

Appendix: Measuring Actual Privacy of Obfuscated Queries in Information Retrieval

Francesco Luigi De Faveri 

Guglielmo Faggioli 

Nicola Ferro 

1 Appendix: Formal Similarity Analysis

In this Appendix, we report Table 1, the complete version of the similarity table in the paper submitted. The results show the complete formal privacy analysis of the privacy provided to the queries of the collection studied. In addition, the Table reported here demonstrates the need for a more comprehensive privacy analysis of the obfuscated texts, not limiting such analysis to a formal parameters study but also towards an actual one.

Table 1: Average Lexical and Sentence similarity measured using the Jaccard between queries’ terms and Cosine similarity of the contextual embeddings computed by the Sentence-BERT MiniLM model. The mechanisms are organised by the obfuscation strategy they use, i.e., *Sampling*, *Embedding Perturbation*, and *Heuristics*.

	Obfuscation Strategy	Mechanism	Lexical Similarity (<i>Jaccard Similarity</i>)										Sentence Similarity (<i>Cosine Similarity</i>)												
													ϵ - Privacy Budget												
			1	5	10	12.5	15	17.5	20	25	30	50	1	5	10	12.5	15	17.5	20	25	30	50			
Deep Learning '19	Sampling	CusText	0.093	0.390	0.809	0.881	0.910	0.922	0.926	0.927	0.927	0.927	-	0.331	0.587	0.836	0.870	0.883	0.888	0.889	0.888	0.888	0.889	-	
		SanText	0.000	0.926	0.927	0.927	0.927	0.927	0.927	0.927	0.927	0.927	-	0.021	0.890	0.890	0.889	0.889	0.890	0.889	0.892	0.893	0.896	-	
		TEM	0.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-	0.024	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-	
		WBB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-	0.463	0.466	0.464	0.462	0.465	0.464	0.465	0.466	0.462	0.467	-	
	Embedding Perturbation	CMP	0.000	0.002	0.104	0.279	0.503	0.700	0.813	0.901	0.921	0.927	-	0.021	0.031	0.200	0.420	0.639	0.782	0.848	0.882	0.888	0.890	-	
		Mhl	0.000	0.002	0.054	0.141	0.281	0.459	0.611	0.804	0.885	0.927	-	0.022	0.029	0.115	0.242	0.409	0.589	0.722	0.842	0.876	0.891	-	
		VickreyCMP	0.000	0.002	0.046	0.100	0.140	0.157	0.163	0.169	0.173	0.192	-	0.021	0.029	0.123	0.214	0.297	0.361	0.406	0.466	0.496	0.553	-	
		VickreyMhl	0.000	0.002	0.026	0.060	0.098	0.126	0.145	0.164	0.172	0.193	-	0.023	0.026	0.075	0.132	0.208	0.273	0.324	0.407	0.460	0.538	-	
	SoTA	AEA	-	-	-	-	-	-	-	-	-	-	0.332	-	-	-	-	-	-	-	-	-	-	0.479	
		FEA	-	-	-	-	-	-	-	-	-	-	0.001	-	-	-	-	-	-	-	-	-	-	0.172	
	Deep Learning '20	Sampling	CusText	0.094	0.384	0.823	0.900	0.929	0.939	0.941	0.942	0.942	0.942	-	0.356	0.623	0.887	0.926	0.937	0.942	0.942	0.943	0.943	0.943	-
			SanText	0.000	0.939	0.942	0.942	0.942	0.942	0.942	0.942	0.942	0.942	-	0.024	0.936	0.938	0.938	0.938	0.939	0.938	0.938	0.937	0.937	-
TEM			0.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-	0.025	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-	
WBB			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-	0.419	0.418	0.419	0.419	0.417	0.422	0.421	0.421	0.418	0.421	-	
Embedding Perturbation		CMP	0.000	0.003	0.098	0.271	0.490	0.686	0.809	0.911	0.935	0.942	-	0.023	0.037	0.210	0.438	0.664	0.816	0.890	0.931	0.937	0.938	-	
		Mhl	0.000	0.003	0.053	0.140	0.271	0.436	0.591	0.802	0.889	0.941	-	0.023	0.034	0.124	0.256	0.433	0.609	0.746	0.886	0.921	0.939	-	
		VickreyCMP	0.000	0.001	0.047	0.102	0.137	0.158	0.164	0.167	0.167	0.185	-	0.022	0.030	0.138	0.247	0.329	0.403	0.45	0.502	0.532	0.584	-	
		VickreyMhl	0.000	0.002	0.027	0.061	0.099	0.127	0.147	0.166	0.168	0.188	-	0.022	0.031	0.087	0.159	0.240	0.307	0.363	0.446	0.492	0.569	-	
SoTA		AEA	-	-	-	-	-	-	-	-	-	-	0.319	-	-	-	-	-	-	-	-	-	-	0.450	
		FEA	-	-	-	-	-	-	-	-	-	-	0.001	-	-	-	-	-	-	-	-	-	-	0.148	
Robust '04		Sampling	CusText	0.103	0.407	0.810	0.873	0.896	0.904	0.906	0.908	0.909	0.909	-	0.486	0.691	0.888	0.912	0.921	0.923	0.924	0.926	0.926	0.924	-
			SanText	0.000	0.908	0.909	0.909	0.909	0.909	0.909	0.909	0.909	0.909	-	0.076	0.924	0.925	0.926	0.926	0.926	0.927	0.927	0.928	0.928	-
	TEM		0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-	0.074	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-	
	WBB		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-	0.509	0.511	0.510	0.509	0.510	0.509	0.509	0.509	0.509	0.509	-	
	Embedding Perturbation	CMP	0.000	0.006	0.202	0.465	0.707	0.839	0.887	0.906	0.908	0.909	-	0.072	0.095	0.362	0.617	0.806	0.891	0.917	0.925	0.925	0.925	-	
		Mhl	0.000	0.004	0.095	0.242	0.448	0.651	0.789	0.889	0.904	0.909	-	0.075	0.090	0.228	0.399	0.596	0.719	0.855	0.917	0.924	0.926	-	
		VickreyCMP	0.000	0.004	0.085	0.144	0.176	0.183	0.185	0.189	0.195	0.220	-	0.072	0.092	0.255	0.377	0.475	0.537	0.581	0.635	0.664	0.706	-	
		VickreyMhl	0.000	0.003	0.044	0.092	0.139	0.167	0.180	0.190	0.196	0.222	-	0.074	0.088	0.175	0.266	0.364	0.445	0.504	0.585	0.632	0.700	-	
	SoTA	AEA	-	-	-	-	-	-	-	-	-	-	0.104	-	-	-	-	-	-	-	-	-	-	0.453	
		FEA	-	-	-	-	-	-	-	-	-	-	0.000	-	-	-	-	-	-	-	-	-	-	0.239	