**OffensEval: Identifying and Categorizing Offensive Language in Social Media (SemEval 2019 - Task 6)**

This is the website for the OffensEval 2019 shared task organized at SemEval 2019.

**Test Data Released for Task C (29-Jan-2019)**

The testing period for task C has begun! You can download the test set by going to the 'Participate' tab above, then go to the "Download Datasets" section from the menu on the left. The zip file is available by clicking the GREEN "Public Data" button.

You have until 01 Feb (12:00pm) UTC to make your submissions here. Instructions are available in the "Submission Instructions" page.

**Test Data Released for Task A (15-Jan-2019)**

The testing period for task A has begun! You can download the test set by going to the 'Participate' tab above, then go to the "Download Datasets" section from the menu on the left. The zip file is available by clicking the "Public Data" button.

You have 72 hours to make your submissions here. Instructions are available in the "Submission Instructions" page.

**Training Data Release**

The OffensEval training set is now available. You can find it in the tab Participate under Files there is a link Public Data in which you can download the file training-v1.zip.

This file contains 1) the file offenseval-training-v1.tsv with 13,240 training instances, 2) the file readme-trainingset-v1.txt with important information about the training set and practice submissions, and finally 3) the file offenseval-annotation.txt with some information about the annotation.

The practice submissions should be carried out in the PRACTICE CodaLab installation. The link is included in the README file. This official CodaLab competition will be used ONLY for test submissions in January. Do not try to upload practice submissions in the official CodaLab. Use the PRACTICE CodaLab installation instead.

If you have any questions please write to semeval-2019-task-6@googlegroups.com (general mailing list) or semeval-2019-task-6-organizers@googlegroups.com (organizer's mailing list). The organizer responsible for the mailing lists is Preslav Nakov.

**Evaluation**

Please note that the NULL labels in the training set columns sub-task B and sub-task C are just placeholders. This class will not be part of the evaluation. Do not train models to predict this class.

Each sub-task will be evaluated independently. There will be three test sets released (one for each sub-task) and three submission deadlines (one for each sub-task) (see Dates for all the competition dates). Here is how the test sets will look like:

* On 15 Jan 2019 the test data for sub-task A will be released. It will contain all test instances. Systems should categorize instances into OFF and NOT. You will have **72 hours** to upload your predictions.
* On 22 Jan 2019 the test data for sub-task B will be released. It will contain a sub-set of the complete test set excluding instances with gold labels NOT in sub-task A. In sub-task B systems should categorize instances into TIN and UNT. You will have **72 hours** to upload your predictions.
* On 29 Jan 2019, the test data for sub-task C will be released. It contain a sub-set of the complete test set excluding instances with gold labels NOT in sub-task A and instances with gold labels UNT in sub-task B. In this sub-task C systems should categorize instances into IND, GRP, and OTH. You will have **72 hours** to upload your predictions.

**Motivation**

Offensive language is pervasive in social media. Individuals frequently take advantage of the perceived anonymity of computer-mediated communication, using this to engage in behavior that many of them would not consider in real life. Online communities, social media platforms, and technology companies have been investing heavily in ways to cope with offensive language to prevent abusive behavior in social media.

One of the most effective strategies for tackling this problem is to use computational methods to identify offense, aggression, and hate speech in user-generated content (e.g. posts, comments, microblogs, etc.). This topic has attracted significant attention in recent years as evidenced in recent publications (Waseem et al. 2017; Davidson et al., 2017, Malmasi and Zampieri, 2018, Kumar et al. 2018) and workshops such as [ALW](https://sites.google.com/site/abusivelanguageworkshop2017/)and [TRAC](https://sites.google.com/view/trac1/home).

In OffensEval we break down offensive content into three sub-tasks taking the **type** and **target** of offenses into account.

**Sub-tasks**

* Sub-task A - Offensive language identification;
* Sub-task B - Automatic categorization of offense types;
* Sub-task C - Offense target identification.

**Data**

The data is retrieved from social media and distributed in tab separated format. The trial and traininga data are available in the "Participate" tab. Please register to the competition to download the files.

Participants are allowed to use external resources and other datasets for this task. Please indicate which resources were used when submitting your results.

**Dates**

* 28 Nov 2018: Training Data Release
* 15 Jan 2019: Sub-task A test data release (00:00 UTC)
* 17 Jan 2019: Submission deadline sub-task A (23:59 UTC)
* 22 Jan 2019: Sub-task B test data release (00:00 UTC)
* 24 Jan 2019: Submission deadline sub-task B (23:59 UTC)
* 29 Jan 2019: Sub-task C test data release (00:00 UTC)
* 31 Jan 2019: Submission deadline sub-task C (23:59 UTC)
* 5 Feb 2019: Results announced
* 23 Feb 2019: System description paper submissions due
* 29 Mar 2019: Author notifications
* 5 Apr 2019: Camera ready submissions due

The system will be open for **72 hours** during each phase. PS: The dataset in CodaLab for sub-task B are earlier. We will release the data on the 22nd.

**Paper Submission**

To submit your paper, please log in to <https://www.softconf.com/naacl2019/SemEval/> click "make a new submission", then in "submission categories" select "\*\*system description\*\*" as the submission type and your task number as the task.

**Task Organizers**

* Marcos Zampieri (University of Wolverhampton, UK)
* Shervin Malmasi (Harvard Medical School, USA)
* Preslav Nakov (Qatar Computing Research Institute, Qatar)
* Sara Rosenthal (IBM Research, USA)
* Noura Farra (Columbia University, USA)
* Ritesh Kumar (Bhim Rao Ambedkar University, India)

**References**

Davidson, T., Warmsley, D., Macy, M. and Weber, I. (2017) Automated Hate Speech Detection and the Problem of Offensive Language. Proceedings of ICWSM.

Kumar, R., Ojha, A.K., Malmasi, S. and Zampieri, M. (2018) Benchmarking Aggression Identification in Social Media. In Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC). pp. 1-11.

Malmasi, S., Zampieri, M. (2018) Challenges in Discriminating Profanity from Hate Speech. Journal of Experimental & Theoretical Artificial Intelligence. Volume 30, Issue 2, pp. 187-202. Taylor & Francis.

Waseem, Z., Davidson, T., Warmsley, D. and Weber, I. (2017) Understanding Abuse: A Typology of Abusive Language Detection Subtasks. Proceedings of the Abusive Language Online Workshop.