Data Sets of Eating Disorders by Categorizing Reddit and Tumblr Posts: A Multilingual Comparative Study Based on Empirical Findings of Texts and Images

**Christina Baskal, Amelie Elisabeth Beutel, Jessika Keberlein, Malte Ollmann,
Esra Üresin, Jana Vischinski, Janina Weihe, Linda Achilles, Christa Womser-Hacker**
Department of Information Science and Natural Language Processing
University of Hildesheim, Universitätsplatz 1, 31141 Hildesheim, GER
{baskal, beutel, keberlei, ollmann, ueresin, vischins, weihej, achilles, womser}@uni-hildesheim.de

Abstract

Research has shown the potential negative impact of social media usage on body image. Various platforms present numerous medial formats of possibly harmful content related to eating disorders. Different cultural backgrounds, represented, for example, by different languages, are participating in the discussion online. Therefore, this research aims to investigate eating disorder specific content in a multilingual and multimedia environment. We want to contribute to establishing a common ground for further automated approaches. Our first objective is to combine the two media formats, text and image, by classifying the posts from one social media platform (Reddit) and continuing the categorization in the second (Tumblr). Our second objective is the analysis of multilingualism. We worked qualitatively in an iterative valid categorization process, followed by a comparison of the portrayal of eating disorders on both platforms. Our final data sets contained 960 Reddit and 2 081 Tumblr posts. Our analysis revealed that Reddit users predominantly exchange content regarding disease and eating behaviour, while on Tumblr, the focus is on the portrayal of oneself and one's body.

Keywords: Tumblr, Reddit, data set, social media analysis, content analysis, multilinguality, eating disorders, multimedia, language resource

1. Introduction

Eating disorders (ED) are a major health issue affecting many adolescents and young adults. The National Eating Disorder Association (NEDA), for instance, reported that approximately 20 million women and 10 million men in America will suffer from an eating disorder at some point in their lives (NEDA, 2021). A review of the prevalence and incidence of eating disorders (Hoek and van Hoeken, 2003) also reveals that only one out of three people in the general population with stringent diagnostic criteria receives treatment. Some sufferers declare their illness to be a legitimate, alternative lifestyle choice (Hoek and van Hoeken, 2003; Fox et al., 2005; Norris et al., 2006).

Therefore, the main purpose of this research is to explore the challenges that multimedia and multilingual social media texts and images pose for categorization and automated processing. We pursue the goal of extracting texts and images from two social media platforms Reddit[1] (text in English) and Tumblr[2] (text and image in German, English, Russian and Turkish). While the most diverse age groups from various countries increasingly spend time on social networks, it simultaneously gains interest to investigate factors such as grammar structure, content and the combination of text and image. Hence, we provided two data sets, in-

cluding the media formats, text and image, that form a basis for automatic analyses. The first data set is based on the eRisk data from 2018[3], consists of Reddit posts concerning ED. We further extended this data by means of a categorization procedure and referred to it as Reddit data set (RDS). The second, the Tumblr data set (TDS), was crafted by us by collecting images and their descriptions from Tumblr[4] We contribute to the ongoing research by enriching both data sets with our categorization, respectively.

The present study will (1) compare the topics discussed in ED communities of the two Social Media platforms, Reddit and Tumblr, by classifying the posts based on a qualitative content analysis approach and (2) investigate the differences of the four languages that are subject to our analyses.

2. Related Work

Past research has shown the impact of media consumption (magazines and TV) on disordered eating (Grabe et al., 2008). Also the influence of social media engagement on dysfunctional eating habits was investigated. One study found that already a short Facebook use of 20 minutes is associated with body weight and shape concerns in their study participants (Mabe et al., 2014). Another shows that Internet exposure correlates signif-

---

[1]Website of the social network Reddit: https://www.reddit.com/
[2]Website of the social network Tumblr: https://www.tumblr.com/

[3]Access to the research collection can be granted by following the instruction found on the website https://tec.citius.usc.es/ir/code/eRisk.html
[4]For access to the TDS, please contact the authors.

icantly with the internalisation of beauty ideals, body surveillance, and the drive for thinness (Tiggemann and Slater, 2013).

Automatic approaches of social media text analyses were used to measure the mental illness severity of anorectic internet users (Chancellor et al., 2016a) or to examine lexical variations of hashtags that derived after the banning of specific pro eating disorder (pro-ED) tags (Chancellor et al., 2016c). Other researchers have examined YouTube comments in different ED communities (pro-ED and the opposing anti-pro-ED community) based on their sentiments (Oksanen et al., 2015), Twitter tweets and how ED symptoms are discussed there (Cavazos-Rehg et al., 2019) and Reddit ED-community differences (Fettach and Benhiba, 2019).

Reddit is also in the focus of the eRisk Lab (early Risk Detection on the Internet)[5] that is held in conjunction with the CLEF Initiative (Conference and Labs of the Evaluation Forum)[6]. The main objective of eRisk is to provide a forum for the evaluation methodologies, performance metrics, and building of test collections concerning issues of health and safety on the internet (Losada et al., 2019). For that purpose, the organisers arranged shared tasks and provide associated data sets. In 2018 and 2019, the early detection of anorexia nervosa by sequentially processing Reddit posts was part of the challenge. The data set of 2018 was used in this paper also and is further described in section 3.1. Another study utilized likewise the anorexia data set of eRisk 2019 to analyse topical trends in anorectic Reddit users (Masood et al., 2020). The lab also puts emphasis on other mental disorders such as self-harm and depression, which patients of eating disorders are considered to be engaging with also (Hudson et al., 2007; Turner et al., 2015; Wang et al., 2017). Our institute also researched the early signs of self-harm (Achilles et al., 2020) and the severity of depression (Bandyopadhyay et al., 2019) of Reddit users in the past using the eRisk data sets.

Anorexia and its depiction on Tumblr was also the target of research in the past (Choudhury, 2015; Wick and Harriger, 2018). Other work studied the differences in the communication about it on Twitter and Tumblr (Branley and Covey, 2017) and more research on anorexia related imagery showed that pictures showing body parts (thin thighs/legs, flat stomachs, protruding hip bones, ribs or collar bones) are most common in the online discussions (Cavazos-Rehg et al., 2019). Another image-based study collected Instagram pictures and qualitatively classified them (Ging and Garvey, 2018). More work on differentiating the imagery of ED-content, represented by the hashtags *thinspiration* and *bonespiration* from the concept of *fitspiration* was done by Talbot and colleagues (Talbot et al., 2017).

All studies presented here were investigating either linguistic phenomena in the English language, or were retrieving ED-related imagery by utilizing English hashtags.

3. Methodological Approach

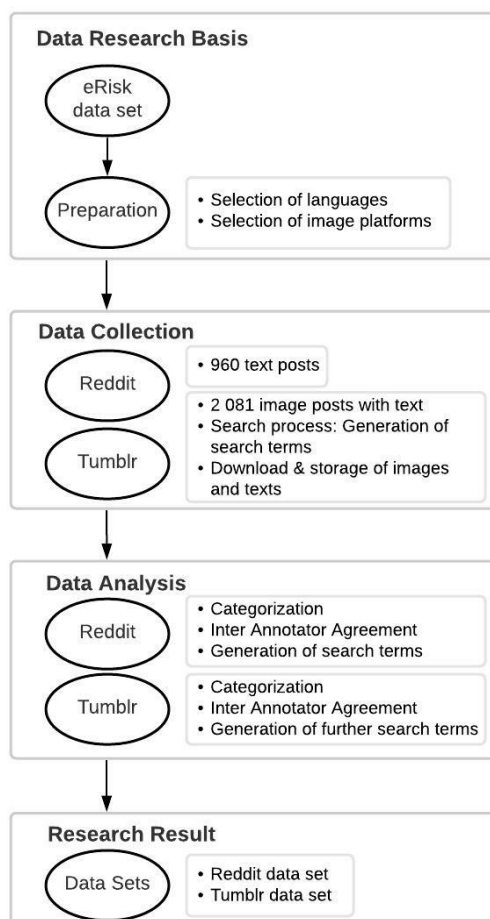Figure 1 represents our workflow. Each individual project step is discussed in more detail in the following sections.

Figure 1: The four stages of our methodological approach, with a focus on qualitative data categorization

3.1. Data Research Basis

Our starting point was the eRisk data set[7], that we defined as the baseline for our methodological approach (see Fig. 1). The data collection process is further described in the work of Losada and Crestani (Losada and Crestani, 2016). The data set was published as part of the CLEF eRisk workshop 2018 (Losada et al., 2018). A distinction is made between training and test data.

---

[5]Website of the eRisk Lab: https://erisk.irlab.org/

[6]Website of the CLEF initiative: http://www.clef-initiative.eu/

[7]Information on the eRisk data collection of 2018 and how to access it can be found here: https://tec.citius.usc.es/ir/code/eRisk.html

We referred solely to the training files. They were subdivided into positive and negative examples. In the context of our research, we worked exclusively with the positive examples, i.e., Reddit posts that explicitly came from users diagnosed with anorexia. Of the total of 152 participants, who were designated as subjects and anonymized by means of a subjectID, 20 people suffered from anorexia. We only analysed the posts of 18 since our review showed that two of them did not write about anorectic content. The Reddit posts were XML files presented in the form of 10 chunks. Each chunk consisted of 20 XML files, which summed up a total of 200 XML files. Furthermore, a chunk provides a chronological sequence: Therefore, chunk 1 contains posts that are further away in time than, for example, chunk 10. For our research, we went through all the posts manually.

Preparation

We selected four languages to investigate differences in the representation of ED given the images, image descriptions or hashtags. The literature review revealed that English is discussed predominantly. However, Russian, Turkish, and German are under-represented in the scientific research so far. We decided to use German, English, Russian and Turkish to examine the ED discourse on social media platforms. To accomplish our research aim, we defined four criteria for selecting an appropriate social media portal: 1) focus on mutual exchange of content, 2) multilingual searches that display results, 3) visibility of sensitive content and 4) download option of images and meta data. Tumblr thus covered all the criteria we previously decided on.

3.2. Data Collection & Data Analysis

First, we started the categorization by looking into the eRisk data set to gain an insight into the architecture of the data set. We defined the goal of the categorization at the beginning in the form of a list of categories that can describe the content of the Reddit posts. Then, we determined to use both main and sub-categories. Each post was assigned one or more main categories and an arbitrary number of matching sub-categories. The sub-categories serve to define the main categories and describe the context of the posts in more detail. The creation of the category list was an iterative process inspired by the taxonomy generation method of Nickerson et al. (2013). Nickerson et al. (2013) accentuate the importance to record end criteria when working iteratively. We defined both subjective and objective end criteria for our approach at the outset. The objective end criterion was that every Reddit post could be classified with at least one main category. For the subjective one, we determined that the categorization must be useful and able to describe the Reddit posts. We all worked together on the first chunk of Reddit posts in the first iteration and thus generated an initial list of main and sub-categories. Then we divided the remaining chunks and worked independently, meeting period-

ically to discuss new category suggestions. All in all, 6 492 posts were dropped during this process due to the fact that those contained content unrelated to our research, for instance, discussions about online games. 960 posts were reviewed and classified qualitatively by the group.

The **Inter Annotator Agreement (IAA)**, we performed to ensure an objective categorization, shows that we achieved a Fleiss' Kappa values of 0.86 for the Reddit posts and 0.83 for the Tumblr posts. Referring to Table 1, both values show near-perfect agreement between the annotators at over 80% (Landis and Koch, 1977). Each annotator independently assigned categories for the first 10% of the RDS and TDS, while only one main category could be assigned to each post.

Fleiss' Kappa	Interpretation
<0.00	Poor agreement
0.00 to 0.20	Slight agreement
0.21 to 0.40	Fair agreement
0.41 to 0.60	Moderate agreement
0.61 to 0.80	Substantial agreement
0.81 to 1.00	Almost perfect

Table 1: Interpretation of Fleiss' Kappa value thresholds

While we categorized the Reddit posts, we also extracted search terms for the later usage on Tumblr. Those would come from prominent words or topics which had to be related to ED and written in the posts. Besides, we found some ED-specific names of brands and new words related to the ED-culture, for instance, *thinspo*. This finding also overlaps with the study results presented in our literature review. *Thinspiration*, of which thinspo is the abbreviation, was the subject of study in different research settings (Wick and Harriger, 2018; Ging and Garvey, 2018; Talbot et al., 2017). While we initially generated those individually, we would draft the first list later. In this process, duplicates and search terms that were too broad would be deleted. Meanwhile, all of them were translated into the previously selected languages by the native speakers in our team. Furthermore, we collected exceptional ED-specific search terms because they could not be translated. Those were words used by the ED community and abbreviations we found.

After generating various search items based on the Reddit posts, we started the first search process on Tumblr to check which terms could be considered further. We were looking specifically for posts with (moving) images that can be saved or an image in the form of text, not a text-only post, which correlated with the Reddit categories. During the initial search, we generated more items by looking at the hashtags and texts under the posts. At the same time, we suspected that these new words were relevant because we frequently saw them during our Tumblr searches. After this process, we found 56 new search items, 26 of which were

Language	Search Terms
English	anorexia relapse, restriction
ED-specific	Ana, an0rex1a
Russian	голод, анорексичка
German	Abführmittel, fasten
Turkish	yeme bozukluğu

Table 2: Example search terms out of the final 127

Language	Search Terms %	Posts %
English	57 (44.9%)	902 (43.3%)
ED-specific	33 (26%)	639 (30.7%)
Russian	21 (16.5%)	350 (16.8%)
German	15 (11.8%)	182 (8.7%)
Turkish	1 (0.8%)	8 (0.4%)
Total	127 (100%)	2 081 (100%)

Table 3: Absolute numbers and probability distribution of final search terms per language

new ED-specific terms.

Following the initial search process on Tumblr, we colour-coded the words into relevant, irrelevant, and no hits to filter out the relevant search terms.

Table 2 shows example search terms of our final list after the completion of the above-mentioned processing steps.

We decided to look at the first 20 relevant image posts per relevant term during the second search process.

We downloaded these images and assigned them an appropriate ID. Additionally, we extracted descriptive metadata, such as captions and hashtags. Since a post could contain multiple images, the maximum number of images was therefore not 20 but could be exceeded. If a post contained more than one image, we marked that in the ID by simply extending it by a new number for the sub-post. Every item with less than three posts was irrelevant and deleted. Furthermore, we noticed that previously relevant ones were no longer relevant because they were textual. According to our criteria, we had to deliberately exclude such contributions, even if they would have been relevant in terms of content. The final composition of the successful ones can be taken from Table 3.

3.3. Research Result

In the following, we present the finalized categorization of both data sets.

Reddit Data Set

In this part we will go into more detail on how we proceeded analytically. For a complete list of all the possible main and sub-categories, please refer to Figure 2. According to the user agreement with the eRisk organizers, it is not permitted to show example posts of the RDS. Therefore, we describe here our general approach to analyze the posts. For instance, if a user referred to food and mealtime, this post would be categorized as

eating behaviour (as represented in Figure 2 with the grey highlighted row) because its main content revolves around what and when to eat. Furthermore, the words *meal plan* and *calories* are mentioned explicitly (likewise highlighted in the white boxes). These are indicators for the respective sub-categories. If a poster writes about their wish to get better and sustain a healthier lifestyle, it implies the main category of disease and the sub-category desire for recovery. A combination of several main and sub-categories is also possible. If a user referred to several sports and an exercise plan in their post, we would classify it with the main category urge to exercise and the corresponding sub-categories question about physiology and sports activities.

Tumblr Data Set

As already mentioned, it is impossible to construct a complete data set of Tumblr posts, meaning the total period of ED-related writings starting from the beginning of Tumblr itself. In this case, an opportunistic selection was chosen. This means that the amount of content determines data availability. It has been noted that in a couple of search terms, the same posts reoccurred after the approximate number of 20 posts. Therefore, we decided to focus on this specific amount to rightfully manage an appropriate number of posts for the random sample and provide no duplicate posts within a search term. Furthermore, it is important to define the amount of selection specifically. In this part, we consolidated our whole retrieved Tumblr data into one data set, namely TDS. Like we did for the Reddit posts, we created a table for the writings we retrieved from Tumblr. The table contains multiple columns, each concentrated explicitly on a specific topic. The first column was generated to retrieve and organize our search terms more effectively, which were already explained in the previous section 3.2.

The next column showed the Reddit main categories that were explained in the previous section 3.3.

In the following column we specified the types of the Tumblr images. Prior to that, we have given some definitions for some types to ensure an unanimous understanding within the team. Our findings on Tumblr showed that the following eight categories occurred the most: *drawing*, *food*, *person*, *meme*, *medication*, *fashion*, *text* and also allowing the option *other* if none of the other types were suitable.

Sometimes the pictures on their own were not identifiable and therefore required some of the metadata like the caption and hashtags for a better understanding. Consequently, we concluded that creating separate columns was necessary.

One of these columns is regarding the multilingualism of the posts. Here, we marked if the hashtags and/or the caption is written in more than one language. In Fig. 5, we marked that there is no multilingual text/hashtags through 'no' and coded the identified language English as '2'.

main category	1) urge to exercise	2) eating behaviour	3) state of anxiety	4) perception of body and weight	5) disease	6) community
sub-category	a) question about physiology	a) call for help	a) call for help	a) comparison with other people	a) biographical content	a) comparison within the community
	b) sport activities	b) calories	b) fear of non-self-control	b) external perception	b) daily routine	b) negative support
		c) fasting	c) fear of rejection/bad feedback	c) inspiration	c) desire for recovery	c) negative tips
		d) loss of control	d) fear of relapse	d) own body perception	d) diagnosis	d) positive support
		e) meal plan	e) fear that acquaintances will find out	e) positive feedback to own looks/body	e) emotionality	e) positive tips
		f) recipe	f) question about physiology/fear of loss of control	f) revulsion	f) fear of recovery	
		g) rejection of specific products		g) weight gain/loss/data	g) medication	
		h) supporting products			h) mental illness	
		i) time of the meal			i) relapse	
					j) side effects of disease	
					k) therapy	

Figure 2: Complete list of main and sub-categories of the RDS and TDS

search term	mapping of the main category	picture classification	multilingual text/hashtags	languages	text	hashtags
en_body_check_09	perception of body and weight	person	no	2	My body checks are so different	#bodycheck#thinspo#thinspiration#myphoto#skinny#slim#skinny girl
en_body_check_10	perception of body and weight	person	no	2	I don't like how my body looks.	#ana#pro ana#anorexia#not pro just for myself#ana tips#ana diary
en_body_check_11	perception of body and weight	person	no	2	My body check!	#anamiia#anamia#anamiia#anarexia#anarexix

Figure 3: An excerpt from the Tumblr data set using the search term "body check" and the main category "perception of body and weight" as an example

To get an idea of the TDS and to retrace our analysis and categorization process, we prepared a graphic with six sample images for each main category (see Fig. 5). We deliberately chose examples for this paper that do not contain inappropriate images of body parts such as thin legs or collar bones which may be disturbing for the readers, or pictures that violate the anonymity of the users.

4. Findings and Discussion

With our work, we contribute to the scientific community by combining a Tumblr and Reddit data set to examine how eating disorders are discussed and how the use of multilingualism is distributed. By applying a common category list, we could compare both data sets. The posts can be analysed in three ways: one and only one term was stated (*single*), the examined term and additionally, one or several terms were stated (*multiple*) and several terms are combined (*combination*).

One of the main findings is that the categories perception of body and weight, eating behaviour and disease are the most frequent ones for both platforms. The finding that the image category person was found the most in both single and multiple distribution supports the idea that Tumblr is a photo-based platform. Here, both self-expression, as well as the portrayal of the body, are at the centre of users. Writing about one's behaviour
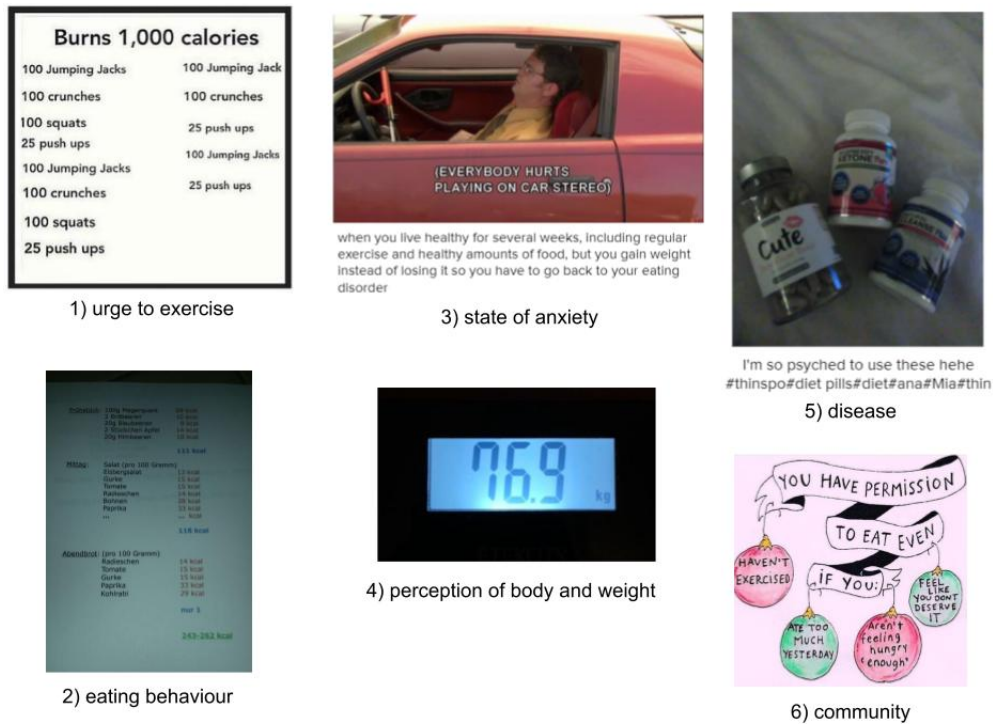
Figure 4: Example images from the Tumblr data set, sorted by different main categories. 1) Shows a workout plan, 2) a meal plan including calories, 3) a meme on ED, 4) the display of a scale, 5) a set of diet pill bottles, 6) an encouraging community post against ED

is probably easier than communicating the same content via image and a short description with hashtags, which reasons that Reddit as a text-based social media platform mainly discusses topics such as disease and eating behaviour. As Cavazos-Rehg et al. (2019) discussed in their paper, the categories body shape, eating concerns and weight concerns appeared in descending order. We cannot compare these categories one-to-one with our category list, as we combined posts regarding body shape and weight concern in the category perception of body and weight. Furthermore, just as Cavazos-Rehg et al. (2019) and Wick and Harriger (2018), we also found that the search term *thinspiration* led to images of body parts such as thin legs and stomachs as well as before-after images. Compared to the image-based study conducted by Ging and Garvey (2018), we found similar categories but distinguished them differently. For instance, our main category disease includes the sub-category mental illness, which contains depression, self-harm and suicide, which all come up in their study. Further, their categories pro-recovery and selfie pictures are also represented in our study with variable names. In contrast to their categorization, we subdivided the images categories in a more incremental approach relevant to the content displayed.

The discussion of Mental Illness Severity topics on both social media platforms focuses on the self-portrayal, eating behaviour and disease of a user. These findings overlap with Chancellor et al. (2016a): Here, one of the three main markers is excessive weight control. Specifically, on Tumblr, we found that (pro-)ED-specific hashtags were applied. This supports Chancellor et al. (2016c) and Chancellor et al. (2016b) findings. Prior to the Tumblr data extraction, the team shifted awareness to the sensitive content that is shared on the social media platform. Whenever a team member felt overwhelmed, it was shared and another team member took over going through the texts and images.

The language distribution can only be considered on Tumblr as the Reddit posts were all in English. Over 85% of all Tumblr posts were monolingual. The multiple distribution of languages indicates that the most commonly used language was English, followed by Russian, German, other languages and Turkish in descending order. German and English was the most common language combination, followed by English and Russian. That indicates that English functions as a lingua franca to reach a large audience, either solely or in combination with other stated languages.

The language distribution of successful search terms also shows that English with 41% was the language with the most search results. However, as the classification of ED-specific search terms is not an official language but rather a set of ED-specific vocabulary, it is not surprising that they led to the second most oc-
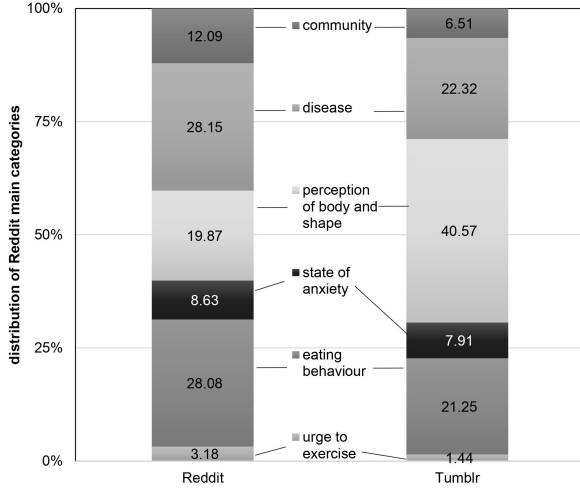
Figure 5: Comparison of Reddit main categories on Reddit and Tumblr in percentage

Categories	EN	ED	RU	DE	TR
no Reddit main Category	11 (1.5%)	24 (4.2%)	0 (0%)	6 (4%)	0 (0%)
urge to exercise	9 (1.2%)	2 (0.3%)	0 (0%)	2 (0.3%)	0 (0%)
eating behavior	184 (25.4%)	110 (19.4%)	37 (13.9%)	18 (11.9%)	0 (0%)
state of anxiety	12 (1.7%)	31 (5.5%)	14 (5.2%)	3 (2%)	0 (0%)
perception of body & weight	323 (44.6%)	310 (54.6%)	167 (62.8%)	69 (45.7%)	3 (100%)
disease	158 (21.9%)	78 (13.7%)	39 (14.7%)	46 (30.5%)	0 (0%)
community	27 (3.7%)	13 (2.3)	9 (3.4%)	7 (4.6%)	0 (0%)
Total	724 (100%)	568 (100%)	266 (100%)	151 (100%)	3 (100%)

Table 4: Category distribution for the total amount and percentage of languages of posts on Tumblr

curred results with 35%.

The category distribution of each language (see Table 4) shows that perception of body and weight was the most discussed topic in all languages indicating that this is a dominant subject regarding ED. For English and ED-specific terms, eating behaviour was the second most commonly used category, while for Russian and German posts, the most prevailing category was disease.

We observed that ED-specific terms were frequently combined with the given languages German, English, Russian, Turkish and/or other languages. For instance, one of the 33 ED-specific terms is thinspiration and often occurs in combination with our selected languages. These combinations were visible either in the post description, in hashtags and/or, in some cases, as text on an image. However, we did not quantitatively anal-

yse the distribution of ED-specific terms and other languages. That would be interesting research to resume on our current findings.

5. Limitations and Future Directions

As the data validation showed, the respective languages' results were not balanced, as Turkish had only one relevant search term. Despite the lack of Turkish-language data, it was possible to conclude that in the ED context on Tumblr, English hashtags are predominantly used by users speaking other languages as well. This can be justified by the fact that users can better express their sense of belonging in this particular community this way, and users may find a larger community with common interests. Regarding our small Turkish data set, we found a study from Bulut and Doğan (2017) showing that Tumblr is one of the more unpopular social media networking sites in Turkey. This is in accordance to a statistic from Clement (2022) published in March 2022, which shows, network-traffic from Turkey to Tumblr.com being almost non-existent. Further research can follow up on our study by considering additional languages and other image platforms. In addition, more Reddit and Tumblr posts could be analyzed to improve the data sets. The data collection can be used for machine processing in further steps to use automatic methods. Furthermore, our collected data can be trained for image recognition: The system learns the defined categories and can match them with the hashtags and images used. This could be helpful for the early detection of eating disorders. Moreover, the texts can be examined linguistically. For instance, the distribution of ED-specific terms in posts could be examined regarding the language combination of monolingual and/or multilingual posts. The special terms as a language-independent construct can be further explored in more in-depth research.

6. Ethical Considerations

The Tumblr posts used in this study were publicly available. The names used by blog authors are fictional. However, in the data set the names were removed and only images, texts and hashtags of the posts were kept. The data set is saved on university servers behind password protection. Quotes have been slightly altered to further protect the individuals who have written these social media contributions. Therefore, a jurisdiction of our university's Ethics Commission is not required for this study.

7. References

Achilles, L., Kisselew, M., Schäfer, J., and Kölle, R. (2020). Using Surface and Semantic Features for Detecting Early Signs of Self-Harm in Social Media Postings. In Linda Cappellato, et al., editors, *Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020*, volume 2696 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Bandyopadhyay, A., Achilles, L., Mandl, T., Mitra, M., and Saha, S. K. (2019). Identification of Depression Severity for Users of Online Platforms. In Robert Jäschke et al., editors, *Proceedings of the Conference on "Lernen, Wissen, Daten, Analysen", Berlin, Germany, September 30 - October 2, 2019*, volume 2454 of *CEUR Workshop Proceedings*, pages 331–342. CEUR-WS.org.

Branley, D. B. and Covey, J. (2017). Pro-ana versus pro-recovery: A content analytic comparison of social media users' communication about eating disorders on Twitter and Tumblr. *Frontiers in Psychology*, 8:1356.

Bulut, Z. A. and Doğan, O. (2017). The abcd typology: Profile and motivations of turkish social network sites users. *Computers in Human Behavior*, 67:73–83.

Cavazos-Rehg, P. A., Krauss, M. J., Costello, S. J., Kaiser, N., Cahn, E. S., Fitzsimmons-Craft, E. E., and Wilfley, D. E. (2019). "I just want to be skinny.": A content analysis of tweets expressing eating disorder symptoms. *PloS one*, 14(1):e0207506.

Chancellor, S., Lin, Z., Goodman, E. L., Zerwas, S., and Choudhury, M. D. (2016a). Quantifying and Predicting Mental Illness Severity in Online Pro-Eating Disorder Communities. In Darren Gergle, et al., editors, *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW 2016, San Francisco, CA, USA, February 27 - March 2, 2016*, pages 1169–1182. ACM.

Chancellor, S., Lin, Z. J., and Choudhury, M. D. (2016b). "This Post Will Just Get Taken Down": Characterizing Removed Pro-Eating Disorder Social Media Content. In Jofish Kaye, et al., editors, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, May 7-12, 2016*, pages 1157–1162. ACM.

Chancellor, S., Pater, J. A., Clear, T. A., Gilbert, E., and Choudhury, M. D. (2016c). #thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. In Darren Gergle, et al., editors, *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW 2016, San Francisco, CA, USA, February 27 - March 2, 2016*, pages 1199–1211. ACM.

Choudhury, M. D. (2015). Anorexia on Tumblr: A Characterization Study. In Patty Kostkova et al., editors, *Proceedings of the 5th International Conference on Digital Health 2015, Florence, Italy, May 18-20, 2015*, pages 43–50. ACM.

Clement, J. (2022). May.

Fettach, Y. and Benhiba, L. (2019). Pro-Eating Disorders and Pro-Recovery Communities on Reddit: Text and Network Comparative Analyses. In *Proceedings of the 21st International Conference on In-formation Integration and Web-based Applications & Services, iiWAS 2019, Munich, Germany, December 2-4, 2019*, pages 277–286. ACM.

Fox, N., Ward, K., and O'rourke, A. (2005). Pro-anorexia, Weight-loss Drugs and the Internet: an 'Anti-recovery' Explanatory Model of Anorexia. *Sociology of health & illness*, 27(7):944–971.

Ging, D. and Garvey, S. (2018). 'Written in these scars are the stories I can't explain': A Content Analysis of Pro-ana and Thinspiration Image Sharing on Instagram. *New Media Soc.*, 20(3):1181–1200.

Grabe, S., Ward, L. M., and Hyde, J. S. (2008). The Role of the Media in Body Image Concerns among Women: A meta-analysis of Experimental and Correlational Studies. *Psychological Bulletin*, 134(3):460.

Hoek, H. W. and van Hoeken, D. (2003). Review of the Prevalence and Incidence of Eating Disorders. *The International Journal of Eating Disorders*, 34:838–396.

Hudson, J. I., Hiripi, E., Pope Jr, H. G., and Kessler, R. C. (2007). The Prevalence and Correlates of Eating Disorders in the National Comorbidity Survey Replication. *Biological psychiatry*, 61(3):348–358.

Landis, J. and Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.

Losada, D. E. and Crestani, F. (2016). A Test Collection for Research on Depression and Language Use. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 28–39. Springer.

Losada, D. E., Crestani, F., and Parapar, J. (2018). Overview of eRisk: Early Risk Prediction on the Internet. In Patrice Bellot, et al., editors, *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings*, volume 11018 of *Lecture Notes in Computer Science*, pages 343–361. Springer.

Losada, D. E., Crestani, F., and Parapar, J. (2019). Overview of eRisk 2019 Early Risk Prediction on the Internet. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 340–357. Springer.

Mabe, A. G., Forney, K. J., and Keel, P. K. (2014). Do you "like" my photo? Facebook use maintains eating disorder risk. *International Journal of Eating Disorders*, 47(5):516–523.

Masood, R., Hu, M., Fabregat, H., Aker, A., and Fuhr, N. (2020). Anorexia Topical Trends in Self-declared Reddit Users. In Iván Cantador, et al., editors, *Proceedings of the First Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020), Samatan, Gers, France, July 6-9, 2020*, volume 2621 of *CEUR Workshop Proceedings*. CEUR-WS.org.

NEDA. (2021). What Are Eating Disorders? Available online at: https://www.nationaleatingdisorders.org/what-are-eating-disorders (last retrieved: 17.08.2021).

Nickerson, R. C., Varshney, U., and Muntermann, J. (2013). A Method for Taxonomy Development and its Application in Information Systems. *European Journal of Information Systems*, 22:336–359.

Norris, M. L., Boydell, K. M., Pinhas, L., and Katzman, D. K. (2006). Ana and the Internet: A Review of Pro-anorexia Websites. *International Journal of Eating Disorders*, 39(6):443–447.

Oksanen, A., Garcia, D., Sirola, A., Näsi, M., Kaakinen, M., Keipi, T., and Räsänen, P. (2015). Pro-Anorexia and Anti-Pro-Anorexia Videos on YouTube: Sentiment Analysis of User Responses . *Journal of Medical Internet Research*, 17(11):e256.

Talbot, C. V., Gavin, J., Van Steen, T., and Morey, Y. (2017). A Content Analysis of Thinspiration, Fitspiration, and Bonespiration Imagery on Social Media. *Journal of eating disorders*, 5(1):1–8.

Tiggemann, M. and Slater, A. (2013). NetGirls: The Internet, Facebook, and Body Image concern in adolescent girls. *International Journal of Eating Disorders*, 46(6):630–633.

Turner, B. J., Yiu, A., Layden, B. K., Claes, L., Zaitsoff, S., and Chapman, A. L. (2015). Temporal Associations Between Disordered Eating and Nonsuicidal Self-injury: Examining Symptom Overlap Over 1 Year. *Behavior Therapy*, 46(1):125–138.

Wang, Y., Tang, J., Li, J., Li, B., Wan, Y., Mellina, C., O'Hare, N., and Chang, Y. (2017). Understanding and Discovering Deliberate Self-harm Content in Social Media. In Rick Barrett, et al., editors, *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 93–102. ACM.

Wick, M. R. and Harriger, J. A. (2018). A Content Analysis of Thinspiration Images and Text Posts on Tumblr. *Body Image*, 24:13–16.