企業プレスリリースからのサイバーインシデント情報の自動収集と分析

明治大学総合数理学部
池上　和輝

研究背景

- 不正アクセスや内部犯行などによる情報漏洩の増加

2014年ベネッセ

ベネッセ 赤字136億円

4～6月 情報漏洩で特損260億円

ベネッセホールディングスが31日発表した2014年4～6月期の連結決算は、136億円の最終赤字になった。前年同期は26億円の最終黒字。通信講座などの顧客情報の漏洩で、おわびにかかる費用など260億円の特別損失を計上。問題が業績直結するわけではなく、業績や資金繰りが改善すれば解消される。今回、不採算の地方空港からの撤退や金融機関からの借り入れなどで対応する方針もあわせて発表した。同社は11年、6機のA380を総額1915億円（現在価格）で購入する契約をエアバスと結んだ。だが支払いが滞ったことなどを理由にエアバスが契約解除を通告し、約700億円の違約金を求める意向とみられる。スカイマークは「違約金は合理性がなく、法的手段も視野に入れて対応する」としている。

後、エアバスへの違約金を「相当額負担せざるを得ない可能性がある」としていた。今回、リスクを開示して今回、リスクを開示した。エアバスからの解約通知を受け、29日に開いた会見で西久保慎一社長は「資金繰りに問題はない」としていた。

必ずしも破綻リスクに

特別損失の内訳は顧客への補償に200億円、おわび文書の発送や事件

- 企業の被害
  - データの損失・破損
  - 事業妨害
  - 損害賠償
  - 広報活動

- これらの脅威への対策を行うために,
  多くのインシデント情報を収集し分析する必要がある

先行研究・問題点

先行研究

- 昨年一つのメディアから手動で収集(2017)

JNSA	本調査	共通
788	279	145

JNSA(日本ネットワークセキュリティ協会)

問題点

- 新聞やネット記事を元にすると，網羅的に収集できない可能性がある.

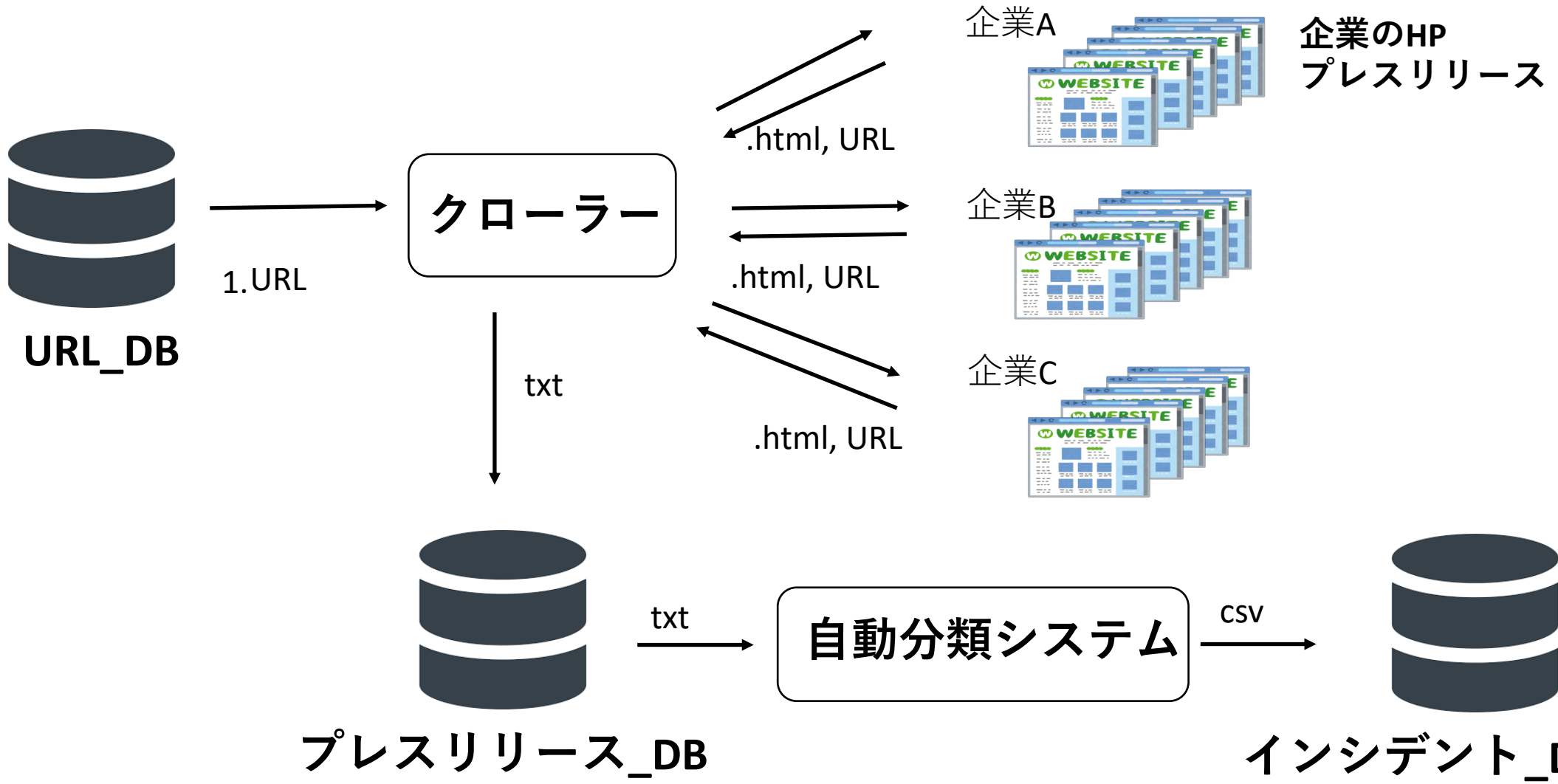
- 人手を使った収集は，コストと時間がかかる.

研究目的・解決方法

- 研究目的
  - メディアなどの偏りなく、網羅的にインシデントを自動収集・分類すること

- 解決方法
  - クローラー・自動分類システムの開発
  - 既存のインシデントデータセットとの比較による評価

システムの全体構成図


企業A
企業のHP
プレスリリース
.html, URL
URL_DB
1.URL
クローラー
企業B
.html, URL
txt
企業C
.html, URL
txt
自動分類システム
csv
プレスリリース_DB
インシデント_DB

自動分類システム(例)

記事(**入力**)

2019/2/2
A社の所有するパソコンに不正アクセスがあったことを確認した.
約3000件の個人情報が流出した可能性がある.

正規表現により抽出

特徴語の出現頻度比較

	入力	人的ミス	不正アクセス	内部犯行
紛失	0	3	0	0
不正アクセス	1	0	2	0
パソコン	1	1	1	4
委託	0	1	1	3
メールアドレス	0	2	0	0
cos類似度		0.33	0.87*	0.45

出力：日付：2019/2/2，規模：3000，原因：不正アクセス

JNSAデータセットの取得項目比較

インシデントリリース(入力)

株式会社ディー・エヌ・エー

2016/04/01

DeNAが運営している「Mobage」において、第三者が「Mobage」ユーザに成りすまし、不正にログインしたと思われる事象が判明した。
不正ログインの確認されたID件数
最大104,847件

取得項目(出力)

	抽出結果	正確(JNSA)
企業名	ディー・エヌ・エー	ディー・エヌ・エー
業種	情報通信・サービスその他	情報通信業
日付	2016/04/01	2016/04/01
被害人数	104847	104847
漏洩原因	不正アクセス	不正アクセス
インシデント内容要約	未定義	有
参照記事のURL		
社会的責任度		普通
漏洩情報区分		個人情報
漏洩経路		インターネット
事後対応		普通

収集結果・JNSAとの比較

- **収集結果**

企業数	期間	取得記事数	インシデント数	インシデント記事割合
537	2004/10/1~ 2018/11/2	17,957	191	(0.01)

- **比較**

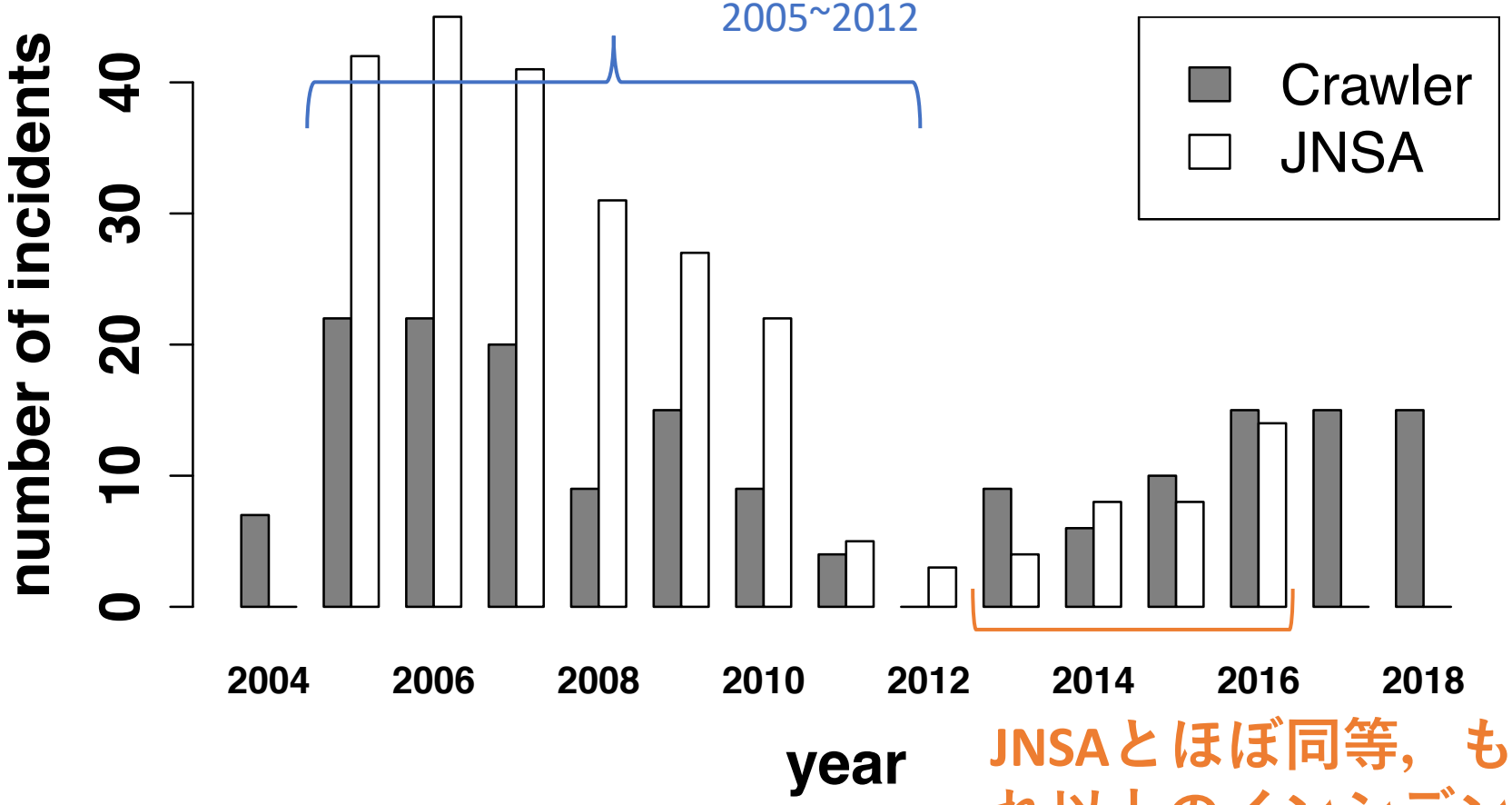
	JNSA	本調査	共通
企業数	65	34	23
インシデント数	251	141	80

インシデント数はJNSAの約６割

経年変化の比較

JNSAの約半分のインシデント数

2005~2012



2013~2016

JNSAとほぼ同等，もしくはそれ以上のインシデント数

共通のインシデントと独自のインシデント例

サイバーエージェント

2010/01/01

不正アクセスにより450件のID、パスワードが流出した可能性がある

JNSA独自

グリー株式会社

2016/12/27

グリーが運営する「GREE」およびスマートフォン向けゲームアプリを管理しているシステムに対して、不正なアクセスがあったことが判明

本調査独自

株式会社ディー・エヌ・エー

2016/04/01

不正にログインと思われる事象が判明した。
不正ログインの確認されたID件数
最大104,847件

共通

自動分類の結果

$$\text{適合率} = \frac{\text{正しく要素を抽出できたインシデント}}{\text{収集した全インシデント}}$$

	日付	被害人数	漏洩原因	日付&規模&原因
適合率	0.882	0.792	0.719	0.505

各要素はそれぞれ**7割以上**の適合率で分類できた
すべての要素を合わせると約5割

まとめ

- クローラーにより191のインシデント記事を収集した
- 2013年以降ではJNSAと同等のインシデントを収集できた
- 独自のインシデントには，不正アクセスのような悪意のあるインシデントも含まれていた
- 要素別の分類では全て７割を超えた適合率で抽出した

収集できなかったリリースに関して

未収集インシデント	リリース有	リリース無
171	37	134

・HPに存在するものは、クロールする階層が足りなかった。
・収集できなかった約８割はウェブサイト情報になかった。

原因推定の誤判定について

結果＼推定	紛失・置忘れ	管理ミス	盗難	誤操作	不正アクセス	ワーム・ウイルス	その他
紛失・置忘れ	53	9	1	0	0	0	2
管理ミス	5	11	1	0	0	0	0
盗難	3	2	29	0	0	0	4
誤操作	0	0	0	12	2	0	1
不正アクセス	0	0	0	1	10	1	2
ワーム・ウイルス	0	0	0	0	1	5	1
その他	0	1	0	0	4	2	8

手動とクローラーの時間比較

- 10企業について，プレスリリースから2016年のインシデント情報(企業名，業種，日付，規模，原因)を抽出しcsvに落とす

手動	クローラー
２６分３７秒	３０分８秒

- クローラーの方が時間がかかった。。。。