

QA system	Lang	Type	Total	P	R	F	Runtime	Ref
QALD-3								
OUR-QA-SYSTEM	en	full	100	0.64	0.42	0.51	1.01	-
OUR-QA-SYSTEM	en	key	100	0.71	0.37	0.48	0.79	-
OUR-QA-SYSTEM	de	key	100	0.79	0.31	0.45	0.22	-
OUR-QA-SYSTEM	de	full	100	0.79	0.28	0.42	0.30	-
OUR-QA-SYSTEM	fr	key	100	0.83	0.27	0.41	0.26	-
gAnswer [51]*	en	full	100	0.40	0.40	0.40	≈ 1 s	[51]
OUR-QA-SYSTEM	fr	full	100	0.70	0.26	0.38	0.37	-
OUR-QA-SYSTEM	es	full	100	0.77	0.24	0.37	0.27	-
OUR-QA-SYSTEM	it	full	100	0.79	0.23	0.36	0.30	-
OUR-QA-SYSTEM	it	key	100	0.84	0.23	0.36	0.24	-
OUR-QA-SYSTEM	es	key	100	0.80	0.23	0.36	0.23	-
RTV [19]	en	full	99	0.32	0.34	0.33	-	[6]
Intui2 [13]	en	full	99	0.32	0.32	0.32	-	[6]
SINA [37]*	en	full	100	0.32	0.32	0.32	≈ 10-20s	[37]
DEANNA [47]*	en	full	100	0.21	0.21	0.21	≈ 1-50 s	[51]
SWIP [35]	en	full	99	0.16	0.17	0.17	-	[6]
Zhu et al. [50]*	en	full	99	0.38	0.42	0.38	-	[50]
QALD-4								
Xser [46]	en	full	50	0.72	0.71	0.72	-	[42]
OUR-QA-SYSTEM	en	key	50	0.76	0.40	0.52	0.32s	-
OUR-QA-SYSTEM	en	full	50	0.56	0.30	0.39	0.46s	-
gAnswer [51]	en	full	50	0.37	0.37	0.37	0.973 s	[42]
CASIA [24]	en	full	50	0.32	0.40	0.36	-	[42]
OUR-QA-SYSTEM	de	key	50	0.92	0.20	0.33	0.04s	-
OUR-QA-SYSTEM	fr	key	50	0.92	0.20	0.33	0.06s	-
OUR-QA-SYSTEM	it	key	50	0.92	0.20	0.33	0.04s	-
OUR-QA-SYSTEM	es	key	50	0.92	0.20	0.33	0.05s	-
OUR-QA-SYSTEM	de	full	50	0.90	0.20	0.32	0.06s	-
OUR-QA-SYSTEM	it	full	50	0.92	0.20	0.32	0.16s	-
OUR-QA-SYSTEM	es	full	50	0.90	0.20	0.32	0.06s	-
OUR-QA-SYSTEM	fr	full	50	0.86	0.18	0.29	0.09s	-
Intui3 [14]	en	full	50	0.23	0.25	0.24	-	[42]
ISOFT [33]	en	full	50	0.21	0.26	0.23	-	[42]
Hakimov [23]*	en	full	50	0.52	0.13	0.21	-	[23]
QALD-5								
Xser [46]	en	full	50	0.74	0.72	0.73	-	[43]
UTQA [34]	en	full	50	-	-	0.65	-	[34]
UTQA [34]	es	full	50	0.55	0.53	0.54	-	[34]
UTQA [34]	fs	full	50	0.53	0.51	0.52	-	[34]
OUR-QA-SYSTEM	en	full	50	0.56	0.41	0.47	0.62s	-
OUR-QA-SYSTEM	en	key	50	0.60	0.27	0.37	0.50s	-
AskNow[15]	en	full	50	0.32	0.34	0.33	-	[15]
QAnswer[36]	en	full	50	0.34	0.26	0.29	-	[43]
OUR-QA-SYSTEM	de	full	50	0.92	0.16	0.28	0.20s	-
OUR-QA-SYSTEM	de	key	50	0.90	0.16	0.28	0.19s	-
OUR-QA-SYSTEM	fr	full	50	0.90	0.16	0.28	0.19s	-
OUR-QA-SYSTEM	fr	key	50	0.90	0.16	0.28	0.18s	-
OUR-QA-SYSTEM	it	full	50	0.88	0.18	0.30	0.20s	-
OUR-QA-SYSTEM	it	key	50	0.90	0.16	0.28	0.18s	-
OUR-QA-SYSTEM	es	full	50	0.88	0.14	0.25	0.20s	-
OUR-QA-SYSTEM	es	key	50	0.90	0.14	0.25	0.20s	-
SemGraphQA[2]	en	full	50	0.19	0.20	0.20	-	[43]
YodaQA[1]	en	full	50	0.18	0.17	0.18	-	[43]
QuerioDali[27]	en	full	50	?	?	?	?	[27]
QALD-6								
OUR-QA-SYSTEM	en	full	100	0.55	0.34	0.42	1.28s	-
OUR-QA-SYSTEM	de	full	100	0.73	0.29	0.41	0.41s	-
OUR-QA-SYSTEM	de	key	100	0.85	0.27	0.41	0.30s	-
OUR-QA-SYSTEM	en	key	100	0.51	0.30	0.37	1.00s	-
SemGraphQA [2]	en	full	100	0.70	0.25	0.37	-	[44]
OUR-QA-SYSTEM	fr	key	100	0.78	0.23	0.36	0.34s	-
OUR-QA-SYSTEM	fr	full	100	0.57	0.22	0.32	0.46s	-
OUR-QA-SYSTEM	es	full	100	0.69	0.19	0.30	0.45s	-
OUR-QA-SYSTEM	es	key	100	0.83	0.18	0.30	0.35s	-
OUR-QA-SYSTEM	it	key	100	0.75	0.17	0.28	0.34s	-
AMUSE [22]	en	full	100	-	-	0.26	-	[22]
OUR-QA-SYSTEM	it	full	100	0.62	0.15	0.24	0.43s	-
AMUSE [22]	es	full	100	-	-	0.20	-	[22]
AMUSE [22]	de	full	100	-	-	0.16	-	[22]

QA system	Lang	Type	Total	P	R	F	Runtime	Ref
QALD-7								
OUR-QA-SYSTEM	en	full	100	0.25	0.28	0.25	1.24s	-
OUR-QA-SYSTEM	fr	key	100	0.18	0.16	0.16	0.32s	-
OUR-QA-SYSTEM	en	key	100	0.14	0.16	0.14	0.88s	-
OUR-QA-SYSTEM	de	full	100	0.10	0.10	0.10	0.34s	-
OUR-QA-SYSTEM	de	key	100	0.12	0.10	0.10	0.28s	-
OUR-QA-SYSTEM	fr	full	100	0.12	0.10	0.10	0.42s	-
OUR-QA-SYSTEM	it	key	100	0.06	0.06	0.06	0.28s	-
OUR-QA-SYSTEM	it	full	100	0.04	0.04	0.04	0.34s	-

Table 3: This table (left column and upper right column) summarizes the results obtained by the QA systems evaluated with QALD-3 (over DBpedia 3.8), QALD-4 (over DBpedia 3.9), QALD-5 (over DBpedia 2014), QALD-6 (over DBpedia 2015-10), QALD-7 (2016-04). We indicated with “*” the systems that did not participate directly in the challenges, but were evaluated on the same benchmark afterwards. We indicate the average running times of a query for the systems where we found them. Even if the runtime evaluations were executed on different hardware, it still helps to give an idea about the scalability.

QA System	Lang	Type	Total	P	R	F	Runtime	Ref
QALD-7 task 4, training dataset								
OUR-QA-SYSTEM	en	full	100	0.37	0.39	0.37	1.68s	-
OUR-QA-SYSTEM	en	key	100	0.35	0.38	0.35	0.80s	-
OUR-QA-SYSTEM	es	key	100	0.31	0.32	0.31	0.45s	-
Sorokin et al. [39]	en	full	100	-	-	0.29	-	[39]
OUR-QA-SYSTEM	de	key	100	0.27	0.28	0.27	1.13s	-
OUR-QA-SYSTEM	fr	key	100	0.27	0.30	0.27	1.14s	-
OUR-QA-SYSTEM	fr	full	100	0.27	0.31	0.27	1.05s	-
OUR-QA-SYSTEM	es	full	100	0.24	0.26	0.24	0.65s	-
OUR-QA-SYSTEM	de	full	100	0.18	0.20	0.18	0.82s	-
OUR-QA-SYSTEM	it	full	100	0.19	0.20	0.18	1.00s	-
OUR-QA-SYSTEM	it	key	100	0.17	0.18	0.16	0.44s	-

Table 4: The table shows the results of OUR-QA-SYSTEM over the QALD-7 task 4 training dataset. We used Wikidata (dated 2016-11-28).

QA System	Lang	Type	Total	Accuracy	Runtime	Ref
OUR-QA-SYSTEM*	en	full	21687	0.571	2.1 s	-
Dai et al.*	en	full	21687	0.626	-	[7]
Bordes et al.	en	full	21687	0.627	-	[4]
Yin et al.	en	full	21687	0.683	-	[48]
Golub and He	en	full	21687	0.709	-	[21]
Lukovnikov et al.	en	full	21687	0.712	-	[29]

Table 5: This table summarizes the QA systems evaluated over SimpleQuestions. Every system was evaluated over FB2M except the ones marked with (*) which were evaluated over FB5M.

Benchmark	Lang	Type	Total	P	R	F	Runtime
LC-QuAD	en	full	5000	0.59	0.38	0.46	1.5 s
WDAquaCore0Questions	mixed	mixed	689	0.79	0.46	0.59	1.3 s

Table 6: This table summarizes the results of OUR-QA-SYSTEM over some newly appeared benchmarks.

Dataset	Lang	Type	Total	P	R	F	Runtime
DBpedia	en	full	100	0.55	0.34	0.42	1.37 s
All KBs supported	en	full	100	0.49	0.39	0.43	11.94s

Table 7: Comparison on QALD-6 when querying only DBpedia and multiple KBs at the same time.