Report Types	Partners	Beth-Israel Deaconess Med Center (BIDMC)	University of Pittsburgh Med Center (UPMC)		Total
	Discharge Summaries	Discharge Summaries	Discharge Summaries (UPMCD)	Patient Notes (UPMCP)	
Training	97	73	98	81	349
Testing	133	123	102	119	477

Concept Type	Count in Partners		Count in BIDMC		Count in UPMCD		Count in UPMCP	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Tests	1571	3258	3036	5967	1217	1357	1544	2317
Treatments	1768	3784	3073	5560	2308	2118	1347	2098
Problems	2885	5545	4187	7047	2728	2653	2167	3305

Assertion Type	Count in Partners		Count in BIDMC		Count in UPMD		Count in UPMP	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Absent	658	1115	938	1479	378	408	561	607
Associated with someone else	31	45	57	86	3	13	0	1
Conditional	38	74	35	74	16	16	14	7
Hypothetical	165	156	217	289	242	257	26	15
Present	1901	3887	2712	4735	1974	1855	1452	2548
Possible	91	268	218	384	114	104	112	127

Relation Type	Count in Partners		Count in BIDMC		Count in UPMD		Count in UPMP	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
PIP	236	638	519	810	319	293	165	245
TeCP	68	171	98	167	82	70	55	180
TeRP	429	1132	564	928	398	402	343	571
TrAP	302	773	583	959	349	319	189	436
TrCP	67	179	117	163	80	61	32	41
TrIP	17	60	34	92	43	38	13	8
TrNAP	21	44	41	68	23	30	21	49
TrWP	7	60	17	49	23	22	9	12

Table 1: Number of reports and samples in each class of concepts, assertions, and relations in the 2010 i2b2/VA challenge data.

	Jiang et al.	Kang et al.	Gurulingappa et al.	Patrick et al.	Torii & Liu	Jonnalagadda & Gonzalez	Sasaki et al.	Roberts et al.	Pai et al.
deBruijn et al.	*	*	*	*	*	*	*	*	*
Jiang et al.		*	*	*	*	*	*	*	*
Kang et al.			NS	NS	*	*	*	*	*
Gurulingappa et al.				NS	*	*	*	*	*
Patrick et al.					*	*	*	*	*
Torii & Liu						NS	*	*	*
Jonnalagadda & Gonzalez							*	*	*
Sasaki et al.								*	*
Roberts et al.									*

Table 3: Statistical significance of exact F-measures in the concept extraction task. * indicates that the system along the row is significantly different from the system along the column at p= 0.05. NS means "not significant difference" at p=0.05. Only the upper diagonal is marked.

System	Medical Problems			Treatments			Tests		
	R	P	F	R	P	F	R	P	F
deBruijn et al.	0.847	0.868	0.857	0.825	0.862	0.843	0.834	0.877	0.855
Jiang et al.	0.834	0.861	0.847	0.800	0.859	0.828	0.801	0.881	0.839
Kang et al.	0.831	0.811	0.821	0.792	0.833	0.812	0.801	0.862	0.830
Gurulingappa et al.	0.807	0.839	0.823	0.771	0.842	0.805	0.799	0.853	0.826
Patrick et al.	0.799	0.835	0.817	0.775	0.856	0.814	0.789	0.862	0.824
Torii and Liu	0.784	0.860	0.821	0.758	0.859	0.806	0.761	0.863	0.809
Jonnalagadda and Gonzalez	0.792	0.824	0.808	0.776	0.832	0.803	0.791	0.844	0.817
Sasaki et al.	0.786	0.814	0.799	0.738	0.834	0.783	0.790	0.863	0.825
Roberts et al.	0.787	0.813	0.800	0.754	0.807	0.779	0.804	0.810	0.807
Pai et al.	0.776	0.805	0.790	0.738	0.805	0.770	0.770	0.836	0.802

Table 4: Exact metrics.

System	Medical Problems			Treatments			Tests		
	R	P	F	R	P	F	R	P	F
deBruijn et al.	0.915	0.938	0.926	0.887	0.927	0.906	0.884	0.930	0.906
Jiang et al.	0.905	0.934	0.919	0.863	0.927	0.894	0.848	0.933	0.888
Kang et al.	0.909	0.887	0.898	0.861	0.906	0.883	0.854	0.919	0.886
Gurulingappa et al.	0.890	0.925	0.907	0.840	0.918	0.877	0.859	0.917	0.887
Patrick et al.	0.878	0.917	0.897	0.834	0.921	0.875	0.839	0.916	0.876
Torii and Liu	0.871	0.955	0.911	0.831	0.942	0.883	0.828	0.939	0.880
Jonnalagadda and Gonzalez	0.880	0.914	0.897	0.849	0.910	0.878	0.856	0.913	0.884
Sasaki et al.	0.866	0.897	0.881	0.799	0.903	0.848	0.838	0.915	0.875
Roberts et al.	0.877	0.906	0.891	0.832	0.890	0.860	0.872	0.879	0.876
Pai et al.	0.870	0.902	0.886	0.819	0.894	0.855	0.826	0.897	0.860

Table 5: Inexact metrics.

	Clark et al	Demner-Fushman et al.	Jiang et al.	Grouin et al.	Divita et al.	Cohen et al.	Roberts et al.	Anick et al.	Chang et al.
deBruijn et al.	NS	NS	NS	*	*	*	*	*	*
Clark et al.		NS	NS	NS	NS	*	*	*	*
Demner-Fushman et al.			NS	NS	NS	*	*	*	*
Jiang et al.				NS	NS	NS	NS	*	*
Grouin et al.					NS	NS	NS	*	*
Divita et al.						NS	NS	*	*
Cohen et al.							NS	NS	*
Roberts et al.								NS	*
Anick et al.									NS

Table 7: Statistical significance of F-measures in the assertion classification task. * indicates that the system along the row is significantly different from the system along the column at p= 0.05. NS means "not significant difference" at p=0.05. Only the upper diagonal is marked.

	deBruijn et al.	Grouin et al.	Patrick et al.	Jonnalagadda & Gonzalez et al.	Divita et al.	Solt et al.	Demner-Fushman et al.	Anick et al.	Cohen et al.
Roberts et al.	NS	*	*	*	*	*	*	*	*
deBruijn et al.		*	*	*	*	*	*	*	*
Grouin et al.			NS	NS	*	*	*	*	*
Patrick et al.				NS	NS	*	*	*	*
Jonnalagadda & Gonzalez					NS	*	*	*	*
Divita et al.						*	*	*	*
Solt et al.							NS	NS	*
Demner-Fushman et al.								NS	NS
Anick et al.									NS

Table 8: Statistical significance of F-measures in the relation classification task. * indicates that the system along the row is significantly different from the system along the column at p= 0.05. NS means "not significant difference" at p=0.05. Only the upper diagonal is marked.

System	Present			Absent			Possible			Conditional			Hypothetical			Associated with someone else		
	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F
deBruijn et al.	0.981	0.939	0.959	0.934	0.951	0.942	0.530	0.818	0.643	0.152	0.963	0.263	0.861	0.909	0.884	0.724	0.955	0.824
Clark et al.	0.980	0.937	0.958	0.920	0.955	0.937	0.532	0.772	0.630	0.287	0.803	0.422	0.859	0.924	0.890	0.779	0.983	0.869
Demner-Fushman et al.	0.983	0.932	0.957	0.923	0.958	0.940	0.509	0.815	0.626	0.257	0.759	0.384	0.792	0.937	0.859	0.766	0.917	0.835
Jiang et al.	0.980	0.933	0.956	0.917	0.959	0.938	0.529	0.792	0.634	0.269	0.780	0.400	0.798	0.885	0.839	0.779	0.942	0.853
Grouin et al.	0.970	0.942	0.956	0.947	0.931	0.939	0.538	0.738	0.622	0.240	0.746	0.363	0.830	0.928	0.876	0.779	0.856	0.816
Divita et al.	0.979	0.931	0.954	0.930	0.943	0.936	0.453	0.800	0.579	0.222	0.905	0.357	0.830	0.928	0.876	0.786	1.000	0.880
Cohen et al.	0.975	0.934	0.954	0.926	0.944	0.935	0.514	0.759	0.613	0.298	0.850	0.442	0.816	0.930	0.869	0.510	0.673	0.580
Roberts et al.	0.978	0.934	0.955	0.908	0.950	0.929	0.519	0.769	0.619	0.263	0.672	0.378	0.801	0.868	0.833	0.738	0.856	0.793
Anick et al.	0.964	0.940	0.952	0.934	0.935	0.935	0.515	0.696	0.592	0.216	0.487	0.300	0.824	0.810	0.817	0.793	0.913	0.849
Chang et al.	0.966	0.933	0.949	0.924	0.937	0.930	0.579	0.753	0.654	0.252	0.494	0.333	0.707	0.842	0.769	0.807	0.813	0.810

Table 9: Per class results on assertion classification .

System	Medical Problem—Treatment Relations			Medical Problem—Test Relations			Medical Problem—Medical Problem Relations		
	R	P	F	R	P	F	R	P	F
Roberts et al.	0.686	0.672	0.679	0.833	0.798	0.815	0.726	0.664	0.694
deBruijn et al.	0.583	0.750	0.656	0.789	0.843	0.815	0.712	0.691	0.701
Grouin et al.	0.646	0.647	0.647	0.801	0.792	0.797	0.645	0.670	0.657
Patrick et al.	0.599	0.671	0.633	0.774	0.813	0.793	0.627	0.677	0.651
Jonnalagadda and Gonzalez	0.679	0.581	0.626	0.828	0.765	0.795	0.730	0.586	0.650
Divita et al.	0.582	0.704	0.637	0.782	0.794	0.788	0.534	0.710	0.610
Solt et al.	0.629	0.621	0.625	0.779	0.801	0.790	0.711	0.469	0.565
Demner-Fushman et al.	0.612	0.642	0.626	0.677	0.835	0.748	0.533	0.662	0.591
Anick et al.	0.619	0.596	0.608	0.787	0.744	0.765	0.502	0.631	0.559
Cohen et al.	0.578	0.606	0.591	0.781	0.750	0.765	0.492	0.627	0.552

Table 10: Per relation type results on relation classification.

System	TrIP			TrWP			TrCP			TrAP			TrNAP		
	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F
Roberts et al.	0.298	0.562	0.389	0.035	0.278	0.062	0.565	0.542	0.554	0.814	0.707	0.757	0.199	0.432	0.272
deBruijn et al.	0.177	0.833	0.292	0.000	NaN	0.000	0.327	0.747	0.455	0.730	0.748	0.739	0.126	0.774	0.216
Grouin et al.	0.414	0.458	0.435	0.168	0.774	0.276	0.435	0.550	0.486	0.760	0.676	0.715	0.251	0.495	0.333
Patrick et al.	0.157	0.861	0.265	0.028	0.800	0.054	0.480	0.495	0.487	0.725	0.699	0.712	0.131	0.556	0.212
Jonnalagadda & Gonzalez	0.207	0.612	0.309	0.007	0.200	0.014	0.457	0.537	0.494	0.835	0.589	0.691	0.147	0.400	0.215
Divita et al.	0.197	0.780	0.315	0.035	0.833	0.067	0.367	0.715	0.485	0.719	0.702	0.710	0.105	0.690	0.182
Solt et al.	0.313	0.591	0.409	0.056	0.667	0.103	0.493	0.389	0.435	0.743	0.685	0.713	0.220	0.316	0.259
Demner-Fushman et al.	0.369	0.635	0.467	0.126	0.346	0.185	0.491	0.536	0.512	0.712	0.675	0.693	0.199	0.376	0.260
Anick et al.	0.237	0.528	0.328	0.014	0.182	0.026	0.561	0.442	0.495	0.731	0.632	0.678	0.157	0.517	0.241
Cohen et al.	0.096	0.576	0.165	0.007	0.200	0.014	0.356	0.608	0.449	0.729	0.608	0.663	0.052	0.435	0.094

Table 11: Per class results on medical problem—treatment relations.

System	TeRP			TeCP		
	R	P	F	R	P	F
Roberts et al.	0.906	0.825	0.864	0.456	0.594	0.516
deBruijn et al.	0.880	0.842	0.861	0.316	0.857	0.462
Grouin et al.	0.881	0.813	0.846	0.391	0.612	0.477
Patrick et al.	0.840	0.840	0.840	0.430	0.614	0.506
Jonnalagadda and Gonzalez	0.911	0.784	0.843	0.400	0.596	0.479
Divita et al.	0.886	0.793	0.837	0.245	0.818	0.377
Solt et al.	0.826	0.842	0.834	0.536	0.577	0.556
Demner-Fushman et al.	0.733	0.872	0.796	0.393	0.594	0.473
Anick et al.	0.848	0.765	0.804	0.475	0.597	0.529
Cohen et al.	0.861	0.766	0.810	0.369	0.599	0.457

Table 12: Per class results on medical problem—test relations.

System	PIP		
	R	P	F
Roberts et al.	0.726	0.664	0.694
deBruijn et al.	0.712	0.691	0.701
Grouin et al.	0.645	0.670	0.657
Patrick et al.	0.627	0.677	0.651
Jonnalagadda and Gonzalez	0.730	0.586	0.650
Divita et al.	0.534	0.710	0.610
Solt et al.	0.711	0.469	0.565
Demner-Fushman et al.	0.533	0.662	0.591
Anick et al.	0.502	0.631	0.559
Cohen et al.	0.492	0.627	0.552

Table 13: Per class results on medical problem—medical problem relations.