

卒業論文2016年度

嘘データの用途と需要その応用例に関する研究（仮）

慶應義塾大学総合政策学部 4 年
野村隼斗

嘘データの用途と需要その応用例に関する研究（仮）

概要

現代において、嘘データは様々な場面で必要になってくるものである。

本論文は、嘘データの用途と需要において調査し、既存の嘘データ作成法を紹介したものである。

さらに、その応用例について考察した。

目次

第1章 序論

1.1 研究背景

1.2 研究目的

1.3 嘘データの用途

1.3.1 EpisoPass

1.3.2 テスト

1.3.3 小説,漫画,成りすまし

第2章 作成方法

2.1 既存の嘘データ生成サービス

2.1.1 なんちゃって個人情報

2.1.2 疑似個人情報データ生成サービス

2.2 既存の類語検索

2.2.1 ymrl式

2.2.2 らいばるサーチ やーやら

第3章 考察

3.1 応用例

3.2 結論

謝辞

参考文献

第一章 序論

本章では研究の背景と目的について説明する。

1.1背景

嘘データは様々な場面において求められる。小説を書くときの設定、信頼の置けないサービスに登録するときの入力するための個人情報、会員制webサービスを作ったときのテスト用大量の個人情報、強力なパスワード生成に用いるため、現代において嘘データは多様な場面で必要とされていると考えられる。

1.2目的

嘘データはしばしば生成が煩雑である。

小説に用いるための登場人物の氏名や住所、webサービスのテストに用いるための氏名、アカウント名、電話番号、メールアドレスなど手作業で大量に作るのは苦行であると考えられる。

偽データの需要と既存の偽データ生成サービスの調査、そしてその応用例の調査が本論文の目的である。

1.3用途

1.3.1 EpisoPass[1] [2]

EpisoPassとは慶應義塾大学環境情報学部 増井俊之教授が開発し、運用しているパスワードを自動生成するシステムだ。人間は新しく覚えた情報を必ず忘れるものであるならば、新しいパスワードを考えたり覚えたりする努力は不毛である。一方、忘れることがないエピソード記憶を誰もが持っているのであれば、そのようなものを利用してパスワードを生成する方が妥当であると思われる。そういうアイデアに基づいて増井教授は「EpisoPass」を開発された。

[[画像]]

EpisoPassにおいて重要な要素の一つが認証の際の偽データによる選択肢の生成である。

これが生成されないことにはこのサービスは成立しない。

また、同じような重みの情報が上手に生成されれば、うれしい。姓であれば同程度の世帯姓、住所であれば似たような緯度・経度の住所.....

調べる

1.3.2テスト(会員制サービス...)

会員制webサービスのテスト用に一定量のまとまった個人情報があると嬉しいということがある。

なんちゃて個人情報[2]や疑似個人情報データ生成サービス[3]を用いる事ができる。

1.3.3小説,漫画,記者のなりすまし等

小説や漫画などのフィクション作品における登場人物の個人情報や住所等が必要になる場合がある。

これらは架空のものである場合が多くわざわざ生み出さなくてはならないものだ。

また、新聞記者の友人は危険人物に取材を行う際、しばしば市民等になりすまし取材を行うようである。このとき行う個人情報生成が煩雑なようなので生成サービスがあれば、便利である。

1.3.4試験問題の嘘選択肢

あらゆる全て人間が生成しているが、四択問題などはテスト問題データベースなどを利用して同カテゴリ・頻度・重みの情報を簡単に生み出すことができると嬉しい人も少なくはないはずである。

第二章 作成方法

本章では既存の嘘データ生成法を紹介する。

2.1 既存サービス

2.1.1 なんちゃって個人情報[3]

ダミーの個人情報を大量に出力できるwebサービス。

名前、ふりがな、メールアドレス、性別、年齢、誕生日、婚姻の有無、血液型、都道府県、電話番号、携帯番号、カレーの食べ方の13項目のうち任意の項目を最大5000人分生成し、HTML、XML、CSV、タブ区切りテキストの形式でダウンロードすることができる。これは、webサービスやアプリケーションを開発する人間にとって、あればうれしい場面があるwebサービスである。

2.1.2 疑似個人情報データ生成サービス[4]

これもなんちゃって個人情報[2]と同じようにwebサービスやアプリケーションを開発する人間にとってあればうれしい場面があるwebサービスであるが、こちらは生成項目が、連番、氏名、性別、電話番号（一般）、電話番号（FAX）、電話番号（携帯電話）、メールアドレス、住所、出身地、血液型、乱数、パスワードとなっており、なんちゃって個人情報とは異なっている。

2.2 既存の類語検索

2.2.1 @ymrl式[5]

@ymrl氏が本名っぽい文字列を作るために使った手法
名字データベース等からクロールしてデータを加工、twitterやmixiで
使ったものである。

2.2.2 らいばるサーチ やーやら[6]

「らいばるサーチ：やーやら」は、京都大学情報学研究科の田中
克己教授、大島裕明特定助教等のグループが開発したものであり、
ユーザーが指定した特定の人物・地名・物の同位語（＝ライバルや
仲間にあたる語）や話題語をウェブから検索してグラフに表示する
ものだ。例えば「織田信長」と入力すれば、その同位語や話題語が
具体的にグラフ表示されるというものである。

第三章 考察

本論では今回の研究を踏まえての考察を述べる。

3.1応用例

同じ頻度・重みの情報をもっと簡単に生成するサービスがあれば
嘘データが必要な場面において非常に有用であると考えられる。
殊、EpisoPassにおいて名前を生成する場合に姓であれば、DBから
同程度の世帯数である姓を抽出し、選択肢として生成する。

住所の嘘選択肢を生成する場合は同じような緯度経度から
名前・住所・似た画像(特徴量,deeplearning)

EpisoPass画像版

3.2結論

本論文において、嘘データの需要・用途について調査し、嘘データ生成のための既存サービスを紹介した。また応用例についても述べた。

EpisoPassにおける認証の選択肢という高尚なものから、webサービスにおけるテスト用、また小説、漫画や成りすまし用のデータというもののまでその需要は多岐にわたると考えられる。

これからも嘘データがの需要はあり続けると考えられるし、よりよい嘘データを簡単に生成することができるサービスが生まれる必要があると推察する。

謝辞

本論文を執筆するにあたり、担当の増井俊之教授には多大なるご指導、ご支援を頂きました。

また、研究室の皆様も多くのアドバイスを下さいました。この場を借りて感謝の意を表します。

参考文献

- [1]EpisoPass:エピソード記憶にもとづくパスワード管理
https://www.wiss.org/WISS2013Proceedings/oral/S6_019.pdf
- [2] EpisoPass <http://episopass.com>
- [3]なんちゃって個人情報 <http://kazina.com/dummy/index.html>
- [4]疑似個人情報データ生成サービス<http://hoge hoge.tk/personal/>
- [5]<http://mkdir.g.hatena.ne.jp/ymrl/20110417/1303013060>
- [6]らいばるサーチ やーやら <http://www.kyoiku-press.com/modules/smartsection/item.php?itemid=10767> (2017年1月閲覧)

大島 裕明, 小山 聡, 田中 克己. Web 検索エンジンのインデックスを用いた同位語とそのコンテキストの発見. 情報処理学会論文誌. データベース, 47(19):98–112, December 2006.