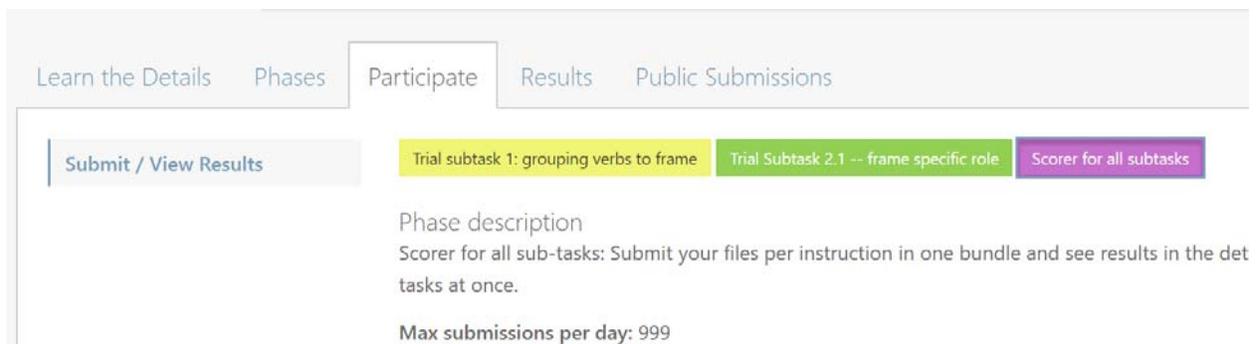


How to use the online scorer?

Prepare your submission file

Put system generated files in one zip file named **system.zip**, and then *zip this zip-file* again! (we need a zip of zip file as an input). Your **system.zip** file can contain up to 3 files, one for each task. These files must be named **as task-1.txt (for sub task 1), task-2.1.txt (for sub task 2.1), and task-2.2.txt (for sub task 2.2)**. Be reminded that choosing other file names prevents the scoring program to function properly. If you do not have a submission file for a subtask, that is fine: You do not need to include any file for it in your zip file. But, if you have results, say for sub-task 1 and sub-task 2.1, then your zip file can contain two files, one for each of the tasks, i.e. your submission is a zip-of(zip-of([task-1.txt , task-2.1.txt])). See attached an example zip file that shows the expected file names (for all tasks) and structure for zip-file submission. **NB:** Submissions for sub-tasks must contain all the records from the test set, otherwise they will be rejected by the scorer. This is a measure to prevent bias and error when computing scores.

Please submit your files to **Scorer for all subtasks** (by clicking on it under the participate tab) shown in image below.



The screenshot shows a web interface with a navigation bar at the top containing five tabs: 'Learn the Details', 'Phases', 'Participate', 'Results', and 'Public Submissions'. The 'Participate' tab is currently selected. Below the navigation bar, there is a 'Submit / View Results' button on the left. To its right, there are three colored buttons: a yellow one for 'Trial subtask 1: grouping verbs to frame', a green one for 'Trial Subtask 2.1 -- frame specific role', and a purple one for 'Scorer for all subtasks'. Below these buttons, the text 'Phase description' is followed by the instruction: 'Scorer for all sub-tasks: Submit your files per instruction in one bundle and see results in the det tasks at once.' At the bottom, it states 'Max submissions per day: 999'.

Once a submission file is submitted (by filling the form and clicking on submit), in a few second a detailed report of the obtained scores can be accessed by expanding the appended entry in the leaderboard table and clicking on the **View detailed results**. The two are shown in the figure below: first expand the entry (marked by 1 in the figure), then click on the link marked by 2 in the figure below.

19	23.67	system (2).zip	12/16/2018 20:52:46	Finished	✓	-
----	-------	----------------	---------------------	----------	---	---

Description:

1

update description

Team name: test

Method name: test

Method description:
testing scorer

Download your submission
View scoring output log
View scoring error log
View predict output log
View predict error log
Download output from prediction step
View detailed results
Download output from scoring step

Leaderboard modifying disabled

Make your submission public

2

If a submission is successful, detailed result are provided in a page such as https://competitions.codalab.org/my/competition/submission/514212/detailed_results/ in which one can see a list of performance measures for the submitted results (shown below, too)

Sub-Task 1 Results								
Instance	#goldClusterNum	#sysClusterNum	purity	inversePurity	pulpuf1	BCubed-Precision	BCubed-Recall	BCubed-f1
41	1	42.33	100	59.48	21.89	100	35.92	

Sub-Task 2.1 Results								
Instance	#goldClusterNum	#sysClusterNum	purity	inversePurity	pulpuf1	BCubed-Precision	BCubed-Recall	BCubed-f1
0	102	1	20.5	100	34.02	9.62	100	17.55

Sub-Task 2.2 Results								
Instance	#goldClusterNum	#sysClusterNum	purity	inversePurity	pulpuf1	BCubed-Precision	BCubed-Recall	BCubed-f1
0	20	1	39.75	100	56.89	29.64	100	45.73

If a submission fails, the best starting point to investigate the reason/s for it is the CodaLab's logs (marked by number 3 in the figure below).

19	23.67	system (2).zip	12/16/2018 20:52:46	Finished	✓	-
----	-------	----------------	---------------------	----------	---	---

Description:

[update description](#)

Team name: test

Method name: test

Method description:
testing scorer

[Download your submission](#)

[View scoring output log](#)

[View scoring error log](#)

[View predict output log](#)

[View predict error log](#)

[Download output from prediction step](#)

[View detailed results](#)

[Download output from scoring step](#)

Leaderboard modifying disallowed

[Make your submission public](#)

3

The two usual suspects for a failed submission are: a) using incorrect file names or structures (e.g., placing files in a directory and then zipping them, forgetting the zip in zip structure and so on); and, b) missing records/items in submission files: the above “view scoring error log” can help to identify these two problems (file not found exceptions vs. record/item mismatch exception); these logs, despite what they may look like, contain helpful information for solving these types of problem.

Last but not least, please note that the leaderboard table shows the average of the performances computed for a given submission across subtasks. As a result, it cannot be used to compare results from one submission to another (in any case the leaderboard won't be public up until the post-evaluation period). For the post-evaluation period, we report results from submissions in a dedicated page with a format similar to the “detailed results” page.