

RU-ADEPT: Russian Anonymized Dataset with Eight Personality Traits

C. Anton Rytting*, Valerie Novak*, James R. Hull*,
Victor M. Frank*, Paul R. Rodrigues†, Jarrett G.W. Lee*, Laurel G. Miller-Sims*

* University of Maryland, College Park, MD, U.S.A.
† Accenture, Washington, D.C., U.S.A

{crying, vnovak, jhull1, vmfrank, jgwl, lmillers}@umd.edu, paul.rodrigues@accenture.com



Overview

RU-ADEPT – a dataset of Russian authors' personality trait scores, demographic information, with associated corpus of the authors' cross-contributions to (up to) four different social media platforms





Personality Survey Data

- OCEAN: Big Five Inventory (BFI-2)
- Dark Triad: Short Dark Triad (SD3)

Demographics Info

Age, Gender, Urban/Rural

Social Media Text

- VKontakte (VK) 
- LiveJournal 
- Blogger 
- Moi Mir (My World) 



Motivation

- Train machine learned classifiers for the application of Big Five and Dark Triad labels on Russian authors.
- Understand the differences in individuals' exhibited personality across social media platforms.
- Understand the impact text length and formality have on personality identification



Novel Contributions

- 1st anonymized dataset pairing demographic and personality trait data with Russian-language social media content.
- 1st paper to describe collection of Dark Triad scores across multiple Russian-language social media platforms.
- 1st dataset of personality traits with participant-generated content sampled across several different social media sites (for a subset of participants, n=149)

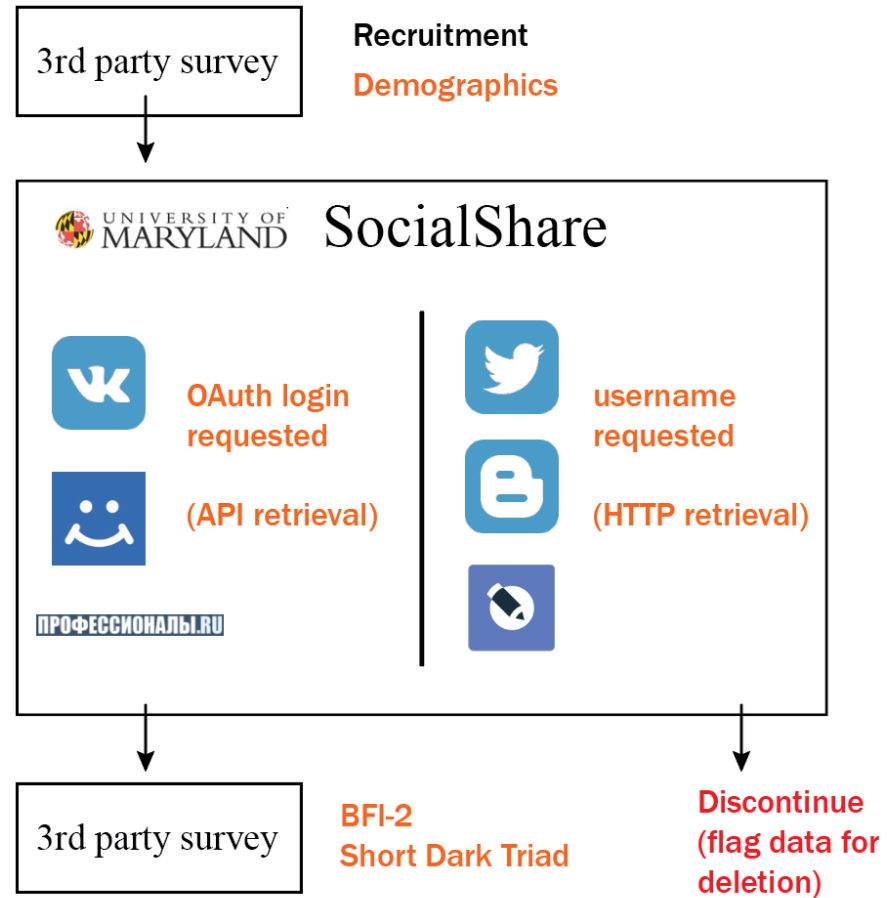


Prior Russian Text Studies of Personality Traits

Authors & Year	Traits	Text Genre/Platform	# of Participants
Bogolyubova et al. (2018)	Dark Triad	Facebook	6,724
Stankevich et al. (2019)	Big Five	Vkontakte	1,020
Litvinova et al. (2016)	Various (inc. Big Five)	Prompted picture descriptions, letters, essays, etc.	1,145 (BF for 192)
Litvinova & Ryzhkova (2018)	Various (inc. Big Five)	Prompted picture descriptions, letters	555 (BF for 209)



SocialShare – Social Media Collection Platform



Five Factor Model (“Big Five”)

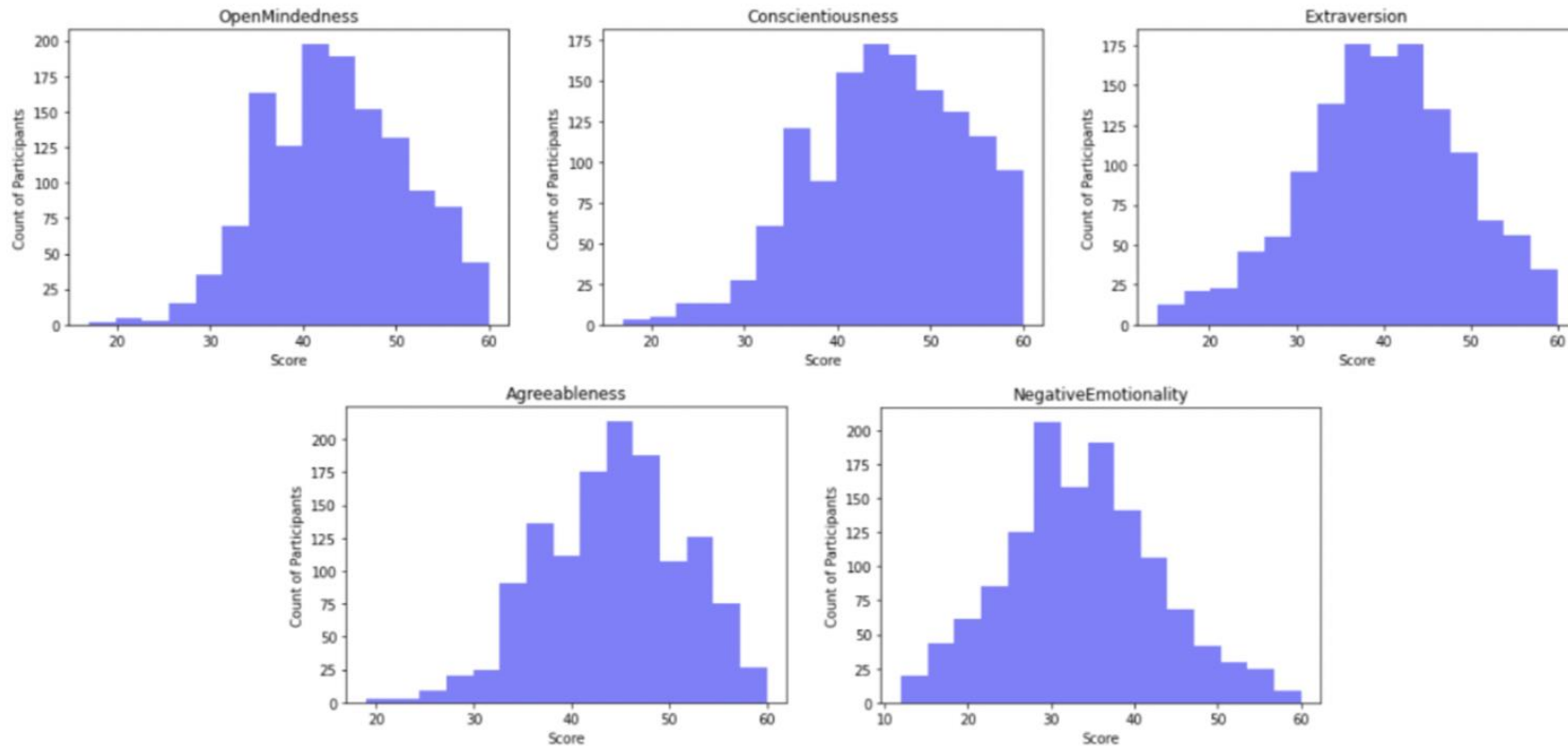
- Five Factors of Personality (a.k.a. OCEAN)
 - Extraversion
 - Agreeableness
 - Conscientiousness
 - Open-Mindedness
 - Negative Emotionality (sometimes called Neuroticism).

We used the *Big Five Inventory-2 (BFI-2)* – Soto et al. (2017)

- 60-item self-report questionnaire; 5-point Likert scale
- Russian translation by Shchebetenko et al. (2020)



Big Five Inventory-2 (BFI-2) Results



Scores are similar to those reported in Shchebetenko et al. (2020).



Dark Triad

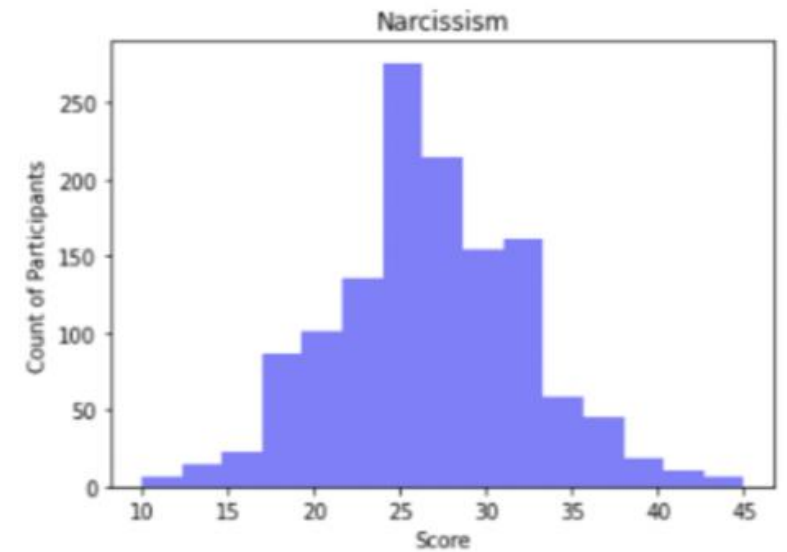
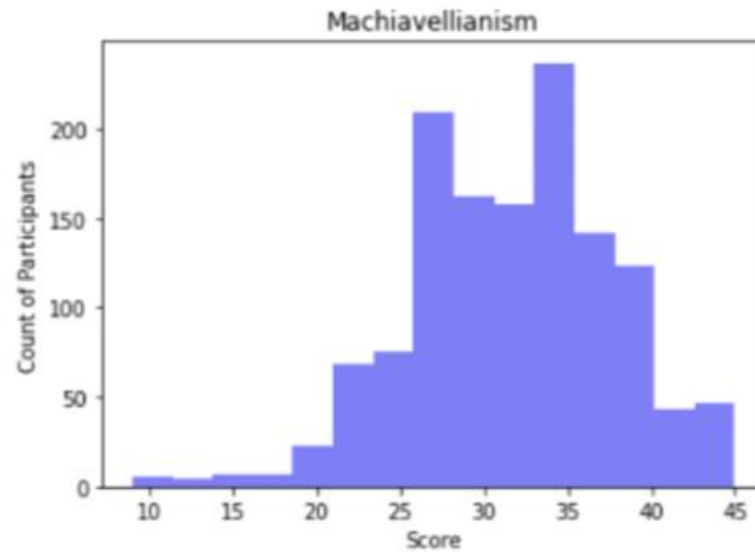
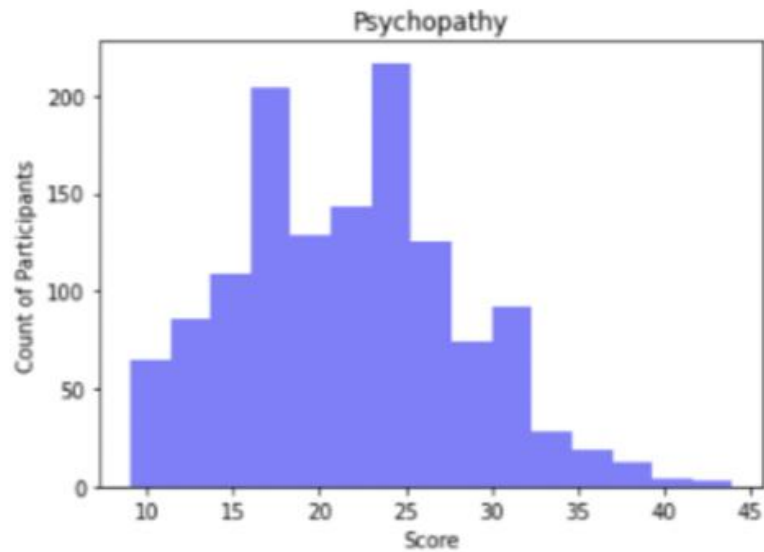
- Dark Triad (Paulhus and Williams, 2002)
 - Machiavellianism - essentially a manipulative personality
 - Narcissism - involves thoughts and behaviors that espouse entitlement, superiority, and dominance
 - Psychopathy - characterized by four subtraits: high impulsivity and thrill-seeking, combined with low anxiety, low empathy for others

We used the *Short Dark Triad (SD3)* (Jones, 2014)

- 27-item self-report questionnaire; 5-point Likert scale
- Russian translation by Egorova et al. (2016)



Short Dark Triad (SD3) Results



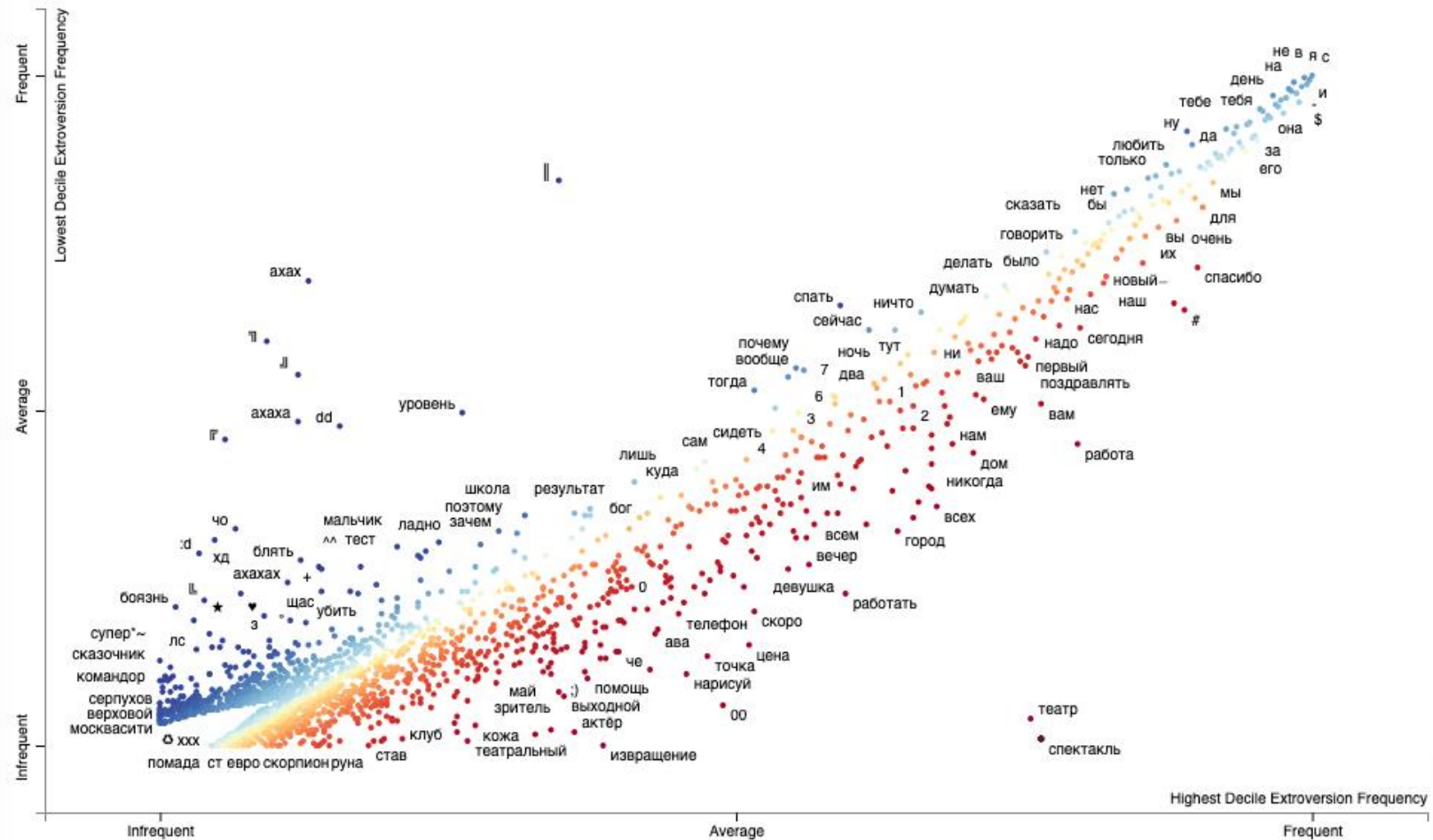
RU-ADEPT Corpus Statistics

Platform	# of Participants	# of Posts	# of Word Tokens
VK Post	1,292	371,290	6,739,054
VK Repost	984	434,979	28,164,325
Live Journal	27	9,854	2,572,032
Blogger	17	888	11,590,630
Moi Mir (My World)	55	5,590	26,359



Lexical Differences (Extraversion)

Scattertext
(Kessler, 2017)
illustrates
differences
between VK
text of low &
high levels of
extraversion



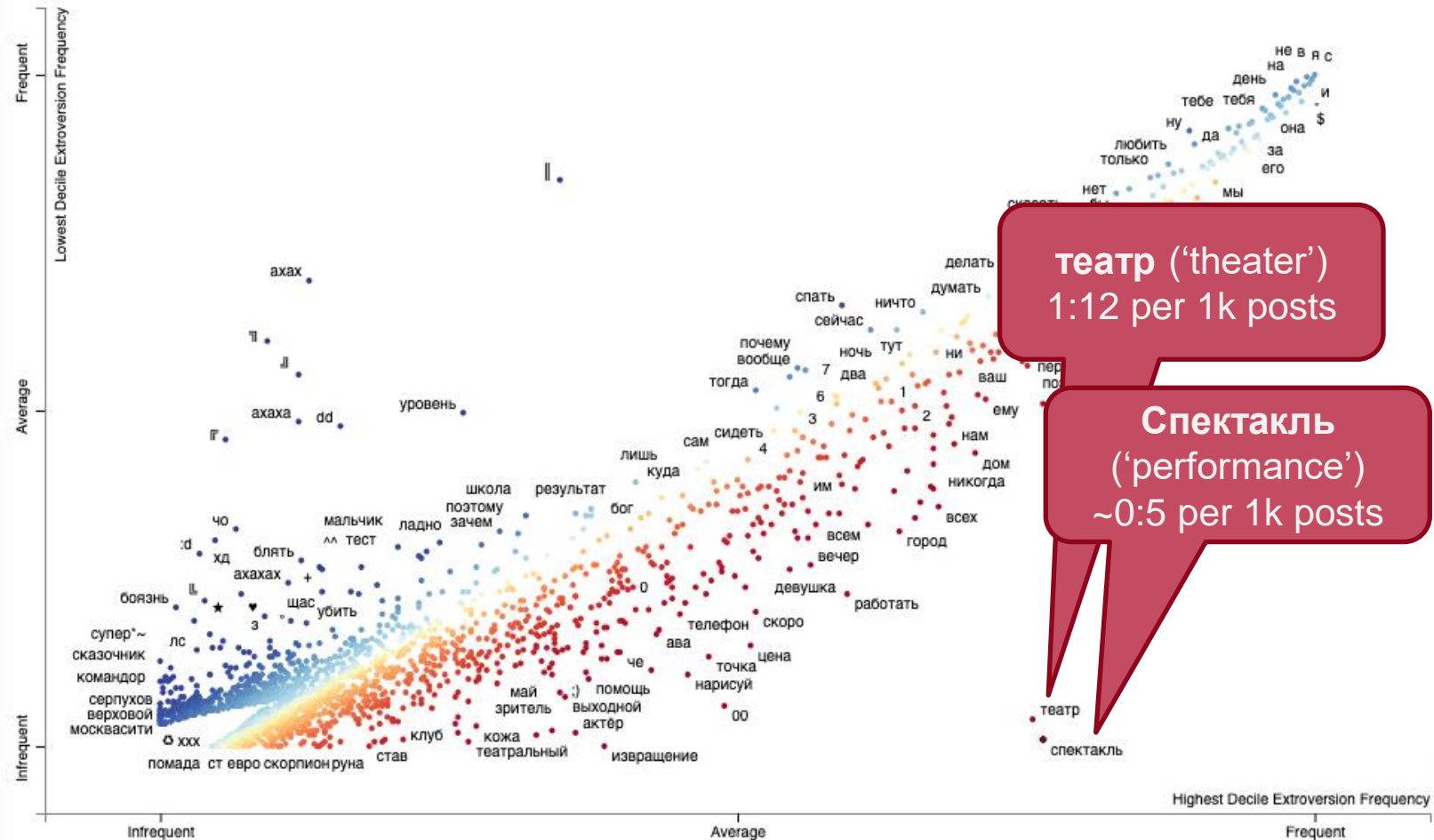
Lexical Differences (Extraversion)

Lowest Decile:

- уровень ('grade', 'level')
- ахах (*laughter*)

Highest Decile:

- театр ('theater')
- спектакль ('performance')



Discussion, Limitations, and Future Work

- Data Collection and Interpretation
 - User-authored vs. “auto-authored” text (see Hull et al. 2021)
 - Limitation: Verifying account IDs for Blogger, Live Journal
- Open Science and Data Privacy
 - Human data creates ethical issues for open science
 - Future work: Pseudonymization for more natural de-identification



Conclusions

- A greater diversity of datasets can make for better social science.

Which textual correlates of personality traits are universal?

Which vary by language, platform, time of collection, etc.?

- RU-ADEPT is a starting place for studying personality traits in a less-Western language.

First large pairing of Big Five + Dark Triad in a Russian language dataset.



How to Get the RU-ADEPT Dataset

- RU-ADEPT is *not* available for direct download
 - Scholars must submit reuse requests for ethics review.
 - Send your request to research_support@umd.edu.
- Please include in the request:
 - institutional and/or funding information,
 - a description of your research plans,
 - what portion of the dataset you need, and
 - any relevant human subjects research review determinations from your internal review board (IRB) or equivalent.



Thank you!

C. Anton Rytting*, Valerie Novak*, James R. Hull*,
Victor M. Frank*, Paul R. Rodrigues†, Jarrett G.W. Lee*, Laurel G. Miller-Sims*

* University of Maryland, College Park, MD, U.S.A.
† Accenture, Washington, D.C., U.S.A

{crying, vnovak, jhull1, vmfrank, jgw1, lmillers}@umd.edu, paul.rodriques@accenture.com

For requests or questions about the corpus, write:

research_support@umd.edu

