QALD-6: Evaluation results

http://www.sc.cit-ec.uni-bielefeld.de/qald

Column descriptions

- The first column specifies the system name and the language it processed. Systems marked with a ∗ constitute late submissions (i.e. after the official end of the test phase).

- *Processed* states for how many of the questions the system provided an answer.

- *Recall*, *Precision* and *F-1* report the macro-measures with respect to the number of processed questions.

- *F-1 Global* in addition reports the macro F-1 measure with respect to the total number of questions.

Hybrid question answering

Total: 25

	Processed	Recall	Precision	F-1	F-1 Global
Xser	23	**0.38**	0.43	**0.41**	**0.39**
askDBpedia	21	0.33	0.51	0.40	0.35
FirstRun LIMSI	24	0.04	**0.61**	0.08	0.08

Statistical question answering over RDF datacubes

Total: 50

	Processed	Recall	Precision	F-1	F-1 Global
SPARKLIS (expert)	50	**0.94**	**0.96**	**0.95**	**0.95**
SPARKLIS (beginner)	50	0.76	0.88	0.82	0.82
QA^3	44	0.62	0.59	0.60	0.53
CubeQA	49	0.41	0.49	0.45	0.44

Multilingual question answering over DBpedia

Total: 100

Sorted by global F-1 score (over all questions):

		Processed	Recall	Precision	F-1	F-1 Global
CANaLI	(en)	100	**0.89**	**0.89**	**0.89**	**0.89**
UTQA	(en)	100	0.69	0.82	0.75	0.75
KWGAnswer	(en)	100	0.59	0.85	0.70	0.70
UTQA	(es)	100	0.62	0.76	0.68	0.68
UTQA	(fa)	100	0.61	0.70	0.65	0.65
NbFramework	(en)	63	0.85	0.87	0.86	0.54
SemGraphQA	(en)	100	0.25	0.70	0.37	0.37
PersianQA∗	(fa)	100	0.19	0.91	0.31	0.31
UIQA (with manual)	(en)	44	0.63	0.54	0.58	0.25
UIQA (without manual)	(en)	36	0.53	0.43	0.48	0.17

Sorted by local F-1 score (over processed questions):

		Processed	Recall	Precision	F-1	F-1 Global
CANaLI	(en)	100	**0.89**	**0.89**	**0.89**	**0.89**
NbFramework	(en)	63	0.85	0.87	0.86	0.54
UTQA	(en)	100	0.69	0.82	0.75	0.75
KWGAnswer	(en)	100	0.59	0.85	0.70	0.70
UTQA	(es)	100	0.62	0.76	0.68	0.68
UTQA	(fa)	100	0.61	0.70	0.65	0.65
UIQA (with manual)	(en)	44	0.63	0.54	0.58	0.25
UIQA (without manual)	(en)	36	0.53	0.43	0.48	0.17
SemGraphQA	(en)	100	0.25	0.70	0.37	0.37
PersianQA∗	(fa)	100	0.19	0.91	0.31	0.31