質問者のプライバシーを保護する特許データベース検索（研究紹介）

中川研究室 修士 2 年 胡 瀚林
指導教員： 中川 裕志 教授

2016 年 7 月 1 日

概要

テキスト検索の質問から個人情報や企業情報が漏洩する可能性があります．既存研究 [1] では質問のある種の匿名性（否認可能性）を保証し，検索の精度と再現率を保持できるテキスト検索システムを提案した．その検索システムはユーザーの真の質問に混ざるデミー単語を生成するメカニズムと真の質問の単語だけの暗号化した関連性スコアを計算できる検索スキームから成る．

質問が持つ主要な意味がデミーから真の質問の単語を分別できる鍵である．本発表では [1] に提案したテキスト検索システムに対して新たな攻撃手法，主意味攻撃を提案し，特許データベースとを用いて提案手法を評価する．

1 はじめに

テキスト検索をするとき，検索質問をサーバー側に渡さなければならない．しかし，検索質問からユーザーの情報が漏洩する危険があることが AOL 事件 [2] より証明された．特に特許検索の場合は質問が研究開発動向などは企業秘密を含んでいるため，一般的なウェブ検索のユーザーよりテキスト検索のプライバシー問題を重視している．

今テキスト検索エンジンの大半が類似検索である．全ての質問単語を含んでいる文章しか検索できないキーワード検索と違い，類似検索は文章と質問の関連性を計算し文章にランクをつける [3]．毎回全ての文章との関連性を計算しないために検索エンジンが単語と文章の関連値を転置ファイルに保存し，質問の単語と文章の関連値の和を質問とその文章の関連性とする．このような計算が必要であるため，[4],[5],[6],[7] などキーワード検索だけ対応できる研究は類似検索に応用できない．また質問のある種の匿名性より質問者のプライバシーを保護する手法がある．[8] では事前的に静的な質問セットを作り，真の質問 q の代わりに真のと一番類似な質問 q' を含んでいる質問セットをサーバーに送る．q' が真の質問の大半な結果を検索できると考えられ，質問セット中の他の質問をダミー質問にする．そのため，このメカニズムは検索の精度と再現率に影響を大きく与える．また，質問の長さの増加と伴って質問の可能な組み合わせが指数的に増加するため，実践的には長い質問が多いテキスト検索 [] と質問拡張 [9, 10] に対応できない．

2 既存研究

3 プライバシー分析

4 メイントピック攻撃

5 実験結果

参考文献

[1] H. Pang, X. Ding and X. Xiao: “Embellishing Text Search Queries to Protect User Privacy”, Proc. VLDB Endow., **3**, 1-2, pp. 598–607 (2010).

[2] M. Barbaro and T.Z.Jr: “A Face Is Exposed for AOL Searcher No. 4417749 - New York Times” (2006).

[3] J. Zobel and A. Moffat: “Inverted Files for Text Search Engines”, ACM Comput. Surv., **38**, 2 (2006).

[4] J. Bethencourt, D. Song and B. Waters: “New constructions and practical applications for private stream searching”, 2006 IEEE Symposium on Security and Privacy (S&P’06), IEEE, pp. 6–pp (2006).

[5] M. J. Freedman, Y. Ishai, B. Pinkas and O. Reingold: “Keyword search and oblivious pseudorandom functions”, Theory of Cryptography Conference, Springer, pp. 303–324 (2005).

[6] D. Boneh, G. Di Crescenzo, R. Ostrovsky and G. Persiano: “Public key encryption with keyword search”, International Conference on the Theory and Applications of Cryptographic Techniques, Springer, pp. 506–522 (2004).

[7] D. X. Song, D. Wagner and A. Perrig: “Practical techniques for searches on encrypted data”, Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on, IEEE, pp. 44–55 (2000).

[8] M. Murugesan and C. Clifton: “Providing Privacy through Plausibly Deniable Search”, Proceedings of the 2009 SIAM International Conference on Data Mining, Proceedings, Society for Industrial and Applied Mathematics, pp. 768–779 (2009).

[9] Y. Qiu and H.-P. Frei: “Concept based query expansion”, Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 160–169 (1993).

[10] J. Xu and W. B. Croft: “Query expansion using local and global document analysis”, Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 4–11 (1996).