Mass Collaboration on the Web: Textual Content Analysis by Means of Natural Language Processing

Ivan Habernal, Johannes Daxenberger, and Iryna Gurevych

Abstract This chapter describes perspectives for utilizing natural language processing (NLP) to analyze artifacts arising from mass collaboration on the web. In recent years, the amount of user-generated content on the web has grown drastically. This content is typically noisy, un- or at best semi-structured, so that traditional analysis tools cannot properly handle it. To discover linguistic structures in this data, manual analysis is not feasible due to the large quantities of data. In this chapter, we explain and analyze web-based resources of mass collaboration, namely wikis, web forums, debate platforms and blog comments. We introduce recent advances and ongoing efforts to analyze textual content in two of these resources with the help of NLP. This includes an approach to discover flows of knowledge in online mass collaboration as well as methods to mine argumentative structures in natural language text. Finally, we outline application scenarios of the previously discussed techniques and resources within the domain of education.

Ivan Habernal
Ubiquitous Knowledge Processing Lab, Department of Computer Science, Technische Universität Darmstadt, Germany, and German Institute for Educational Research, Frankfurt am Main, Germany

Johannes Daxenberger
Ubiquitous Knowledge Processing Lab, Department of Computer Science, Technische Universität Darmstadt, Germany

Iryna Gurevych
Ubiquitous Knowledge Processing Lab, Department of Computer Science, Technische Universität Darmstadt, Germany, and German Institute for Educational Research, Frankfurt am Main, Germany
www.ukp.informatik.tu-darmstadt.de

1 Introduction

Mass collaboration on the web, as a practice with far-reaching implications for a knowledge society, promotes communication, collaborative authoring, and information sharing (Cress et al, 2013). This phenomenon has been investigated from many perspectives, such as knowledge management and collaborative learning (Cress and Kimmerle, 2008; Su and Beaumont, 2010), information quality (Kane, 2011; Ferschke, 2014), knowledge construction (Moskaliuk et al, 2012; Stahl et al, 2014), or design processes (Kim et al, 2011). The main four principles of the mass collaboration paradigm are *openness, peering, sharing, and acting globally* (Tapscott and Williams, 2008, pp. 20). In contrast to existing works, this chapter presents a novel viewpoint that targets mass collaboration from the Natural Language Processing (NLP) perspective and explores corresponding methods that are able to cope with the current information overload. As an example domain, we focus on education as its breadth attracts not only researchers, but also practitioners or policy-makers. We discuss specific NLP methods and their suitability and reliability within that domain.

Collaborative learning and mass collaboration are popular educational strategies that encourage learners to engage not only in social activities and knowledge sharing, but also to actively construct new knowledge (Eryilmaz et al, 2013). Onrubia and Engel (2009) examined phases of collaborative knowledge construction and found that each phase represents a higher level of cognitive complexity than the previous one, more in-depth study, more convergence and more shared understanding of the meanings constructed by the members of a group. In a review of research examining social, cognitive, and teaching presence in online learning environments, Garrison and Arbaugh (2007) conclude that collaborative learning can help learners to retain learned information longer and foster their higher-order thinking skills.

In the field of computer-supported collaborative learning, asynchronous online discussions (Eryilmaz et al, 2013) and wikis (Wheeler et al, 2008; Larusson and Alterman, 2009) are widely used tools. They can facilitate a natural setting for collaborative knowledge construction, e.g., by offering students the opportunity to reflect on peers' contributions and analyze their own ideas before articulating them (Pena-Shaff and Nicholls, 2004). According to Lund and Rasmussen (2010), for teachers it is becoming increasingly important to develop competence in designing technology-mediated and collaborative tasks. Their findings reveal the need to examine the complex relationships between methods, tasks, activities, and assessment in order to develop teaching with the help of Web 2.0 applications.

Among the knowledge-oriented platforms, there are numerous scenarios that approach collaboration from other directions. For instance, computer-supported argumentation facilitates communication and argumentation between multiple, and perhaps distant, participants (Scheuer et al, 2010). Debate platforms are tailored for the purpose of education, but serve a wide audience beyond traditional classrooms and across regional borders.

One of the main challenges that education-related mass collaboration has to face is the huge amount of textual content generated by users. As a consequence, learners

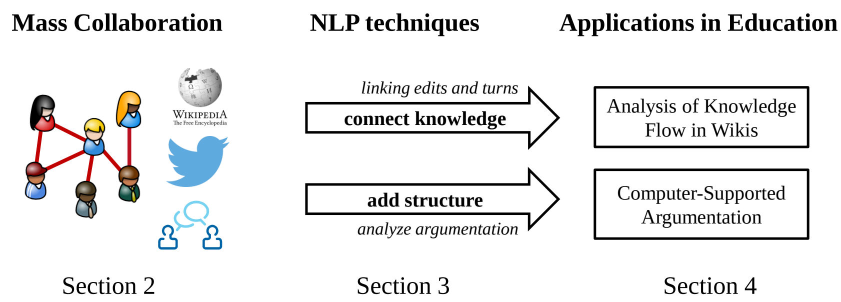


Fig. 1 Outline of this chapter.

may not be able to effectively process the massive load of textual material in which to look for relevant information, and the work overload of instructors increases. Information scattered across multiple locations, difficulty to keep an overview, or abundance of non-relevant or low-quality content are among the main obstacles to easily access and make use of the required information. Furthermore, current platforms for mass collaboration in education do not offer intelligent tools that would support users in their information needs and help to overcome the information overload. To tackle this issue, NLP is the key technology that enables extracting, analyzing, and utilizing valuable information from textual data. This article presents NLP perspectives for the field of mass collaboration in education and educational research.

The main trends in the current NLP research can be characterized with the following key phrases: (1) *data-driven*, meaning that the methods learn from human-annotated data using various statistical models from the machine learning area (Smith, 2011), (2) *semi-/un-supervised*, so the costly and error-prone human annotations are minimized by employing methods that utilize large amounts of unlabeled data (Søgaard, 2013), and (3) *resource-driven*, which means that various existing resources are combined together in order to boost the prior knowledge of the methods (Gurevych et al, 2012). Whereas the performance of some methods achieves almost human-like results, such as in part-of-speech tagging (Moore, 2014), syntax parsing (Krishnamurthy and Mitchell, 2014), or named entity recognition (Che et al, 2013), more complex tasks remain challenging. These are, for instance, discourse processing (Ji and Eisenstein, 2014), sentiment analysis (Habernal et al, 2014b), question answering (Yih et al, 2013), or argumentation mining (Stab and Gurevych, 2014b), among others. One limitation of many current NLP methods is *task dependency* in terms of, e.g., task-specific features, limited domains, or language-dependent resources. To tackle these issues, recent attempts have tried to rely solely on the data without any prior task-specific information (Collobert et al, 2011).

Figure 1 gives an overview of the contents of this chapter and explains our view on mass collaboration on the web. We have selected two use cases of NLP techniques which can be applied to a range of resources for online mass collaboration. Some of these resources are discussed in §2. As displayed in Figure 1 and explained in detail in §3, we use NLP techniques to (i) connect knowledge by linking edits and discussion turns, and (ii) to add structure to natural language texts by analyzing argumentation. Finally, in §4, we show how these techniques can be applied to mass collaboration in the educational domain.

2 Web-based resources of mass collaboration and their properties

This section discusses several types of web-based resources which we think are particularly useful for the study of mass collaboration. For each of the resources, we also refer to related work. Later, in §3, we demonstrate the usage of our recently developed NLP techniques which process data from two of these resources. Typical properties of the presented resources will be summarized in a tabular form at the end of this section.

2.1 Wikis

Wikis are a popular tool to present content which should be easily accessible and editable (Leuf and Cunningham, 2001). The open encyclopedia Wikipedia, which can be edited by anybody, is probably the best known example for a wiki. However, wikis are not necessarily open to everybody. Companies often use wikis to maintain internal documentation, as many wikis allow a fine-granular access right management. Independent of whether they are closed to the public or open to everybody, wikis are always designed to facilitate collaboration on content which is shared among many or few editors and readers. Hence, wikis typically offer technologies to support online collaboration. One helpful tool is the revision history which is maintained for each page of the wiki, so that everybody can follow its entire development (Ferschke et al, 2011). To enable open discussion about the content of a page, many wikis additionally offer a dedicated discussion forum, in the form of a normal wiki page, used exclusively to discuss issues about the associated main content page (Viegas et al, 2007).

Data extracted from wikis is usually rather clean in a grammatical sense. Most editors are eager to add content with a certain level of quality, as other readers and editors are able to track back each change to its author. However, wikis offer limited possibilities to structure content below the page level. Large projects such as Wikipedia often have developed guidelines and best practices to ensure coherent structures across pages, e.g. infoboxes (Wu and Weld, 2010). Nevertheless, such practices are not enforced technically and consequently ignored by many users.

As mentioned above, an important piece of information which is usually made available by wikis is the revision history of a page. In addition, the *discussion pages* which are bound to main content pages and are available in some wikis, offer valuable information not just about the development of an article, but also about the – potentially controversial – discourse with respect to the page content (Ferschke et al, 2012). As a consequence, discussion pages might contain implicit knowledge about a topic which is not visible within the article itself.

The size of data extracted from wikis obviously varies a lot depending on the project. By the end of 2014, the English Wikipedia, as one of the largest wikis, contains 4,7 million content (article) pages, which approximately receive 3 million edits each month. Whereas the number of pages is growing rather slowly, the ever-growing revision history for the English Wikipedia is a very large resource for NLP researchers. All of Wikipedia's content is open and licensed under the permissive Creative Commons License.

To manage the size of the data, in particular the revision history of the larger Wikipedias, sophisticated data structures and algorithms are required. One disadvantage of the openness of many wikis and the fact that anyone can edit its contents is the lack of quality control and the danger of vandalism (Potthast et al, 2008; Priedhorsky et al, 2007). The many-eyes principle only works for pages with a minimum number of readers and editors, whereas many unpopular pages remain untouched for years. Many edits with malicious intentions can be detected by automatic programs and are deleted quickly. However, some vandalism may go unnoticed for a long time, so that users cannot be fully sure about the quality of what they are reading (Priedhorsky et al, 2007).

Wikipedia's revision history data has been used for several NLP applications, including spelling error correction, paraphrasing or information retrieval (Ferschke et al, 2013). Since the revision history stores all edits including metadata such as the names of the authors, comments and timestamps, it is a very promising resource to analyze collaborative writing processes (Daxenberger and Gurevych, 2012). Addionally, Wikipedia covers more than 250 languages, and is thus a valuable resource for research in languages which otherwise offer little user-generated content on the web. Ferschke et al (2013) present an extensive survey summarizing studies and applications about the dynamic contents in Wikipedia.

In §3.1, we will show how the Wikipedia revision history and the discussion pages can be linked with each other, enabling a detailed analysis of the knowledge flow from the discussion to the article contents. Based on the findings from §3.1, we will discuss applications of wiki mass collaboration in education and educational research in §4.1.

2.2 Discussion forums

Online forums belong to the family of social media sites whose main purpose is to mediate discussions within a certain community. In the educational domain, forums can also facilitate a natural setting for collaborative knowledge construction (Eryilmaz et al, 2013). In contrast to wikis, where the emphasis is put on creating knowledge in the form of consistent and coherent articles, information in forums is usually implicitly spread across the discourse within a thread.

Taking into account the educational domain, discussion forums have played a dominant role when exploring collaborative learning and critical thinking in the past decade (Guzdial and Turns, 2000; Gilbert and Dabbagh, 2005; Niu and Van Aalst, 2009; Perkins and Murphy, 2006; Du et al, 2008; Hrastinski, 2008).

Whereas in threaded discussions the flow of dialog can be explicitly followed, many forums rely on a linear order of entries. This results into implicitly encoded flow of simultaneous discussions (i.e., using quotations). Restoring the context must rely on, e.g., thread disentanglement techniques (Elsner and Charniak, 2010; Jamison and Gurevych, 2013).

The content, as in many social media platforms, is usually very noisy. This may involve unusual spelling, irregular capitalization, and idiosyncratic abbreviations (Bontcheva and Rout, 2014) as well as non-dictionary slang, wordplay, or censor avoidance (Clark and Araki, 2011). Moreover, the temporal nature of the data, the social context, and implicit information about the participants represent an under-researched problem (Bontcheva and Rout, 2014).

The above-mentioned properties of non-standard language in social texts pose challenges to many NLP methods. For example, automatic tokenization is difficult because the majority of tokenizers are trained on newswire texts and perform poorly on social media, where punctuation plays a different role, e.g. in emoticons, hashtags, etc. (O'Connor et al, 2010). Consider the following example by Yang and Eisenstein (2013): *"gimme suttin 2 beleive innnn."* These custom abbreviations, phonetic substitution, or slang affect the vocabulary size and introduce many infrequent words. Saif et al (2012) show that 93% of words in their Twitter corpus (1.6 million Tweets) occur less than ten times. This causes data sparsity problems in many machine learning approaches.

Therefore text normalization is often considered as one of the first tasks when dealing with social media texts. The previous example would be normalized to *"Give me something to believe in."* (Yang and Eisenstein, 2013). Recent work on text normalization handles the problem by mapping the noisy words to their normalized counterparts in a dynamically generated lexicon in an unsupervised manner. Han et al (2012) create the lexicon using distributional and string similarity. Hassan and Menezes (2013) acquire the lexicon from unlabeled data using random walks on a contextual similarity graph which is constructed from n-gram sequences obtained from large unlabeled corpora. Yang and Eisenstein (2013) propose a log-linear unsupervised model to capture the relationship between standard (normalized) and non-standard tokens, reaching state-of-the-art F_1 score of about 0.73–0.82.

2.3 Debate platforms

A specific type of discussion forums are debate platforms that explicitly deal with one (mostly controversial) topic and allow users to add their opinions either on the *for* side or the *against* side. The debate topic is usually expressed in the form of a statement and is accompanied by a final voting of all pros and cons. From the collaboration perspective, users contribute their arguments on the issue and the final product is usually a two-sided summary of the main arguments. The scope of discussed topics range from very general ones (including religion[1], vegetarianism[2], etc.) to very narrowed ones, e.g., for particular policy-making (such as 'weapons inspectors leaving Syria'[3]).

The degree of moderation and involvement of editors varies. Whereas many portals do not put any restrictions on the content and structure, some of them rely on heavy post-editing and even involve academics from the field to curate the final debate (such as idebate.org). Each discussion is then also provided with objective background information about the issue and justifies the propositions by linking to their respective sources (comparable to the Wikipedia citation conventions).

Similar to a practice used in general online discussions, some debate portals provide mechanisms for voting (positive and negative votes for particular posts; see e.g. www.createdebate.org). The votes are then used to rank best arguments as well as to evaluate the dispute and display the winning position.

Although these portals provide mostly structured and high-quality content in terms of topic relatedness or argumentativeness, they have not yet been heavily exploited in NLP approaches. Gottipati et al (2013) try to predict positions of posts and external articles towards the proposed topic on Debatepedia.[4] Using the same source, Cabrio and Villata (2012) analyze relations between pairs of posts from the argumentation perspective. They automatically examine whether a particular post either supports or attacks another post, which later results into instantiation of a full argumentation graph over the topic debate.

Argumentation mining is becoming an emerging sub-field of NLP (Habernal et al, 2014a; Stab and Gurevych, 2014a). Since creating annotated resources for argumentation mining is costly and error prone, debate portals may serve as an additional data source and thus facilitate semi-supervised scenarios, such as active learning (Settles, 2009) or co-learning (Zhu and Goldberg, 2009).

[1] http://undergod.procon.org/

[2] http://vegetarian.procon.org/

[3] http://idebate.org/debatabase/debates/international/house-would-have-weapons-inspectors-leave-syria

[4] Now acessible under idebate.org

2.4 Blogosphere and microblogging

Blogging and especially *microblogging* platforms, a sub-field of social media, represent another growing field for the study of mass collaboration (Zhao et al, 2011; Carroll et al, 2011). Blogs might not be considered as resources to study collaboration as they are usually written by only one author. However, popular blogs often receive several hundred comments, which either refer to the text initially posted by the author of the blog, or to other comments. Interaction among bloggers facilitates networking with unique characteristics, where individuals experience a sense of community (Agarwal and Liu, 2009). This kind of collaboration has become even more popular with the rise of microblogs such as Twitter. Twitter users retweet posts from others and respond to tweets.

From the educational and mass-collaboration research point of view, blogging and microblogging have drawn much attention in recent years (Kim, 2008; Ebner et al, 2010; Robertson, 2011; Deng and Yuen, 2011; Chu et al, 2012; Cakir, 2013).

All phenomena of language common to discussion forums (cf. §2.2) can be also applied to blogosphere. An additional property is the brevity of texts published in microblogging platforms, as the content is usually limited to a few hundreds of characters (140 for Twitter). Also, the trend of *following* particular authors is present both in blogosphere (using, e.g., RSS subscriptions) and microblogging platforms (e.g., following mechanism on Twitter) (Kwak et al, 2010).

Given the amount of constantly growing content, one of the main challenges is viable processing of this kind of data, for such purposes as information retrieval, blog post search, information extraction, or network analysis (Santos et al, 2012; Agarwal and Liu, 2009). Parallel computing architectures which can handle such massive data volumes (Tjong Kim Sang and van den Bosch, 2013) are required to do so.

Table 1 summarizes typical properties of all resources introduced in this section.

3 Recent advances in NLP for mass collaboration

Given the growing number of resources to study online mass collaboration presented in the last section, we will now turn to discuss some of our recent advances and ongoing efforts in NLP operating on these resources. To a limited extent, the approaches presented in the following are dependent on the type of resource they are applied to (e.g. wikis, or debate platforms). This can be seen as a side effect of the type and aspect of mass collaboration inherent to each of the discussed resources, see Table 1. In wikis, the collaboration is made explicit by the revision history of articles, whereas in other resources such as blogs or debate platforms, collaboration is rather implicit in the sense that the attribution of contributions to individual users or the order of contributions might be harder to reproduce. Different kinds of collaborative aspects require different web environments and technologies, and thus different methodologies to be analyzed. For example, the revision history of a wiki

	Wikis	Discussion Forums	Debate Platforms	Blogs	Microblogging
Collaborative type	Explicit	Implicit	Implicit	Implicit	Implicit
Collaborative aspect	Knowledge construction	Information exchange	Argumentation	Information exchange	Information exchange
Text Quality	Edited	Noisy	Noisy, Edited	Noisy	Noisy
Challenges for NLP	Vandalism	Noise, Implicit discussion flow	Noise, Argument relevance	Meta-Texts/ Comments	Noise, Little context
Morphological and syntax processing	Easy	Medium	Medium	Medium	Hard
Licence	Creative Commons	Various	Creative Commons, Copyrighted	Copyrighted	Copyrighted
NLP applications	Information extraction, Text classification	Discourse analysis, Information extraction	Argumentation mining, Stance classification	Information extraction, Text classification	Opinion mining

Table 1 Language resources discussed in this chapter and their values of selected properties.

page can be used to analyze the collaborative construction of knowledge, whereas online debate platforms can facilitate better understanding of collaborative argumentation. An analysis of knowledge construction requires different methodologies as compared to the analysis of argumentation.

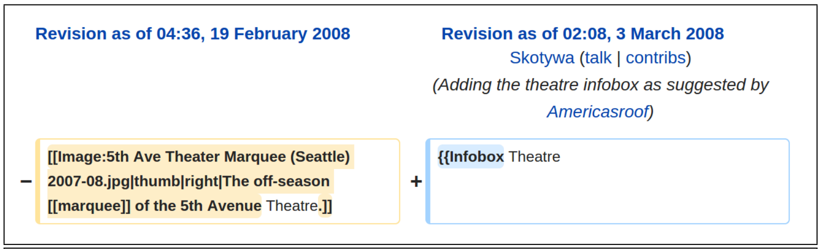
NLP methods typically deal with this problem in two steps. First, the text at hand needs to be prepared in order to serve as input for more sophisticated processing tools. This step is often referred to as *linguistic preprocessing*. Linguistic preprocessing typically involves basic NLP tasks such as the cleaning of noisy text or normalization (cf. § 2.2), the segmentation of text into smaller units such as sentences or words, and syntactical parsing. In Table 1, the properties referred to as text quality, challenges for NLP, and morphological and syntax processing are important parameters for linguistic preprocessing. Second, once the raw text has been prepared, it can be further processed with NLP methods targeted towards specific research questions. This is where the collaborative aspects and potentially higher-level NLP applications need to be considered.

We have discussed the properties of mass collaboration resources which need to be considered for linguistic preprocessing in §2 (see Table 1 for an overview). In the following, we turn to two novel approaches which reflect the usage of NLP methodology to answer questions about collaborative aspects in mass collaboration. Two particular examples are: (i) how to connect the knowledge in Wikipedia articles and discussion pages, and (ii) how to deal with argumentation and controversies in online discussion forums. For both use cases, we apply supervised machine learning classifiers, i.e. we use human-labeled data for training a model which can be used to automatically label further data. To do so, we have to define a set of features, tailored for the task at hand. These features abstract over the actual document contents, and thus enable the model to generalize and identify the relevant pieces of information, depending on the classification task. For example, a textual document can be represented by the frequency of each word it contains (the so called bag-of-words model).

3.1 Connecting knowledge in wikis

Many platforms used for collaborative writing, including wikis, do not explicitly allow their users to interact directly, so that the implicit effort of coordination behind the actual writing is not documented. As explained in §2.1, Wikipedia offers its users a platform to coordinate their writing, called discussion pages. Numerous studies have analyzed the nature of collaborative knowledge construction using wikis and their various properties. These include, e.g., the role of discussion pages (Hadjerrouit, 2013; Meishar-Tal and Gorsky, 2010), the impact of redundancy and polarity (Moskaliuk et al, 2012), or scripting for supporting collaborative writing (Wichmann and Rummel, 2013). In our recent study (Daxenberger and Gurevych, 2014), we analyzed links between edits in Wikipedia articles and turns (discourse segments) from their discussion pages. Our motivation is to better understand implicit details about the writing process and the knowledge flow in collaboratively created resources. Direct links between the article text (e.g. a controversial paragraph) and the discussion going on in the background of the article text can help the readers to understand the development of an article up to the point of time when they are accessing it. If new Wikipedia editors were able to directly access a past discussion thread about the piece of text they are about to edit, the organizational overload for senior editors pointing new users to the relevant parts of the discussion might be lowered.

We use the concepts of *edits* and *turns* developed in previous work (Daxenberger and Gurevych, 2012; Ferschke et al, 2012). Edits are local modifications extracted from consecutive Wikipedia revisions, e.g. spelling corrections, content additions or referencing. Each revision of a Wikipedia article consists of one or more edits and may be accompanied by a comment, in which the author of this revision explains the edit(s). Turns are segments from Wikipedia discussion pages. A turn is part of a topic, similar to a thread in a discussion forum, and can be attributed to a unique author. An *edit-turn-pair* is defined as a pair of an edit from the article's revision history and a turn from the discussion page bound to the same article. We consider an edit-turn-pair to be corresponding if the turn contains an explicit performative and the edit corresponds to this performative. For example, a Wikipedia user might suggest adding information to an article and announce the lack of information on the discussion page as displayed in Figure 2 (lower box, first turn). Another user adds the missing information to the part of the article in question (upper box) and leaves a report about the action on the discussion page (lower box, second turn). Both the turn which announces the missing link and the edit which adds the link, as well as the turn which reports the addition of the link and the edit to the article page are corresponding edit-turn-pairs.



Thanks. I'm fine with the wording. I basically wanted to Skinners name attached to the article. I ventured in because I did some biographies on the original owners of the Space Needle. Now that Ive taken an interest in the article one other item that might be useful in this very good article is the theatre infobox. -Americasroof

> Infobox added. Anyone else think we should try for featured article? I want to nominate it at a time when there can be more than just me paying attention to respond to feedback. -Skotywa

Fig. 2 An edit (top) as displayed on an English Wikipedia diff page along with two corresponding turns (bottom).

We collected and annotated a corpus of 636 edit-turn-pairs using the crowd-sourcing platform Amazon Mechanical Turk. For mature Wikipedia articles, combining each edit with each turn results in a very large search space for corresponding edit-turn-pairs. We therefore limited the time span between edits and turns considered for correspondence to 24 hours. Despite this limitation, the number of non-corresponding edit-turn-pairs still outscores the number of corresponding pairs by far, resulting in great class imbalance. To tackle this problem, we manually picked about 250 pairs of corresponding turns and revisions from a random sample of English Wikipedia articles. The resulting edit-turn-pairs were used as positive seeds in the Mechanical Turk annotation study, to keep the workers from labeling all given pairs as non-corresponding. We collected five crowd-sourced votes for each edit-turn-pair, and created the final labeling via majority voting. On a randomly selected subset of 100 pairs, the agreement with expert annotations is Cohen's $\kappa = 0.72$, showing that the corpus can be used to draw conclusions (Artstein and Poesio, 2008). The resulting corpus contains 128 corresponding and 508 non-corresponding edit-turn-pairs.

We used the DKPro TC framework (Daxenberger et al, 2014) to train a model on the annotated data. The model was trained on various features including textual features (such as similarity features between the turn and the edit) and meta-data features (such as the user name or the time difference). With the help of this model, a machine learning classifier can automatically recognize corresponding and non-corresponding edit-turn-pairs in Wikipedia articles. Despite the small size of our corpus, a Random Forest classifier (Breiman, 2001), achieved 0.79 macro F_1 score. One particular application of our system is that a possibly controversial discussion about e.g. the neutrality of an article can be associated with the edits that were triggered by this particular discussion.

3.2 Argumentation mining in online media

Argumentation mining deals with automatically identifying argumentative structures within natural language texts. Despite its strong background and long history in philosophy and logic (Toulmin, 1958; Walton, 2012), practical NLP approaches to argumentation have gained attention just recently (Feng and Hirst, 2011; Mochales and Moens, 2011). In its very simplistic and abstract form, an argument consists of a claim that the author wants to persuade the readers about, accompanied by one or more reasons that are put forward to support the claim.

In an ongoing study, we investigate how argumentation is conveyed across online media, such as forums, blogs, or comments to newswire articles. With a focus on controversies in education (such as single-sex schools, mainstreaming, or home-schooling), we collected a dataset containing 5,444 documents from various web sources. A subset of this collection (990 documents) was manually labeled by three independent annotators with respect to its persuasiveness and argumentativeness, discarding non-relevant documents on the document level. We achieved moderate agreement (Fleiss' $\kappa = 0.59$) based on three expert annotators.

In the next step, the structure of the argumentation is being further investigated in a more fine-grained manner. Relying on an adapted argumentation model by Toulmin (1958), each document is annotated on the statement level with its corresponding functional argumentation concepts, such as the claim, the grounds, the backing, etc. Consider the following actual example from a discussion forum that argues about public versus private schools.

> [The public schooling system is not as bad as some may think.]$_{\text{CLAIM}}$ [Some mentioned that those who are educated in the public schools are less educated,]$_{\text{REBUTTAL}}$ [well I actually think it would be in the reverse.]$_{\text{REFUTATION}}$ [Student who study in the private sector actually pay a fair amount of fees to do so and I believe that the students actually get let off for a lot more than anyone would in a public school. And its all because of the money.
> In a private school, a student being expelled or suspended is not just one student out the door, its the rest of that students schooling life fees gone. Whereas in a public school, its just the student gone.]$_{\text{GROUNDS}}$ [I have always gone to public schools and when I finished I got into University. I do not feel disadvantaged at all.]$_{\text{BACKING}}$

Fig. 3 Example of annotated text with argument components.

The annotation spans correspond to the argument components in the scheme, which can be also demonstrated in a diagram as in Figure 4, where the content of individual components was manually rephrased and simplified.

Such a detailed discourse annotation brings many challenges. First, proper boundaries of argument components (where the component begins and ends) must be identified. The boundaries might occur on the sentence level or an arbitrary phrase level. Second, the granularity of the argument components must be considered, for instance, whether the GROUNDS component in the previous example (Fig. 3) should be kept as a single component or split into multiple components. Third, assigning

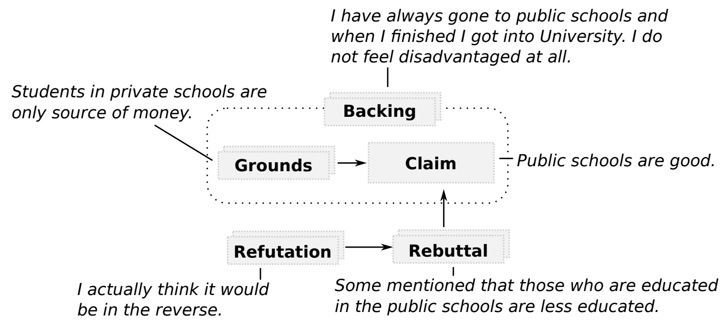


Fig. 4 Extended Toulmin's scheme used for annotation of arguments with an instantiated example from a single *public vs. private schools* discussion forum post.

the proper role of a particular component depends on the overall context, such as the case of REFUTATION in the example in Fig 3. Furthermore, not only is the user-generated discourse noisy (in terms of grammatical errors and other social media-related phenomena), but the main difficulty is that the argumentation is not well articulated. This means that the argumentation structures are often implicit and require complex inference. In contrast to argumentative essays (Stab and Gurevych, 2014a) or the legal domain (Mochales and Moens, 2011), user-generated texts often lack qualities typical to proper argumentation (Schiappa and Nordin, 2013). Particular challenges are, e.g., implicit claims, unclear stance towards the controversy, off-topic text unrelated to the argument, or appeal to emotions and other fallacies.

Using the extended Toulmin's model, three independent annotators labeled 340 documents (total 84,673 tokens, average 249.04 tokens per document, total 3,890 sentences). Annotations were performed in three steps with discussions, updating the annotation guidelines, and clarifying unclear cases. In the final step, the inter-annotation agreement reached 0.481 Krippendorff's unitized alpha α_U (Krippendorff, 2004) across all registers (blogs, newswire articles, forum posts, and article comments). However, when considering only article comments and forum posts, the agreement was significantly higher (0.603 α_U). This difference has multiple reasons. First, we observed that the obtained agreement negatively correlates with the length of the document (p-value ≤ 0.05) and blog posts and articles tend to be much longer than comments or forum posts. Second, some topics were inherently challenging to annotate. We observed that in the case of the *private vs. public schools* domain, the agreement negatively correlates with the text readability (tested on four different readability measures, p-value ≤ 0.05). Third, newswire articles and blogs employ various literary devices, such as quotations, narratives, or interviews, which cannot be easily modeled by the Toulmin's scheme, given its inherent limitations (see, e.g., (van Eemeren et al, 2014, pp. 233) for a theoretical discussion of applications of the model).

Using the annotated corpora, we developed and evaluated a supervised machine learning system. We treat the problem of finding argument component spans in the text as sequence labeling, where each token is labeled either as (1) COMPONENT-

B for beginning of the component (for example CLAIM-B), as (2) COMPONENT-I where the token is inside the component span (for example CLAIM-I), or (3) 'O' as other, meaning that the token is not part of the argument. An excerpt of such BIO coding is shown in Fig. 5, which correspond to the previous example in Fig. 3.

The$^{\text{CLAIM-B}}$ public$^{\text{CLAIM-I}}$ schooling$^{\text{CLAIM-I}}$ system$^{\text{CLAIM-I}}$ is$^{\text{CLAIM-I}}$ not$^{\text{CLAIM-I}}$ as$^{\text{CLAIM-I}}$ bad$^{\text{CLAIM-I}}$ as$^{\text{CLAIM-I}}$ some$^{\text{CLAIM-I}}$ may$^{\text{CLAIM-I}}$ think$^{\text{CLAIM-I}}$.$^{\text{CLAIM-I}}$ Some$^{\text{REBUTTAL-B}}$ mentioned$^{\text{REBUTTAL-I}}$ that$^{\text{REBUTTAL-I}}$ ...
... disadvantaged$^{\text{BACKING-I}}$ at$^{\text{BACKING-I}}$ all$^{\text{BACKING-I}}$.$^{\text{BACKING-I}}$

Fig. 5 BIO annotation of the example text shown in Fig. 3.

As a classifier, we employ SVM^{hmm} framework for sequence labeling (Joachims et al, 2009), DKPro TC for feature extraction and experiment setup (Daxenberger et al, 2014), and DKPro Core for linguistic annotations (Eckart de Castilho and Gurevych, 2014).

We experimented with many different types of features. The baseline feature set contains only binary features denoting presence of unigrams, bigrams, and trigrams in a sentence. The system trained using these features and default hyper-parameter settings yields 0.156 macro F_1 score in the 10-fold cross validation scenario. The richest feature set incorporates morphological features, syntactic features, coreference features, features obtained from semantic frames, features based on sentiment analysis, features exploiting unsupervised models (LDA and word embedding), and features produced by a discourse parser. With this configuration, the performance reaches macro F_1 score of 0.220 and significantly outperforms the baseline setting (p-value < 0.001, exact Liddell's test (Liddell, 1983)).

One of the causes explaining the low macro F_1 score is the skewed distribution of the classes in the labeled data. As REBUTTAL-B, REBUTTAL-I, REFUTATION-B, and REFUTATION-I represent only 3.7% of the data, the model cannot learn these classes and their F_1 score is mostly zero. This negatively affects the macro-averaged overall F_1 score. Furthermore, the evaluation on the token level is very strict as it penalizes also wrongly identified boundaries of the argument component. If the results are measured using Krippendorff's α_U, the system achieves 0.30 score, which is in the middle between the baseline (0.11) and the human performance (0.48). Further investigation of the error types is currently in progress.

4 Towards NLP for mass collaboration in the educational domain

The NLP methods presented in §3 can be utilized to foster intelligent and informed mass collaboration in education. They either directly cover the educational domain (argumentation mining in §3.2), or can be adapted to it (wiki-based collaboration in

§3.1). In this section, we will discuss the benefits emerging from incorporating such methods into mass collaboration in educational scenarios.

4.1 Analyzing collaboration and knowledge creation in wikis

The use of wikis in education, and in particular in teaching (Forte and Bruckman, 2006), has several advantages. Due to the nature of wikis, editing one's own or other people's text is simple and very well documented. This enables a detailed analysis of the collaborative writing process.

Several studies have analyzed the user network to get insights about collaboration in wikis (Brandes et al, 2009; Laniado and Tasso, 2011). The networks in these studies are made up of nodes representing authors and edges representing collaboration. Collaboration can be defined in various ways, e.g. as editing the same article or the same sentence within an article. Such networks are also known as coauthorship networks (Newman, 2004). Coauthorship networks typically only record the existence of interaction between users, but do not take into account the context of edits. Like this, information about whether an edit modifies the text base (i.e. a change which has an effect on the meaning of the text, e.g. addition of information) or the text surface (a change which does not change the meaning, e.g. a spelling correction) is lost (Faigley and Witte, 1981).

Daxenberger and Gurevych (2013) present a system to automatically classify edit operations such as grammatical error corrections or additions of information to Wikipedia pages. This tool can be used to add more specific information about the collaboration of users in wikis. In large scale scenarios, groups of users based on edit behavior (content adders, cleaners, all-round editors) can be identified (Liu and Ram, 2011). Using additional information about the quality of the final text product (which in educational settings is often available through grading), computers could assist writers to find more successful ways of collaboration, based on the order and preference of different kinds of revision. The intuition behind this is explained in the following example: after several iterations of content addition, it might be necessary to backup the existing information in the text with references and apply factual or grammatical corrections, rather than adding more information to the article text.

The information about the writing process which is implicitly documented in wikis becomes even more useful when additional resources such as underlying discussion pages are taken into account. By automatically linking issues raised on discussion pages to edit operations in wikis (Daxenberger and Gurevych, 2014), fine-grained information can enrich the data under investigation as well as bring completely new insights when these methods are applied to large data collections (cf. § 3.1). While analyzing the revision history helps to understand *which* changes were made, linking discussion segments to edits helps to understand *why* changes were made. This information can be very valuable to teachers in wiki-based education (both in classroom as well as in mass collaboration settings), as it helps to understand potentially hidden collaborative processes and communication not documented in the revision history of the final product. For the same reason, linking edits and discussion segments can also be a useful tool for educational research, as it might reveal deeper insights about the success and failure of collaborative tasks. It also helps to address possibly controversial issues (Kittur et al, 2007). Linking discussions and revision history makes it possible to understand which knowledge in the main text has been created through discussion in the background (Cress and Kimmerle, 2008).

4.2 Computer-supported argumentation

Apart from the NLP approaches to argumentation mining, research on computer-supported argumentation has been also very active, as shown by Scheuer et al (2010) in their recent survey of various models and argumentation formalisms from the educational perspective. Noroozi et al (2013) describe collaborative argumentation as engaging a group of learners in dialogical argumentation, critical thinking, elaboration, and reasoning so that they can build up a shared understanding of the issue at stake instead of merely convincing or changing their own and each other's beliefs.

Existing tools for collaborative argumentation rely on scripts to support student discussions by way of dialogue models that describe desirable discussion moves and sequences (Dillenbourg and Hong, 2008; Scheuer et al, 2014). Fischer et al (2013) outline a script theory of guidance in computer-supported collaborative learning. Many studies employ extensions or modification of the argument model proposed by Toulmin (1958). Noroozi et al (2013) investigate the formal-argumentative dimension of computer-supported collaborative learning by letting learners construct single arguments and exchange them in argumentation sequences to solve complex problems. Weinberger and Fischer (2006) analyze asynchronous discussion boards in which learners engage in an argumentative discourse with the goal to acquire knowledge. For coding the argument dimension, they created a set of argumentative moves based on Toulmin's model. Stegmann et al (2007) experiment with template-based methods that allowed to enter a claim, grounds and qualifications.

The above-mentioned tools and approaches to computer-supported argumentation can eminently benefit from NLP techniques for an automatic argument analysis, classification, and summarization. Instead of relying on, e.g., scrips (Dillenbourg and Hong, 2008; Scheuer et al, 2010; Fischer et al, 2013) or explicit argument diagramming (Scheuer et al, 2014), collaborative platforms can provide scholars with a summary of the whole argumentation to the topic, reveal the main argumentative patterns, provide the weaknesses of other's arguments, as well as identify shortcomings that need to be improved in the argumentative knowledge construction. Automatic analysis of micro-arguments can also help to overcome the existing trade-off between freedom (free-text option) and guidance (scripts) (Dillenbourg and Hong, 2008). Moreover, discovering fallacies in arguments (Schiappa and Nordin, 2013) might also have a positive impact on the learner's ability to construct reasonable argumentative discourse. Visualization of argumentation, e.g., using graphical connections that indicate arguments and the corresponding counterarguments, may further support learners to refine their argumentation (Kirschner et al, 2003).

5 Conclusions

As mass collaboration is shifting rapidly towards the big data paradigm, the massive amount of unstructured, textual data being produced represents one of the main challenges. In order to utilize this knowledge and information, new techniques are required that are capable of processing, extracting, and understanding the content in an automatic and intelligent manner. We believe that natural language processing is a key technology to support mass collaboration and the research based on the resulting content. The benefits are wide-ranging. In the educational domain, for example, learners can be directly supported as they are provided with access to automatically generated, structured information and feedback about the knowledge creation process of themselves and their fellow learners. Furthermore, with the help of deep analysis, new patterns and behavior in mass collaboration platforms can be explored and might foster further research in the field.

This chapter demonstrated how recent advances and ongoing efforts in NLP can boost research into mass collaboration. We have presented NLP methods capable of linking discussions in wikis with the actual undertaken actions (§ 3.1) and approaches to analyzing argumentation in user-generated web discussions (§ 3.2). Apart from the examples shown here, there exist other scenarios dealing with the information overload that benefit from utilizing NLP, e.g., question-answering in online communities (Gurevych et al, 2009), or MOOCs – massive open online courses (Shatnawi et al, 2014).

In § 2, we reviewed several types of mass-collaborative resources and their properties. Such a wide range of text registers, genres, and quality pose challenges to NLP in terms of domain adaptation. The majority of data-driven NLP approaches (and their underlying machine learning models) are trained and tested under bias by sampling from a particular restricted domain or register (Søgaard, 2013). Experiments show that, for instance, applying a part-of-speech tagging or named entity recognition model (traditionally trained on newswire corpora) to Twitter significantly degrades the performance (Gimpel et al, 2011; Finin et al, 2010). Although the approaches presented in § 3 have been tested in a cross-domain setting to some extent, drawing hard conclusions about their adaptation to a very different register (refer to Table 1) would require additional experiments. Apart from adapted linguistic preprocessing, adapting the presented approaches to different domain might involve creating new annotated resources in order to evaluate the models in the target domains. However, given the labour-intensity and difficulty of annotating e.g. argument components, the task of a broad domain-independent evaluation remains fairly challenging. Adapting existing models to other domains with different distribution of features and/or classes is current research in NLP.

There are several directions for the future work. Apart from the obvious one (for example domain adaptation as discussed above and other NLP-specific research questions), one particular area worth investigating is how the presented methods could be applied in other non-expert communities, for instance computer-supported collaborative learning. Bridging the gap between NLP and other communities requires innovative technical solutions with emphasis on usability, reproducibility, flexibility, and interactivity (Eckart de Castilho, 2014). One successful example of this endeavor is the DKPro framework (Eckart de Castilho and Gurevych, 2014) which integrates a multitude of linguistic tools, yet provides a user-friendly API and facilitates its adoption by non-expert users.

Acknowledgments

This work has been supported by the Volkswagen Foundation as part of the Lichtenberg-Professorship Program under grant № I/82806 and by the German Institute for Educational Research (DIPF).

References

Agarwal N, Liu H (2009) Modeling and Data Mining in Blogosphere. Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan & Claypool Publishers

Artstein R, Poesio M (2008) Inter-Coder Agreement for Computational Linguistics. Computational Linguistics 34(4):555–596

Bontcheva K, Rout D (2014) Making sense of social media streams through semantics: A survey. Semantic Web 5(5):373–403

Brandes U, Kenis P, Lerner J, van Raaij D (2009) Network analysis of collaboration structure in Wikipedia. In: Proceedings of the 18th international World Wide Web conference, pp 731–740

Breiman L (2001) Random forests. Machine Learning 45(1):5–32

Cabrio E, Villata S (2012) Combining textual entailment and argumentation theory for supporting online debates interactions. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2, pp 208–212

Cakir H (2013) Use of blogs in pre-service teacher education to improve student engagement. Computers & Education 68:244–252

Carroll JM, Jiang H, Rosson MB, Shih SI, Wang J, Xiao L, Zhao D (2011) Supporting activity awareness in computer-mediated collaboration. In: 2011 International Conference on Collaboration Technologies and Systems (CTS), pp 1–12

Eckart de Castilho R (2014) Natural language processing: Integration of automatic and manual analysis. Dissertation, Technische Universität Darmstadt

Eckart de Castilho R, Gurevych I (2014) A broad-coverage collection of portable NLP components for building shareable analysis pipelines. In: Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT, pp 1–11

Che W, Wang M, Manning CD, Liu T (2013) Named entity recognition with bilingual constraints. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp 52–62

Chu SKW, Chan CKK, Tiwari AFY (2012) Using blogs to support learning during internship. Computers & Education 58(3):989–1000

Clark E, Araki K (2011) Text normalization in social media: Progress, problems and applications for a pre-processing system of casual English. Procedia - Social and Behavioral Sciences 27:2–11

Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P (2011) Natural language processing (almost) from scratch. Journal of Machine Learning Research 12:2493–2537

Cress U, Kimmerle J (2008) A systemic and cognitive view on collaborative knowledge building with wikis. International Journal of Computer-Supported Collaborative Learning 3(2):105–122

Cress U, Barron B, Halatchliyski I, Oeberst A, Forte A, Resnick M, Collins A (2013) Mass collaboration - an emerging field for CSCL research. In: CSCL 2013 Conference Proceedings, Volume 1, pp 557–563

Daxenberger J, Gurevych I (2012) A Corpus-Based Study of Edit Categories in Featured and Non-Featured Wikipedia Articles. In: Proceedings of the 24^{th} International Conference on Computational Linguistics, pp 711–726

Daxenberger J, Gurevych I (2013) Automatically Classifying Edit Categories in Wikipedia Revisions. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp 578–589

Daxenberger J, Gurevych I (2014) Automatically detecting corresponding edit-turn-pairs in Wikipedia. In: Proceedings of the 52^{st} Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp 187–192

Daxenberger J, Ferschke O, Gurevych I, Zesch T (2014) DKPro TC: A Java-based Framework for Supervised Learning Experiments on Textual Data. In: Proceedings of the 52^{nd} Annual Meeting of the Association for Computational Linguistics. System Demonstrations, pp 61–66

Deng L, Yuen AH (2011) Towards a framework for educational affordances of blogs. Computers & Education 56(2):441–451

Dillenbourg P, Hong F (2008) The mechanics of CSCL macro scripts. International Journal of Computer-Supported Collaborative Learning 3(1):5–23

Du J, Zhang K, Olinzock A, Adams J (2008) Graduate students' perspectives on the meaningful nature of online discussions. Journal of Interactive Learning Research 19:21–36

Ebner M, Lienhardt C, Rohs M, Meyer I (2010) Microblogs in higher education - a chance to facilitate informal and process-oriented learning? Computers & Education 55(1):92–100 van Eemeren FH, Garssen B, Krabbe ECW, Snoeck Henkemans AF, Verheij B, Wagemans JHM (2014) Handbook of Argumentation Theory. Springer, Berlin/Heidelberg

Elsner M, Charniak E (2010) Disentangling Chat. Computational Linguistics 36(3):389–409

Eryilmaz E, Pol J, Ryan T, Clark P, Mary J (2013) Enhancing student knowledge acquisition from online learning conversations. International Journal of Computer-Supported Collaborative Learning 8(1):113–144

Faigley L, Witte S (1981) Analyzing revision. College composition and communication 32(4):400–414

Feng VW, Hirst G (2011) Classifying arguments by scheme. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, pp 987–996

Ferschke O (2014) The quality of content in open online collaboration platforms: Approaches to NLP-supported information quality management in Wikipedia. Dissertation, Technische Universität Darmstadt

Ferschke O, Zesch T, Gurevych I (2011) Wikipedia Revision Toolkit: Efficiently Accessing Wikipedia's Edit History. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. System Demonstrations, pp 97–102

Ferschke O, Gurevych I, Chebotar Y (2012) Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp 777–786

Ferschke O, Daxenberger J, Gurevych I (2013) A Survey of NLP Methods and Resources for Analyzing the Collaborative Writing Process in Wikipedia. In: Gurevych I, Kim J (eds) The Peoples Web Meets NLP: Collaboratively Constructed Language Resources, Theory and Applications of Natural Language Processing, Springer

Finin T, Murnane W, Karandikar A, Keller N, Martineau J, Dredze M (2010) Annotating named entities in Twitter data with crowdsourcing. In: Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, pp 80–88

Fischer F, Kollar I, Stegmann K, Wecker C (2013) Toward a script theory of guidance in computer-supported collaborative learning. Educational Psychologist 48(1):56–66

Forte A, Bruckman A (2006) From Wikipedia to the Classroom: Exploring Online Publication and Learning. In: Proceedings of the 7th International Conference on Learning Sciences, International Society of the Learning Sciences, pp 182–188

Garrison DR, Arbaugh B (2007) Researching the community of inquiry framework: Review, issues, and future directions. The Internet and Higher Education 10(3):157–172

Gilbert PK, Dabbagh N (2005) How to structure online discussions for meaningful discourse: A case study. British Journal of Educational Technology 36(1):5–18

Gimpel K, Schneider N, O'Connor B, Das D, Mills D, Eisenstein J, Heilman M, Yogatama D, Flanigan J, Smith NA (2011) Part-of-speech tagging for Twitter: annotation, features, and experiments. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp 42–47

Gottipati S, Qiu M, Sim Y, Jiang J, Smith NA (2013) Learning topics and positions from Debatepedia. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp 1858–1868

Gurevych I, Bernhard D, Ignatova K, Toprak C (2009) Educational question answering based on social media content. In: Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling, pp 133–140

Gurevych I, Eckle-Kohler J, Hartmann S, Matuschek M, Meyer CM, Wirth C (2012) UBY - A Large-Scale Unified Lexical-Semantic Resource Based on LMF. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp 580–590

Guzdial M, Turns J (2000) Effective discussion through a computer-mediated anchored forum. Journal of the Learning Sciences 9:437–469

Habernal I, Eckle-Kohler J, Gurevych I (2014a) Argumentation mining on the Web from information seeking perspective. In: Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing, pp 26–39

Habernal I, Ptáček T, Steinberger J (2014b) Supervised sentiment analysis in Czech social media. Information Processing & Management 50(5):693–707

Hadjerrouit S (2013) Collaborative writing with wikis: Pedagogical and technological implications for successful implementation in teacher education. In: Sampson DG, Isaias P, Ifenthaler D, Spector JM (eds) Ubiquitous and Mobile Learning in the Digital Age, Springer New York, pp 173–189

Han B, Cook P, Baldwin T (2012) Automatically constructing a normalisation dictionary for microblogs. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp 421–432

Hassan H, Menezes A (2013) Social text normalization using contextual graph random walks. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp 1577–1586

Hrastinski S (2008) The potential of synchronous communication to enhance participation in online discussions: A case study of two e-learning courses. Information and Management 45:499–506

Jamison E, Gurevych I (2013) Headerless, quoteless, but not hopeless? Using pairwise email classification to disentangle email threads. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing, pp 327–335

Ji Y, Eisenstein J (2014) Representation learning for text-level discourse parsing. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp 13–24

Joachims T, Finley T, Yu CNJ (2009) Cutting-plane training of structural SVMs. Machine Learning 77(1):27–59

Kane GC (2011) A multimethod study of information quality in wiki collaboration. ACM Transactions on Management Information Systems 2(1):4:1–4:16

Kim D, Lee Ss, Maeng S, Lee KP (2011) Developing idea generation for the interface design process with mass collaboration system. In: Marcus A (ed) Design, User Experience, and Usability. Theory, Methods, Tools and Practice, Lecture Notes in Computer Science, vol 6769, Springer Berlin Heidelberg, pp 69–76

Kim HN (2008) The phenomenon of blogs and theoretical model of blog use in educational contexts. Computers & Education 51(3):1342–1352

Kirschner P, Shum SB, Carr C (eds) (2003) Visualizing Argumentation. Springer London

Kittur A, Suh B, Pendleton BA, Chi EH (2007) He says, she says: Conflict and coordination in Wikipedia. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp 453–462

Krippendorff K (2004) Measuring the reliability of qualitative text analysis data. Quality and Quantity 38(6):787–800

Krishnamurthy J, Mitchell TM (2014) Joint syntactic and semantic parsing with combinatory categorial grammar. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp 1188–1198

Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web, pp 591–600

Laniado D, Tasso R (2011) Co-authorship 2.0: Patterns of collaboration in Wikipedia. In: Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia, HT ’11, pp 201–210

Larusson JA, Alterman R (2009) Wikis to support the “collaborative” part of collaborative learning. International Journal of Computer-Supported Collaborative Learning 4(4):371–402

Leuf B, Cunningham W (2001) The Wiki Way: Quick Collaboration on the Web. Addison-Wesley Longman Publishing Co., Inc.

Liddell FD (1983) Simplified exact analysis of case-referent studies: matched pairs; dichotomous exposure. Journal of Epidemiology & Community Health 37(1):82–84

Liu J, Ram S (2011) Who does what: Collaboration patterns in the Wikipedia and their impact on article quality. ACM Transactions on Management Information Systems 2(2):11:1–11:23

Lund A, Rasmussen I (2010) Tasks 2.0: Education meets social computing and mass collaboration. In: Proceedings of Society for Information Technology & Teacher Education International Conference 2010, pp 4058–4065

Meishar-Tal H, Gorsky P (2010) Wikis: what students do and do not do when writing collaboratively. Open Learning: The Journal of Open, Distance and e-Learning 25(1):25–35

Mochales R, Moens MF (2011) Argumentation mining. Artificial Intelligence and Law 19(1):1–22

Moore R (2014) Fast high-accuracy part-of-speech tagging by independent classifiers. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp 1165–1176

Moskaliuk J, Kimmerle J, Cress U (2012) Collaborative knowledge building with wikis: The impact of redundancy and polarity. Computers & Education 58(4):1049–1057

Newman MEJ (2004) Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences of the United States of America 101(suppl 1):5200–5205

Niu H, Van Aalst J (2009) Participation in knowledge-building discourse: An analysis of online discussions in mainstream and honours social studies courses. Canadian Journal of Learning and Technology 35:1–18

Noroozi O, Weinberger A, Biemans HJ, Mulder M, Chizari M (2013) Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. Computers & Education 61(0):59–76

O'Connor B, Krieger M, Ahn D (2010) TweetMotif: exploratory search and topic summarization for Twitter. In: International AAAI Conference on Weblogs and Social Media, pp 384–385

Onrubia J, Engel A (2009) Strategies for collaborative writing and phases of knowledge construction in CSCL environments. Computers & Education 53(4):1256–1265

Pena-Shaff JB, Nicholls C (2004) Analyzing student interactions and meaning construction in computer bulletin board discussions. Computers & Education 42(3):243–265

Perkins C, Murphy E (2006) Identifying and measuring individual engagement in critical thinking in online discussions: An exploratory case study. Educational Technology & Society 9:298–307

Potthast M, Stein B, Gerling R (2008) Automatic vandalism detection in Wikipedia. In: Advances in Information Retrieval: Proceedings of the 30th European Conference on IR Research, Springer, pp 663–668

Priedhorsky R, Chen J, Lam STK, Panciera K, Terveen L, Riedl J (2007) Creating, destroying, and restoring value in Wikipedia. In: Proceedings of the 2007 International ACM Conference on Supporting Group Work, pp 259–268

Robertson J (2011) The educational affordances of blogs for self-directed learning. Computers and Education 57(2):1628–1644

Saif H, He Y, Alani H (2012) Alleviating Data Sparsity for Twitter Sentiment Analysis. In: 2nd Workshop on Making Sense of Microposts (#MSM2012): Big things come in small packages at the 21st International Conference on the World Wide Web (WWW'12), CEUR Workshop Proceedings, Lyon, France, pp 2–9

Santos RLT, Macdonald C, McCreadie R, Ounis I, Soboroff I (2012) Information retrieval on the blogosphere. Foundations and Trends in Information Retrieval 6(1):1–125

Scheuer O, Loll F, Pinkwart N, McLaren BM (2010) Computer-supported argumentation: A review of the state of the art. International Journal of Computer-Supported Collaborative Learning 5(1):43–102

Scheuer O, McLaren B, Weinberger A, Niebuhr S (2014) Promoting critical, elaborative discussions through a collaboration script and argument diagrams. Instructional Science 42(2):127–157

Schiappa E, Nordin JP (2013) Argumentation: Keeping Faith with Reason, 1st edn. Pearson UK

Settles B (2009) Active learning literature survey. Tech. Rep. 1648, University of Wisconsin-Madison

Shatnawi S, Gaber MM, Cocea M (2014) Automatic content related feedback for MOOCs based on course domain ontology. In: Corchado E, Lozano JA, Quintin H, Yin H (eds) Intelligent Data Engineering and Automated Learning – IDEAL 2014, vol 8669, Springer International Publishing, pp 27–35

Smith NA (2011) Linguistic Structure Prediction. Morgan & Claypool Publishers

Søgaard A (2013) Semi-Supervised Learning and Domain Adaptation in Natural Language Processing. Morgan & Claypool Publishers

Stab C, Gurevych I (2014a) Annotating argument components and relations in persuasive essays. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp 1501–1510

Stab C, Gurevych I (2014b) Identifying argumentative discourse structures in persuasive essays. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp 46–56

Stahl G, Cress U, Law N, Ludvigsen S (2014) Analyzing the multidimensional construction of knowledge in diverse contexts. International Journal of Computer-Supported Collaborative Learning 9(1):1–6

Stegmann K, Weinberger A, Fischer F (2007) Facilitating argumentative knowledge construction with computer-supported collaboration scripts. International Journal of Computer-Supported Collaborative Learning 2(4):421–447

Su F, Beaumont C (2010) Evaluating the use of a wiki for collaborative learning. Innovations in Education and Teaching International 47(4):417–431

Tapscott D, Williams AD (2008) Wikinomics: How Mass Collaboration Changes Everything. Portfolio

Tjong Kim Sang E, van den Bosch A (2013) Dealing with big data: The case of Twitter. Computational Linguistics in the Netherlands Journal 3:121–134

Toulmin SE (1958) The Uses of Argument. Cambridge University Press

Viegas FB, Wattenberg M, Kriss J, van Ham F (2007) Talk before you type: Coordination in Wikipedia. In: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, pp 78–88

Walton D (2012) Using argumentation schemes for argument extraction: A bottom-up method. International Journal of Cognitive Informatics and Natural Intelligence 6(3):33–61

Weinberger A, Fischer F (2006) A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers & Education 46(1):71–95

Wheeler S, Yeomans P, Wheeler D (2008) The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. British Journal of Educational Technology 39(6):987–995

Wichmann A, Rummel N (2013) Improving revision in wiki-based writing: Coordination pays off. Computers & Education 62(0):262–270

Wu F, Weld DS (2010) Open information extraction using Wikipedia. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp 118–127

Yang Y, Eisenstein J (2013) A log-linear model for unsupervised text normalization. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp 61–72

Yih Wt, Chang MW, Meek C, Pastusiak A (2013) Question answering using enhanced lexical semantic models. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Sofia, Bulgaria, pp 1744–1753

Zhao D, Rosson MB, Matthews T, Moran T (2011) Microblogging's impact on collaboration awareness: A field study of microblogging within and between project teams. In: 2011 International Conference on Collaboration Technologies and Systems, pp 31–39

Zhu X, Goldberg AB (2009) Introduction to Semi-Supervised Learning, vol 6, Synthesis lectures on artificial intelligence and machine learning edn. Morgan & Claypool Publishers