pyPANTERA: A Python PAckage for Natural language obfuscaTion Enforcing pRivacy & Anonymization

CIKM 2024 (October 21 – 25) - Boise, Idaho, USA

Francesco L. De Faveri, Guglielmo Faggioli, and Nicola Ferro

Department of Information Engineering, University of Padova, Italy

Intelligent Interactive Information Access Hub

Introduction

Textual data which contains ***private user information*** are everywhere.

Social Networks

Medical Records

Search Engines

Privacy Problem

How do we provide **privacy** to textual data?

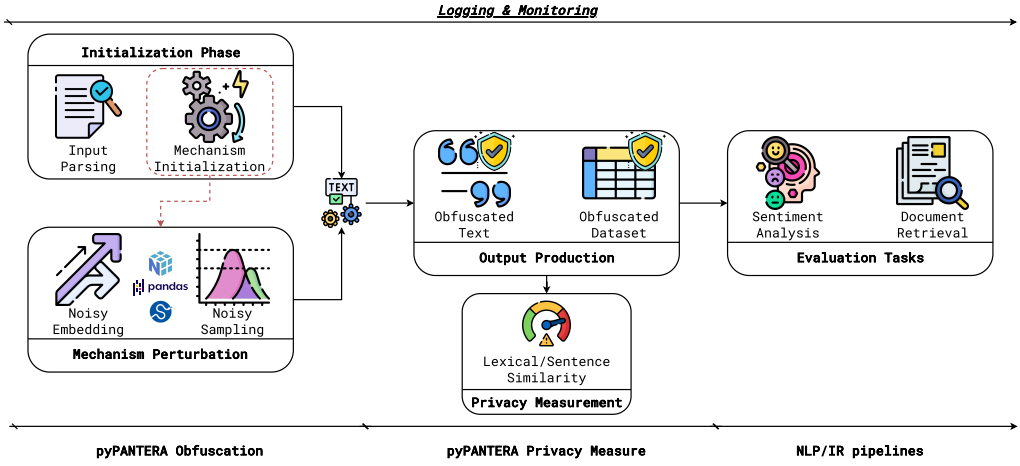
Research Gap

The State-of-the-Art framework for privacy is ε-Differential Privacy (DP).

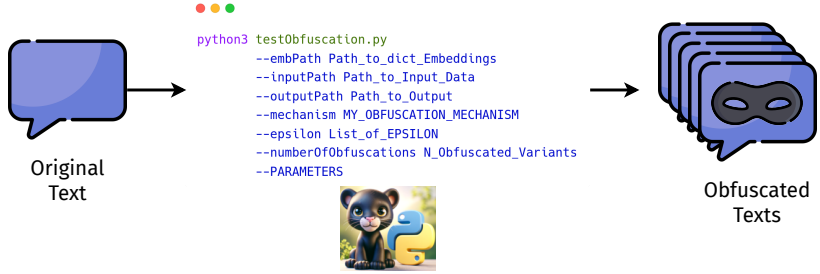
However, textual obfuscation mechanisms:

- Studied for **different** datasets and tasks.
- **No** unified framework for *implementing*, *studing* and *testing* new DP mechanisms.

Pipeline of pyPANTERA



pyPANTERA 1.0.1



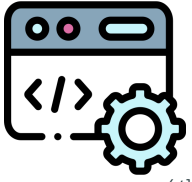
ε-DP obfuscation Mechanisms available in pyPANTERA:

- *Embedding Perturbation*: Cumulative Multiv. Perturbation (**CMP**) [3], Mahalanobis (**Mhl**) [4], **VickreyCMP/Mhl** [5].
- *Sampling Obfuscation*: Customized Text (**CusText**) [2], Sanitization Text (**SanText**) [6], and Truncated Exponential Mechanism (**TEM**) [1].

Development Insights

The entire package is developed using Python (Version 3.10).

- Input can be given using the **CLI** (checked and validated).
- Computational efficiency: **multiprocessing** on available CPUs.
- Constant monitoring of operations via **logs** information.



Experiments - Tasks & Setup

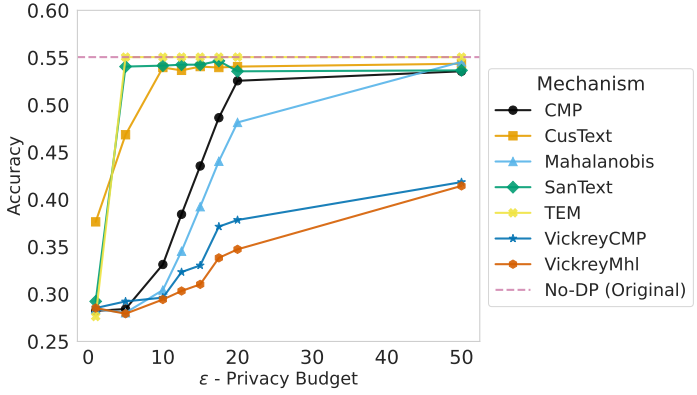
Tasks:

- **Sentiment Analysis** on Tweets (Kaggle Twitter dataset).
- **Documents Retrieval** with obfuscated queries (DL'19 passage corpus).
- **Privacy Evaluation** of Lexical and Semantic Similarities between texts.

Setup:

- Conda environment and code in the pyPANTERA **GitHub Repository**!

Experiments Results i



Evaluation Task: Sentiment Analysis

Experiments Results ii

		Semantic Similarity (*MiniLM*)							
		ε - Privacy Budget							
Perturbation	**Mechanism**	1.0	5.0	10.0	12.5	15.0	17.5	20.0	50.0
Embedding	CMP	0.025	0.037	0.225	0.429	0.628	0.836	0.847	0.902
	Mahalanobis	0.016	0.027	0.088	0.242	0.435	0.587	0.730	0.908
	VickreyCMP	0.018	0.045	0.103	0.169	0.348	0.382	0.435	0.596
	VickreyMhl	0.037	0.030	0.078	0.109	0.202	0.264	0.303	0.498
Sampling	CusText	0.357	0.627	0.881	0.900	0.908	0.908	0.909	0.910
	SanText	0.031	0.902	0.906	0.910	0.917	0.900	0.902	0.907
	TEM	0.037	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Evaluation Task: Privacy Measurement

Conclusion & Future Work

pyPANTERA offers an ***open-source***, ***unified*** and ***flexible*** framework for text obfuscation.

- **User-friendly** Web Application.
- **New DP Mechanism just in!** Words Blending Boxes mechanism (**WBB**)[1].
- **Extend** the framework to non-DP mechanisms.
- **Efficiency** optimization for large datasets.

[1] Minor Revision in the Information Sciences Journal

CIKM 2024 BOISE OCTOBER 21-25

pyPANTERA: A Python PAckage for Natural language obfuscaTion Enforcing pRivacy & Anonymization

Thanks for the attention! Question time :)

Francesco L. De Faveri, Guglielmo Faggioli, and Nicola Ferro

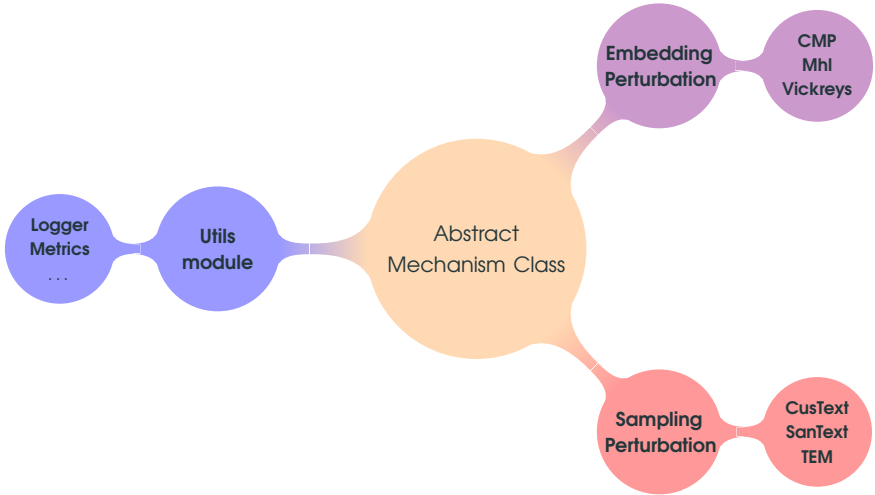
Department of Information Engineering, University of Padova, Italy

Full Text Available Here!

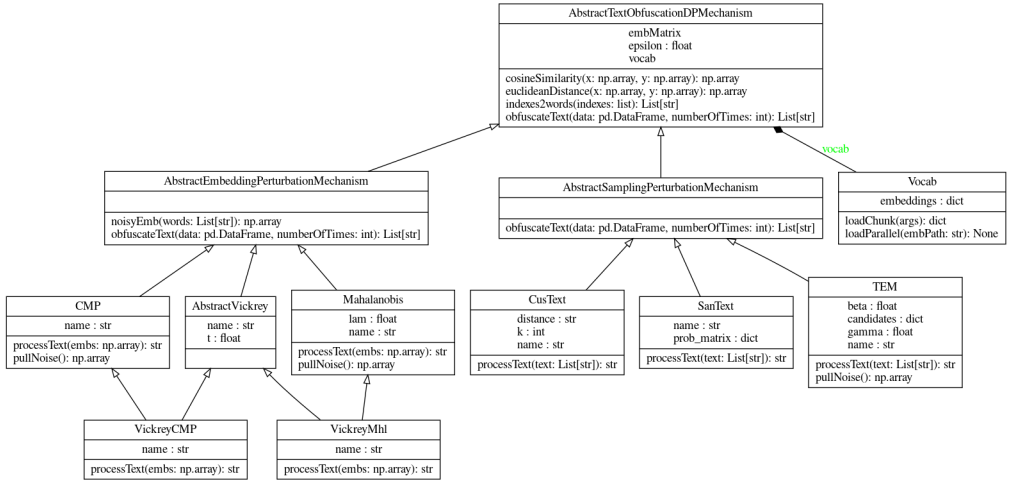
Framework Source Code

Backup

Backup: pyPANTERA Structure



Backup: pyPANTERA Class Diagram



Backup: Mechansims Configuration Parameters

Perturbation Families			
Embedding		*Sampling*	
Mechanism	**Parameters**	**Mechanism**	**Parameters**
CMP	-	CusText	$K = 10$
Mahalanobis	$\lambda = 1$	SanText	-
VickreyCMP	$t = 0.75$	TEM	$\beta = 0.001$
VickreyMhl	$t = 0.75$, $\lambda = 1$		

Table of the parameters of the mechanism used for performing the experimental showcase of the different tasks.

Backup: Experiments Results i

		Recall								nDCG@10							
		ε - Privacy Budget								ε - Privacy Budget							
Perturbation	**Mechanism**	1.0	5.0	10.0	12.5	15.0	17.5	20.0	50.0	1.0	5.0	10.0	12.5	15.0	17.5	20.0	50.0
Embedding	CMP	0.000	0.000	0.028	0.174	0.292	0.403	0.430	0.444	0.000	0.000	0.052	0.277	0.544	0.546	0.535	0.564
	Mahalanobis	0.000	0.000	0.001	0.077	0.134	0.290	0.368	0.447	0.000	0.000	0.003	0.103	0.262	0.455	0.494	0.565
	VickreyCMP	0.000	0.000	0.020	0.016	0.048	0.053	0.165	0.235	0.000	0.000	0.031	0.016	0.166	0.159	0.221	0.372
	VickreyMhl	0.000	0.001	0.002	0.002	0.029	0.042	0.122	0.191	0.000	0.005	0.007	0.004	0.062	0.097	0.158	0.293
Sampling	CusText	0.053	0.245	0.430	0.442	0.444	0.443	0.443	0.443	0.143	0.439	0.576	0.571	0.569	0.569	0.569	0.569
	SanText	0.000	0.444	0.447	0.448	0.444	0.450	0.447	0.444	0.000	0.564	0.569	0.570	0.568	0.559	0.568	0.562
	TEM	0.000	0.498	0.498	0.498	0.498	0.498	0.498	0.498	0.000	0.636	0.636	0.636	0.636	0.636	0.636	0.636
None	Original	-	-	-	-	-	-	-	-	0.498	-	-	-	-	-	-	-	-	0.636

Evaluation Task: Document Retrieval

Backup: Experiments Results ii

		Lexical Similarity (*Jaccard Similarity*)							
		ε - Privacy Budget							
Perturbation	**Mechanism**	1.0	5.0	10.0	12.5	15.0	17.5	20.0	50.0
Embedding	CMP	0.000	0.000	0.119	0.274	0.460	0.735	0.785	0.935
	Mahalanobis	0.000	0.002	0.047	0.140	0.302	0.457	0.590	0.935
	VickreyCMP	0.000	0.000	0.039	0.061	0.180	0.191	0.164	0.212
	VickreyMhl	0.000	0.013	0.028	0.038	0.098	0.134	0.117	0.151
Sampling	CusText	0.089	0.374	0.816	0.880	0.925	0.929	0.929	0.935
	SanText	0.000	0.935	0.935	0.935	0.935	0.935	0.935	0.935
	TEM	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Evaluation Task: Lexical similarity on Privacy Measurement

References

[1] R. S. Carvalho, T. Vasiloudis, O. Feyisetan, and K. Wang. TEM: high utility metric differential privacy on text. In S. Shekhar, Z. Zhou, Y. Chiang, and G. Stiglic, editors, *Proceedings of the 2023 SIAM International Conference on Data Mining, SDM 2023, Minneapolis-St. Paul Twin Cities, MN, USA, April 27-29, 2023*, pages 883–890. SIAM, 2023.

References ii

[2] S. Chen, F. Mo, Y. Wang, C. Chen, J.-Y. Nie, C. Wang, and J. Cui. A customized text sanitization mechanism with differential privacy. In A. Rogers, J. Boyd-Graber, and N. Okazaki, editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5747–5758, Toronto, Canada, July 2023. Association for Computational Linguistics.

[3] O. Feyisetan, B. Balle, T. Drake, and T. Diethe. Privacy- and utility-preserving textual analysis via calibrated multivariate perturbations. In J. Caverlee, X. B. Hu, M. Lalmas, and W. Wang, editors, *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining, Houston, TX, USA, February 3-7, 2020*, pages 178–186. ACM, 2020.

References iii

[4] Z. Xu, A. Aggarwal, O. Feyisetan, and N. Teissier. A differentially private text perturbation method using regularized mahalanobis metric. In O. Feyisetan, S. Ghanavati, S. Malmasi, and P. Thaine, editors, *Proceedings of the Second Workshop on Privacy in NLP*, pages 7–17, Online, Nov. 2020. Association for Computational Linguistics.

[5] Z. Xu, A. Aggarwal, O. Feyisetan, and N. Teissier. On a utilitarian approach to privacy preserving text generation. In O. Feyisetan, S. Ghanavati, S. Malmasi, and P. Thaine, editors, *Proceedings of the Third Workshop on Privacy in Natural Language Processing*, pages 11–20, Online, June 2021. Association for Computational Linguistics.

References iv

[6] X. Yue, M. Du, T. Wang, Y. Li, H. Sun, and S. S. M. Chow. Differential privacy for text analytics via natural text sanitization. In C. Zong, F. Xia, W. Li, and R. Navigli, editors, *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3853–3866, Online, Aug. 2021. Association for Computational Linguistics.