

修士論文

# 特許検索における質問意図の曖昧化

48-156229 胡瀚林

指導教員 中川裕志 教授

2017 年 1 月

東京大学大学院情報理工学系研究科数理情報学専攻



## 概要

企業が特許を取る前に、類似な特許が既に存在するかを確かめるために特許データベースを検索する必要がある。しかし、検索の質問から企業秘密が漏洩する可能性がある。ウェブテキスト検索の質問からユーザーの興味を守る手法が多数存在している。その中真の質問と同時にダミー質問を提出する手法が一番効率的、現実的である。一般的なウェブテキスト検索と違い、特許データベース検索は長い検索質問を用い、検索の精度と再現率を重視しているため、既存手法を直接特許検索に適用することはできない。本発表では既存手法を破られる攻手法を提案し、その攻撃手法に対応できるダミー質問生成システムを提案する。



# 目次

第 1 章	はじめに	1
第 2 章	特許	2
2.1	特許文章 . . . . .	2
2.2	国際特許分類 . . . . .	2
2.3	特許検索 . . . . .	2
第 3 章	曖昧化検索	3
3.1	否認可能検索を利用したプライバシー保護 [1] . . . . .	3
3.2	質問者のプライバシーを保護する質問加工法 [2] . . . . .	3
3.3	質問意図を曖昧化するキーワード検索 [3] . . . . .	3
第 4 章	語意分析	4
4.1	tf-idf . . . . .	4
4.2	潜在意味解析 . . . . .	4
4.3	潜在的ディリクレ配分法 . . . . .	4
第 5 章	プライバシー分析 (攻撃手法)	5
5.1	メイントピック攻撃 . . . . .	5
5.2	類似度攻撃 [4](事前情報あり) . . . . .	5
5.3	類似度攻撃 2(事前情報なし) . . . . .	5
第 6 章	質問曖昧化 (提案手法)	6
6.1	単語ベクトル . . . . .	6
6.2	質問曖昧化 . . . . .	6
第 7 章	データベース分割	7
第 8 章	実験	8
8.1	データベース . . . . .	8
8.2	tfidf vs lda vs lsa . . . . .	8

iv 目次

8.3	データベース分割 . . . . .	8
8.4	検索結果分析 (真の質問が当たられる確率 vs ダミー質問と真の質問の検索結果の類似度) . . . . .	8
第 9 章	おわりに	9
	謝辞	10
	参考文献	11
	付録 A	12

## 第 1 章

# はじめに

## 第 2 章

# 特許

2.1 特許文章

2.2 国際特許分類

2.3 特許検索



## 第 3 章

# 曖昧化検索

- 3.1 否認可能検索を利用したプライバシー保護 [1]
- 3.2 質問者のプライバシーを保護する質問加工法 [2]
- 3.3 質問意図を曖昧化するキーワード検索 [3]

## 第 4 章

# 語意分析

### 4.1 tf-idf

### 4.2 潜在意味解析

### 4.3 潜在的ディリクレ配分法

## 第 5 章

# プライバシー分析 (攻撃手法)

- 5.1 メイントピック攻撃
- 5.2 類似度攻撃 [4](事前情報あり)
- 5.3 類似度攻撃 2(事前情報なし)

## 第 6 章

# 質問曖昧化 (提案手法)

### 6.1 単語ベクトル

### 6.2 質問曖昧化

## 第 7 章

# データベース分割

## 第 8 章

# 実験

### 8.1 データベース

### 8.2 tfidf vs lda vs lsa

### 8.3 データベース分割

### 8.4 検索結果分析 (真の質問が当たられる確率 vs ダミー質問と真の質問の検索結果の類似度)

## 第9章

# おわりに

## 謝辭



## 参考文献

- [1] “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).
- [2] “Embellishing Text Search Queries to Protect User Privacy”, Proc. VLDB Endow., **3**, 1-2, pp. 598–607 (2010).
- [3] “On masking topical intent in keyword search”, 2014 IEEE 30th International Conference on Data Engineering, IEEE, pp. 256–267 (2014).
- [4] “SimAttack: private web search under fire”, Journal of Internet Services and Applications, **7**, 1, p. 1 (2016).

## 付録 A