

Web-Based Annotation Tool for Instant Messaging Conversations

Jaromír Plhák, Michaela Lebedíková, Michal Tkaczyk, and David Šmahel

Faculty of Informatics, Masaryk University, Brno, Czech Republic

{xplhak, x450458, x245062, davs}@fi.muni.cz

Goal

- Develop an application that can detect online risk phenomena and social support
 - Get data from adolescents
 - Annotate data
 - Generate models
 - Involve OCR to previously developed application for acquiring data
 - And ...

Goal ... or a Dream?

- Develop an application that can detect online risk phenomena and social support
 - Get data from adolescents
 - **Annotate data**
 - Generate model
 - Involve OCR to previously developed application for acquiring data
 - And ...



IRTIS Annotation Tool – Challenges

- Anonymize the data
- Parse exported data to suitable units of text that would be large enough to convey the meaning of the conversation and its topic
- Annotation of complete utterances representing one user prompt



IRTIS Annotation Tool – Functionality (1)

- Data can be uploaded to the server, anonymized, and divided into conversations
- Tagsets can be specified and include selected tags
- Conversations are grouped in a series of messages called bursts
- Annotators can choose the burst (from the assigned bursts) they want to work on

IRTIS Annotation Tool – Functionality (2)

- Conversations can be annotated by annotators using tags from tagset:
 - Tags can be selected using key pressing
 - Annotators can select up to three additional tags
 - Previous and subsequent conversations can be loaded on demand
 - Interactive annotation manual is integrated into GUI
- Supervisor mode allows supervisors to decide ambiguous cases
- Statistics about work progress will be displayed in the GUI
- Annotated data can be exported into MS Excel sheet or .csv files

User: [REDACTED]

Please select your assignment:

Active Unfinished Finished

Assignment	Burst	TagSet	Status	Conversations	Start time	Finish time
0	0_1_unique	0	Finished	3808/3808	18/01/2022, 14:04:27	04/07/2023, 17:57:54
1	0_1_unique_q_mark	0	Finished	578/578	18/01/2022, 08:37:45	12/07/2023, 20:37:44
2	0_1_same_q_mark	0	Finished	316/318	28/01/2022, 10:01:58	28/01/2022, 12:31:20
3	0_1_same_two_q_marks	0	Finished	31/31	28/01/2022, 09:50:58	28/01/2022, 10:01:41
4	0_2_unique	0	Unfinished	35/4065	13/07/2023, 12:48:28	NaN
5	0_2_unique_q_mark	0	Active	0/549	NaN	NaN

Conversation: 2024 - 69 - 210 (36/4065)

- 0: NO TAG (0)
- 1: Informační podpora (1)
- 2: Emocionální podpora (2)
- 3: Začlenění do skupiny (3)
- 4: Uznání (4)
- 5: Nabídka pomoci (5)
- 21: Cizí jazyk (6)

#	Line	Ann 1	Ann 2	Tag
9	09.10.2019 19:04:47 - Kateřina Lendrová: Už meli laborky			
10	09.10.2019 19:05:39 - Pavel Ashby: říkala písemku na příklady	1	1	
11	09.10.2019 19:06:23 - Darja Poljansky <OWNER>: Hej podle mě fakt ne ale jak chcete	1	1	
12	09.10.2019 19:06:37 - Pavel Ashby: tak ono je to easy	2		2 <input type="text"/> <input type="text"/>
13	09.10.2019 19:22:07 - Darja Poljansky <OWNER>: Hej ten případ se sanema			
14	09.10.2019 19:22:13 - Darja Poljansky <OWNER>: U té práce			
15	09.10.2019 19:22:19 - Darja Poljansky <OWNER>: To kdy nastava			
16	09.10.2019 19:22:50 - Darja Poljansky <OWNER>: ?			
17	09.10.2019 19:24:32 - Pavel Ashby: kdyz jedes na sanich?			
18	09.10.2019 19:24:46 - Darja Poljansky <OWNER>: No ale jaký pohyb			
19	09.10.2019 19:24:51 - Darja Poljansky <OWNER>: Šikmo dolů?			
20	09.10.2019 19:25:23 - Pavel Ashby: wtffff			
21	09.10.2019 19:26:41 - Darja Poljansky <OWNER>: Se uklidni			
22	09.10.2019 19:26:49 - Darja Poljansky <OWNER>: Jen se snažím něco zjistit			
23	09.10.2019 19:26:58 - Pavel Ashby: vsak jaa jsem v klidu			
24	09.10.2019 19:27:04 - Pavel Ashby: jen nechapu jaky pohyb			
25	09.10.2019 19:27:07 - Pavel Ashby: dolu			

- Previous
- Reset
- Following
- Go back
- Stop Annotation
- Finish burst
- Save this conversation and load next one

50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200 1250 1300 1350 1400 1450 1500
1550 1600 1650 1700 1750 1800 1850 1900 1950 2000 2050 2100 2150 2200 2250 2300 2350 2400 2450 2500 2550 2600 2650 2700 2750 2800 2850 2900 2950 3000
3050 3100 3150 3200 3250 3300 3350 3400 3450 3500 3550 3600 3650 3700 3750 3800 3850 3900 3950 4000 4050

- Enlarge
- Hide

ANOTAČNÍ MANUÁL: PODPŮRNÉ INTERAKCE

Verze 2. 23. 11, zkrácená verze CO JE TO PODPORA (PODPŮRNÉ INTERAKCE ONLINE)?

Podpůrné interakce online mezi vrstevníky, my se zaměříme konkrétně na **poskytování podpory** ve smyslu komunikace, o které můžeme předpokládat, že u příjemce vyvolá pocit/přesvědčení, že někomu na něm záleží, je milován, vážen a oceňován, že někdo má o něj starost, je pro něj oporu, že mu někdo poskytne radu nebo užitečnou informaci, bude-li potřeba, pomůže mu, a také, že je součástí skupiny, s jejíž členy tráví čas a má společné plány.

Pro každý typ podpory platí, že pokud výpověď v daném řádku nemá tento účel, neměla by být označena jako podpora.

1. Informační podpora

Je obsahem řádku něco z následujícího:

[dávání rad/tipů | učení | zpětné vazby | předávání znalostí/informací/zkušenosti, které druhý potřebuje a jsou mu nápomocné v řešení problému/pochopení situace, v které se ocitl]

- Z kontextu konverzace musí být zřejmé, že informace jsou pro příjemce PODPORU v nějakém konkrétním jednání/akci/situaci | jsou nápomocné v řešení problému

Statistics about annotations

Show entries

Search:

User	Hash	Real data	Burst	TagSet	Status	Start time	Finish time	Lines	Work time	Fully annotated	
Jana	████████████████████	False	0	0	Unfinished	03/12/2020, 11:19:34	NaN	449/4142	2:04:56	18/259	Details
Jarek	████████████████████	False	0	0	Unfinished	04/12/2020, 16:18:39	NaN	246/4142	1:21:20	15/259	Details
Jarek	████████████████████	False	47	1	Unfinished	14/03/2021, 22:11:46	NaN	0/10088	0:06:10	0/182	Details
Karolína	████████████████████	True	10	0	Finished	21/12/2020, 09:48:49	29/12/2020, 11:28:06	8003/8003	8:59:24	391/391	Details
Karolína	████████████████████	True	11	0	Finished	29/12/2020, 11:48:14	02/01/2021, 18:01:23	8008/8008	9:53:19	447/447	Details
Karolína	████████████████████	True	12	0	Finished	04/01/2021, 16:09:57	05/01/2021, 21:26:23	8145/8145	9:09:17	437/437	Details
Karolína	████████████████████	False	5	0	Finished	09/01/2021, 13:22:55	NaN	4004/4004	0:01:44	243/243	Details
Karolína	████████████████████	True	13	0	Finished	07/01/2021, 11:35:50	09/01/2021, 13:23:58	8148/8148	8:37:30	425/425	Details

Annotation Statistics (1)

- 272,465 utterances with online risky behavior tagset (ORB)
- 196,772 utterances with social support tagset (SocS)

Table 1: Overview of the number of annotated utterances and IAA (Cohen's κ)

Category	Tagset	Annotated by at least one annotator	κ
Aggression, harassment, hate	ORB	5393 (1.98%)	.470
Mental health problems	ORB	3101 (1.14%)	.460
Alcohol, drugs	ORB	2301 (0.85%)	.609
Sexual content	ORB	3550 (1.30%)	.485
Informational support	SocS	9967 (5.07%)	.685
Emotional support	SocS	9669 (4.92%)	.639
Social companionship	SocS	3331 (1.70%)	.604
Appraisal	SocS	2524 (1.28%)	.650
Instrumental support	SocS	5317 (2.70%)	.599

Annotation Statistics (2)

Table 2: Times of annotations for selected bursts with 64,452 utterances.

Annotator's id	Tagset	Total time (h:mm:ss)	Per utterance
1	SocS	68:47:33	3.84 sec.
2	SocS	77:13:49	4.31 sec.
3	ORB	25:51:13	1.44 sec.
4	ORB	22:27:41	1.25 sec.

Limitations and Future Work

- We developed IRTIS AT for a specific task
- Some functionality has to be solved manually or outside the tool

- User accounting with roles
- Store data directly in the database
- Automatic division and anonymization of uploaded data
- Additional functionality like automatic IAA calculation

Conclusions

- User-friendly annotation tool for annotation of texts from online communicators like Messenger
- Used in practice for annotating social support and online risky behavior in anonymized data of adolescents
- Outputs can be practically usable in many applications like parental control applications provided by social networking sites