

3.2. Task 2

The predictions for this task were made in form of triplets (protein-paper-GO) plus a piece of evidence text. More than 30.000 of these individual results were submitted and had to be reviewed by the GO curators. Because of lack of time (three curators dedicated several months to the revision of the results) not all results could get checked. In that cases the entire proteins were skipped and not individual participants, so everyone was affected in the same way.

The curator evaluated:

- the GO terms (if they are correctly predicted and the evidence text corresponds to the term)
- the proteins (if the protein appears in the evidence text)
- and they made additional comments

The scheme for both GO and proteins was "high" (meaning that the GO term or the protein were correct), "generally" (for GO terms this means that they are not totally wrong but too general to be really useful for annotation, for protein this means that the specific protein is not there but a homologue from another organism or a reference to the protein family), and "low" (which basically means that the prediction was wrong).

A summary with the following information (per participant) was distributed to all participants:

- The number of proteins that the evaluators checked (evaluated results); low numbers means most likely that the participant did not send more results and not that the results had not been evaluated
- Absolute and relative numbers of "perfect" results. This means that for both GO and protein the evaluation was "high".
- The same numbers for the cases where the protein was correct ("high") and the GO term evaluation was "generally"

3.2.1 Results by users

Results for task 2.1:

user	run	evaluated results	"perfect" predictions	correct protein, "general" GO
user4	1	1048	268 (25.57%)	74 (7.06%)
user5	1	1053	166 (15.76%)	77 (7.31%)
	2	1050	166 (15.81%)	90 (8.57%)
	3	1050	154 (14.67%)	86 (8.19%)
user7	1	1057	272 (25.73%)	154 (14.57%)
	2	1864	43 (2.31%)	40 (2.15%)
	3	1703	66 (3.88%)	40 (2.35%)
user9	1	251	125 (49.80%)	13 (5.18%)
	2	70	33 (47.14%)	5 (7.14%)
	3	89	41 (46.07%)	7 (7.87%)
user10	1	45	36 (80.00%)	3 (6.67%)
	2	59	45 (76.27%)	2 (3.39%)
	3	64	50 (78.12%)	4 (6.25%)

user	run	evaluated results	"perfect" predictions	correct protein, "general" GO
user14	1	1050	303 (28.86%)	69 (6.57%)
user15	1	524	59 (11.26%)	28 (5.34%)
	2	998	125 (12.53%)	69 (6.91%)
user17	1	413	83 (20.10%)	19 (4.60%)
	2	458	7 (1.53%)	(0.00%)
user20	1	1048	301 (28.72%)	57 (5.44%)
	2	1048	280 (26.72%)	60 (5.73%)
	3	1050	239 (22.76%)	59 (5.62%)

Results of task 2.2:

user	run	evaluated results	"perfect" predictions	correct protein, "general" GO
user4	1	661	78 (11.80%)	49 (7.41%)
user7	1	156	1 (0.64%)	1 (0.64%)
	2	384	19 (4.95%)	9 (2.34%)
	3	263	2 (0.76%)	10 (3.80%)
user9	1	28	9 (32.14%)	3 (10.71%)
	2	41	14 (34.15%)	1 (2.44%)
	3	41	14 (34.15%)	1 (2.44%)
user10	1	120	35 (29.17%)	8 (6.67%)
	2	86	24 (27.91%)	6 (6.98%)
	3	116	37 (31.90%)	11 (9.48%)
user15	1	502	3 (0.60%)	8 (1.59%)
	2	485	16 (3.30%)	26 (5.36%)
user17	1	247	52 (21.05%)	23 (9.31%)
	2	55	1 (1.82%)	(0.00%)
	3	99	1 (1.01%)	1 (1.01%)
user20	1	673	20 (2.97%)	30 (4.46%)
	2	672	38 (5.65%)	26 (3.87%)
	3	673	58 (8.62%)	27 (4.01%)

In addition to the summary every participant got back the detailed evaluations for each protein-paper-GO triplet indicating the protein (Swiss-Prot/TREMBL identifier), the paper (directory + name in the way it was distributed), the GO code (GO id), then if that specific record was evaluated by the curators, and the evaluations themselves (GO term, protein and comments).