

公募シンポジウム

## 公募シンポジウム3

### セキュアかつフェアなデータ流通プラットフォームとセキュリティ基盤技術

2019年11月23日(土) 14:40 ~ 16:40 C会場 (国際会議場 2階国際会議室)

#### [3-C-2-03] 文書データに対するプライバシーリスク評価

○三本 知明<sup>1</sup>、清本 晋作<sup>1</sup>、北村 光司<sup>2</sup>、宮地 充子<sup>3</sup>（1. 株式会社KDDI総合研究所、2. 産業技術総合研究所、3. 大阪大学）

キーワード：Privacy, Privacy Metric, Document Data

文書データは個人の属性情報と同様、様々なシーンで公開、提供されることがある。例えば学校で発生した事故レポートは、事故防止のための研究に用いられることがあるが、そういった文書データには個人名やセンシティブな情報が含まれていることがあり、取り扱いには十分に気をつける必要がある。本発表ではこれまであまり議論されてこなかった文書データ、特に学校で発生した事故に関する文書に対するプライバシーリスクについて議論を行い、想定される攻撃者モデルやリスク評価指標の提案を行う。我々は実際の学校で発生した事故レポートの分析および評価実験を実施し、サニタイズ処理の課題の発見や提案する手法の有効性を確認しており、本発表ではその結果を元に試作した文書データのリスク評価ツールの紹介をする。

# 文書データに対するプライバシーリスク評価

三本知明<sup>\*1</sup>、清本晋作<sup>\*1</sup>、  
北村光司<sup>\*2</sup>、宮地充子<sup>\*3</sup>

\*1 株式会社 KDDI 総合研究所、  
\*2 産業技術総合研究所、\*3 大阪大学

## Privacy Evaluation Method for Documents

Tomoaki Mimoto<sup>\*1</sup>, Shinsaku Kiyomoto, Koji Kitamura<sup>\*2</sup>, Atsuko Miyaji<sup>\*3</sup>

\*1 KDDI Research, Inc.,

\*2 National Institute of Advanced Industrial Science and Technology,

\*3 Osaka University

A document is sanitized when the document is open to the public or provided to an outsider. Court documents, medical records, and accident reports are used for social development, such as new drug development and accident prevention, and some processes for privacy preservation are needed when they are published. We discuss the privacy risk for document, which has been seldom discussed so far, and propose an adversary model and a metric of privacy risk for documents. Our main target is accident reports in school and we analyze actual accident reports and evaluate the risk of them to find conditions of sanitizing and confirm the efficiency of proposal method. Furthermore, we have developed a prototype of a risk evaluation tool for documents. We have confirmed that the tool enables to find risks which are not found manually.

Keywords: Privacy, Privacy metric, Document

## 1. まえがき

### 1.1 はじめに

ドキュメントの墨塗り等の加工(以後、サニタイズ処理)は、ドキュメントの機密レベルを下げて、公開できるようなドキュメントにすることを目的としている。例えば府省は「行政機関の保有する情報の公開に関する法律」や「独立行政法人等の保有する情報の公開に関する法律」などに基づき、保有する文書を公開している。また、医療機関では患者の医療記録をセンシティブな情報、例えば患者を特定する情報や生命に関わる病名などをサニタイズ処理し、研究に役立てている。このようにドキュメントのサニタイズ処理は様々な分野で必要とされているが、体系化された処理方法は存在しておらず、多くは人手で行っている。そのためサニタイズ処理の程度がドキュメントによって異なる、大量のドキュメントを処理するのに非常に時間がかかってしまうという課題がある。さらに、特に専門的なドキュメントに関しては、どの情報を削除するかを決定するために専門家の知識が必要となる場合があり、そういった面でも膨大なコストがかかってしまう。このようにドキュメントのサニタイズ処理が求められる分野が多い一方で、ドキュメントのプライバシーリスクに関する研究は多くはない。

本稿ではドキュメント、特に学校で起きた事故レポートのサニタイズ処理の効率化に向けて検討する。具体的には日本スポーツ振興センターが公開している実際の事故レポートを分析し、センシティブな情報や文書データにおけるリスクの検討を行う。また、今後のドキュメントのリスク評価の研究に向けて、単語の情報量をもとにしたリスクとなる単語の検出実験、FastText を用いた事故レポートの判定実験を実施し、それらの結果をもとに簡易的なアプリケーションの試作開発を行う。

### 1.2 既存研究

ドキュメントにおけるリスク評価やサニタイズ処理に関する既存研究として、[1, 2, 3]などが挙げられる。[1]では、あらかじめ高リスクな単語が決められており、単語の情報量と自己相互情報量をもとに高リスクな単語との関連性が一定以上ある

単語を墨塗りする。[2]ではドキュメントはエンティティとタームから構成されていることを前提としており、各エンティティには関連するタームがセットされている。また、あらかじめ保護すべきエンティティが決められている。[2]では $k$ -匿名性[6]と似た K-safety という概念を取り入れており、ドキュメントに含まれるタームから推測されるエンティティが  $K$  個以上となるようにタームを墨塗りする。[3]では各単語が人や場所、住所等何を意味しているかといった内容理解や、データマイニングアルゴリズムを利用して特定の単語に関連する単語の判定、「He」などの代名詞がどの単語を表しているかの判定などの機能を含むサニタイズツールの試作を行っている。いずれの既存研究でも完全な自動化は不可能であるとの見解を示しており、有用性の評価も人がサニタイズ処理したドキュメントを正解とした評価を行っている。

### 1.3 貢献

本稿の貢献は以下の三点である。まず、これまで明確にならなかつたドキュメント(特に本稿では、学校の事故レポート)に含まれるプライバシーリスクについて検討し、攻撃者モデルおよびプライバシーリスクの指標を提案する。次に実際の加工前後の事故レポートを分析することで、提案する攻撃者モデルやリスク評価指標の妥当性を確認し、プライバシー向上に向けた検討を行う。最後に、以上の検討を踏まえて実際の事故レポート作成時および公開時に利用することを想定したドキュメントのリスク評価ツールを試作する。ドキュメントの匿名化やリスク評価に関する既存研究は少なく、また日本語に対応したリスク評価ツールは、本稿での試作がはじめての試みとなる。

## 2. 事故レポート分析

### 2.1 事故レポートについて

本章では、過去に発生した 119 件の死亡事故レポートの分析を行う。これらの事故レポートは加工してプライバシーリスクを低減した上で、日本スポーツ振興センターが学校事故事例

検索データベースとして公開している ([https://www.jpnsport.go.jp/anken/anken\\_school/anken\\_school/tabid/822/Default.aspx](https://www.jpnsport.go.jp/anken/anken_school/anken_school/tabid/822/Default.aspx))。加工前の事故レポートはドキュメント(発生状況)のほか、死亡障害種、発生年月日、学年種、被災学年、性別、発生場所、既往歴、災害共済給付金額などの情報がメタデータとして付随しているが、日本スポーツ振興センターが公開する情報はドキュメントのサニタイズ処理以外に、発生年月日を発生年に汎化する、既往歴や災害共済給付金額の情報などセンシティブな情報を削除する、といった処理がされている。

## 2.2 分析結果

事故レポートを分析した結果を本章で述べる。本稿で調査した事故レポートは 119 件であり、実施した内容は次のとおりである。まず、実際の事故レポートと公開用に加工した事故レポートとを比較し、加工の対象となる単語や表現を抽出したところ、加工された単語は次のように分類することができた。

- 変更なし、あるいは細かな日本語の修正

ドキュメントにプライバシー性の高い情報が含まれていない場合、サニタイズ処理は行われない。一部のドキュメントに関しては日本語の文法的な誤りがある場合や、一文が長い場合があり、そのようなドキュメントは読みやすいように修正されることがあるが、本質的にはドキュメントの内容は変わっていない。今回調査したドキュメントについては 119 件中 31 件が該当した。

- 一意に特定可能な単語の処理

加工前の事故レポートには病院名や活動を行っていた施設名がそのまま記載されているものがあつた。これらの情報は当然プライバシー性の高い情報であり、例外なく〇〇総合病院→総合病院や△△公園→公園というように一般化、あるいは削除されていた。今回調査したドキュメントには 119 件中 24 件に個人や場所が一意に特定可能な単語が含まれていた。

- 特徴的な単語の処理

加工前の事故レポートには校舎名や委員会名、階数などそれだけでは特定できないが、その事件・事故を特定するための有効な手がかりとなる補助的な情報を持つ表現が見られた。これらの情報は加工における明確な基準はなく、サニタイズ処理者の主観により対応(処理を行わない、汎化する、削除する)が異なっていた。しかし加工後の事故レポートを確認すると、特徴的な単語が残っている場合があり、実際にその単語と事故・事件の状況をキーワードとして Web 検索を行うと、事故・事件の推測が可能になるケースが存在した。このように特徴的な単語をどのように扱うかの検討が今後必要となる。今回調査したドキュメントではこのような単語の一般化および削除は 119 件中 14 件で行われていた。

- 医療措置の処理

加工前の事故レポートには輸血、CT など医療措置の詳細が記載されているものが複数確認できた。これらの情報は、事故・事件の詳細を知る知人や友人、その家族であれば事

故レポートと実際の事故・事件とを紐付ける手がかりとなりうる。また加工後の事故レポートを確認しても、措置内容が全く処理されていないものや、全て削除されているものなどサニタイズ処理者によって対応が大きく異なっており、これらの情報も加工における明確な基準はないことがわかる。今回調査したドキュメントには 119 件中 12 件に医療措置に関する単語が含まれていた。

- 児童の様体説明の処理

加工前の事故レポートには意識不明、心肺停止など児童の様体に関する記述があるものが複数確認できた。これらの情報も、医療措置に関する単語と同様、事故・事件の詳細を知る知人や友人、その家族であれば事故レポートと実際の事故・事件とを紐付ける手がかりとなりうる。また加工後の事故レポートを確認しても、全く処理されていないものや、全て削除されているものなどサニタイズ処理者によって対応が大きく異なっており、これらの情報も加工における明確な基準はないことがわかる。今回調査したドキュメントには 119 件中 17 件に児童の様体に関する単語が含まれていた。

- 日時の処理

加工前の事故レポートには日時や経過時間など時間に関する単語が 119 件中 67 件と多数含まれていた。これらの情報も、医療措置や児童の様体に関する単語と同様、事故・事件の詳細を知る知人や友人、その家族であれば事故レポートと実際の事故・事件とを紐付ける手がかりとなりうる。また加工後の事故レポートを確認したところ、大部分が何かしらの処理を行っていたが、サニタイズ処理者によってその程度は様々であり、これらの情報も加工における明確な基準はないことがわかる。例えば 15 時に部活動の練習が始まり、16 時に異変があつた際、サニタイズ処理者によっては時間に関する表現を全て削除するケースや、練習開始 1 時間後というように経過時間のみを残すケースがあつた。

- 死因の処理

加工前の事故レポートには多発外傷、心不全など死因に関する記述が複数確認できた。これらの情報も、医療措置や児童の様体に関する単語と同様、事故・事件の詳細を知る知人や友人、その家族であれば事故レポートと実際の事故・事件とを紐付ける手がかりとなりうる。また加工後の事故レポートを確認しても、全く処理されていないものや、全て削除されているものがあり、更には死因が追記されている事故レポートもあり、これらの情報も加工における明確な基準はないことがわかる。今回調査した事故レポートには 119 件中 8 件に死因が記載されていた。

実際の事故レポートを確認したところ、特徴的な単語でも加工されずに残っている場合があり、これらの単語に対する処理が必要であることがわかる。また、ドキュメントのサニタイズ処理は従来手作業で行われているため時間がかかり、サニタイズ処理者の知識や考え方によって処理の程度が異なる。したがって、定量的なリスク評価指標とドキュメントの自動加工、あるいは加工を補助するツールが必要である。

### 3. 文書データにおける攻撃者とリスク評価

事故レポートのリスク評価を行うにあたり、ドキュメントに含まれるリスクとは何かを検討する必要がある。主な従来研究では保護すべき対象があらかじめ決められており、ドキュメントとそのドキュメントから識別され得る自然人(データ主体)との結びつきについては考慮していない。例えば医療データの場合、ドキュメントに含まれるエイズなどのセンシティブな情報が保護すべき対象であり、そのドキュメントが誰について書かれたものかは考慮していない。

事故レポートに関しては考え方が異なり、ドキュメントには事故の状況等が記載され、それ自体がセンシティブな情報となることは少ない。しかし、本研究で検討する事故レポートには付随するメタ情報に既往歴や災害共済給付金額などのセンシティブな情報が含まれる。そのため事故レポートについては、事故から個人が結びつき、さらにそこからセンシティブな情報が漏れることが考えられる。したがって、データを第三者に提供する場合にセンシティブなメタ情報を削除する、事故レポートに該当する児童を推測できないように加工するといった必要がある。

日本スポーツ振興センターが保有する一般に公開する情報には既往歴、災害共済給付金額等のセンシティブな情報が含まれていないが、こういったデータは事故予防等を目的として、研究機関などに提供する場合がある。研究目的である場合や同意を取得している場合であればデータの提供は可能であるが、安全管理措置の一つとしてサンタイズ処理を施してから提供することが考えられる。このとき、故意でなくともドキュメントから当該児童が特定され、センシティブな情報と紐付くことがプライバシーリスクの一つとして挙げられる。

以上のことから、事故レポートに含まれるリスクとは、事故レポートと実際の事故とが紐づき、結果として当該児童が特定されることとして考えることができる。またドキュメントをそのまま扱うことが難しいため、MeCab (<https://github.com/neologd/mecab-ipadic-neologd>) を用いて形態素解析を実施し、助詞、助動詞を除いた単語の集合として扱う。このとき次のような攻撃者モデルを想定することができる。

**攻撃者モデル:** 攻撃者は事故 $A_i$ に関するドキュメント $R_i = \{r_1, \dots, r_m\}$ を持つ。攻撃者は Web にアクセスすることができ、 $R_i \subseteq R_i$ を検索キーワードとして $A_i$ を推測する。

なお、[2]のようにドキュメントを単語の集合として $k$ -匿名化と同様に考えることも可能だが、特に学校の事故においては、たとえ $k = 2$ 程度であってもドキュメントをそこまで一般化すると有用性が著しく低下することが予想される。

このような攻撃者モデルを想定した場合、リスクの指標として次の二つを想定する。まず一つ目は自己情報量を用いた評価指標である。ドキュメントに含まれる単語の情報量が多いほど、検索により実際の事故とドキュメントが紐づく可能性が高まるため、単語の持つ情報量はリスクの指標として扱うことができる。情報理論ではある事柄 $x$ の情報量 $I(x)$ は次のように定義される。

$$I(x) = -\log P(x)$$

ここで $P(x)$ は $x$ が起こる確率を表す。本稿では全ドキュメントを $D = \{R_1, \dots, R_n\}$ とし、 $\#D$ を $D$ に含まれる全単語の個数、 $\#D(x)$ を $D$ に含まれる $x$ の個数とすると、

$$P(x) = \#D(x)/\#D$$

として表される。したがって最も単純なドキュメントのリスク評価指標として、 $\forall x \in R_i$ に対して $I(R_i) = \sum I(x)$ とすることができる。特に本稿で扱う事故レポートについては、ドキュメントの他に様々なメタ情報が付随しているため、精度を上げるためにすべての事故レポートにおける単語の数を分母とするのではなく、被災学年や性別、発生場所等で絞ることもできる。[1]では共起確率も用いているが、事故レポートに関してはほとんど効果がないと考えられる。例えば、医療データの場合、病名と共起する治療法や薬剤は病名を特定するリスクとなりうるが、事故レポートについては個人名や地名などが主なリスクとして考えられるため、それに共起する単語を評価要素に含めることによるメリットは考えにくい。

二つ目のリスク評価指標では実際に攻撃を行う、つまりドキュメントに含まれる単語 $r \in R_i$ を組み合わせて検索ワードとし、Web 検索を実施することを想定する。定量的な評価としては、検索結果上位のページのうちの当該事故・事件の割合などを用いることができる。ドキュメントに含まれる単語の数と Web 検索のキーワードが複数利用できることを考えると、その組み合わせは非常に大きくなることが予想される。そのためこの評価指標を用いる場合、情報量の大きい単語だけを用いる、検索キーワードの数を制限するなどの対応が必要となる。

例えばドキュメント $R_i$ に出現する情報量の大きい上位 $n$ 個の単語を $\max^n(I(r \in R_i))$ とし、 $r \in R_i$ をキーワードとして、 $A_i$ に関する記事が Web 検索により取得できる確率、例えば上位 10 件中 $A_i$ に関する記事が何件あるか、を $P(B(r) = A_i)$ とすると、最も単純な評価指標としては、 $P(B(\max^n(I(r_j \in R_i))) = A_i)$ などで表すことができる。

## 4. 評価実験

### 4.1 情報量によるリスク評価

我々が提案するリスク評価手法の有効性の評価を実施した。具体的には、加工前の事故レポートに含まれる単語の情報量を計測し、情報量が高い単語が加工後の事故レポートではどのように扱われているかを確認した。

実験では約 30 万件の加工済みの事故レポートをデータセットの全体 $D$ として、 $D$ に対して形態素解析を実施し、単語のリストを得る。次に加工前の事故レポート119件に対して形態素解析を実施、各単語の情報量を計測した。単語の総数 $\#D$ は約 1,500 万語であり、最も情報量が高い単語 $x$ は出現回数 $\#D(x) = 1$ で $I(x) = 23.8$ であった。この単語は 93 個存在し、そのうち加工後の事故レポートでも加工されなかった単語は 35 個あった。これらの単語を含む事故レポート 35 件に対して、各単語と各レポートに出現する別の単語をキーワードとして Web 検索を実施したところ、35 件中 9 件で該当する事故、あるいは事件の記事を確認することができた。続いて情報量が高い単語は出現回数が 2 回、3 回の単語で、それぞれ情報量は 22.79、22.21 であった。出現回数が 2 回の単語は 40 個存在し、そのうち加工後のレポートでも加工されなかった単語は 20 個あった。これらの単語を含む事故レポート 20 件に対して、各単語と各レポートに出現する別の単語をキーワードとして Web 検索を実施したところ、20 件中 3 件で該当する事故、あるいは事件の記事を確認することができた。出現回数が 3 回の単語は 36 個存在しており、そのうち加工後のレポートでも加工されなかった単語は 21 個あった。これらの単語を含む事故レポート 21 件に対して、同様に Web 検索を実施したが、

1 件も該当する事故、あるいは事件の記事を確認することができなかった。

情報量が多いが加工対象とならなかった単語を確認したところ、「北進」のような普段あまり使わない表現や、「外掛け」のように特定の競技で使われる技の名前や道具の名前、またサニタイズ処理者によって対応が異なる時間など数字に関する情報や、病名、児童の様体に関する単語であることが分かった。次に加工対象とならなかったが、それをキーワードとすることで当該事故・事件と紐づくような単語を確認したところ、車の車種や特殊な部活、数字に関する情報などがあつた。このことから情報量を元にしたリスク評価手法は有効である一方で、情報量だけでは実際はリスクとはならない単語であっても高リスクと判定してしまう可能性があることが分かる。

情報量	単語数	実際には加工されなかった単語数	実際の記事と結びついた単語数
23.80	93	35	9
22.79	40	20	3
22.21	36	21	0

表 1: 実データのリスク評価実験

また、部活内で発生した怪我などに比べて、殺人事件や大きな被害があつた事故、死亡事故は容易に検索できることから、提案する攻撃者モデルにおいては、事故・事件の話題性や被害規模も評価指標の要素として考えることができる。

#### 4. 2 FastText によるレポート種別の判別

次に、事故レポートの種別判定(加工前と加工後のレポート判定、および死亡事故と非死亡事故の判定)が可能であることを確認するため、FastText[5]による実験を実施した。FastTextとはFacebook AI Researchがオープンソースプロジェクトとして公開している自然言語処理ライブラリ(<https://github.com/facebookresearch/fastText/>)であり、Word2Vec (<https://code.google.com/archive/p/word2vec/>) [7]と同様単語をベクトルで表現し、高速なテキストの分類を可能とする。加工前の死亡事故レポートの数が少なく、十分な学習ができなかったため、加工前の非死亡事故レポートと加工後の死亡事故レポートを利用して以下の実験を実施した。

最初の実験では約 650 件ずつの加工前の非死亡事故レポートと加工後の死亡事故レポートにそれぞれ異なるラベルを付与して教師データとし、別の約 650 件ずつの加工前の非死亡事故レポートと加工後の死亡事故レポートに対してラベルの推測を行った。実験の結果を表 2 に記す。実験結果から、ほぼ確実に加工前の非死亡事故レポートと加工後の死亡事故レポートとの判別ができていることが分かる。しかし、これでは加工前と加工後の判定ができているのか、死亡事故と非死亡事故の判定ができているのかの区別ができないため、追加で以下の実験を実施した。

	Positive	Negative
True	650	646
False	10	6

表 2: 加工前の非死亡事故と加工後の死亡事故の判定

先の学習モデルを用いて 116 件の加工前の死亡事故レポートと 115 件の加工後の死亡事故レポートに対してラベルの推測を行った。実験の結果を表 3 に記す。実験結果では、ほ

ぼすべてのレポートが加工後の死亡事故レポートとして判定されたことが分かる。この実験結果から先の実験では死亡事故と非死亡事故の区別ができていることが分かる。先にも述べたとおり、我々が提案する攻撃者モデルを想定する場合、死亡事故と非死亡事故ではそもそもドキュメントが持つリスクが異なると考えられる。したがって、この評価結果はリスク評価の一つの要素として扱うことができると考える。

	Positive	Negative
True	1	115
False	0	115

表 3: 加工前後の死亡事故の判定

#### 5. 試作ツール

これまで述べた通り、攻撃者モデルやリスクの定義はドキュメントの種類によって大きく異なる。本研究で試作したツールは学校における事故レポートの作成者およびサニタイズ処理者を利用者として想定している。また従来研究にもあるようにドキュメントの完全な自動化は困難であるため、サニタイズ処理の支援ツールとして開発した。試作したツールの外観を図 1, 2 に記す。図 1 では、事故レポートに付随するメタデータを入力する。プライバシリスクは情報量を用いた評価が行われるため、チェックボックスにチェックを入れることでデータのクラスタリングを行うことができ、そのクラスタにおける情報量の評価が行える。こうすることで大量のデータが必要になるが、例えば「沖縄」に限ると「雪」という単語の情報量が大きくなるというように、クラスタごとに特異な単語を見つけることができる。図 2 では実際のレポートを記述する。レポートを記述後、評価ボタンを押すことで入力文書を形態素解析し、各単語の情報量を計算する。しさい値を用いて、クラスタをデータセット全体とした場合の情報量と、全ドキュメントをデータセット全体とした場合の情報量から高リスクとなる単語は赤、中リスクとなる単語は青で表示し、ユーザに修正を求める仕組みとなっている。初期版では Web 検索による評価は行っていないが、FastText によるドキュメント自体の危険度(死亡事故か非死亡事故か)の判定は行っており、特にプライバシリスクが高いドキュメントについては左側の丸の色が青色から赤色に変更され、より注意が必要であることを表す。

現段階では「北進」や「外掛け」といった情報量が多いがリスクが低い単語も中リスクとして判定されるが、表 1 にあるようなリスクのある単語を高リスクとして判定できることを確認している。

条件抽出	識別番号	属性名	内容
<input type="checkbox"/>	1	発行年度	2014
<input type="checkbox"/>	2	都道府県名	東京都
<input type="checkbox"/>	3	市町村名	AAAA市
<input type="checkbox"/>	4	被災学校種別名	小学校
<input type="checkbox"/>	5	被災学校区分名	公立
<input type="checkbox"/>	6	被災学校分類名	小学校
<input type="checkbox"/>	7	被災学年	3
<input type="checkbox"/>	8	被災年齢	9
<input type="checkbox"/>	9	性別表記	男
<input type="checkbox"/>	10	被災年月日	20140101

図 1: 試作ツール画面 1

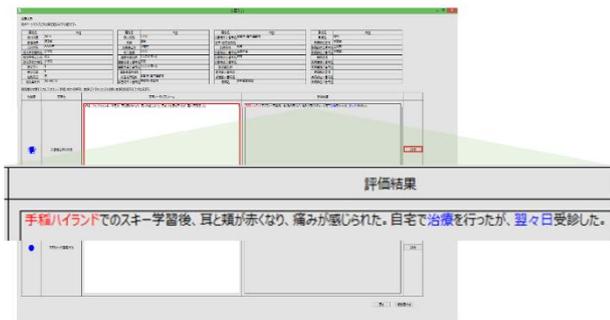


図 2: 試作ツール画面 2

## 6.1 成果と貢献

本稿ではドキュメント、特に学校で発生した事故レポートにおけるプライバシーリスクについて検討し、攻撃者モデルおよびリスク評価手法を提案した。さらに実際の事故レポートを分析し評価実験を行うことで、提案する手法の有効性を確認した。また、分析結果および実験結果を元にドキュメントにおけるリスク評価ツールを試作した。試作したツールは実際にリスクのある単語の判定が行えることを確認している。

## 6.2 今後の展望

今後はより多くの実データを用いてプライバシーリスクの検討を進め、ドキュメントの話題性なども考慮したリスク評価指標の確立を目指す。また実際の事故レポートでは知人、友人、またはその家族などある程度その事故・事件の情報を持つ人間を想定した加工を行っているため、今回提案した攻撃者モデルよりも強い攻撃者モデルの定義も検討する必要がある。ツールについては実際にユーザからのフィードバックを受けて UI などのユーザビリティを改善するとともに、WordNet (<http://compling.hss.ntu.edu.sg/wnja/>)を活用するなどしてリスクのある単語の一般化を行うといった自動化を進める。さらに、情報量が大きい但实际上はリスクの小さい単語の判別方法を確立することで、リスク評価の精度向上を目指す。

## 謝辞

本研究の一部は JST, CREST (JPMJCR1404) の支援を受けたものである

## 参考文献

- 1) S{¥}anchez, David and Batet, Montserrat, "C-sanitized: A privacy model for document redaction and sanitization," Journal of the Association for Information Science and Technology, 148--163, 2016, Wiley Online Library.
- 2) Chakaravarthy, Venkatesan T. and Gupta, Himanshu and Roy, Prasan and Mohania, Mukesh K, "Efficient techniques for document sanitization," Proceedings of the 17th ACM conference on Information and knowledge management, 843--852, 2008, ACM.
- 3) Bier, Eric and Chow, Richard and Golle, Philippe and King, Tracy Holloway and Staddon, Jessica, "The rules of redaction: Identify, protect, review (and repeat)," IEEE Security & Privacy, Vol. 7, No. 6, 46--53, 2009.
- 4) [https://www.jpnsport.go.jp/anzaen/anzen¥\\_school/anzaen¥\\_school/anzaen¥\\_school/tabid/822/Default.aspx](https://www.jpnsport.go.jp/anzaen/anzen¥_school/anzaen¥_school/anzaen¥_school/tabid/822/Default.aspx)
- 5) Joulin, Armand and Grave, Edouard and Bojanowski, Piotr and Douze, Matthijs and J{¥}eou, H{¥}erve and Mikolov, Tomas,

"FastText.zip: Compressing text classification models," arXiv preprint arXiv:1612.03651, 2016.

- 6) Sweeney, Latanya, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol.10, No. 05, 557--570, 2002.
- 7) Mikolov, Tomas and Chen, Kai and Corrado, Greg and Dean, Jeffrey, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.