

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検索意図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質問曖昧化手法が一番効率的、現実的である。本稿では特許検索における既存な質問曖昧化手法 [1, 2, 3] を実装し、類似度攻撃 [4] で既存手法の安全性を評価した。

また、類似度攻撃 [4] を含め、多くの既存な質問曖昧化に対する攻撃手法は攻撃者が質問者の事前情報を持つと仮定する。本稿では事前情報なしの攻撃手法を提案し、その攻撃手法に対応できる既存な質問曖昧化の改良と新たな質問曖昧化手法を提案する。

目次

第 1 章	はじめに	1
第 2 章	特許	2
2.1	特許分類	2
2.2	特許検索	2
第 3 章	曖昧化検索	5
3.1	否認可能検索を利用したプライバシー保護 [1]	5
3.2	質問者のプライバシーを保護する質問加工法 [2]	5
3.3	質問意図を曖昧化するキーワード検索 [3]	5
第 4 章	意味分析	6
4.1	tf-idf	6
4.2	潜在意味解析	6
4.3	潜在的ディリクレ配分法	6
第 5 章	プライバシー分析 (攻撃手法)	7
5.1	メイントピック攻撃	7
5.2	類似度攻撃 [4](事前情報あり)	7
5.3	類似度攻撃 2(事前情報なし)	7
第 6 章	質問曖昧化 (提案手法)	9
6.1	単語ベクトル	9
6.2	質問曖昧化	9
第 7 章	データベース分割	10
第 8 章	評価実験	11
8.1	データベース	11
8.2	tfidf vs lda vs lsa	11
8.3	データベース分割	11

iv 目次

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)	11
第 9 章	おわりに	12
	謝辞	13
	参考文献	14
	付録 A	15

第 1 章

はじめに

テキスト検索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 [?] より証明された。特に特許検索の場合は検索質問が研究開発動向など企業秘密を含んでいるため、一般的なウェブ検索の質問者より質問のプライバシー問題を重視している。

本稿の構成は次の通りである。第二章では特許文章と特許検索の特徴を述べる。第三章では既存な質問曖昧化メカニズム [1, 2, 3] を述べる。第四章では曖昧化メカニズムがよく用いる意味分析手法を述べる。第五章では既存な攻撃手法 [4] を述べ、[4] の改良と新たな攻撃手法を提案する。第六、七章では新たな質問曖昧化手法を提案する。最後に、第八章で評価実験を述べ、第九章で全体をまとめる。

第 2 章

特許

特許検索質問のプライバシーを保護する手法を説明する前に特許検索と特許そのものを簡単に紹介する必要がある。特許法第 1 条には、「この法律は、発明の保護及び利用を図ることにより、発明を奨励し、もつて産業の発達に寄与することを目的とする」とある。特許制度は、発明者には一定期間、一定の条件のもとに特許権という独占的な権利を与えて発明の保護を図る一方、その発明を公開して利用を図ることにより新しい技術を人類共通の財産としていくことを定めて、これにより技術の進歩を促進し、産業の発達に寄与しようというものである。[?] 特許を取るには以下の条件を満たさなければならない: 新規性: 公知の発明と同様の発明は特許を受けることができない; 進歩性: 先行技術に基づいて容易に発明をすることができる発明は特許を受けることができない。単一性: 発明の単一性の要件を満たさない二以上の発明は一つの願書で出願することができない。

特許を受けようとする発明を特定するために特許請求の範囲を記載する必要がある。

図 2.1 で表した例のように、特許の請求項は特定の書き方がある。誤解を招かないように技術用語は、学術用語を用いる。また、一般的な文章は単語をなるべく重複しないようにする一方、特許文章は単語を全体を通じて統一して使用する。

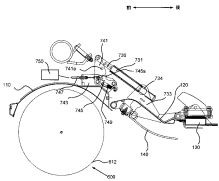
2.1 特許分類

特許の一つ特徴は全ての特許が人の手によって分類されている。特許分類を用いることより検索する特許文章が減り、似たようなキーワードを含むが分類が違う特許文章を排除することができる。今最も使われている特許分類が世界知的所有権機関 (WIPO) による管理されている国際特許分類 (IPC) である。国際特許分類は階層構造であり、一番上の階層は A から H までの 8 個のセクションである。セクション以下は??に表したように四つの階層に分類されている。

2.2 特許検索

JP 2016-208844 A 2016.12.15		JP 2016-208844 A 2016.12.15	
(19) 日本国特許庁(JP)		(12) 公開特許公報(A)	
		(11) 特許出願公開番号 特開2016-208844 (P2016-208844A) (43) 公開日 平成28年12月15日(2016.12.15)	
(51) Int. Cl.		F 1	
A 01 B 35/04 (2006.01)		A O 1 B 35/04 E 2 B O 3 4	
		テーマコード (参考)	
		2 B O 3 4	
		審査請求 未請求 請求項の数 4 O L (全 13 頁)	
(21) 出願番号 特開2015-92186 (P2015-92186)		(71) 出願人 390010836	
(22) 出願日 平成27年4月28日 (2015.4.28)		小機工業株式会社	
		岡山県岡山市南区中睦684番地	
		(74) 代理人 111000408	
		特許業務法人高橋・林アンドパートナーズ	
		(72) 発明者 河原 文雄	
		岡山県岡山市南区中睦684番地 小機工業株式会社内	
		Fターム(参考) 2B034 AA03 BA06 BB01 BB02 EA02	
		EB06 EB33 JA06	
(54) 【発明の名称】 農作業機			

(57) 【要約】
【課題】代かき作業機を昇降させる必要がない場面において、オート装置が代かき作業機を昇降させることを防止する。
【解決手段】本発明の一実施形態に係る農作業機は、耕耘作業を行うロータリ作業部を回転自在に支持する機体と、機体に設けられ、ロータリ作業部の上部を覆うカバー部と、カバー部の後端部に回転可能に支持されたエプロンと、エプロンの背面に取り付けられた支持部材と、カバー部に取り付けられ、エプロンのロック状態とフリー状態を切り替え可能なエプロン回転制御部と、を備え、エプロン回転制御部は、後端部が支持部材に対し取り付けられたロッド部と、ロッド部の前端部を回転自在に支持し、被係合部を有する第1アーム部と、カバー部に回転自在に支持され、係合部を有する第2アーム部と、第2アーム部を回転させる駆動部とを有し、係合部は、第1アーム部が回転するときに、被係合部の回転を規制するように構成されてよい。
【選択図】図1



【特許請求の範囲】
【請求項1】
耕耘作業を行うロータリ作業部を回転自在に支持する機体と、
前記機体に設けられ、前記ロータリ作業部の上部を覆うカバー部と、
前記カバー部の後端部に回転可能に支持されたエプロンと、
前記エプロンの背面に取り付けられた支持部材と、
前記カバー部に取り付けられ、前記エプロンが自由に回転できない状態であるロック状態と前記エプロンが自由に回転できる状態であるフリー状態とに切り替えることができるエプロン回転制御部と、を備え、
前記エプロン回転制御部は、後端部が前記支持部材に対して、摺動可能に取り付けられたロッド部と、
前記ロッド部の前端部を回転自在に支持し、前記カバー部に対して回転自在に支持されて、被係合部を有する第1アーム部と、
前記カバー部に回転自在に支持され、係合部を有する第2アーム部と、
前記第2アーム部を回転させる駆動部とを有し、
前記係合部は、前記ロッド部が前方に移動するに伴い前記第1アーム部が回転するときに、前記被係合部の回転を規制することを特徴とする農作業機。
【請求項2】
前記被係合部は、ピン部材であることを特徴とする請求項1に記載の農作業機。
【請求項3】
前記駆動部は、ワイヤとワイヤ制御部を含むことを特徴とする請求項1又は請求項2に記載の農作業機。
【請求項4】
耕耘作業を行うロータリ作業部を回転自在に支持する機体と、
前記機体に設けられ、前記ロータリ作業部の上部を覆うカバー部と、
前記カバー部の後端部に回転可能に支持されたエプロンと、
前記エプロンの背面に取り付けられた支持部材と、
前記カバー部に取り付けられ、前記エプロンが自由に回転できない状態であるロック状態と前記エプロンが自由に回転できる状態であるフリー状態とに切り替えることができるエプロン回転制御部と、を備え、
前記エプロン回転制御部は、後端部が前記支持部材に対して、摺動可能に取り付けられたロッド部と、
前記ロッド部の前端部を回転自在に支持し、前記カバー部に対して回転自在に支持されて、係合部を有する第1アーム部と、
前記カバー部に回転自在に支持され、被係合部を有する第2アーム部と、
前記第2アーム部を回転させる駆動部とを有し、
前記被係合部は、前記ロッド部が前方に移動するに伴い前記第1アーム部が回転するときに、前記係合部の回転を規制することを特徴とする農作業機。
【発明の詳細な説明】
【技術分野】
【0001】
本発明は、農作業機に関する。特に、本発明は、エプロンが自由に回転できない状態であるロック状態とエプロンが自由に回転できる状態であるフリー状態とに切り替えることができるエプロン回転制御部を備える農作業機に関する。
【背景技術】
【0002】
耕耘ロータにより耕耘された耕土を整地するエプロン（第1整地板）とエプロンの後部に上下方向に回転自在に設けられて耕土表面を均平にするレベラ（第2整地板）を備える農作業機、例えば、代かき作業機は、一般に、走行可能な走行機体の後部に三点リンク連結機構を介して昇降可能に連結されて、走行機体の前進走行とともに進行しながら代かき

図 2.1. 特許文章例

セクション:A
サブセクション : 61
クラス: C
メイングループ:5
サブグループ:08

健康および娯楽
医学または獣医学:衛生学
歯科:口腔または歯科衛生
歯の充填または被覆
歯冠:その製造; 口中での歯冠固定

表 2.1. 国際特許分類例:A61C 5/08

検索タイプ	検索対象 (specification)	検索目的
技術水準調査 (State of the Art Search)	アイデア	自分の発明に関連する背景知識を得る
新規性調査 (Novelty Search)	特許文章	特許登録の可能性を判断する
侵害調査 (Infringement Search)	商品と 商品に関連する技術	権利侵害とならないかを判断する

表 2.2. 特許検索タイプ

符号	意味
N	辞書中の単語の数
$W = \{1, 2, 3, \dots, N\}$	単語集合
M	コーパス中の文書の数
$D = \{1, 2, 3, \dots, M\}$	文章集合
K	トピック数
$T = \{1, 2, 3, \dots, K\}$	トピック集合
$\ell_i = \{t_1, t_2, \dots, K\}$	単語 i のトピックベクトル
ℓ	質問のトピックベクトル

表 2.3. 表記法

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]

Algorithm 1 HDGA(On Masking Topical Intent in Keyword Search)

Input: 質問: q_1

```

1:  $Q = \{q_1\} \delta_{q_1} = \underset{t \in T}{\operatorname{argmax}} Pr[t|q_1]$ 
2: for all  $t \in T \setminus \{\delta_{q_1}\}$  do
3:    $e_t = h(\delta_{q_1} || t || s)$ 
4: end for
5:  $T_D = \{t_{q_1}^1, t_{q_1}^2, \dots, t_{q_1}^2 | \forall t_1 \in T_D, \forall t_2 \in T \setminus T_D, e_{t_1} > e_{t_2}\}$ 
6: for all  $t \in T_D$  do
7:   while  $\underset{t \in T}{\operatorname{argmax}} Pr[t|q'] \neq t$  do
8:     randomly select  $|q_1|$  keywords for  $t$  based on  $Pr[w|t]$ , to form a dummy query
        $q'$ 
9:   end while
10:   $Q = Q \cup \{q'\}$ 
11: end for
12: Shuffle queries in  $Q$  return  $Q$ 

```

第 4 章

意味分析

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

Algorithm 2 メイントピック攻撃

Input: 質問: $q = \{t_i\}$, 単語のトピックベクトル集合 $L = \{\ell_i\}$

```

1:  $R = \phi, \ell = 0$ 
2:  $\ell = \sum_{t_i \in Q} \ell_{t_i}$ 
3:  $maintopic = \operatorname{argmax}_j \ell[j]$ 
4: for all  $bk_k \in q$  do
5:    $R = R \cup \{\max_{t_i} \ell_{t_i}[maintopic]\}$ 
6: end for
7: return  $R$ 

```

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)

Algorithm 3 類似攻撃

Input: 質問集合: $Q = \{q_i^r \mid i \in \{1, 2, 3, 4\}, r \in \{1, 2, \dots, R\}\}$, 単語のトピックベクトル集合 $L = \{\ell_i\}$ 1: $p_i = q_i^1 \ i \in \{1, 2, 3, 4\}, result = \phi$ 2: **for** $r = 2, 3, \dots, R$ **do**3: **for** $i = 1, 2, 3, 4$ **do**4: $j = \operatorname{argmax}_j \frac{p_i \cdot q_j^r}{|p_i| |q_j^r|}$ 5: $d_i = \frac{p_i \cdot q_j^r}{|p_i| |q_j^r|}$ 6: $temp_i = \frac{1}{r}(p_i(r-1) + q_j)$ 7: **end for**8: **for** $i = 1, 2, 3, 4$ **do**9: $p_i = temp_i$ 10: **end for**11: $result = result \cup \{\operatorname{argmax}_i d_i\}$ 12: **end for**13: **return** $result$

第 6 章

質問曖昧化 (提案手法)

事前に単語をグループにする [2], 質問をグループにする [1] と同じようにトピックをグループにすることよりトピック出現頻度で質問者が興味あるトピックを特定することが防ぐと考えられる. [3] ではハッシュ関数でトピックをグループにしているが, 各トピック間の関係を配慮していない. また, [3] では各ダミートピックから単語をランダムに選ぶので, 新の質問に含まれている単語が違ってても真の質問が同じトピックに属するならダミー質問が同じような性質を持つ. 一方, 同じ真の質問に対して同じダミー質問を生成することができない. 真の質問に含まれている単語という情報を用いてないことが [3] に紹介した手法が Simattack に弱い原因だと考えられる. simattack から真の質問を守るために真の質問が同じ単語を含むとき, ダミー質問も同様に同じ単語を含んでほしい. それを実現するため提案手法では単語ベクトルを用いた.

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

特許分類を用いることにより特許データベースを分割することができる．分割したデータベース各々に対して同じような信憑性を持つ質問を提出すると真に検索したいデータベースを隠すことができると考えられる．

第 8 章

評価実験

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