

匿名加工情報公表サイト調査

(2)加工対象情報の抽出

菊池研究室 B4 金子侑紀

研究背景

○「こころのサポートシステム 面談カウンセリング内容」を基に作成した匿名加工情報に含まれる個人情報（要配慮個人情報を含む。）の項目

- ・ カウンセリング実施月
- ・ 年齢
- ・ 利用者区分
- ・ カウンセリング実施回数
- ・ 実施区分
- ・ 勤続年数
- ・ 職種

○「ハロー健康相談 24 などの相談内容」を基に作成した匿名加工情報に含まれる個人情報（要配慮個人情報を含む。）の項目

- ・ 年齢
- ・ 相談内容
- ・ 回答内容
- ・ 病名

1. 匿名加工情報に含まれる情報項目について

以下の項目について、一部削除や置き換え等の措置を施した上で、匿名加工情報の作成および第三者提供を行います。

- (1)HOPカード番号
- (2)統合ID番号
- (3)会員コード
- (4)世帯ID番号
- (5)年齢
- (6)住所コード

※提供データには「氏名」「生年月日」「電話番号」は一切含まれておりません

2. 匿名加工情報の第三者への提供方法について

匿名加工情報についてデータファイル化を行い、セキュリティが確保されたインターネット経由または電子媒体等の方法により情報提供を行います。

<http://www.t-pec.co.jp/privacy/t-privacy3.html>

https://www.heiwado.jp/release/2018/0828_anonymous.html

研究目的

- 「(1)自動クローラーシステムの開発」で収集したデータから必要な情報を取得する
 - 個人に関する情報の項目
 - 提供手法

同業他社はどんな情報を
公開しているんだろう?



手動収集データに対する抽出

手作業で該当部分の抜き出しを行った後、5パターンの正規表現を用いて抽出する

(次のとおり 下記 以下)+([\s\S]+)\n([\s\S]+提供[の])*方法	53
DPC+([\s\S]+)\n([\s\S]+提供[の])*方法	19
情報[の]*項目([\s\S]+)\n([\s\S]+提供[の])*方法	5
項目は([\s\S]+)です	4
([\s\S]+)(上記項目 提供[の])*方法 第三者に提供)	115
計	196
抽出成功	196
抽出失敗	40
抽出対象が含まれていない(ため除外)	72
計	308

抽出結果の一部

企業名	個人に関する情報の項目	提供方法
株式会社日本医薬総合研究所	年齢(生年), 性別, 処方せん情報, 調剤情報, 各種アンケート回答	電子メール, CD-ROM, USB等の外部記憶媒体, HTTPS
セキ薬局	氏名, 生年月日, 被保険者記号番号, 公費受給者番号, 医師氏名, 処方日, 調剤日, 性別, 生年, 処方・調剤履歴	セキュリティが確立された転送方式
イオン銀行	性別, 年代, 申込手段, 当行普通預金口座の有無, 現在の借入の有無, 契約から初回借入までの経過日数, 現在の返済実績等	パスワードで保護し, CD-ROMで手交

抽出に失敗した例

具体的にはオーディエンスデータから個人を特定できる特異なデータ、個人を特定できる可能性のあるデータ等を排除または置き換え、同じ属性を持つデータが同一データセット内においてデータ件数に応じ複数件以上ある状態になるまで匿名加工する等を実施します。

株式会社マイデータ・インテリジェンス https://www.meyportal.com/policy/user_protect

抽出したい情報

- オーディエンスデータ
- 個人を特定できる特異なデータ
- 個人を特定できる可能性のあるデータ

抽出対象が含まれていない例

匿名加工情報の作成および第三者提供について

(1) 当行は、匿名加工情報を作成した場合には、匿名加工情報に含まれる個人に関する情報の項目を公表いたします。

(2) 当行は、匿名加工情報を第三者提供する場合には、提供する匿名加工情報に含まれる個人に関する情報の項目および提供の方法について公表するとともに、提供先に、提供される情報が匿名加工情報である旨を明示いたします。

株式会社 阿波銀行 <http://www.awabank.co.jp/policy/privacy/>

抽出システムの精度評価

(次のとおり 下記 以下)+([¥s¥S]+)¥n[¥s¥S]+提供[の]*方法	44
DPC+([¥s¥S]+)¥n[¥s¥S]+提供[の]*方法	14
情報[の]*項目([¥s¥S]+)¥n[¥s¥S]+提供[の]*方法	4
項目は([¥s¥S]+)です	3
([¥s¥S]+)(上記項目 提供[の]*方法 第三者に提供)	60
計	125
抽出成功	125
抽出したデータが過剰	27
抽出失敗	167
計	319

抽出したデータが過剰な例

...サイト内検索(中略)取り組み

匿名加工情報の作成及び第三者提供について

匿名加工情報の取り組み

匿名加工情報に関する問合せ窓口

匿名加工情報の取り組み

DPC制度の導入の影響評価及び今後のDPC制度の見直しを図る目的で、厚生労働省が収集し管理する情報となるデータ（DPCデータ）を作成しております。また、審査支払機関への請求のため診療に係る費用を診療報酬明細書（レセプト）として作成しております。

DPCデータは、診療録からの情報および診療報酬明細書からの情報で構成されており、レセプトデータは、医療機関情報・保険者情報・診療行為情報・医薬品情報・特定器材情報等から構成されております。

DPCデータ並びにレセプトデータを利活用することで、医療の質向上および病院経営の改善に役立てる事が可能になるため、匿名加工後のデータを第三者へ提供しております。第三者提供するこれらのデータは氏名、住所、電話番号は含みません。なお、地域傾向や受診年齢層等进行分析するため、郵便番号（上3桁のみ）、生年月日(生年月及び入院時年齢に変換を行い100歳以上は100歳に一括り)、各種保険証に関する情報については保険者番号（健康保険事業の各運営主体を指す番号）のみを含みます。

当院は上述の通り、診療情報から匿名加工情報を作成（毎月継続）し、第三者に提供しております。

匿名加工情報の提供の方法

取得に失敗したファイル

原因	件数
PDFファイルの破損	75
サーバーエラー・ファイルの破損	25
提供しないと明記しているサイト	5
公表サイトでない	3
項目と方法は別ファイルに記載	11
抽出対象が含まれていない	42
抽出失敗	8
計	169

まとめ

- 公表サイトの表記パターンを5種類に分類した
- 公表サイトから項目と個人に関する情報の自動抽出を実現した
 - 手動収集データ: 196/308
 - 自動収集データ: 125/319
- 個人に関する情報の項目と提供方法を明示していない公表サイトが多く存在する
 - 手動収集データ: 72/308
 - 自動収集データ: 42/319