. In *Proceedings of the 14th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 866–874. ACM, 2008.
- Thorsten Brants, Francine Chen, and Ayman Farahat. A System for New Event Detection. In *Proceedings of the 26th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 330–337. ACM, 2003.
- Chris Brockett, William B. Dolan, and Michael Gamon. Correcting ESL Errors Using Phrasal SMT Techniques. In *Proceedings of the 21st International Conference on Computational Linguistics (COLING) and the 44th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 249–256. ACL, 2006.
- Jane Bromley, James W. Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. Signature Verification Using a “Siamese” Time Delay Neural Network. *PRAI*, 7(04):669–688, 1993.
- Peter F. Brown, Stephen A. Della-Pietra, Vincent J. Della-Pietra, and Robert L. Mercer. The Mathematics of Statistical Machine Translation. *Computational Linguistics*, 19(2):263–313, 1993.

- Razvan Bunescu and Yunfeng Huang. Learning the Relative Usefulness of Questions in Community QA. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 97–107. ACL, 2010a.
- Razvan Bunescu and Yunfeng Huang. A Utility-driven Approach to Question Ranking in Social QA. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pages 125–133. ACL, 2010b.
- Grégoire Burel. *Community and Thread Methods for Identifying Best Answers in Online Question Answering Communities*. PhD thesis, The Open University, 2016.
- Grégoire Burel, Yulan He, and Harith Alani. Automatic Identification of Best Answers in Online Enquiry Communities. In *Proceedings of the Extended Semantic Web Conference (ESWC)*, pages 514–529. Springer, 2012.
- Gregoire Burel, Paul Mulholland, and Harith Alani. Structural Normalisation Methods for Improving Best Answer Identification in Question Answering Communities. In *Proceedings of the 25th International World Wide Web Conference*, pages 673–678. ACM, 2016.
- Moira Burke, Elisabeth Joyce, Tackjin Kim, Vivek Anand, and Robert Kraut. Introductions and Requests: Rhetorical Strategies that Elicit Response in Online Communities. In *Proceedings of the 3rd Communities and Technologies Conference*, pages 21–39. Springer, 2007.
- Robin D. Burke, Kristian J. Hammond, Vladimir Kulyukin, Steven L. Lytinen, Noriko Tomuro, and Scott Schoenberg. Question Answering from Frequently Asked Question Files: Experiences with the FAQ Finder System. *AI*, 18(2):57, 1997.
- Li Cai, Guangyou Zhou, Kang Liu, and Jun Zhao. Learning the Latent Topics for Question Retrieval in Community QA. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 273–281. ACL, 2011.
- Fabio Calefato, Filippo Lanubile, and Nicole Novielli. Moving to Stack Overflow: Best-Answer Prediction in Legacy Developer Forums. In *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, page 13. ACM, 2016.
- Xin Cao, Gao Cong, Bin Cui, Christian Søndergaard Jensen, and Ce Zhang. The Use of Categorization Information in Language Models for Question Retrieval. In *Proceedings of the 18th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 265–274. ACM, 2009.

- Xin Cao, Gao Cong, Bin Cui, and Christian S. Jensen. A Generalized Framework of Exploring Category Information for Question Retrieval in Community Question Answer Archives. In *Proceedings of the 19th International World Wide Web Conference*, pages 201–210. ACM, 2010.
- Xin Cao, Gao Cong, Bin Cui, Christian S. Jensen, and Quan Yuan. Approaches to Exploring Category Information for Question Retrieval in Community Question-Answer Archives. *TOIS*, 30(2):7, 2012.
- Yunbo Cao, Wen-Yun Yang, Chin-Yew Lin, and Yong Yu. A Structural Support Vector Method for Extracting Contexts and Answers of Questions from Online Forums. *IPM*, 47(6):886–898, 2011.
- Giuseppe Carenini, Raymond T. Ng, and Xiaodong Zhou. Summarizing Email Conversations with Clue Words. In *Proceedings of the 16th International World Wide Web Conference*, pages 91–100. ACM, 2007.
- David Carmel, Avihai Mejer, Yuval Pinter, and Idan Szpektor. Improving Term Weighting for Community Question Answering Search Using Syntactic Analysis. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 351–360. ACM, 2014.
- Rose Catherine, Amit Singh, Rashmi Gangadharaiah, Dinesh Raghu, and Karthik Visweswariah. Does *Similarity* Matter? The Case of Answer Extraction from Technical Discussion Forums. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 175–184. ACL, 2012.
- Rose Catherine, Rashmi Gangadharaiah, Karthik Visweswariah, and Dinesh Raghu. Semi-Supervised Answer Extraction from Discussion Forums. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pages 1–9. ACL, 2013.
- Huseyin Cavusoglu, Zhuolun Li, and Ke-Wei Huang. Can Gamification Motivate Voluntary Contributions?: The Case of StackOverflow Q&A Community. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW)*, pages 171–174. ACM, 2015.
- Pedro Chahuara, Thomas Lampert, and Pierre Gancarski. Retrieving and Ranking Similar Questions from Question-Answer Archives Using Topic Modelling and Topic Distribution Regression. In *Proceedings of the 20th International Conference on Theory and Practice of Digital Libraries (TPDL): Research and Advanced Rechnology for Digital Libraries*, pages 41–53. Springer, 2016.

- Kevin Chai, Pedram Hayati, Vidyasagar Potdar, Chen Wu, and Alex Talevski. Assessing Post Usage for Measuring the Quality of Forum Posts. In *Proceedings of the 4th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, pages 233–238. IEEE, 2010.
- Wen Chan, Xiangdong Zhou, Wei Wang, and Tat-Seng Chua. Community Answer Summarization for Multi-Sentence Question with Group L1 Regularization. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, volume Volume 1: Long Papers, pages 582–591. ACL, 2012.
- Wen Chan, Jintao Du, Weidong Yang, Jinhui Tang, and Xiangdong Zhou. Term Selection and Result Reranking for Question Retrieval by Exploiting Hierarchical Classification. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 141–150. ACM, 2014.
- Guibin Chen, Chunyang Chen, Zhenchang Xing, and Bowen Xu. Learning a Dual-Language Vector Space for Domain-Specific Cross-Lingual Question Retrieval. In *Proceedings of the 31st IEEE/ACM International Conference On Automated Software Engineering (ASE)*, pages 744–755. ACM, 2016a.
- Long Chen, Dell Zhang, and Levene Mark. Understanding User Intent in Community Question Answering. In *Proceedings of the 21st International World Wide Web Conference*, pages 823–828. ACM, 2012.
- Long Chen, Joemon M. Jose, Haitao Yu, and Fajie Yuan. A Hybrid Approach for Question Retrieval in Community Question Answering. *The Computer Journal, Section C: Computational Intelligence, Machine Learning and Data Analytics*, pages 1–13, 2016b.
- Zhi Chen, Li Zhang, and Weihua Wang. PostingRank: Bringing Order to Web Forum Postings. In *Proceedings of the Asia Information Retrieval Symposium (AIRS)*, pages 377–384. Springer, 2008.
- Jason H.D. Cho, Parikshit Sondhi, Chengxiang Zhai, and Bruce R. Schatz. Resolving Healthcare Forum Posts via Similar Thread Retrieval. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB)*, pages 33–42. ACM, 2014.
- Erik Choi. Motivations and Expectations for Asking Questions within Online Q&A. *TCDL*, 9(2):29–35, 2013.
- Erik Choi and Chirag Shah. Asking for More than an Answer: What do Askers Expect in Online Q&A Services? *JIS*, pages 1–12, 2016.

- Erik Choi, Vanessa Kitzie, and Chirag Shah. Developing a Typology of Online Q&A Models and Recommending the Right Model for Each Question Type. *JASIST*, 49(1):1–4, 2012.
- Erik Choi, Vanessa Kitzie, and Chirag Shah. Investigating Motivations and Expectations of Asking a Question in Social Q&A. *First Monday*, 19(3), 2014.
- Alton Y.K. Chua and Snehasish Banerjee. Measuring the Effectiveness of Answers in Yahoo! Answers. *Online Information Review*, 39(1):104–118, 2015a.
- Alton Y.K. Chua and Snehasish Banerjee. Answers or No Answers: Studying Question Answerability in Stack Overflow. *JIS*, pages 720–731, 2015b.
- Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. In *Proceedings of the 2014 Workshop on Deep Learning and Representation (held at NIPS 2014)*, pages 1–9. MIT Press, 2014.
- Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and Diversity in Information Retrieval Evaluation. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 659–666. ACM, 2008.
- Charles L.A. Clarke, Nick Craswell, Ian Soboroff, and Azin Ashkan. A Comparative Analysis of Cascade Measures for Novelty and Diversity. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 75–84. ACM, 2011.
- Derrick Coetzee, Armando Fox, Marti A. Hearst, and Björn Hartmann. Should your MOOC Forum use a Reputation System? In *Proceedings of the 17th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW)*, pages 1176–1187. ACM, 2014.
- Gao Cong, Long Wang, Chin-Yew Lin, Young-In Song, and Yueheng Sun. Finding Question-Answer Pairs from Online Forums. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 467–474. ACM, 2008.
- Gregorio Convertino, Massimo Zancanaro, Tiziano Piccardi, and Felipe Ortega. Toward a Mixed-Initiative QA system From Studying Predictors in Stack Exchange to Building a Mixed-Initiative Tool. *International Journal of Human-Computer Studies*, 99:1–20, 2017.

- Denzil Correa and Ashish Sureka. Fit or Unfit: Analysis and Prediction of ‘Closed Questions’ on Stack Overflow. In *Proceedings of the 1st ACM Conference on Online Social Networks*, pages 201–212. ACM, 2013.
- Denzil Correa and Ashish Sureka. Chaff from the Wheat: Characterization and Modeling of Deleted Questions on Stack Overflow. In *Proceedings of the 23rd International World Wide Web Conference*, pages 631–642. ACM, 2014.
- Steve Cronen-Townsend, Yun Zhou, and W. Bruce Croft. Predicting Query Performance. In *Proceedings of the 25th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 299–306. ACM, 2002.
- Giovanni Da San Martino, Alberto Barrón Cedeño, Salvatore Romeo, Antonio Uva, and Alessandro Moschitti. Learning to Re-Rank Questions in Community Question Answering Using Advanced Features. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1997–2000. ACM, 2016.
- Daniel Hasan Dalip, Marcos André Gonçalves, Marco Cristo, and Pavel Calado. Exploiting User Feedback to Learn to Rank Answers in Q&A Forums: A Case Study with Stack Overflow. In *Proceedings of the 36th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 543–552. ACM, 2013.
- Arpita Das, Manish Shrivastava, and Manoj Chinnakotla. Mirror on the Wall: Finding Similar Questions with Deep Structured Topic Modeling. In *Proceedings of the 2016 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 454–465. Springer, 2016a.
- Arpita Das, Harish Yenala, Manoj Chinnakotla, and Manish Shrivastava. Together We Stand: Siamese Networks for Similar Question Retrieval. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 378–387. ACL, 2016b.
- John Davies, York Sure, Holger Lausen, Ying Ding, Michael Stollberg, Dieter Fensel, Rubén Lara Hernández, and Sung-Kook Han. Semantic Web Portals: State-of-the-Art Survey. *Journal of Knowledge Management*, 9(5): 40–49, 2005.
- David Dearman and Khai N. Truong. Why Users of Yahoo! Answers do not Answer Questions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 329–332. ACM, 2010.
- P. Deepak. MixKMeans: Clustering Question-Answer Archives. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1576–1585. ACL, 2016.



- Shilin Ding, Gao Cong, Chin-Yew Lin, and Xiaoyan Zhu. Using Conditional Random Fields to Extract Contexts and Answers of Questions from Online Forums. *ACL*, 8:710–718, 2008.
- Byron Dom and Deepa Paranjpe. A Bayesian Technique for Estimating the Credibility of Question Answerers. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pages 399–409. SIAM, 2008.
- Hualei Dong, Jian Wang, Hongfei Lin, Bo Xu, and Zhihao Yang. Predicting Best Answerers for New Questions: An Approach Leveraging Distributed Representations of Words in Community Question Answering. In *Proceedings of the 9th International Conference on Frontier of Computer Science and Technology*, pages 13–18. IEEE, 2015.
- Cicero dos Santos, Luciano Barbosa, Dasha Bogdanova, and Bianca Zadrozny. Learning Hybrid Representations to Retrieve Semantically Equivalent Questions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL) and the 7th International Joint Conference on Natural Language Processing*, volume 2, pages 694–699. ACL, 2015.
- Gideon Dror, Yehuda Koren, Yoelle Maarek, and Idan Szpektor. I want to Answer; who has a Question?: Yahoo! Answers Recommender System. In *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1109–1117. ACM, 2011.
- Gideon Dror, Yoelle Maarek, and Idan Szpektor. Will my Question be Answered? Predicting “Question Answerability” in Community Question-Answering Sites. In *Proceedings of the 2013 Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, pages 499–514. Springer, 2013.
- Lan Du, Wray Buntine, and Huidong Jin. A Segmented Topic Model Based on the Two-Parameter Poisson-Dirichlet Process. *Machine Learning*, 81(1): 5–19, 2010.
- Huizhong Duan and Chengxiang Zhai. Exploiting Thread Structures to Improve Smoothing of Language Models for Forum Post Retrieval. In *Proceedings of the 33rd Annual European Conference on Information Retrieval Research (ECIR): Advances in Information Retrieval*, pages 350–361. Springer, 2011.
- Huizhong Duan, Yunbo Cao, Chin-Yew Lin, and Yong Yu. Searching Questions by Identifying Question Topic and Question Focus. In *Proceedings of the 2008 Annual Meeting of the Association for Computational Linguistics (ACL)-hlt*, pages 156–164. ACL, 2008.

- Pablo Ariel Duboue. Extractive Email Thread Summarization: Can we do Better than He Said She Said? In *Proceedings of the 7th International Conference on Natural Language Generation (INLG)*, pages 85–89. ACL, 2012.
- Jonathan L. Elsas and Jaime G. Carbonell. It Pays to be Picky: an Evaluation of Thread Retrieval in Online Forums. In *Proceedings of the 32nd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 714–715. ACM, 2009.
- Micha Elsner and Eugene Charniak. You Talking to Me? A Corpus and Algorithm for Conversation Disentanglement. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL)-hlt*, pages 834–842. ACL, 2008.
- Anthony Fader, Luke S. Zettlemoyer, and Oren Etzioni. Paraphrase-Driven Learning for Open Question Answering. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1608–1618. ACL, 2013.
- Ch. Muhammad Shahzad Faisal, Ali Daud, Faisal Imran, and Seungmin Rho. A Novel Framework for Social Web Forums’ Thread Ranking Based on Semantics and Post Quality Features. *The Journal of Supercomputing*, pages 1–20, 2016.
- Weiguo Fan. *Effective Search in Online Knowledge Communities: A Genetic Algorithm Approach*. PhD thesis, Virginia Polytechnic Institute and State University, 2009.
- Robert Farrell, Peter G. Fairweather, and Kathleen Snyder. Summarization of Discussion Groups. In *Proceedings of the 10th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 532–534. ACM, 2001.
- Donghui Feng, Jihie Kim, Erin Shaw, and Eduard Hovy. Towards Modeling Threaded Discussions Using Induced Ontology Knowledge. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence*, pages 1289–1294. AAAI, 2006a.
- Donghui Feng, Erin Shaw, Jihie Kim, and Eduard Hovy. Learning to Detect Conversation Focus of Threaded Discussions. In *Proceedings of the 2006 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 208–215. ACL, 2006b.

- Minwei Feng, Bing Xiang, Michael R. Glass, Lidan Wang, and Bowen Zhou. Applying Deep Learning to Answer Selection: A Study and an Open Task. In *Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pages 813–820. IEEE, 2015.
- Alejandro Figueroa and Günter Neumann. Context-Aware Semantic Classification of Search Queries for Browsing Community Question-Answering Archives. *KBS*, 96:1–13, 2016.
- Simone Filice, Danilo Croce, Alessandro Moschitti, and Roberto Basili. KeLP at SemEval-2016 Task 3: Learning Semantic Relations between Questions and Answers. *Proceedings of the 10th Conference on Semantic Evaluation (SemEval)*, pages 1116–1123, 2016.
- Therese Firmin and Michael J. Chrzanowski. An Evaluation of Automatic Text Summarization Systems. *AATS*, 325:336, 1999.
- Blaz Fortuna, Eduarda Mendes Rodrigues, and Natasa Milic-Frayling. Improving the Classification of Newsgroup Messages Through Social Network Analysis. In *Proceedings of the 16th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 877–880. ACM, 2007.
- Daniel Fried, Peter Jansen, Gustave Hahn-Powell, Mihai Surdeanu, and Peter Clark. Higher-order Lexical Semantic Models for Non-Factoid Answer Reranking. *TACL*, 3:197–210, 2015.
- Hongping Fu, Zhendong Niu, Chunxia Zhang, Hanchao Yu, Jing Ma, Jie Chen, Yiqiang Chen, and Junfa Liu. ASELM: Adaptive Semi-Supervised ELM with Application in Question Subjectivity Identification. *Neurocomputing*, 207:599–609, 2016a.
- Min Fu, Min Zhu, Yabo Su, Qihui Zhu, and Mingzhao Li. Modeling Temporal Behavior to Identify Potential Experts in Question Answering Communities. In *Proceedings of the 2016 International Conference on Cooperative Design, Visualization and Engineering (CDVE)*, pages 51–58. Springer, 2016b.
- Bojan Furlan, Bosko Nikolic, and Veljko Milutinovic. A Survey of Intelligent Question Routing Systems. In *Proceedings of the 6th IEEE International Conference Intelligent Systems (IS)*, pages 014–020. IEEE, 2012.
- Adabriand Furtado, Nazareno Andrade, Nigini Oliveira, and Francisco Brasileiro. Contributor Profiles, their Dynamics, and their Importance in Five Q&A Sites. In *Proceedings of the 2013 ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW)*, pages 1237–1252. ACM, 2013.

- Kuzman Ganchev, João Graça, Jennifer Gillenwater, and Ben Taskar. Posterior Regularization for Structured Latent Variable Models. *JourMLR*, 11 (Jul):2001–2049, 2010.
- Li Gao, Yao Lu, Qin Zhang, Hong Yang, and Yue Hu. Query Expansion for Exploratory Search with Subtopic Discovery in Community Question Answering. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 4715–4720. IEEE, 2016.
- Nikesh Garera, Chris Callison-Burch, and David Yarowsky. Improving Translation Lexicon Induction from Monolingual Corpora via Dependency Contexts and Part-of-Speech Equivalences. In *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL)*, pages 129–137. ACL, 2009.
- Rich Gazan. Social Q&A. *JASIST*, 62(12):2301–2312, 2011.
- S. Geerthik, S. Venkatraman, and Rajiv Gandhi. AnswerRank: Identifying Right Answers in QA system. *IJECE*, 6(4):1889, 2016.
- T. Georgiou, M. Karvounis, and Y. Ioannidis. Extracting Topics of Debate Between Users on Web Discussion Boards. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*. ACM, 2010.
- Andreas Girgensohn and Alison Lee. Making Web Sites Be Places for Social Interaction. In *Proceedings of the 2002 ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW)*, pages 136–145. ACM, 2002.
- George Gkotsis, Karen Stepanyan, Carlos Pedrinaci, John Domingue, and Maria Liakata. It’s All in the Content: State of the Art Best Answer Prediction Based on Discretisation of Shallow Linguistic Features. In *Proceedings of the 2014 Web Science Conference (WebSci)*, pages 202–210. ACM, 2014.
- Swapna Gottipati, David Lo, and Jing Jiang. Finding Relevant Answers in Software Forums. In *Proceedings of the 26th IEEE/ACM International Conference On Automated Software Engineering (ASE)*, pages 323–332. IEEE, 2011.
- Barbara J. Grosz and Candace L. Sidner. Attention, Intention and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204, 1986.
- Vladislav A. Grozin, Natalia F. Gusarova, and Natalia V. Dobrenko. Feature Selection for Language Independent Text Forum Summarization. In *Proceedings of the 6th International Conference on Knowledge Engineering and the Semantic Web (KESW)*, pages 63–71. Springer, 2015.

- Toni Gruetze, Ralf Krestel, and Felix Naumann. Topic Shifts in StackOverflow: Ask it Like Socrates. In *Natural Language Processing and Information Systems: Proceedings of the 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016*, pages 213–221. Springer, 2016.
- Jeanette K. Gundel and Thorstein Fretheim. Topic and Focus. *The Handbook of Pragmatics*, 175:196, 2004.
- Jinwen Guo, Shengliang Xu, Shenghua Bao, and Yong Yu. Tapping on the Potential of Q&A Community by Recommending Answer Providers. In *Proceedings of the 17th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 921–930. ACM, 2008.
- Lifan Guo and Xiaohua Hu. Identifying Authoritative and Reliable Contents in Community Question Answering with Domain Knowledge. In *Proceedings of the 2013 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 133–142. Springer, 2013.
- Iryna Gurevych, Delphine Bernhard, Kateryna Ignatova, and Cigdem Toprak. Educational Question Answering Based on Social Media Content. In *Proceedings of the International Conference on Artificial Intelligence in Education (IJAIED)*, pages 133–140. Springer, 2009.
- Zoltan Gyongyi, Georgia Koutrika, Jan Pedersen, and Hector Garcia-Molina. Questioning Yahoo! Answers. Technical Report Technical Report, Stanford Infolab, 2007.
- Xiaohui Han, Jun Ma, Yun Wu, and Chaoran Cui. A Novel Machine Learning Approach to Rank Web Forum Posts. *Soft Computing*, 18(5):941–959, 2014.
- Benjamin V. Hanrahan, Gregorio Convertino, and Les Nelson. Modeling Problem Difficulty and Expertise in StackOverflow. In *Proceedings of the 2012 ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW)*, pages 91–94. ACM, 2012.
- Tianyong Hao and Eugene Agichtein. Finding Similar Questions in Collaborative Question Answering Archives: Toward Bootstrapping-based Equivalent Pattern Learning. *Information Retrieval*, 15(3):332–353, 2012a.
- Tianyong Hao and Eugene Agichtein. Bootstrap-based Equivalent Pattern Learning for Collaborative Question Answering. In *Proceedings of the 13th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, pages 318–329. Springer, 2012b.

- Sanda M. Harabagiu, Dan I. Moldovan, Marius Paşca, Rada Mihalcea, Mihai Surdeanu, Răzvan Bunescu, Corina R Gîrju, Vasile Rus, and Paul Morărescu. Falcon: Boosting Knowledge for Answer Engines. In *Proceedings of the 9th Text REtrieval Conference (TREC)*, pages 479–488. NIST, 2000.
- F. Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, and Joseph A. Konstan. Predictors of Answer Quality in Online Q&A Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 865–874. ACM, 2008.
- F. Maxwell Harper, Daniel Moy, and Joseph A. Konstan. Facts or Friends?: Distinguishing Informational and Conversational Questions in Social Q&A Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 759–768. ACM, 2009.
- Mithak I. Hashem. Improvement of Email Summarization Using Statistical Based Method. *International Journal of Computer Science and Mobile Computing (IJCSMC)*, 3(2):382–388, 2014.
- Ahmed Hassan, Vahed Qazvinian, and Dragomir Radev. What’s with the Attitude?: Identifying Sentences with Attitude in Online Discussions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1245–1255. ACL, 2010.
- Taher H. Haveliwala. Topic-Sensitive PageRank. In *Proceedings of the 11th International World Wide Web Conference*, pages 517–526. ACM, 2002.
- Jing He and Decheng Dai. Summarization of Yes/No Questions Using a Feature Function Model. In *Proceedings of the 3rd Asian Conference on Machine Learning (ACML)*, pages 351–366. Springer, 2011.
- Ulf Hermjakob. Parsing and Question Classification for Question Answering. In *Proceedings of the ACL Workshop on Open-Domain Question Answering (QA)*, volume 12, pages 1–6. ACL, 2001.
- Atefeh Heydari, Mohammadali Tavakoli, Zuriati Ismail, and Naomie Salim. Leveraging Quality Metrics in Voting Model Based Thread Retrieval. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 10(1):117–123, 2016.
- Geoffrey E. Hinton. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14(8):1771–1800, 2002.
- Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7):1527–1554, 2006.

- Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.
- Thomas Hofmann. Probabilistic Latent Semantic Indexing. In *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 50–57. ACM, 1999.
- Liangjie Hong and Brian D. Davison. A Classification-Based Approach to Question Answering in Discussion Boards. In *Proceedings of the 32nd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 171–178. ACM, 2009.
- Doris Hoogeveen, Karin M. Verspoor, and Timothy Baldwin. CQADup-Stack: A Benchmark Data Set for Community Question-Answering Research. In *Proceedings of the 20th Australasian Document Computing Symposium (ADCS)*, pages 3–9. ACM, 2015.
- Eduard Hovy and Chin-Yew Lin. Automated Text Summarization and the SUMMARIST System. In *Proceedings of the workshop on TIPSTER held at Baltimore, Maryland: October 13-15, 1998*, pages 197–214. ACL, 1998.
- Eduard Hovy, Laurie Gerber, Ulf Hermjakob, Chin-Yew Lin, and Deepak Ravichandran. Toward Semantics-based Answer Pinpointing. In *Proceedings of the 2001 Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 1–7. ACL, 2001.
- Wei-Ning Hsu, Yu Zhang, and James Glass. Recurrent Neural Network Encoder with Attention for Community Question Answering. *CoRR*, arXiv preprint arXiv:1603.07044, 2016.
- Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme Learning Machine: Theory and Applications. *Neurocomputing*, 70(1):489–501, 2006.
- Jizhou Huang, Ming Zhou, and Dan Yang. Extracting Chatbot Knowledge from Online Discussion Forums. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 423–428. Morgan Kaufmann Publishers, 2007.
- Zhiheng Huang, Marcus Thint, and Zengchang Qin. Question Classification Using Head Words and Their Hypernyms. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 927–936. ACL, 2008.
- Rodney Huddleston. *English Grammar: An Outline*. Cambridge University Press, 1988.

- Daisuke Ishikawa, Tetsuya Sakai, and Noriko Kando. Overview of the NTCIR-8 Community QA Pilot Task (Part I): The Test Collection and the Task. In *Proceedings of the 8th NTCIR Workshop on Research in Information Access Technologies Information Retrieval, Question Answering and Summarization*, pages 421–432. ACM, 2010.
- Abraham Ittycheriah, Martin Franz, Wei-Jing Zhu, Adwait Ratnaparkhi, and Richard J. Mammone. Question Answering Using Maximum Entropy Components. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 1–7. ACL, 2001.
- Peter Jansen, Mihai Surdeanu, and Peter Clark. Discourse Complements Lexical Semantics for Non-Factoid Answer Reranking. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 977–986. ACL, 2014.
- Kalervo Järvelin and Jaana Kekäläinen. Cumulated Gain-based Evaluation of IR Techniques. *TOIS*, 20(4):422–446, 2002.
- Jiwoon Jeon, W. Bruce Croft, and Joon Ho Lee. Finding Semantically Similar Questions Based on Their Answers. In *Proceedings of the 28th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 617–618. ACM, 2005a.
- Jiwoon Jeon, W. Bruce Croft, and Joon Ho Lee. Finding Similar Questions in Large Question and Answer Archives. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 84–90. ACM, 2005b.
- Jiwoon Jeon, W. Bruce Croft, Joon Ho Lee, and Soyeon Park. A Framework to Predict the Quality of Answers with Non-textual Features. In *Proceedings of the 29th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 228–235. ACM, 2006.
- Minwoo Jeong, Chin-Yew Lin, and Gary Geunbae Lee. Semi-Supervised Speech Act Recognition in Emails and Forums. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1250–1259. ACL, 2009.
- Zongcheng Ji, Fei Xu, Bin Wang, and Ben He. Question-Answer Topic Model for Question Retrieval in Community Question Answering. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2471–2474. ACM, 2012.



- Jian Jiao. *A Framework for Finding and Summarizing Product Defects, and Ranking Helpful Threads from Online Customer Forums Through Machine Learning*. PhD thesis, Virginia Polytechnic Institute and State University, 2013.
- Blooma Mohan John, Dion Hoe Lian Goh, Alton Yeow Kuan Chua, and Nilmini Wickramasinghe. Graph-based Cluster Analysis to Identify Similar Questions: A Design Science Approach. *JAIS*, 17(9):590, 2016.
- Shafiq Joty, Alberto Barrón-Cedeno, Giovanni Da San Martino, Simone Filice, Lluís Màrquez, Alessandro Moschitti, and Preslav Nakov. Global Thread-Level Inference for Comment Classification in Community Question Answering. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 573–578. ACL, 2015.
- Pawel Jurczyk and Eugene Agichtein. Discovering Authorities in Question Answer Communities by Using Link Analysis. In *Proceedings of the 16th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 919–922. ACM, 2007a.
- Pawel Jurczyk and Eugene Agichtein. Hits on Question Answer Portals: Exploration of Link Analysis for Author Ranking. In *Proceedings of the 30th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 845–846. ACM, 2007b.
- Wei-Chen Kao, Duen-Ren Liu, and Shiu-Wen Wang. Expert Finding in Question-Answering Websites: A Novel Hybrid Approach. In *Proceedings of the 25th ACM Symposium on Applied Computing (SAC)*, pages 867–871. ACM, 2010.
- Pairin Katerattanakul and Keng Siau. Measuring Information Quality of Web Sites: Development of an Instrument. In *Proceedings of the 20th International Conference on Information Systems*, pages 279–285. AIS, 1999.
- Boris Katz, Sue Felshin, Deniz Yuret, Ali Ibrahim, Jimmy Lin, Gregory Marton, Alton Jerome McFarland, and Baris Temelkuran. Omnibase: Uniform Access to Heterogeneous Data for Question Answering. In *Proceedings of the 2002 International Conference on Application of Natural Language to Information Systems*, pages 230–234. Springer, 2002.
- Jihie Kim, Grace Chern, Donghui Feng, Erin Shaw, and Eduard Hovy. Mining and Assessing Discussions on the Web Through Speech Act Analysis. In *Proceedings of the 5th International Semantic Web Conference (ISWC), Workshop on Web Content Mining with Human Language Technologies*. Springer, 2006.

- Jong Wook Kim, K Selçuk Candan, and Mehmet E. Dönderler. Topic Segmentation of Message Hierarchies for Indexing and Navigation Support. In *Proceedings of the 14th International World Wide Web Conference*, pages 322–331. ACM, 2005.
- Soojung Kim and Sanghee Oh. Users’ Relevance Criteria for Evaluating Answers in a Social Q&A Site. *JASIST*, 60(4):716–727, 2009.
- Soojung Kim, Jung Sun Oh, and Sanghee Oh. Best-Answer Selection Criteria in a Social Q&A Site from the User-oriented Relevance Perspective. *JASIST*, 44(1):1–15, 2007.
- Su Nam Kim, Lawrence Cavdon, and Timothy Baldwin. Classifying Dialogue Acts in One-on-one Live Chats. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 862–871. ACL, 2010a.
- Su Nam Kim, Lawrence Cavdon, and Timothy Baldwin. Classifying Dialogue Acts in One-on-one Live Chats. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 862–871. ACL, 2010b.
- Su Nam Kim, Li Wang, and Timothy Baldwin. Tagging and Linking Web Forum Posts. In *Proceedings of the 14th Conference on Computational Natural Language Learning (CoNLL)*, pages 192–202. ACL, 2010c.
- Vanessa Kitzie, Erik Choi, and Chirag Shah. Analyzing Question Quality Through Intersubjectivity: World Views and Objective Assessments of Questions on Social Question-Answering. *JASIST*, 50(1):1–10, 2013.
- Mike Klaas. Toward Indicative Discussion Fora Summarization. Technical Report TR-2005-04, UBC CS, 2005.
- Jon M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5):604–632, 1999.
- Philipp Koehn and Kevin Knight. Learning a Translation Lexicon from Monolingual Corpora. In *Proceedings of the ACL-02 Workshop on Unsupervised Lexical Acquisition*, pages 9–16. ACL, 2002.
- Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical Phrase-based Translation. In *Proceedings of the 2003 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 48–54. ACL, 2003.
- Giridhar Kumaran and James Allan. Text Classification and Named Entities for New Event Detection. In *Proceedings of the 27th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 297–304. ACM, 2004.

- John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, pages 282–289. JMLR, 2001.
- Derek Lam. *Exploiting E-mail Structure to Improve Summarization*. PhD thesis, MIT, 2002.
- Andrew Lampert, Robert Dale, and Cécile Paris. The Nature of Requests and Commitments in Email Messages. In *Proceedings of the AAAI Workshop on Enhanced Messaging (WS-08-04)*, pages 42–47. AAAI, 2008.
- Man Lan, Guoshun Wu, Chunyun Xiao, Yuanbin Wu, and Ju Wu. Building Mutually Beneficial Relationships between Question Retrieval and Answer Ranking to Improve Performance of Community Question Answering. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 832–839. IEEE, 2016.
- Mirella Lapata. Automatic Evaluation of Information Ordering: Kendall’s Tau. *Computational Linguistics*, 32(4):471–484, 2006.
- Thomas D. LaToza and Brad A. Myers. Hard-to-Answer Questions about Code. In *Evaluation and Usability of Programming Languages and Tools*, page 8. ACM, 2010.
- Jey Han Lau and Timothy Baldwin. An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. In *Proceedings of the 1st Workshop on Representation Learning for NLP (RepL4NLP)*, pages 78–86. ACL, 2016.
- Long T. Le, Chirag Shah, and Erik Choi. Evaluating the Quality of Educational Answers in Community Question-Answering. In *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries (JCDL)*, pages 129–138. ACM, 2016.
- Quoc V. Le and Tomas Mikolov. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 1188–1196. JMLR, 2014.
- Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- JunChoi Lee and Yu-N Cheah. Semantic Relatedness Measure for Identifying Relevant Answers in Online Community Question Answering Services. In *Proceedings of the 9th International Conference on IT in Asia (CITA)*. IEEE, 2015.

- Jung-Tae Lee, Sang-Bum Kim, Young-In Song, and Hae-Chang Rim. Bridging Lexical Gaps Between Queries and Questions on Large Online Q&A Collections with Compact Translation Models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 410–418. ACL, 2008.
- Tao Lei, Hrishikesh Joshi, Regina Barzilay, Tommi Jaakkola, Katerina Tymoshenko, Alessandro Moschitti, and Lluís Marquez. Denoising Bodies to Titles: Retrieving Similar Questions with Recurrent Convolutional Models. *CoRR*, abs/1512.05726, 2015.
- Tao Lei, Hrishikesh Joshi, Regina Barzilay, Tommi S. Jaakkola, Katerina Tymoshenko, Alessandro Moschitti, and Lluís Màrquez. Semi-Supervised Question Retrieval with Gated Convolutions. In *Proceedings of the 2016 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 1279–1289. ACL, 2016.
- Oliver Lemon, Alex Gruenstein, and Stanley Peters. Collaborative Activities and Multi-Tasking in Dialogue Systems. *Traitement Automatique des Langues (TAL), Special Issue on Dialogue*, 43(2):131–154, 2002.
- Baichuan Li. *A Computational Framework for Question Processing in Community Question Answering Services*. PhD thesis, Chinese University of Hong Kong, 2014.
- Baichuan Li and Irwin King. Routing Questions to Appropriate Answerers in Community Question Answering Services. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1585–1588. ACM, 2010.
- Baichuan Li, Irwin King, and Michael R. Lyu. Question Routing in Community Question Answering: Putting Category in its Place. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2041–2044. ACM, 2011.
- Baichuan Li, Tan Jin, Michael R. Lyu, Irwin King, and Barley Mak. Analyzing and Predicting Question Quality in Community Question Answering Services. In *Proceedings of the 21st International World Wide Web Conference*, pages 775–782. ACM, 2012.
- Baoli Li, Yandong Liu, and Eugene Agichtein. CoCQA: Co-training over Questions and Answers with an Application to Predicting Question Subjectivity Orientation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 937–946. ACL, 2008a.

- Baoli Li, Yandong Liu, Ashwin Ram, Ernest V. Garcia, and Eugene Agichtein. Exploring Question Subjectivity Prediction in Community QA. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 735–736. ACM, 2008b.
- Shuguang Li and Suresh Manandhar. Improving Question Recommendation by Exploiting Information Need. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)-hlt*, volume 1, pages 1425–1434. ACL, 2011.
- Xin Li and Dan Roth. Learning Question Classifiers. In *Proceedings of the 19th International Conference on Computational Linguistics*, volume 1, pages 1–7. ACL, 2002.
- Yiyang Li, Lei Su, Jun Chen, and Liwei Yuan. Semi-Supervised Learning for Question Classification in CQA. *Natural Computing*, pages 1–11, 2016.
- Chen Lin, Jiang-Ming Yang, Rui Cai, Xin-Jing Wang, Wei Wang, and Lei Zhang. Modeling Semantics and Structure of Discussion Threads. In *Proceedings of the 18th International World Wide Web Conference*, pages 1103–1104. ACM, 2009.
- Chin-Yew Lin. ROUGE: A Package for Automatic Evaluation of Summaries. In *Proceedings of the Text Summarization Branches Out Workshop*, pages 74–81. ACL, 2004.
- Chin-Yew Lin and Eduard Hovy. From Single to Multi-Document Summarization: A Prototype System and its Evaluation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 457–464. ACL, 2002.
- Fei Liu, Alistair Moffat, Timothy Baldwin, and Xiuzhen Zhang. Quit While Ahead: Evaluating Truncated Rankings. In *Proceedings of the 39th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 953–956. ACM, 2016.
- Qiaoling Liu and Eugene Agichtein. Modeling Answerer Behavior in Collaborative Question Answering Systems. In *Proceedings of the 33rd Annual European Conference on Information Retrieval Research (ECIR): Advances in Information Retrieval*, pages 67–79. Springer, 2011.
- Qiaoling Liu, Eugene Agichtein, Gideon Dror, Evgeniy Gabrilovich, Yoelle Maarek, Dan Pelleg, and Idan Szpektor. Predicting Web Searcher Satisfaction with Existing Community-based Answers. In *Proceedings of the 34th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 415–424. ACM, 2011.

- Qiaoling Liu, Eugene Agichtein, Gideon Dror, Yoelle Maarek, and Idan Szpektor. When Web Search Fails, Searchers Become Askers: Understanding the Transition. In *Proceedings of the 35th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 801–810. ACM, 2012.
- Xiaoyong Liu, W. Bruce Croft, and Matthew Koll. Finding Experts in Community-Based Question-Answering Services. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 315–316. ACM, 2005.
- Yandong Liu and Eugene Agichtein. You’ve Got Answers: Towards Personalized Models for Predicting Success in Community Question Answering. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL): Human Language Technologies*, pages 97–100. ACL, 2008a.
- Yandong Liu and Eugene Agichtein. On the Evolution of the Yahoo! Answers QA Community. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 737–738. ACM, 2008b.
- Yandong Liu, Jiang Bian, and Eugene Agichtein. Predicting Information Seeker Satisfaction in Community Question Answering. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 483–490. ACM, 2008a.
- Yuanjie Liu, Shasha Li, Yunbo Cao, Chin-Yew Lin, Dingyi Han, and Yong Yu. Understanding and Summarizing Answers in Community-based Question Answering Services. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING)*, pages 497–504. ACL, 2008b.
- Zhe Liu and Bernard J. Jansen. Identifying and Predicting the Desire to Help in Social Question and Answering. *Information Processing & Management*, 53(2):490–504, 2016.
- Ziming Liu and Xiaobin Huang. Evaluating the Credibility of Scholarly Information on the Web: A Cross Cultural Study. *The International Information & Library Review*, 37(2):99–106, 2005.
- Byron Long and Ronald Baecker. A Taxonomy of Internet Communication Tools. In *Proceedings of the 1997 World Conference on the WWW, Internet & Intranet (WebNet)*, pages 1–15. Association for the Advancement of Computing in Education (AACE), 1997.

- Vanessa Lopez, Michele Pasin, and Enrico Motta. Aqualog: An Ontology-Portable Question Answering System for the Semantic Web. In *Proceedings of the European Semantic Web Conference (ESWC)*, pages 546–562. Springer, 2005.
- Jie Lou, Kai Hin Lim, Yulin Fang, and Jerry Zeyu Peng. Drivers Of Knowledge Contribution Quality And Quantity In Online Question And Answering Communities. In *Proceedings of the Pacific Asia Conference on Information Systems (PACIS)*, page 121. AIS, 2011.
- Jie Lou, Yulin Fang, Kai H. Lim, and Jerry Zeyu Peng. Contributing High Quantity and Quality Knowledge to Online Q&A Communities. *JASIST*, 64(2):356–371, 2013.
- Pamela J. Ludford, Dan Cosley, Dan Frankowski, and Loren Terveen. Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 631–638. ACM, 2004.
- Marco Lui and Timothy Baldwin. You Are What You Post: User-level Features in Threaded Discourse. In *Proceedings of the 14th Australasian Document Computing Symposium (ADCS)*, pages 98–105. ACM, 2009.
- Marco Lui and Timothy Baldwin. Classifying User Forum Participants: Separating the Gurus from the Hacks, and Other Tales of the Internet. In *Proceedings of the 2010 Australasian Language Technology Association Workshop (ALTA)*, pages 49–57. ACL, 2010.
- Steven Lytinen and Noriko Tomuro. The Use of Question Types to Match Questions in FAQFinder. In *Proceedings of the AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases (SS-02-06)*, pages 46–53. AAAI, 2002.
- Craig Macdonald and Iadh Ounis. Voting Techniques for Expert Search. *Knowledge and information systems*, 16(3):259–280, 2008a.
- Craig Macdonald and Iadh Ounis. Key Blog Distillation: Ranking Aggregates. In *Proceedings of the 17th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1043–1052. ACM, 2008b.
- Craig Macdonald and Iadh Ounis. Learning Models for Ranking Aggregates. In *Proceedings of the 33rd Annual European Conference on Information Retrieval Research (ECIR): Advances in Information Retrieval*, pages 517–529. Springer, 2011.
- Preetham Madeti. Using Apache Spark’s MLlib to Predict Closed Questions on Stack Overflow. Master’s thesis, Youngstown State University, 2016.

- Juha Makkonen, Helena Ahonen-Myka, and Marko Salmenkivi. Simple Semantics in Topic Detection and Tracking. *Information retrieval*, 7(3-4): 347–368, 2004.
- Krissada Maleewong. Predicting Quality-Assured Consensual Answers in Community-Based Question Answering Systems. In *Recent Advances in Information and Communication Technology 2016: Proceedings of the 12th International Conference on Computing and Information Technology (IC2IT)*, pages 117–127. Springer, 2016.
- Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. Design Lessons from the Fastest Q&A Site in the West. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2857–2866. ACM, 2011.
- Daniel Marcu and William Wong. A Phrase-based, Joint Probability Model for Statistical Machine Translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 133–139. ACL, 2002.
- Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, 5(4):1093–1113, 2014.
- Donald Metzler and W. Bruce Croft. Analysis of Statistical Question Classification for Fact-based Questions. *Information Retrieval*, 8(3):481–504, 2005.
- Rada Mihalcea and Paul Tarau. TextRank: Bringing Order into Texts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 404–411. ACL, 2004.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *CoRR*, abs/1301.3781, 2013.
- Zhao-Yan Ming, Tat-Seng Chua, and Gao Cong. Exploring Domain-specific Term Weight in Archived Question Search. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1605–1608. ACM, 2010.
- Alistair Moffat and Justin Zobel. Rank-biased Precision for Measurement of Retrieval Effectiveness. *TOIS*, 27(1):2, 2008.
- Piero Molino, Luca Maria Aiello, and Pasquale Lops. Social Question Answering: Textual, User, and Network Features for Best Answer Prediction. *TOIS*, 35(1):4, 2016.



- Dana Movshovitz-Attias, Yair Movshovitz-Attias, Peter Steenkiste, and Christos Faloutsos. Analysis of the Reputation System and User Contributions on a Question Answering Website: StackOverflow. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 886–893. IEEE, 2013.
- Preslav Nakov, Lluís Màrquez, Walid Magdy, Alessandro Moschitti, Jim Glass, and Bilal Randeree. SemEval-2015 Task 3: Answer Selection in Community Question Answering. In *Proceedings of the 9th Conference on Semantic Evaluation (SemEval)*, pages 269–281. ACL, 2015.
- Preslav Nakov, Lluís Màrquez, Alessandro Moschitti, Walid Magdy, Hamdy Mubarak, abed Alhakim Freihat, Jim Glass, and Bilal Randeree. SemEval-2016 Task 3: Community Question Answering. In *Proceedings of the 10th Conference on Semantic Evaluation (SemEval)*, pages 525–545. ACL, 2016.
- Preslav Nakov, Doris Hoogeveen, Lluís Màrquez, Alessandro Moschitti, Hamdy Mubarak, Timothy Baldwin, and Karin M. Verspoor. SemEval-2017 Task 3: Community Question Answering. In *Proceedings of the 11th Conference on Semantic Evaluation (SemEval)*. ACL, 2017.
- Henry Nassif, Mitra Mohtarami, and James Glass. Learning Semantic Relatedness in Community Question Answering Using Neural Models. In *Proceedings of the 1st Workshop on Representation Learning for NLP (Repl4NLP)*, pages 137–147. ACL, 2016.
- Ani Nenkova and Amit Bagga. Facilitating Email Thread Access by Extractive Summary Generation. In *Proceedings of the 2003 International Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 287–296. ACL, 2003.
- Ani Nenkova, Rebecca Passonneau, and Kathleen McKeown. The Pyramid Method: Incorporating Human Content Selection Variation in Summarization Evaluation. *TSLP*, 4(2):1–13, 2007.
- Paula S. Newman. Exploring Discussion Lists: Steps and Directions. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 126–134. ACM, 2002.
- Paula S. Newman and John C. Blitzer. Summarizing Archived Discussions: A Beginning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI '03)*, pages 273–276. ACM, 2003.
- Lan Nie, Brian D. Davison, and Xiaoguang Qi. Topical Link Analysis for Web Search. In *Proceedings of the 29th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 91–98. ACM, 2006.

- Yuanping Nie, Jiuming Huang, Zongsheng Xie, Hai Li, Pengfei Zhang, and Yan Jia. NudtMDP at TREC 2015 LiveQA Track. In *Proceedings of the 24th Text REtrieval Conference (TREC) (LiveQA Track)*. NIST, 2015.
- Michael Niemann. *The Duality of Expertise: Identifying Expertise Claims and Community Opinions within Online Forum Dialogue*. PhD thesis, Monash University, 2015.
- Blair Nonnecke and Jenny Preece. Lurker Demographics: Counting the Silent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 73–80. ACM, 2000.
- Adekunle Isiaka Obasa, Naomie Salim, and Atif Khan. Hybridization of Bag-of-Words and Forum Metadata for Web Forum Question Post Detection. *Indian Journal of Science and Technology*, 8(32):1–12, 2016.
- Franz Josef Och, Christoph Tillmann, and Hermann Ney. Improved Alignment Models for Statistical Machine Translation. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 20–28. ACL, 1999.
- Adi Omari, David Carmel, Oleg Rokhlenko, and Idan Szpektor. Novelty Based Ranking of Human Answers for Community Questions. In *Proceedings of the 39th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 215–224. ACM, 2016.
- Daniel F.O. Onah, Jane E. Sinclair, and Russell Boyatt. Exploring the Use of MOOC Discussion Forums. In *Proceedings of the London International Conference on Education*, pages 1–4. Infonomics Society, 2014.
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999-66, Stanford InfoLab, 1999.
- Aditya Pal and Joseph A. Konstan. Expert Identification in Community Question Answering: Exploring Question Selection Bias. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1505–1508. ACM, 2010.
- Aditya Pal, Shuo Chang, and Joseph A. Konstan. Evolution of Experts in Question Answering Communities. In *Proceedings of the 6e AAAI International Conference on Weblogs and Social Media (ICWSM)*, pages 274–281. AAAI, 2012a.
- Aditya Pal, F. Maxwell Harper, and Joseph A. Konstan. Exploring Question Selection Bias to Identify Experts and Potential Experts in Community Question Answering. *TOIS*, 30(2):10, 2012b.

- C. Pechsiri and R. Piriyaikul. Developing a Why-How Question Answering System on Community Web Boards with a Causality Graph Including Procedural Knowledge. *Information Processing in Agriculture*, 3(1):36–53, 2016.
- Dan Pelleg and Andrew W. Moore. *X*-means: Extending *K*-means with Efficient Estimation of the Number of Clusters. In *Proceedings of the 17th International Conference on Machine Learning (ICML)*, pages 727–734. JMLR, 2000.
- Anselmo Peñas and Alvaro Rodrigo. A Simple Measure to Assess Non-Response. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL): Human Language Technologies*, pages 1415–1424. ACL, 2011.
- Florent Perronnin and Christopher Dance. Fisher Kernels on Visual Vocabularies for Image Categorization. In *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- Boaz Petersil, Avihai Mejer, Idan Szpektor, and Koby Crammer. That’s not my Question: Learning to Weight Unmatched Terms in CQA Vertical Search. In *Proceedings of the 39th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 225–234. ACM, 2016.
- Yuval Pinter, Roi Reichart, and Idan Szpektor. Syntactic Parsing of Web Queries with Question Intent. *Proceedings of the 2016 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 670–680, 2016.
- Jay M. Ponte. *A Language Modeling Approach to Information Retrieval*. PhD thesis, University of Massachusetts Amherst, 1998.
- Jay M. Ponte and W. Bruce Croft. A Language Modeling Approach to Information Retrieval. In *Proceedings of the 21st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 275–281. ACM, 1998.
- Luca Ponzanelli, Andrea Mocci, Alberto Bacchelli, and Michele Lanza. Understanding and Classifying the Quality of Technical Forum Questions. In *Proceedings of the 14th International Conference on Quality Software*, pages 343–352. IEEE, 2014.
- Ana-Maria Popescu, Oren Etzioni, and Henry Kautz. Towards a Theory of Natural Language Interfaces to Databases. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, pages 149–157. ACM, 2003.

- Jagat Sastry Pudipeddi, Leman Akoglu, and Hanghang Tong. User Churn in Focused Question Answering Sites: Characterizations and Prediction. In *Proceedings of the 23rd International World Wide Web Conference*, pages 469–474. ACM, 2014.
- Minghui Qiu and Jing Jiang. A Latent Variable Model for Viewpoint Discovery from Threaded Forum Posts. In *Proceedings of the 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 1031–1040. ACL, 2013.
- Xipeng Qiu and Xuanjing Huang. Convolutional Neural Tensor Network Architecture for Community-based Question Answering. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1305–1311. AAAI, 2015.
- Bo Qu, Gao Cong, Cuiping Li, Aixin Sun, and Hong Chen. An Evaluation of Classification Models for Question Topic Categorization. *JASIST*, 63(5): 889–903, 2012.
- Mingcheng Qu, Guang Qiu, Xiaofei He, Cheng Zhang, Hao Wu, Jiajun Bu, and Chun Chen. Probabilistic Question Recommendation for Question Answering Communities. In *Proceedings of the 18th International World Wide Web Conference*, pages 1229–1230. ACM, 2009.
- Daphne Ruth Raban. The Incentive Structure in an Online Information Market. *JASIST*, 59(14):2284–2295, 2008.
- Daphne Ruth Raban. Self-Presentation and the Value of Information in Q&A Websites. *JASIST*, 60(12):2465–2473, 2009.
- Davood Rafiei, Krishna Bharat, and Anand Shukla. Diversifying Web Search Results. In *Proceedings of the 19th International World Wide Web Conference*, pages 781–790. ACM, 2010.
- Preethi Raghavan, Rose Catherine, Shajith Iqbal, Nanda Kambhatla, and Debapriyo Majumdar. Extracting Problem and Resolution Information from Online Discussion Forums. In *Proceedings of the 16th International Conference on Management of Data (COMAD)*, pages 77–88. Computer Society of India, 2010.
- Owen Rambow, Lokesh Shrestha, John Chen, and Chirsty Lauridsen. Summarizing Email Threads. In *Proceedings of the 2004 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 105–108. ACL, 2004.

- Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra, and Yann LeCun. Efficient Learning of Sparse Representations with an Energy-based Model. *NIPS*, pages 1137–1144, 2007.
- Zhaochun Ren, Hongya Song, Piji Li, Shangsong Liang, Jun Ma, and Maarten de Rijke. Using Sparse Coding for Answer Summarization in Non-Factoid Community Question-Answering. In *Proceedings of the 2016 SIGIR WebQA Workshop*. ACM, 2016.
- Fatemeh Riahi, Zainab Zolaktaf, Mahdi Shafiei, and Evangelos Milios. Finding Expert Users in Community Question Answering. In *Proceedings of the 21st International World Wide Web Conference*, pages 791–798. ACM, 2012.
- Soo Young Rieh. Judgment of Information Quality and Cognitive Authority in the Web. *JASIST*, 53(2):145–161, 2002.
- Soo Young Rieh and Nicholas J. Belkin. Understanding Judgment of Information Quality and Cognitive Authority in the WWW. In *Proceedings of the 61st Annual Meeting of the American Society for Information Science*, volume 35, pages 279–289. ASIS&T, 1998.
- R. Rienks. *Meetings in Smart Environments: Implications of Progressing Technology*. PhD thesis, University of Twente, 2007.
- Stephen E. Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, and Mike Gatford. Okapi at TREC-3. In *Proceedings of the 3rd Text REtrieval Conference (TREC)*, pages 109–126. NIST, 1994.
- Stephen E. Robertson, Hugo Zaragoza, and Michael Taylor. Simple BM25 Extension to Multiple Weighted Fields. In *Proceedings of the 13th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 42–49. ACM, 2004.
- Anthony C. Robinson. Exploring Class Discussions from a Massive Open Online Course (MOOC) on Cartography. *Modern Trends in Cartography*, pages 173–182, 2015.
- Carolyn P. Rosé, Barbara S. Di Eugenio, Lori Levin, and Carol Van Ess-Dykema. Discourse Processing of Dialogues with Multiple Threads. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 31–38. ACL, 1995.
- Daniel E. Rose and Danny Levinson. Understanding User Goals in Web Search. In *Proceedings of the 13th International World Wide Web Conference*, pages 13–19. ACM, 2004.

- Lorenzo A. Rossi and Omprakash Gnawali. Language Independent Analysis and Classification of Discussion Threads in Coursera MOOC Forums. In *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IRI)*, pages 654–661. IEEE, 2014.
- Ripon K. Saha, Avigit K. Saha, and Dewayne E. Perry. Toward Understanding the Causes of Unanswered Questions in Software Information Sites: A Case Study of Stack Overflow. In *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering*, pages 663–666. ACM, 2013.
- Tirath Prasad Sahu, Naresh Nagwani, and Shrish Verma. Multivariate Beta Mixture Model for Automatic Identification of Topical Authoritative Users in Community Question Answering Sites. *IEEE Access*, 4:5343–5355, 2016a.
- Tirath Prasad Sahu, Naresh Kumar Nagwani, and Shrish Verma. TagLDA based User Persona Model to Identify Topical Experts for Newly Posted Questions in Community Question Answering Sites. *International Journal of Applied Engineering Research*, 11(10):7072–7078, 2016b.
- Tirath Prasad Sahu, Naresh Kumar Nagwani, and Shrish Verma. Topical Authoritative Answerer Identification on Q&A Posts using Supervised Learning in CQA Sites. In *Proceedings of the 9th Annual ACM India Conference*, pages 129–132. ACM, 2016c.
- Tetsuya Sakai, Daisuke Ishikawa, Noriko Kando, Yohei Seki, Kazuko Kuriyama, and Chin-Yew Lin. Using Graded-Relevance Metrics for Evaluating Community QA Answer Selection. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 187–196. ACM, 2011.
- Daniel Schall and Florian Skopik. An Analysis of the Structure and Dynamics of Large-scale Q/A Communities. In *Proceedings of the East European Conference on Advances in Databases and Information Systems*, pages 285–301. Springer, 2011.
- Kim Schouten and Flavius Frasinca. Finding Implicit Features in Consumer Reviews for Sentiment Analysis. In *Proceedings of the 2014 International Conference on Web Engineering*, pages 130–144. Springer, 2014.
- Anne Schuth, Maarten Marx, and Maarten de Rijke. Extracting the Discussion Structure in Comments on News-Articles. In *Proceedings of the 9th Annual ACM International Workshop on Web Information and Data Management*, pages 97–104. ACM, 2007.
- John R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, 1969.

- Jangwon Seo and W. Bruce Croft. Blog Site Search Using Resource Selection. In *Proceedings of the 17th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1053–1062. ACM, 2008.
- Jangwon Seo, W. Bruce Croft, and David A. Smith. Online Community Search Using Thread Structure. In *Proceedings of the 18th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1907–1910. ACM, 2009.
- Jangwon Seo, W. Bruce Croft, and David A. Smith. Online Community Search Using Conversational Structures. *Information Retrieval*, 14(6):547–571, 2011.
- Aliaksei Severyn and Alessandro Moschitti. Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks. In *Proceedings of the 38th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 373–382. ACM, 2015.
- Pnina Shachaf. Answer Reliability on Q&A Sites. In *Proceedings of the Americas Conference on Information Systems (AMCIS)*, page 376. AIS, 2010.
- Pnina Shachaf. A Comparative Assessment of Answer Quality on Four Question Answering Sites. *JIS*, 37(5):476–486, 2011.
- Giovanni Da Shafiq Joty, Lluís Màrquez, and Preslav Nakov. Joint Learning with Global Inference for Comment Classification in Community Question Answering. In *Proceedings of the 2016 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 703–713. ACL, 2016.
- Chirag Shah. Building a Parsimonious Model for Identifying Best Answers Using Interaction History in Community Q&A. *JASIST*, 52(1):1–10, 2015.
- Chirag Shah and Vanessa Kitzie. Social Q&A and Virtual Reference – Comparing Apples and Oranges with the Help of Experts and Users. *JASIST*, 63(10):2020–2036, 2012.
- Chirag Shah and Jefferey Pomerantz. Evaluating and Predicting Answer Quality in Community QA. In *Proceedings of the 33rd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 411–418. ACM, 2010.
- Chirag Shah, Sanghee Oh, and Jung Sun Oh. Research Agenda for Social Q&A. *Library & Information Science Research*, 31(4):205–209, 2009.
- Chirag Shah, Marie L. Radford, Lynn Silipigni Connaway, Erik Choi, and Vanessa Kitzie. “How much change do you get from \$40” - Analyzing and Addressing Failed Questions on Social Q&A. *JASIST*, 49(1):1–10, 2012.

- Chirag Shah, Vanessa Kitzie, and Erik Choi. Modalities, Motivations, and Materials – Investigating Traditional and Social Online Q&A Services. *JIS*, pages 1–19, 2014.
- Rebecca Sharp, Peter Jansen, Mihai Surdeanu, and Peter Clark. Spinning Straw into Gold: Using Free Text to Train Monolingual Alignment Models for Non-Factoid Question Answering. In *Proceedings of the 2015 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 231–237. ACL, 2015.
- Libin Shen and Aravind K. Joshi. Ranking and Reranking with Perceptron. *Machine Learning*, 60(1-3):73–96, 2005.
- Yikang Shen, Wenge Rong, Nan Jiang, Baolin Peng, Jie Tang, and Zhang Xiong. Word Embedding based Correlation Model for Question/Answer Matching. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 3511–3517. AAAI, 2015a.
- Yikang Shen, Wenge Rong, Zhiwei Sun, Yuanxin Ouyang, and Zhang Xiong. Question/answer matching for cqa system via combining lexical and sequential information. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 275–281. AAAI, 2015b.
- Lokesh Shrestha and Kathleen McKeown. Detection of Question-Answer Pairs in Email Conversations. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*, pages 889–895. ACL, 2004.
- Anna Shtok, Gideon Dror, Yoelle Maarek, and Idan Szpektor. Learning from the Past: Answering New Questions with Past Answers. In *Proceedings of the 21st International World Wide Web Conference*, pages 759–768. ACM, 2012.
- Joao Silva, Luísa Coheur, Ana Cristina Mendes, and Andreas Wichert. From Symbolic to Sub-Symbolic Information in Question Classification. *Artificial Intelligence Review*, 35(2):137–154, 2011.
- Amit Singh. Entity Based Q&A Retrieval. In *Proceedings of the Joint Meeting of the Conference on Empirical Methods in Natural Language Processing (EMNLP) and the Conference on Computational Natural Language Learning (CoNLL)*, pages 1266–1277. ACL, 2012.
- Amit Singh, P. Deepak, and Dinesh Raghu. Retrieving Similar Discussion Forum Threads: A Structure Based Approach. In *Proceedings of the 35th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 135–144. ACM, 2012.



- Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. Reasoning with Neural Tensor Networks for Knowledge Base Completion. *NIPS*, 26:926–934, 2013.
- Parikshit Sondhi and ChengXiang Zhai. Mining Semi-Structured Online Knowledge Bases to Answer Natural Language Questions on Community QA Websites. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 341–350. ACM, 2014.
- Young-In Song, Chin-Yew Lin, Yunbo Cao, and Hae-Chang Rim. Question Utility: A Novel Static Ranking of Question Search. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, pages 1231–1236. AAAI, 2008.
- Cleyton Souza, Franck Aragão, José Remígio, Evandro Costa, and Joseana Fechine. Using CQA History to Improve Q&A Experience. In *Proceedings of the 2016 International Conference on Computational Science and Its Applications*, pages 570–580. Springer, 2016.
- Ivan Srba. Promoting Sustainability and Transferability of Community Question Answering. *Information Sciences and Technologies Bulletin of the ACM Slovakia*, pages 1–7, 2011.
- Ivan Srba and Maria Bielikova. A Comprehensive Survey and Classification of Approaches for Community Question Answering. *TWEB*, 10(3):18, 2016.
- Nicola Stokes and Joe Carthy. Combining Semantic and Syntactic Document Classifiers to Improve First Story Detection. In *Proceedings of the 24th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 424–425. ACM, 2001.
- Diane M. Strong, Yang W. Lee, and Richard Y. Wang. Data Quality in Context. *Communications of the ACM*, 40(5):103–110, 1997.
- Qi Su, Dmitry Pavlov, Jyh-Herng Chow, and Wendell C. Baker. Internet-Scale Collection of Human-Reviewed Data. In *Proceedings of the 16th International World Wide Web Conference*, pages 231–240. ACM, 2007.
- Sai Praneeth Suggu, Kushwanth N. Goutham, Manoj K. Chinnakotla, and Manish Shrivastava. Deep Feature Fusion Network for Answer Quality Prediction in Community Question Answering. In *Proceedings of the NeurIR 2016 SIGIR Workshop on Neural Information Retrieval*. arXiv, 2016.
- Ke Sun, Yunbo Cao, Xinying Song, Young-In Song, Xiaolong Wang, and Chin-Yew Lin. Learning to Recommend Questions Based on User Ratings. In *Proceedings of the 18th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 751–758. ACM, 2009.

- Mihai Surdeanu, Massimiliano Ciaramita, and Hugo Zaragoza. Learning to Rank Answers on Large Online QA Collections. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL): Human Language Technologies*, pages 719–727. ACL, 2008.
- Mihai Surdeanu, Massimiliano Ciaramita, and Hugo Zaragoza. Learning to Rank Answers to Non-Factoid Questions from Web Collections. *Computational Linguistics*, 37(2):351–383, 2011.
- Maggy Anastasia Suryanto, Ee Peng Lim, Aixin Sun, and Roger H.L. Chiang. Quality-Aware Collaborative Question Answering: Methods and Evaluation. In *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining (WSDM)*, pages 142–151. ACM, 2009.
- Jun Suzuki, Hirotoshi Taira, Yutaka Sasaki, and Eisaku Maeda. Question Classification using HDAG Kernel. In *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering*, volume 12, pages 61–68. ACL, 2003.
- Saori Suzuki, Shin’ichi Nakayama, and Hideo Joho. Formulating Effective Questions for Community-based Question Answering. In *Proceedings of the 34th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 1261–1262. ACM, 2011.
- Ming Tan, Bing Xiang, and Bowen Zhou. LSTM-based Deep Learning Models for Non-Factoid Answer Selection. In *Proceedings of the 2016 International Conference on Learning Representations (ICLR) Workshop Track*. arXiv, 2016.
- Jaime Teevan, Susan T. Dumais, and Daniel J. Liebling. To Personalize or not to Personalize: Modeling Queries with Variation in User Intent. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 163–170. ACM, 2008.
- Qiongjie Tian and Baoxin Li. Weakly Hierarchical Lasso based Learning to Rank in Best Answer Prediction. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 307–314. IEEE, 2016.
- Qiongjie Tian, Peng Zhang, and Baoxin Li. Towards Predicting the Best Answers in Community-based Question-Answering Services. In *Proceedings of the 7th AAAI International Conference on Weblogs and Social Media (ICWSM)*, pages 725–728. AAAI, 2013a.

- Yuan Tian, Pavneet Singh Kochhar, Ee-Peng Lim, Feida Zhu, and David Lo. Predicting Best Answerers for New Questions: An Approach Leveraging Topic Modeling and Collaborative Voting. In *Proceedings of the 5th International Conference on Social Informatics (SocInfo), International Workshops*, pages 55–68. Springer, 2013b.
- Almer S. Tigelaar, Rieks op den Akker, and Djoerd Hiemstra. Automatic Summarisation of Discussion Fora. *Natural Language Engineering*, 16(02): 161–192, 2010.
- Mattia Tomasoni. Metadata-aware Measures for Answer Summarization in Community Question Answering. Master’s thesis, University of Uppsala, Sweden, 2003.
- Mattia Tomasoni and Minlie Huang. Metadata-aware Measures for Answer Summarization in Community Question Answering. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 760–769. ACL, 2010.
- Noriko Tomuro. Question Terminology and Representation for Question Type Classification. In *Proceedings of the Second International Workshop on Computational Terminology (COMPUTERM 2002)*, pages 1–7. ACL, 2002.
- Noriko Tomuro and Steven L. Lytinen. Selecting Features for Paraphrasing Question Sentences. In *Proceedings of the Workshop on Automatic Paraphrasing at Natural Language Processing Pacific Rim Symposium (NL-PRS)*, pages 55–62. National Electronics and Computer Technology Center (NECTC), 2001.
- Quan Hung Tran, Vu Tran, Tu Vu, Minh Nguyen, and Son Bao Pham. JAIST: Combining Multiple Features for Answer Selection in Community Question Answering. In *Proceedings of the 9th Conference on Semantic Evaluation (SemEval)*, pages 215–219. ACL, 2015.
- Christoph Treude, Ohad Barzilay, and Margaret-Anne Storey. How do Programmers Ask and Answer Questions on the Web? (Nier Track). In *Proceedings of the 33rd International Conference on Software Engineering (ICSE)*, pages 804–807. IEEE, 2011.
- Xudong Tu, Xin-Jing Wang, Dan Feng, and Lei Zhang. Ranking Community Answers via Analogical Reasoning. In *Proceedings of the 18th International World Wide Web Conference*, pages 1227–1228. ACM, 2009.
- Kateryna Tymoshenko, Daniele Bonadiman, and Alessandro Moschitti. Learning to Rank Non-Factoid Answers: Comment Selection in Web Forums. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2049–2052. ACM, 2016.

- Jan Ulrich. *Supervised Machine Learning for Email Thread Summarization*. PhD thesis, University of British Columbia, 2008.
- David Vallet and Pablo Castells. Personalized Diversification of Search Results. In *Proceedings of the 35th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 841–850. ACM, 2012.
- Jelica Vasiljevic, Tom Lampert, and Milos Ivanovic. The Application of the Topic Modeling to Question Answer Retrieval. In *Proceedings of the 6th International Conference of Information Society and Technology (ICIST)*, volume 1, pages 241–246. ICIST, 2016.
- Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and Composing Robust Features with Denoising Autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1096–1103. ACM, 2008.
- Ellen M. Voorhees. The TREC-8 Question Answering Track Report. In *Proceedings of the 8th Text REtrieval Conference (TREC)*, pages 77–82. NIST, 1999.
- Stephen Wan and Kathy McKeown. Generating Overview Summaries of Ongoing Email Thread Discussions. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*, pages 549–555. ACL, 2004.
- Nayer Wanas, Motaz El-Saban, Heba Ashour, and Waleed Ammar. Automatic Scoring of Online Discussion Posts. In *Proceedings of the 2nd ACM Workshop on Information Credibility on the Web (WICOW'08)*, pages 19–26. ACM, 2008.
- Baoxun Wang, Bingquan Liu, Chengjie Sun, Xiaolong Wang, and Bo Li. Adaptive Maximum Marginal Relevance Based Multi-Email Summarization. In *Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence*, pages 417–424. Springer, 2009a.
- Baoxun Wang, Bingquan Liu, Chengjie Sun, Xiaolong Wang, and Lin Sun. Extracting Chinese Question-Answer Pairs from Online Forums. In *Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics (SMC'09)*, pages 1159–1164. IEEE, 2009b.
- Baoxun Wang, Xiaolong Wang, Chengjie Sun, Bingquan Liu, and Lin Sun. Modeling Semantic Relevance for Question-Answer Pairs in Web Social Communities. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1230–1238. ACL, 2010a.

- Di Wang and Eric Nyberg. A Long Short-term Memory Model for Answer Sentence Selection in Question Answering. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers)*, pages 707–712. ACL, 2015a.
- Di Wang and Eric Nyberg. CMU OAQA at TREC 2015 LiveQA: Discovering the Right Answer with Clues. In *Proceedings of the 24th Text REtrieval Conference (TREC) (LiveQA Track)*, pages 1–6. NIST, 2015b.
- G. Alan Wang, Jian Jiao, and Weiguo Fan. Searching for Authoritative Documents in Knowledge-base Communities. *Proceedings of the 2009 International Conference on Information Systems (ICIS)*, page 109, 2009c.
- G. Alan Wang, Jian Jiao, Alan S. Abrahams, Weiguo Fan, and Zhongju Zhang. ExpertRank: A Topic-aware Expert Finding Algorithm for Online Knowledge Communities. *Decision Support Systems*, 54(3):1442–1451, 2013a.
- Hongning Wang, Chi Wang, ChengXiang Zhai, and Jiawei Han. Learning Online Discussion Structures by Conditional Random Fields. In *Proceedings of the 34th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 435–444. ACM, 2011a.
- Jian Wang, Jiqing Sun, Hongfei Lin, Hualei Dong, and Shaowu Zhang. Predicting Best Answerers for New Questions: An Approach Leveraging Convolution Neural Networks in Community Question Answering. In *Proceedings of the 2016 Chinese National Conference on Social Media Processing*, pages 29–41. Springer, 2016.
- Kai Wang, Zhaoyan Ming, and Tat-Seng Chua. A Syntactic Tree Matching Approach to Finding Similar Questions in Community-based QA Services. In *Proceedings of the 32nd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 187–194. ACM, 2009d.
- Kai Wang, Zhao-Yan Ming, Xia Hu, and Tat-Seng Chua. Segmentation of Multi-Sentence Questions: Towards Effective Question Retrieval in cQA Services. In *Proceedings of the 33rd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 387–394. ACM, 2010b.
- Li Wang, Su Nam Kim, and Timothy Baldwin. Thread-level Analysis over Technical User Forum Data. In *Proceedings of the 2010 Australasian Language Technology Association Workshop (ALTA)*, pages 27–31. ACL, 2010c.
- Li Wang, Su Nam Kim, and Timothy Baldwin. The Utility of Discourse Structure in Identifying Resolved Threads in Technical User Forums. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 2739–2756. ACL, 2012.

- Li Wang, Su Nam Kim, and Timothy Baldwin. The Utility of Discourse Structure in Forum Thread Retrieval. In *Proceedings of the 9th Asian Information Retrieval Societies Conference (AIRS 2013)*, pages 284–295. Springer, 2013b.
- Richard Y. Wang and Diane M. Strong. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4):5–33, 1996.
- Wei Wang, Baichuan Li, and Irwin King. Improving Question Retrieval in Community Question Answering with Label Ranking. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 349–356. IEEE, 2011b.
- Xin-Jing Wang, Xudong Tu, Dan Feng, and Lei Zhang. Ranking Community Answers by Modeling Question-Answer Relationships via Analogical Reasoning. In *Proceedings of the 32nd International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 179–186. ACM, 2009e.
- Yi-Chia Wang, Mahesh Joshi, and Carolyn P. Rosé. A Feature Based Approach to Leveraging Context for Classifying Newsgroup Style Discussion Segments. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL) Companion Volume Proceedings of the Demo and Poster Sessions*, pages 73–76. ACL, 2007.
- Yida Wang, Jiang-Ming Yang, Wei Lai, Rui Cai, Lei Zhang, and Wei-Ying Ma. Exploring Traversal Strategy for Web Forum Crawling. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 459–466. ACM, 2008.
- Yu Wang and Eugene Agichtein. Query Ambiguity Revisited: Clickthrough Measures for Distinguishing Informational and Ambiguous Queries. In *Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 361–364. ACL, 2010.
- Zhe Wang and Pengyi Zhang. Examining User Roles in Social Q&A: the Case of Health Topics in Zhihu.com. In *Proceedings of the 2016 Annual Meeting of the Association of Information Science and Technology (ASIS&T)*, pages 1–6. Wiley, 2016.
- Wei Wei, ZhaoYan Ming, Liqiang Nie, Guohui Li, Jianjun Li, Feida Zhu, Tianfeng Shang, and Changyin Luo. Exploring Heterogeneous Features for Query-focused Summarization of Categorized Community Answers. *Information Sciences*, 330:403–423, 2016.

- Markus Weimer and Iryna Gurevych. Predicting the Perceived Quality of Web Forum Posts. In *Proceedings of the 2007 Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 643–648. ACL, 2007.
- Markus Weimer, Iryna Gurevych, and Max Mühlhäuser. Automatically Assessing the Post Quality in Online Discussions on Software. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL): Interactive Poster and Demonstration Sessions*, pages 125–128. ACL, 2007.
- Howard T. Welsler, Eric Gleave, Danyel Fisher, and Marc Smith. Visualizing the Signatures of Social Roles in Online Discussion Groups. *Journal of Social Structure (JoSS)*, 8(2):1–32, 2007.
- Miaomiao Wen, Diyi Yang, and Carolyn Rosé. Sentiment Analysis in MOOC Discussion Forums: What does it tell us? In *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, pages 130–137. International Educational Data Mining Society, 2014.
- Florian Wolf and Edward Gibson. Representing Discourse Coherence: A Corpus-based Study. *Computational Linguistics*, 31(2):249–287, 2005.
- Jian-Syuan Wong, Bart Pursel, Anna Divinsky, and Bernard J. Jansen. An Analysis of MOOC Discussion Forum Interactions from the Most Active Users. In *Proceedings of the 2015 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP-BRiMS)*, pages 452–457. Springer, 2015.
- Guoshun Wu and Man Lan. Leverage Web-based Answer Retrieval and Hierarchical Answer Selection to Improve the Performance of Live Question Answering. In *Proceedings of the 24th Text REtrieval Conference (LiveQA Track)*. NIST, 2015.
- Hu Wu, Yongji Wang, and Xiang Cheng. Incremental Probabilistic Latent Semantic Analysis for Automatic Question Recommendation. In *Proceedings of the 2008 ACM Conference on Recommender Systems*, pages 99–106. ACM, 2008.
- Wensi Xi, Jesper Lind, and Eric Brill. Learning Effective Ranking Functions for Newsgroup Search. In *Proceedings of the 27th sigir*, pages 394–401. ACM, 2004.
- Yang Xianfeng and Liu Pengfei. Question Recommendation and Answer Extraction in Question Answering Community. *International Journal of Database Theory and Application (IJDTA)*, 9(1):35–44, 2016.

- Siqi Xiang, Wenge Rong, Yikang Shen, Yuanxin Ouyang, and Zhang Xiong. Multidimensional Scaling Based Knowledge Provision for New Questions in Community Question Answering Systems. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 115–122. IEEE, 2016.
- Sihong Xie, Qingbo Hu, Weixiang Shao, Jingyuan Zhang, Jing Gao, Wei Fan, and Philip S. Yu. Effective Crowd Expertise Modeling via Cross Domain Sparsity and Uncertainty Reduction. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pages 648–656. SIAM, 2016.
- Congfu Xu, Xin Wang, and Yunhui Guo. Collaborative Expert Recommendation for Community-Based Question Answering. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 378–393. Springer, 2016.
- Fei Xu, Zongcheng Ji, and Bin Wang. Dual Role Model for Question Recommendation in Community Question Answering. In *Proceedings of the 35th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 771–780. ACM, 2012.
- Gu Xu and Wei-Ying Ma. Building Implicit Links from Content for Forum Search. In *Proceedings of the 29th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 300–307. ACM, 2006.
- Xiaobing Xue, Jiwoon Jeon, and W. Bruce Croft. Retrieval Models for Question and Answer Archives. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 475–482. ACM, 2008.
- Baoguo Yang and Suresh Manandhar. Exploring User Expertise and Descriptive Ability in Community Question Answering. In *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 320–327. IEEE, 2014.
- Diyi Yang, Mario Piergallini, Iris Howley, and Carolyn P. Rosé. Forum Thread Recommendation for Massive Open Online Courses. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, pages 257–260. International Educational Data Mining Society, 2014a.
- Jiang-Ming Yang, Rui Cai, Yida Wang, Jun Zhu, Lei Zhang, and Wei-Ying Ma. Incorporating Site-Level Knowledge to Extract Structured Data from Web Forums. In *Proceedings of the 18th International World Wide Web Conference*, pages 181–190. ACM, 2009a.



- Jie Yang, Ke Tao, Alessandro Bozzon, and Geert-Jan Houben. Sparrows and Owls: Characterisation of Expert Behaviour in StackOverflow. In *Proceedings of the International Conference on User Modeling, Adaptation, and Personalization*, pages 266–277. Springer, 2014b.
- Lichun Yang, Shenghua Bao, Qingliang Lin, Xian Wu, Dingyi Han, Zhong Su, and Yong Yu. Analyzing and Predicting Not-Answered Questions in Community-based Question Answering Services. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, pages 1273–1278. AAAI, 2011.
- Liu Yang, Minghui Qiu, Swapna Gottipati, Feida Zhu, Jing Jiang, Huiping Sun, and Zhong Chen. CQARank: Jointly Model Topics and Expertise in Community Question Answering. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 99–108. ACM, 2013.
- Wen-Yun Yang, Yunbo Cao, and Chin-Yew Lin. A Structural Support Vector Method for Extracting Contexts and Answers of Questions from Online Forums. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 514–523. ACL, 2009b.
- Yiming Yang, Tom Pierce, and Jaime Carbonell. A Study of Retrospective and On-line Event Detection. In *Proceedings of the 21st International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 28–36. ACM, 1998.
- Yuan Yao, Hanghang Tong, Tao Xie, Leman Akoglu, Feng Xu, and Jian Lu. Want a Good Answer? Ask a Good Question First! *CoRR*, arXiv preprint arXiv:1311.6876, 2013.
- Yuan Yao, Hanghang Tong, Tao Xie, Leman Akoglu, Feng Xu, and Jian Lu. Detecting High-Quality Posts in Community Question Answering Sites. *Information Sciences*, 302:70–82, 2015.
- David M. Zajic, Bonnie J. Dorr, and Jimmy Lin. Single-Document and Multi-Document Summarization Techniques for Email Threads Using Sentence Compression. *Information Processing & Management*, 44(4):1600–1610, 2008.
- Zhongwu Zhai, Bing Liu, Lei Zhang, Hua Xu, and Peifa Jia. Identifying Evaluative Sentences in Online Discussions. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, pages 933–938. AAAI, 2011.
- Dell Zhang and Wee Sun Lee. Question Classification Using Support Vector Machines. In *Proceedings of the 26th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 26–32. ACM, 2003.

- Jingyuan Zhang, Xiangnan Kong, Roger Jie Luo, Yi Chang, and Philip S. Yu. NCR: A Scalable Network-based Approach to Co-ranking in Question-and-Answer Sites. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 709–718. ACM, 2014a.
- Jun Zhang, Mark S. Ackerman, and Lada Adamic. Expertise Networks in Online Communities: Structure and Algorithms. In *Proceedings of the 16th International World Wide Web Conference*, pages 221–230. ACM, 2007a.
- Kai Zhang, Wei Wu, Haocheng Wu, Zhoujun Li, and Ming Zhou. Question Retrieval with High Quality Answers in Community Question Answering. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 371–380. ACM, 2014b.
- Kai Zhang, Wei Wu, Fang Wang, Ming Zhou, and Zhoujun Li. Learning Distributed Representations of Data in Community Question Answering for Question Retrieval. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 533–542. ACM, 2016.
- Kuo Zhang, Juan Zi, and Li Gang Wu. New Event Detection Based on Indexing-Tree and Named Entity. In *Proceedings of the 30th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 215–222. ACM, 2007b.
- Weinan Zhang, Zhaoyan Ming, Yu Zhang, Liqiang Nie, Ting Liu, and Tat-Seng Chua. The Use of Dependency Relation Graph to Enhance the Term Weighting in Question Retrieval. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 3105–3120. ACL, 2012.
- Shiqi Zhao, Haifeng Wang, Chao Li, Ting Liu, and Yi Guan. Automatically Generating Questions from Queries for Community-based Question Answering. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 929–937. ACL, 2011.
- Zhou Zhao, Lijun Zhang, Xiaofei He, and Wilfred Ng. Expert Finding for Question Answering via Graph Regularized Matrix Completion. *TKDE*, 27(4):993–1004, 2015.
- Guangyou Zhou, Li Cai, Jun Zhao, and Kang Liu. Phrase-Based Translation Model for Question Retrieval in Community Question Answer Archives. In *Proceedings of the 2011 Annual Meeting of the Association for Computational Linguistics (ACL): Human Language Technologies*, pages 653–662. ACL, 2011a.

- Guangyou Zhou, Siwei Lai, Kang Liu, and Jun Zhao. Topic-Sensitive Probabilistic Model for Expert Finding in Question Answer Communities. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1662–1666. ACM, 2012a.
- Guangyou Zhou, Kang Liu, and Jun Zhao. Exploiting Bilingual Translation for Question Retrieval in Community-Based Question Answering. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 3153–3170. ACL, 2012b.
- Guangyou Zhou, Kang Liu, and Jun Zhao. Topical Authority Identification in Community Question Answering. In *Proceedings of the Chinese Conference on Pattern Recognition*, pages 622–629. Springer, 2012c.
- Guangyou Zhou, Yubo Chen, Daojian Zeng, and Jun Zhao. Towards Faster and Better Retrieval Models for Question Search. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2139–2148. ACM, 2013a.
- Guangyou Zhou, Fang Liu, Yang Liu, Shizhu He, and Jun Zhao. Statistical Machine Translation Improves Question Retrieval in Community Question Answering via Matrix Factorization. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 852–861. ACL, 2013b.
- Guangyou Zhou, Yang Liu, Fang Liu, Daojian Zeng, and Jun Zhao. Improving Question Retrieval in Community Question Answering Using World Knowledge. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2239–2245. AAAI, 2013c.
- Guangyou Zhou, Yubo Chen, Daojian Zeng, and Jun Zhao. Group Non-negative Matrix Factorization with Natural Categories for Question Retrieval in Community Question Answer Archives. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 89–98. ACL, 2014.
- Guangyou Zhou, Tingting He, Jun Zhao, and Po Hu. Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering. In *Proceedings of the() Joint 53rd Annual Meeting of the Association for Computational Linguistics (ACL) and the 7th International Joint Conference on Natural Language Processing*, volume Volume 1: Long Papers, pages 250–259. ACL, 2015.
- Guangyou Zhou, Zhiwen Xie, Tingting He, Jun Zhao, and X. Hu. Learning the Multilingual Translation Representations for Question Retrieval in Community Question Answering via Non-negative Matrix Factorization. *TASLP*, 24(7):1305–1314, 2016a.

- Guangyou Zhou, Yin Zhou, Tingting He, and Wensheng Wu. Learning Semantic Representation with Neural Networks for Community Question Answering Retrieval. *KBS*, 93:75–83, 2016b.
- Liang Zhou and Eduard Hovy. Digesting Virtual “Geek” Culture: The Summarization of Technical Internet Relay Chats. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 298–305. ACL, 2005.
- Liang Zhou and Eduard Hovy. On the Summarization of Dynamically Introduced Information: Online Discussions and Blogs. In *Proceedings of the AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs (SS-06-03)*, pages 237–242. AAAI, 2006.
- Shu Zhou and Simon Fong. Exploring the Feature Selection-Based Data Analytics Solutions for Text Mining Online Communities by Investigating the Influential Factors: A Case Study of Programming CQA in Stack Overflow. *Big Data Applications and Use Cases*, pages 49–93, 2016.
- Tom Chao Zhou, Chin-Yew Lin, Irwin King, Michael R. Lyu, Young-In Song, and Yunbo Cao. Learning to Suggest Questions in Online Forums. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, pages 1298–1303. AAAI, 2011b.
- Tom Chao Zhou, Xiance Si, Edward Y. Chang, Irwin King, and Michael R. Lyu. A Data-Driven Approach to Question Subjectivity Identification in Community Question Answering. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pages 164–170. AAAI, 2012d.
- Yanhong Zhou, Gao Cong, Bin Cui, Christian S. Jensen, and Junjie Yao. Routing Questions to the Right Users in Online Communities. In *Proceedings of the 25th IEEE International Conference on Data Engineering (ICDE)*, pages 700–711. IEEE, 2009.
- Yun Zhou and W. Bruce Croft. Query Performance Prediction in Web Search Environments. In *Proceedings of the 30th International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 543–550. ACM, 2007.
- Zhi-Min Zhou, Man Lan, Zheng-Yu Niu, and Yue Lu. Exploiting User Profile Information for Answer Ranking in CQA. In *Proceedings of the 21st International World Wide Web Conference*, pages 767–774. ACM, 2012e.
- Hengshu Zhu, Huanhuan Cao, Hui Xiong, Enhong Chen, and Jilei Tian. Towards Expert Finding by Leveraging Relevant Categories in Authority Ranking. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2221–2224. ACM, 2011.

- Mingliang Zhu, Weiming Hu, and Ou Wu. Topic Detection and Tracking for Threaded Discussion Communities. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 77–83. IEEE, 2008.
- Zhemín Zhu, Delphine Bernhard, and Iryna Gurevych. *A Multi-Dimensional Model for Assessing the Quality of Answers in Social Q&A Sites*. PhD thesis, Technische Universität Darmstadt, 2009.
- Zainab Zolaktaf, Fatemeh Riahi, Mahdi Shafiei, and Evangelos Milios. Modeling Community Question-Answering Archives. In *Proceedings of the 2nd Workshop on Computational Social Science and the Wisdom of Crowds (held at NIPS 2011)*, pages 1–5. MIT Press, 2011.