

Search Among Sensitive Content

ECIR 2021 Tutorial

Graham McDonald and Douglas W. Oard

University of Glasgow (UK), University of Maryland (USA)
graham.mcdonald@glasgow.ac.uk, oard@umd.edu

April 16, 2021

Bibliography

- Anandan, B., Clifton, C., Jiang, W., Murugesan, M., Pastrana-Camacho, P. & Si, L. (2012), ‘t-Plausibility: Generalizing words to desensitize text’, *Transactions on Data Privacy* **5**(3), 505–534.
URL: https://www.cs.purdue.edu/homes/lsi/TDP_2012.pdf
- Bagga, A. & Baldwin, B. (1998), Entity-based cross-document coreferencing using the vector space model, in C. Boitet & P. Whitelock, eds, ‘36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, COLING-ACL ’98, August 10-14, 1998, Université de Montréal, Montréal, Quebec, Canada. Proceedings of the Conference’, Morgan Kaufmann Publishers / ACL, pp. 79–85.
URL: <https://www.aclweb.org/anthology/P98-1012/>
- Baron, J. R., Sayed, M. F. & Oard, D. W. (2020), ‘Providing more efficient access to government records: A use case involving application of machine learning to improve FOIA review for the deliberative process privilege’, *arXiv preprint arXiv:2011.07203*.
URL: <https://arxiv.org/abs/2011.07203>
- Beeferman, D., Berger, A. L. & Lafferty, J. D. (1999), ‘Statistical models for text segmentation’, *Mach. Learn.* **34**(1-3), 177–210.
URL: <https://doi.org/10.1023/A:1007506220214>

- Belkin, N. J. & Croft, W. B. (1992), ‘Information filtering and information retrieval: Two sides of the same coin?’, *Commun. ACM* **35**(12), 29–38.
URL: <https://doi.org/10.1145/138859.138861>
- Bell, D. E. (2005), Looking back at the bell-la padula model, *in* ‘21st Annual Computer Security Applications Conference’, IEEE Computer Society, pp. 337–351.
URL: <https://doi.org/10.1109/CSAC.2005.37>
- Berardi, G., Esuli, A., Macdonald, C., Ounis, I. & Sebastiani, F. (2015), Semi-automated text classification for sensitivity identification, *in* ‘Proceedings of the 24th ACM International on Conference on Information and Knowledge Management’, p. 1711–1714.
URL: <https://doi.org/10.1145/2806416.2806597>
- Biega, A. J., Potash, P., Daumé, H., Diaz, F. & Finck, M. (2020), Operationalizing the legal principle of data minimization for personalization, *in* ‘Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval’, pp. 399–408.
URL: <http://users.umiacs.umd.edu/~hal/docs/daume20minimization.pdf>
- Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. (2010), The balanced accuracy and its posterior distribution, *in* ‘20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010’, IEEE Computer Society, pp. 3121–3124.
URL: <https://doi.org/10.1109/ICPR.2010.764>
- Burghardt, T., Böhm, K., Guttman, A. & Clifton, C. (2010), Search-log anonymization and advertisement: are they mutually exclusive?, *in* ‘Proceedings of the 19th ACM International Conference on Information and Knowledge Management’, pp. 1269–1272.
URL: <https://ieeexplore.ieee.org/document/6951353>
- Chakaravarthy, V. T., Gupta, H., Roy, P. & Mohania, M. K. (2008), Efficient techniques for document sanitization, *in* ‘Proceedings of the 17th ACM conference on Information and knowledge management’, pp. 843–852.
URL: <http://cscdb.nku.edu/pais09/pdf/p-pais01.pdf>
- Cormack, G. V. & Grossman, M. R. (2014), Evaluation of machine-learning protocols for technology-assisted review in electronic discovery, *in* ‘The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’14, Gold Coast , QLD, Australia - July

06 - 11, 2014', ACM, pp. 153–162.

URL: <https://doi.org/10.1145/2600428.2609601>

Cormack, G. V. & Grossman, M. R. (2015), 'Autonomy and reliability of continuous active learning for technology-assisted review', *CoRR* **abs/1504.06868**.

URL: <http://arxiv.org/abs/1504.06868>

Cormack, G. V., Grossman, M. R., Hedin, B. & Oard, D. W. (2010), Overview of the TREC 2010 legal track., in 'TREC'.

URL: <https://trec.nist.gov/pubs/trec19/papers/LEGAL10.OVERVIEW.pdf>

Craswell, N., Campos, D., Mitra, B., Yilmaz, E. & Billerbeck, B. (2020), 'ORCAS: 18 million clicked query-document pairs for analyzing search', *CoRR* **abs/2006.05324**.

URL: <https://arxiv.org/abs/2006.05324>

Culpepper, J. S., Diaz, F. & Smucker, M. D. (2018), 'Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in lorne (SWIRL)', *ACM SIGIR Forum* **52**(1), 34–90.

URL: <https://doi.org/10.1145/3274784.3274788>

Cumby, C. & Ghani, R. (2011), A machine learning based system for semi-automatically redacting documents, in 'Twenty-Third IAAI Conference'.

Damessie, T. T., Scholer, F. & Culpepper, J. S. (2016), The influence of topic difficulty, relevance level, and document ordering on relevance judging, in 'Proceedings of the 21st Australasian Document Computing Symposium, ADCS 2016, Caulfield, VIC, Australia, December 5-7, 2016', ACM, pp. 41–48.

URL: <http://dl.acm.org/citation.cfm?id=3015033>

Downie, J. S., West, K., Ehmann, A. F. & Vincent, E. (2005), The 2005 music information retrieval evaluation exchange (MIREX 2005): Preliminary overview, in 'ISMIR 2005, 6th International Conference on Music Information Retrieval, London, UK, 11-15 September 2005, Proceedings', pp. 320–323.

URL: <http://ismir2005.ismir.net/proceedings/xxxx.pdf>

Dwork, C., McSherry, F., Nissim, K. & Smith, A. D. (2016), 'Calibrating noise to sensitivity in private data analysis', *J. Priv. Confidentiality* **7**(3), 17–51.

URL: <https://doi.org/10.29012/jpc.v7i3.405>

- El Emam, K., Jonker, E., Arbuckle, L. & Malin, B. (2011), ‘A systematic review of re-identification attacks on health data’, *PloS one* **6**(12), e28071.
URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0028071>
- Fu, Z., Ren, K., Shu, J., Sun, X. & Huang, F. (2015), ‘Enabling personalized search over encrypted outsourced data with efficiency improvement’, *IEEE transactions on parallel and distributed systems* **27**(9), 2546–2559.
URL: <https://ieeexplore.ieee.org/document/7349214>
- Fu, Z., Ren, K., Shu, J., Sun, X. & Huang, F. (2016), ‘Enabling personalized search over encrypted outsourced data with efficiency improvement’, *IEEE Trans. Parallel Distributed Syst.* **27**(9), 2546–2559.
URL: <https://doi.org/10.1109/TPDS.2015.2506573>
- Gabriel, M., Paskach, C. & Sharpe, D. (2013), The challenge and promise of predictive coding for privilege, in ‘ICAIL 2013 DESI V Workshop’.
URL: <http://users.umiacs.umd.edu/~oard/desi5/research/Gabriel-final2.pdf>
- Gollins, T., McDonald, G., Macdonald, C. & Ounis, I. (2014), On using information retrieval for the selection and sensitivity review of digital public records, in ‘Proceeding of the 1st International Workshop on Privacy-Preserving IR: When Information Retrieval Meets Privacy and Security co-located with 37th Annual International ACM SIGIR conference, PIR@SIGIR’, Vol. 1225 of *CEUR Workshop Proceedings*, CEUR-WS.org, pp. 39–40.
URL: http://ceur-ws.org/Vol-1225/pir2014_submission_9.pdf
- Grossman, M. R., Cormack, G. V. & Roegiest, A. (2016), TREC 2016 total recall track overview, in ‘Proceedings of The Twenty-Fifth Text REtrieval Conference, TREC 2016, Gaithersburg, Maryland, USA, November 15-18, 2016’, Vol. 500-321 of *NIST Special Publication*, National Institute of Standards and Technology (NIST).
URL: <http://trec.nist.gov/pubs/trec25/papers/Overview-TR.pdf>
- Hill, S., Zhou, Z., Saul, L. K. & Shacham, H. (2016), ‘On the (in)effectiveness of mosaicing and blurring as tools for document redaction’, *Proc. Priv. Enhancing Technol.* **2016**(4), 403–417.
URL: <https://doi.org/10.1515/popets-2016-0047>
- Kamara, S. (2015), ‘Encrypted search’, *XRDS* **21**(3), 30–34.
URL: <https://doi.org/10.1145/2730908>

- Krishna, K., Tomar, G. S., Parikh, A. P., Papernot, N. & Iyyer, M. (2020), Thieves on sesame street! model extraction of bert-based apis, *in* ‘8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020’, OpenReview.net.
URL: <https://openreview.net/forum?id=Byl5NREFDr>
- Li, N., Li, T. & Venkatasubramanian, S. (2007), t-closeness: Privacy beyond k-anonymity and l-diversity, *in* ‘Proceedings of the 23rd International Conference on Data Engineering, ICDE 2007, The Marmara Hotel, Istanbul, Turkey, April 15-20, 2007’, IEEE Computer Society, pp. 106–115.
URL: <https://doi.org/10.1109/ICDE.2007.367856>
- Li, N., Qardaji, W. H. & Su, D. (2011), ‘Provably private data anonymization: Or, k-anonymity meets differential privacy’, *CoRR* **abs/1101.2604**.
URL: <http://arxiv.org/abs/1101.2604>
- Machanavajjhala, A., Kifer, D., Gehrke, J. & Venkatasubramanian, M. (2007), ‘L-diversity: Privacy beyond k-anonymity’, *ACM Trans. Knowl. Discov. Data* **1**(1), 3.
URL: <https://doi.org/10.1145/1217299.1217302>
- Martinez, S., Sánchez, D. & Valls, A. (2012), ‘Evaluation of the disclosure risk of masking methods dealing with textual attributes’, *International Journal of Innovative Computing, Information and Control* **8**(7), 4869–4882.
URL: <https://crises-deim.urv.cat/web/docs/publications/nonisijournals/699.pdf>
- McDonald, G. (2019), A Framework for Technology-Assisted Sensitivity Review: Using Sensitivity Classification to Prioritise Documents for Review, PhD thesis, School of Computing Science, University of Glasgow.
URL: <http://theses.gla.ac.uk/id/eprint/41076>
- McDonald, G., Macdonald, C. & Ounis, I. (2017), Enhancing sensitivity classification with semantic features using word embeddings, *in* ‘Advances in Information Retrieval - 39th European Conference on IR Research, ECIR 2017, Aberdeen, UK, April 8-13, 2017, Proceedings’, Vol. 10193 of *Lecture Notes in Computer Science*, pp. 450–463.
URL: https://doi.org/10.1007/978-3-319-56608-5_35
- McDonald, G., Macdonald, C. & Ounis, I. (2018), Active learning strategies for technology assisted sensitivity review, *in* ‘Advances in Information Retrieval - 40th European Conference on IR Research, ECIR 2018, Grenoble,

- France, March 26-29, 2018, Proceedings’, Vol. 10772 of *Lecture Notes in Computer Science*, Springer, pp. 439–453.
URL: https://doi.org/10.1007/978-3-319-76941-7_33
- McDonald, G., Macdonald, C. & Ounis, I. (2020a), Active learning stopping strategies for technology-assisted sensitivity review, in ‘Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020’, ACM, pp. 2053–2056.
URL: <https://doi.org/10.1145/3397271.3401267>
- McDonald, G., Macdonald, C. & Ounis, I. (2020b), ‘How the accuracy and confidence of sensitivity classification affects digital sensitivity review’, *ACM Transactions on Information Systems* **39**(1), 1–34.
URL: <https://doi.org/10.1145/3417334>
- Mojdeh, M. & Cormack, G. V. (2008), A mail client plugin for privacy-preserving spam filter evaluation, in ‘CEAS 2008 - The Fifth Conference on Email and Anti-Spam, 21-22 August 2008, Mountain View, California, USA’.
URL: <http://www.ceas.cc/2008/papers/ceas2008-paper-52.pdf>
- Narvala, H., McDonald, G. & Ounis, I. (2020), Receptor: A platform for exploring latent relations in sensitive documents, in ‘Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval’, pp. 2161–2164.
URL: <https://eprints.gla.ac.uk/218388/1/218388.pdf>
- Nissenbaum, H. (1997), ‘Toward an approach to privacy in public: Challenges of information technology’, *Ethics & Behavior* **7**(3), 207–219.
- Nissenbaum, H. (2009), *Privacy in context: Technology, policy, and the integrity of social life*, Stanford University Press.
- Oard, D. W., Sebastiani, F. & Vinjumur, J. K. (2018), ‘Jointly minimizing the expected costs of review for responsiveness and privilege in e-discovery’, *ACM Transactions on Information Systems* **37**(1), 1–35.
URL: <https://doi.org/10.1145/3268928>
- Oard, D. W., Shilton, K. & Lin, J. J. (2016), Evaluating search among secrets., in ‘EVIA@ NTCIR’.
URL: <https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings12/pdf/evia/05-EVIA2016-OardD.pdf>

- Oard, D. W., Vinjumur, J. & Sebastiani, F. (2017), When is it rational to review for privilege?, *in* 'Proceedings of the Workshop on Using Advanced Data Analysis in eDiscovery & Related Disciplines to Identify and Protect Sensitive Information in Large Collections (DESI'17)'.
URL: <http://users.umiacs.umd.edu/~oard/desi7/papers/DO.pdf>
- Oard, D. W. & Webber, W. (2013), 'Information retrieval for e-discovery', *Foundations and Trends in Information Retrieval* **7**(2-3), 99–237.
URL: <https://doi.org/10.1561/15000000025>
- Pevzner, L. & Hearst, M. A. (2002), 'A critique and improvement of an evaluation metric for text segmentation', *Comput. Linguistics* **28**(1), 19–36.
URL: <https://doi.org/10.1162/089120102317341756>
- Politou, E., Alepis, E. & Patsakis, C. (2018), 'Forgetting personal data and revoking consent under the gdpr: Challenges and proposed solutions', *Journal of Cybersecurity* **4**(1).
URL: <https://academic.oup.com/cybersecurity/article/4/1/tyy001/4954056>
- Pozen, D. E. (2005), 'The mosaic theory, national security, and the freedom of information act', *Yale Law Journal* **115**, 628.
URL: https://scholarship.law.columbia.edu/faculty_scholarship/1383/
- Roegiest, A. & Cormack, G. V. (2015), Total recall track tools architecture overview, *in* 'Proceedings of TREC', National Institute of Standards and Technology (NIST).
- Roegiest, A., Cormack, G. V., Clarke, C. L. A. & Grossman, M. R. (2015), TREC 2015 total recall track overview, *in* 'Proceedings of The Twenty-Fourth Text REtrieval Conference, TREC 2015, Gaithersburg, Maryland, USA, November 17-20, 2015', Vol. 500-319 of *NIST Special Publication*, National Institute of Standards and Technology (NIST).
URL: <https://trec.nist.gov/pubs/trec24/papers/Overview-TR.pdf>
- Roegiest, A., Lipani, A., Beutel, A., Olteanu, A., Lucic, A., Stoica, A.-A., Das, A., Biega, A., Voorn, B., Hauff, C., Spina, D., Lewis, D., Oard, D. W., Yilmaz, E., Hasibi, F., Kazai, G., McDonald, G., Haned, H., Ounis, I., van der Linden, I., Garcia-Gathright, J., Baan, J., Lau, K. N., Balog, K., de Rijke, M., Sayed, M., Panteli, M., Sanderson, M., Lease, M., Ekstrand, M. D., Lahoti, P. & Kamishima, T. (2019), 'FACTS-IR: Fairness, accountability, confidentiality, transparency, and safety in information retrieval', *SIGIR Forum* **53**(2).

- Samarati, P. (2001), ‘Protecting respondents’ identities in microdata release’, *IEEE Trans. Knowl. Data Eng.* **13**(6), 1010–1027.
URL: <https://doi.org/10.1109/69.971193>
- Samarati, P. & Sweeney, L. (1998), ‘Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression’.
- Sánchez, D. & Batet, M. (2016), ‘C-sanitized: A privacy model for document redaction and sanitization’, *Journal of the Association for Information Science and Technology* **67**(1), 148–163.
- Sánchez, D., Batet, M. & Viejo, A. (2012), Detecting sensitive information from textual documents: An information-theoretic approach, in ‘Modeling Decisions for Artificial Intelligence - 9th International Conference, MDAI 2012, Girona, Catalonia, Spain, November 21-23, 2012. Proceedings’, Vol. 7647 of *Lecture Notes in Computer Science*, Springer, pp. 173–184.
URL: https://doi.org/10.1007/978-3-642-34620-0_17
- Sánchez, D., Batet, M. & Viejo, A. (2014), ‘Utility-preserving sanitization of semantically correlated terms in textual documents’, *Inf. Sci.* **279**, 77–93.
URL: <https://doi.org/10.1016/j.ins.2014.03.103>
- Sánchez, D., Batet, M., Viejo, A., Rodríguez-García, M. & Castella-Roca, J. (2018), ‘A semantic-preserving differentially private method for releasing query logs’, *Information Sciences* **460**, 223–237.
URL: <https://crises-deim.urv.cat/web/docs/publications/journals/1078.pdf>
- Sayed, M. F., Cox, W., Rivera, J. L., Christian-Lamb, C., Iqbal, M., Oard, D. W. & Shilton, K. (2020), A test collection for relevance and sensitivity, in ‘Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval’, p. 1605–1608.
URL: <https://doi.org/10.1145/3397271.3401284>
- Sayed, M. F. & Oard, D. W. (2019), Jointly modeling relevance and sensitivity for search among sensitive content, in ‘Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019’, ACM, pp. 615–624.
URL: <https://doi.org/10.1145/3331184.3331256>
- Saygin, Y., Hakkini-Tur, D. & Tur, G. (2006), Sanitization and anonymization of document repositories, in ‘Web and information security’, IGI

Global, pp. 133–148.

URL: <http://people.sabanciuniv.edu/ysaygin/documents/publications/Sanitization.pdf>

Schneider, J., Chan, P., Edwards, G. & Hangal, S. (2017), ‘ePADD: Computational analysis software facilitating screening, browsing, and access for historically and culturally valuable email collections’, *D-Lib Magazine* **23**(5/6).

URL: <http://www.dlib.org/dlib/may17/schneider/05schneider.html>

Shilton, K., Wickner, A., Oard, D. W. & Lin, J. (2017), Protecting sensitive email: Archival views on challenges and opportunities, in ‘Proceedings of PC4DS’.

URL: https://cs.uwaterloo.ca/~jimmylin/publications/Wickner_etal.2017.pdf

Si, L. & Yang, G. H. (2014a), ‘PIR 2014 the first international workshop on privacy-preserving IR: when information retrieval meets privacy and security’, *SIGIR Forum* **48**(2), 83–88.

URL: <https://doi.org/10.1145/2701583.2701593>

Si, L. & Yang, G. H. (2014b), Privacy-preserving IR: when information retrieval meets privacy and security, in ‘The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’14, Gold Coast , QLD, Australia - July 06 - 11, 2014’, ACM, p. 1295.

URL: <https://doi.org/10.1145/2600428.2600737>

Sun, W., Wang, B., Cao, N., Li, M., Lou, W., Hou, Y. T. & Li, H. (2013), Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking, in ‘Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security’, pp. 71–82.

URL: <https://www.cnsr.ictas.vt.edu/publication/06656804.pdf>

Swaminathan, A., Mao, Y., Su, G., Gou, H., Varna, A. L., He, S., Wu, M. & Oard, D. W. (2007), Confidentiality-preserving rank-ordered search, in ‘Proceedings of the 2007 ACM Workshop On Storage Security And Survivability, StorageSS 2007, Alexandria, VA, USA, October 29, 2007’, ACM, pp. 7–12.

URL: <https://doi.org/10.1145/1314313.1314316>

Sweeney, L. (1996), Replacing personally-identifying information in medical records, the scrub system., in ‘Proceedings of the AMIA annual fall symposium’, American Medical Informatics Association, p. 333.

URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233179/>

- Sweeney, L. (2002), ‘k-anonymity: A model for protecting privacy’, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10**(5), 557–570.
URL: <https://doi.org/10.1142/S0218488502001648>
- van Rijsbergen, C. J. (1979), *Information Retrieval*, Butterworth.
- Vinjunur, J. K. (2015), Evaluating expertise and sample bias effects for privilege classification in e-discovery, in ‘Proceedings of the 15th International Conference on Artificial Intelligence and Law’, pp. 119–127.
URL: <https://ediscovery.umiacs.umd.edu/pub/icail15.pdf>
- Vinjunur, J. K., Oard, D. W. & Paik, J. H. (2014), Assessing the reliability and reusability of an e-discovery privilege test collection, in ‘The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’14, Gold Coast , QLD, Australia - July 06 - 11, 2014’, ACM, pp. 1047–1050.
URL: <https://doi.org/10.1145/2600428.2609506>
- Wang, C., Cao, N., Li, J., Ren, K. & Lou, W. (2010), Secure ranked keyword search over encrypted cloud data, in ‘2010 IEEE 30th international conference on distributed computing systems’, IEEE, pp. 253–262.
URL: <https://ieeexplore.ieee.org/document/5541682>
- Wells, C. E. (2006), ‘CIA v. Sims: Mosaic theory and government attitude’, *Admin. L. Rev.* **58**, 845.
URL: <https://scholarship.law.missouri.edu/cgi/viewcontent.cgi?article=1392&context=facpubs>
- Yang, G. H. & Soboroff, I. (2015a), ‘Privacy preserving IR 2015: A SIGIR 2015 workshop’, *SIGIR Forum* **49**(2), 98–101.
URL: <https://doi.org/10.1145/2888422.2888437>
- Yang, G. H. & Soboroff, I. (2015b), Privacy-preserving IR 2015: When information retrieval meets privacy and security, in ‘Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval’, Association for Computing Machinery, New York, NY, USA, p. 1157–1158.
URL: <https://doi.org/10.1145/2766462.2767857>
- Yang, G. H., Soboroff, I., Xiong, L., Clarke, C. L. & Garfinkel, S. L. (2016), Privacy-preserving IR 2016: Differential privacy, search, and social media, in ‘Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval’, Association for

Computing Machinery, New York, NY, USA, p. 1247–1248.

URL: <https://doi.org/10.1145/2911451.2917763>

Yang, G. H. & Zhang, S. (2017), Differential privacy for information retrieval, in ‘Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2017, Amsterdam, The Netherlands, October 1-4, 2017’, ACM, pp. 325–326.

URL: <https://doi.org/10.1145/3121050.3121107>

Yang, G. H. & Zhang, S. (2018), Differential privacy for information retrieval, in ‘Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018’, ACM, pp. 777–778.

URL: <https://doi.org/10.1145/3159652.3162006>

Yu, J., Lu, P., Zhu, Y., Xue, G. & Li, M. (2013), ‘Toward secure multi-keyword top-k retrieval over encrypted cloud data’, *IEEE transactions on dependable and secure computing* **10**(4), 239–250.

URL: <https://ieeexplore.ieee.org/document/6425381>

Zhang, S. & Yang, G. H. (2017), Deriving differentially private session logs for query suggestion, in J. Kamps, E. Kanoulas, M. de Rijke, H. Fang & E. Yilmaz, eds, ‘Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2017, Amsterdam, The Netherlands, October 1-4, 2017’, ACM, pp. 51–58.

URL: <https://doi.org/10.1145/3121050.3121076>

Zhang, S., Yang, G. H. & Singh, L. (2016), Anonymizing query logs by differential privacy, in ‘Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval’, pp. 753–756.

URL: <http://infosense.cs.georgetown.edu/publication/84.pdf>