

(a) A population P w.r.t "UC Berkeley"

	Title	Conference	Year	Citation	University
t_1	CrowdDB: answering queries with crowdsourcing	SIGMOD	2011	120	UC Berkeley
t_2	Opinion space: a scalable tool for browsing online comments	CHI	2010	30	UC Berkeley
t_3	YFilter: efficient and scalable filtering of XML documents	ICDE		17	UC Berkeley
t_4	Online Aggregation	SIGMOD		100	UC Berkeley
t_5	Pay-as-you-go user feedback for dataspace systems	SIGMOD	2008	107	UC Berkeley
t_6	CrowdER: crowdsourcing entity resolution	VLDB	2012	0	UC Berkeley
t_7	Online aggregation and continuous query support in MapReduce	SIGMOD	2010	44	UC Berkeley
...
t_{10000}	Efficient and scalable filtering of XML documents	ICDE	2003	24	UC Berkeley

(b) The cleaning results for P w.r.t **avg**_{citation}(.)

	Correct(t)	Remove(t, P)	Dedup(t, P)
t_1	144	False	{t_1, t_{123}}
t_2	30	False	{ t_2 }
t_3	298	False	{t_3, t_{360}, t_{10000}}
t_4	695	True	{t_4, t_{128}, t_{6780}}
t_5	107	False	{ t_5 }
t_6	34	False	{ t_6 }
t_7	44	False	{ t_7 }
...
t_{10000}	298	False	{t_3, t_{360}, t_{10000}}