Vol.2017-DPS-170 No.22
Vol.2017-CSEC-76 No.22
2017/3/3

OSINT による収集と自動タグ生成システムの提案

天野純一郎[†1]　森滋男[†1]　水越一郎[†1†2]　後藤厚宏[†1]

概要：近年，企業や組織に対する，情報窃取などを目的としたサイバー攻撃が深刻な問題となっている．本稿では，こうしたサイバー攻撃に対処するアナリストにおけるサイバー脅威インテリジェンス分析強化の研究として，OSINT による収集と自動タグ生成システムの提案を行う．

キーワード：サイバーセキュリティ，オープン・ソース・インテリジェンス

Proposal of automated OSINT acquisition and tag generation system

JUNICHIRO AMANO[†1]　SHIGEO MORI[†1]
ICHIRO MIZUKOSHI[†1†2]　ATSUHIRO GOTO[†1]

Abstract: In recent years cyber attacks have increased and threats posed by them have become complicated. Cyber-analysts face this problem day to day. In such a situation, threat intelligence and OSINT(Open Source Intelligence) in order to support the analysis have been empowered.

Keywords: cyber security, open source intelligence

1. はじめに

1.1 サイバー攻撃の発生と増加

サイバー空間は，わが国が経済活動を行う上で重要な場所である．この空間には，既に数多くの組織・個人のコンピュータやスマートフォンなどが接続されている．こうしたサイバー空間の広がりの中で，絶えずサイバー攻撃が発生している．

サイバー攻撃には，情報窃取を目的とした標的型攻撃や，金銭を目的としたランサムウェア攻撃などが存在する．2015 年 5 月には，日本年金機構から 125 万件の個人情報を窃取されるというサイバー攻撃事案が発生した[1]. 2016 年には，サイバー攻撃によって JTB のオンラインサービスから 793 万件の個人情報が流出した恐れがあると報じられた[2]．さらに日本経済団体連合会のコンピュータが標的型攻撃と思われるサイバー攻撃を受けていたことが明らかになった[3]．この経団連の事案では，非公開の委員会の資料や参加者の個人情報が漏洩した可能性があると報じられた[4]．こうした脅威は一向に減ることはない．

1.2 サイバー攻撃の巧妙・複雑化と発見の遅れ

近年の傾向として，JPCERT/CC が報告しているように組織を標的とした「高度サイバー攻撃」(標的型攻撃)が国内でも表面化している[5]．サイバー攻撃を行う脅威主体は目的を達成するために，例えばマルウェアの活動時間を決めるなど[6]，攻撃を隠蔽しながら継続的に攻撃を仕掛けた．また，中小企業や地方の支店初めに狙い，そこを踏み台として最終目的となる組織を攻撃するような複雑化した攻撃も発生していると言われている[7]．その結果からか，FireEye の Mandiant が 2016 年に報告したアジア太平洋地域のレポートでは，企業に対する最初の侵害から発見までに平均 520 日もかかっており，これは世界の平均値 374 日よりも 146 日発見に時間がかかっている[8]．また，表面化した攻撃はごく一部であると考えることもできるため更なるサイバー攻撃対策に積極的に取り組む必要がある．

こうした状況を受け，サイバー攻撃に対抗するためにセキュリティオペレーション支援技術(固有の脅威インテリジェンスの生成など)が研究されている[9]．こうした背景には，インシデントを見つけ出すために大量のログ取得が必要となったことがあげられる．サイバー攻撃は時間との勝負である．脅威主体が情報を持ち出したり，破壊的活動を実行する前に阻止する必要がある．それに対して脅威主体は，攻撃をスクリプトによって自動化，攻撃ツールのカスタマイズ・改良，を行うことで自身の身を悟られないように秘匿化し，活動を隠蔽する．またサイバー攻撃は質との勝負でもある．サイバー攻撃を行う脅威主体は目的を達成するまで活動を継続するといった長期間のオペレーションも報告されている．防護側は，使用する脆弱性や手口の変更など，作戦・戦術の変化を認識し対応する能力が必要となっている．

†1 情報セキュリティ大学院大学
Institute of Information Security
†2 東日本電信電話会社
Nippon Telegraph and Telephone East Corporation

Vol.2017-DPS-170 No.22
Vol.2017-CSEC-76 No.22
2017/3/3

2. サイバーセキュリティとインテリジェンス

2.1 Threat Intelligence(脅威インテリジェンス)

Threat Intelligence に関する統一的な定義はないため，組織ごとに定義している．例えば，日立ソリューションズが提供するセキュリティ用語では，「脅威インテリジェンス(Threat Intelligence)とは、『インテリジェンス』、すなわち、サイバーセキュリティに関する複数の情報の集合体のこと。あるいは、様々な情報をつなぎ合わせ、その集合体の中から新たな脅威に関する知見を導き、セキュリティ対策に活用していく取り組みのことを指す。」と定義している[10]. Gartner は，「Threat intelligence は，その脅威または危険に対する対応の決定に使用できる，既知または新たな脅威の資産に対する危険性に関する，コンテキスト，メカニズム，インジケータ，意味合い，実行可能なアドバイスを含む証拠にもとづいた知識である．」と定義している[11]. このように各社それぞれの定義を持っているが，総じてサイバーセキュリティに関する情報を収集し情報をつなぎ合わせ証拠に基づいたインテリジェンスを生成し，脅威に対する防護に活用することを指している．そのため，本稿では「複雑化するサイバー攻撃を読み解き適切に対処するための知識と活動」と定義する．一般的なインテリジェンス活動は表 2.1 のように分けられる．

表 2.1 一般的なインテリジェンス活動(参考文献[12][13])

収集源の区分	細部情報収集活動	例
公開情報からの情報収集	OSINT (Open Source Intelligence)	新聞・雑誌・テレビ・インターネット等からの情報収集活動
サイバー空間からの情報収集	CYBINT/DNINT (Cyber Intelligence/Digital Network Intelligence)	サイバー空間からの情報収集活動
技術的手段による情報収集	IMINT(Imagery Intelligence)	偵察衛星等による情報収集活動
	MASINT (Measurement and Signature Intelligence)	赤外線や空気中の核物質等科学的な変化をとらえる情報収集活動
	SIGINT(Signal Intelligence)	諜報機関等によるケーブルや信号からの情報収集活動
人的手段による情報収集	HUMIT(Human Intelligence)	パーティー会場など人を介した情報収集活動
その他	TECHINT(Technical Intelligence)	外国軍の装備を研究し、使われている技術や弱点などを見つける情報収集活動
	FINIT(Financial Intelligence)など	金融取引記録などからの情報収集活動

また一般的なインテリジェンス活動は，表 2.1 のサイクルで行われる．

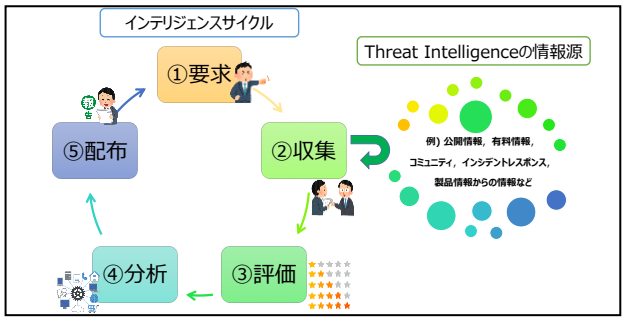

図 2.1 インテリジェンスサイクル (参考文献[14][15]を参考に作成)

サイクルの開始となる要求は要求者によって異なるが，Threat Intelligence に適用しようとした場合，要求(目的)としてサイバー攻撃対処や保護すべき資産を守ることなどを設定することが可能である．また要求があった場合に情報をすぐに取り出すための知識を蓄積することができる．

2.2 OSINT

公開情報を収集しインテリジェンスを生成する一連の活動に OSINT(Open Source Intelligence)がある．OSINT は従来，米国のインテリジェンスコミュニティを中心に利用が進んでいたが，製薬業界の M&A を事例分析用途などでも研究されている[17]. この OSINT は，特定の情報要求のもとに行われ，複数の公開情報をつなぎ合わせることによって情報要求者にとって価値のある情報を作り出す取り組みのことである．はじめは OSD(Open Source Data)や OSIF(Open Source Information)の情報を集め，情報要求に沿う形で情報の選別，抽出，選択，関連付けなど OSINT の目的である価値あるインテリジェンスとなり，最終的にその中でも非常に有効なものは OSINT-V(Validated OSINT)となる[18].

2.2.1 OSINT の活用例

Threat Intelligence に OSINT は利用可能である．実際に

Vol.2017-DPS-170 No.22
Vol.2017-CSEC-76 No.22
2017/3/3

OSINT がどのようなサイバーセキュリティに活用できるかという事例を想定したものを表 2.2 に示す.

表 2.2　個別情報から得られる OSINT([15]を参考に作成)

個別の情報	情報源
新たな Java の脆弱性を組み込んだ Exploit Kit が地下市場フォーラムで販売された	Dark Web 上の地下市場フォーラム
アジアの Java の脆弱性エクスプロイトの感染率が米国よりもかなり高い	セキュリティベンダ A のブログ
ボットネットの新種マルウェアが既に感染が確認されたウイルスに掲載	セキュリティベンダ B のウイルスリスト
時を同じくして大手金融機関による複数の中小地方銀行買収	大手ニュースサイト A のニュース記事
これにより不渡り小切手手数料の値上が値上がりし消費者が抗議	大手ニュースサイト B のニュース記事
ハクティビストグループが SNS 上で米国の銀行システムに対する攻撃を宣言	マイクロブログ (Twitter) , SNS(Facebook)
あるハクティビストのマイクロブログアカウントからボットネット使用指令が出された	マイクロブログ(Twitter)

表 2.2 のように個別情報をつなぎ合わせることで,「米国の銀行システムに対して, ハクティビストが Java 脆弱性をつくボットネットを用いた, アジアの IP アドレスを起点の DDoS 攻撃を実施する」という信憑性のある推測ができる. しかしここで, 人が個別の情報をつなぎ合わせる必要性, 情報源からの情報収集や個別情報の選択・関連付けなどの問題が生じる.

2.2.2 OSINT ツールとサービス

こうした問題を解決する既存の OSINT ツールの調査したところ, 構造化された情報であるインジケータでの相関や手動のタグ付けでインジケータと関連する意味・文脈を付与することが実現されていた. また非構造化テキスト情報についてはインデクシングやタグクラウドを用いてテキスト間での関連付けが実現されている. このため, 非構造化テキスト情報と構造化された情報との関連付けができる. 調査した範囲では, MISP[16]が最もインジケータとテキスト情報を関連付けが実現されていた. これは機能として, インジケータ情報だけでなくテキスト情報も同じイベント内に記録できるためである. ただしインジケータを含まない非構造化テキスト情報があった場合には, インジケータとの関連付けが困難な問題が残る. この点は, タグを用いることで関連付けが可能である. しかし, このタグ付けは手動で行う必要がある. 特にインジケータを含まないテキスト情報は膨大にあるため, OSINT に利用可能とするためには自動化されたタグ付けの仕組みが必要となる.

3. サイバー空間における状況認識

アナリストは常にサイバー攻撃を見逃さないよう取り組んでいる. しかしサイバー攻撃を発見するためには, 個々の過去の経験や知識, 技術力に依存している[19]. こうした経験とサイバー攻撃の状況を認識するために必要となる情報の関連付けを迅速に行うことができれば, 分析や推測のプロセスをより強化することができる.

3.1 状況認識

一般的に状況認識を説明するものとして OODA ループ (OODA Loop; ウォーダ・ループ) がある. OODA ループは、米国空軍のジョン・ボイド大佐によって提唱された意思決定理論である. 朝鮮戦争の航空戦についての洞察を基盤にして, 指揮官のあるべき意思決定プロセスを分かりやすく理論化したものである[20]. OODA ループを図 3.1 に示す.

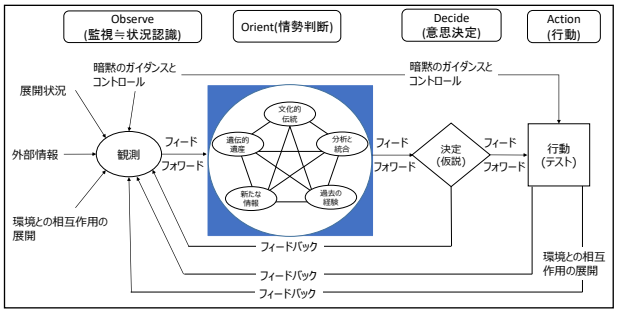

図 3.1　OODA Loop(参考文献[20]を参考に作成)

OODA ループはビジネスの世界でも利用されている. PDCA は, 環境が安定的で変化が少ないビジネスにおいて有効であるが想定できない変化に対応が難しい. これに対して OODA は周囲の状況から自分たちのすべきことを策定し, アクションを取るという意思決定プロセスをとるため不安定な変化が大きい状況でも迅速な対応が可能となる[21] .

3.1.1 センスメイキング

OODA ループはさらに, 細かく分けることができ, OODA ループの OO に該当する部分をセンスメイキング(Sensemaking)という. センスメイキングとは, 人間が経験から意味を与える過程を言う. 想定, 予測, 期待していないことを、感知(observation)し評定(orientation)するプロセスである. Pirolli たちはこれをインテリジェンスの分析に適用している.インテリジェンス分析のためのセンスメイキングループのモデル図 3.2 によると, アナリストは現在の状況と決定するタイプに応じ, 複数のループにわたってト

ップダウンプロセスとボトムアッププロセスを繰り返しセンスメイキングすることが示されている[22].

ここで，センスメイキングを行うアナリストの重要な点は，証拠(Evidence)に基づいた調査を行うという点であり，これらの情報をもとに情報のすくい出し，情報の検索，関連付けを行い，分析，推測を立てる．そのため，センスメイキングの中でこの証拠を共有することができれば，インテリジェンスの生成に役立てることができる．

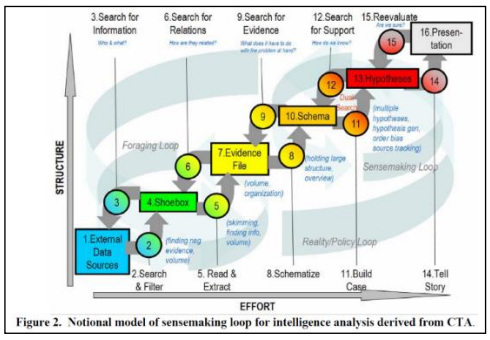

図 3.2　インテリジェンス分析のためのセンスメイキングループのモデル([22]より引用)

3.2 サイバー状況認識

サイバー空間においても状況認識の利用価値があることが示されている．NISC は，サイバーセキュリティ戦略の中で,「守るべき重要な情報や情報システムのサイバー空間への依存が一層高まる中，手法の複雑・巧妙化等によりサイバー攻撃の脅威も高まっている．このような状況では，各主体によるこれまでの取組は継続しつつも，刻々と変化するリスクに対し，社会メカニズムとして、適時適切な資源配分の下で動的に対応していくことが必要である.」として,「例えば、Observe（モニタリング），Orient（情勢判断），Decide（意思決定），Act（行動）を繰り返すことにより，迅速かつ適格な意思決定を行う「動的防御プロセス連携」（OODA ループ）.」を示している[23]．総務省は動的防御プロセスの具体例を示している[24]．OODA ループを利用したサイバー攻撃の動的防御プロセスを図 3.3 に示す．

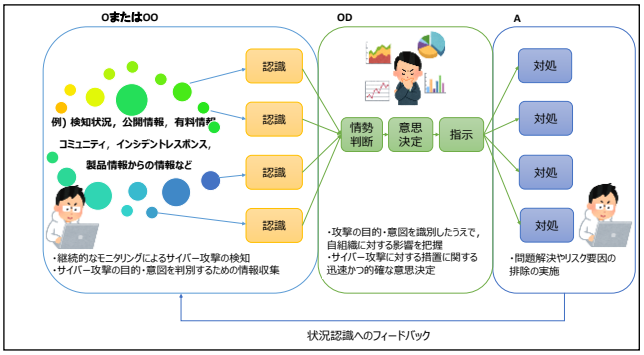

図 3.3　OODA ループを利用した動的防御プロセス(参考文献[24][25]を参考に作成)

(1)　サイバー状況認識の関連研究

状況認識に関連する研究として，小松たちは国内におけるサイバーセキュリティ状況認識についての研究の必要性を指摘している[26]．ここでは，サイバー状況認識の研究目的として，主に 4 つの目的，①サイバー監視の運用効率化(大量の情報から必要な情報の抽出)，②アナリストが情報をもとに攻撃を認識するかを明らかにする，③セキュリティアナリストの訓練，③サイバー監視の指標(SOC の評価)があることが示された．

Ulrik Franke たちは，102 のサイバー状況認識の論文，研究テーマについて①一般的なサイバー状況認識一般的なサイバー状況認識，②産業制御システムのサイバー状況認識（主に電力網），③危機管理のためのサイバー状況認識，④サイバー状況認識のためのツール，アーキテクチャ，アルゴリズム，⑤情報融合，⑥サイバー状況認識のための可視化，⑦サイバー状況認識のためのヒューマンコンピュータインタラクションの設計仕様とワークフロー，⑧国家規模や大規模なサイバー状況認識，⑨サイバー状況認識に関する演習，⑩サイバー状況認識のための情報交換，⑪軍事サイバー状況認識のように分類できることを示した[27]．この論文では，サイバー状況認識のための情報交換，欺瞞の危険性，軍事行動におけるサイバー戦の被害評価の問題などの分野の研究は少ないことが指摘されていたが，論文は 2014 年時点のものであり，2017 年現在では，CARINA のような SOC における情報共有に着目した研究も行われている[19].

3.3 SOC におけるサイバー状況認識

SOC(Security Operation Centre)においても継続的に Threat Intelligence を活用するために，サイバー状況認識し刻々と変化するリスクに対して対応することが必要である．その際アナリストは継続的に OSINT を活用し個々の情報の証拠からセンスメイキングを行うことができる．これを図 3.4 に示す．

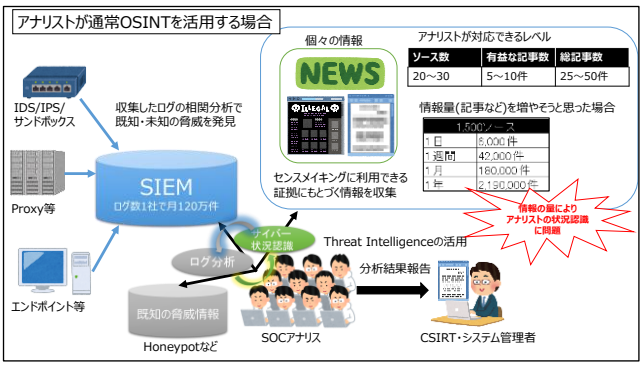

図 3.4　アナリストが通常 OSINT を利用する場合(情報量は参考文献[28]を参考に作成)

SOC アナリストは，サイバー攻撃を検出した場合，上記の個々の情報から得られる証拠にもとづきサイバー攻撃に意味付けや状況認識を行い，判断し対処に必要な情勢判断に役立つ内容を含む分析結果を強化して報告できる．その

際，サイバー攻撃で狙われた標的や攻撃を行ってきた脅威主体の脅威度，目的や影響について分析することで意思決定や対処に結び付けることができる．

しかしながら，今後 IT だけでなく OT や IoT，モバイルなどこれまで認知度が高くなかった脅威や新たな攻撃手法を用いた情報などが増加する恐れがある．セキュリティ人材不足が叫ばれるなか，このような状況に陥るとアナリストの負荷が増加し継続的なサイバー状況認識が難しくなり，経験との関連付けが困難となり攻撃の見逃しや，アナリストの分析に影響が出る可能性がある．今回は，あくまで OSINT に着目したが，それ以外から得られるものと組み合わせることも重要である．

4. 提案方式

4.1 SOC アナリストの分析を強化する OSINT の提案

アナリストがサイバー空間での状況認識を得るために，OSINT として非構造化テキスト情報を収集し状況を認識することは有効な手法である．しかしサイバー空間上にある情報量は膨大であり，それらの情報すべてを SOC のアナリストが認識することが困難であるという課題があった．

図 3.4 での問題を解決するために既存の OSINT ツールである MISP をデータベースとして利用しタグを使うことを提案する．MISP を利用することで，インジケータを含む情報については API を利用し SIEM のログとセンスメイキングに利用できるタグ情報を利用可能となる．またあらかじめ MISP に非構造化テキスト情報を入力しておき用意されているタグを付与することで，アナリストはあらかじめ集められた構造化されたインジケータ情報とともに分析強化のために役立てることができる．さらに Threat Intelligence の作成に役立つサイバー攻撃を行う脅威主体の帰属問題についても OSINT で取り組める範囲で部分的に解決可能である．ただし部分的に解決できない問題として，非構造化テキスト情報を用いる際に，アナリストのセンスメイキングに必要な自動タグ生成に課題が残る．

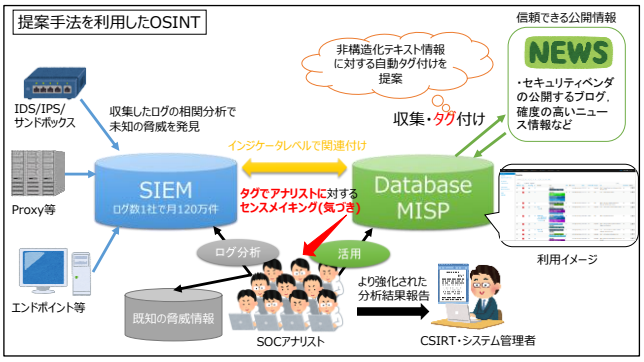

図 4.1 提案手法を利用した OSINT の場合

4.2 センスメイキングに利用できる自動タグ生成の提案

分析を強化できるタグ付の方法として，アナリストが，情報の関連付けを行う際には，証拠に基づいて調査を行うことが示されていた．そのため，今回は証拠にもとづいた表 4.1 で示す 5W2H を抽出することでアナリストのセンスメイキングに役立たせ分析を向上させると提案する．

表 4.1 分析を強化できるタグ付けの方法

5W2H	情報	具体的証拠
Who(誰が)	A 銀行を狙うサイバー犯罪者	人名，組織名
Wen(いつ)	20XX 年 XX 月 XX 日ごろ	時制
Where(どこで)	Z 国の A 銀行の	国名，都市名など
What(なにを)	ATM に	固有名詞
Why(なぜ)	不正アクセスによるマルウェア感染により	Exploit，脆弱性(CVE)など
How(どうやって)	ATM から不正な預金の引き下ろしを許した	攻撃手口(スピアフィッシング，Word 文書)
How much(どのくらい)	30 台の ATM から	数値

自動タグ付けの実現方法については CoreNLP を使用することにより，人名や国名に対してタグ付けが可能なため，上記提案を実現するために CoreNLP での整理および分類を試みる．この CoreNLP は，Java ベースのオープンソースツールであり，民間組織や政府機関でも使用されている．今回自動タグ生成のために使用する CoreNLP は，Stanford 大学が提供している自然言語処理の汎用ツールで，アノテーターと呼ばれる機能を利用することで，品詞の特定，構文木・依存関係の決定，固有名詞解析，共参照の特定などを自動で行うことができる[29]. 今回は，英語のみの機能を利用したが，CoreNLP は多言語対応している．

関連研究として今回用いた CoreNLP は，IRC 上ハクティビスト活動分析の研究でも利用されている[30].

4.3 CoreNLP を使用した自動タグ生成の検証

(1) 検証環境

ハードウェア：

MacBook Pro (Retina 13-inch, Early 2015)
プロセッサ 2.7GHz Intel Core i5
メモリ 8GB 1867 MHz DDR3

仮想ソフトウェア：

VirtualBox バージョン 5.1.10

仮想マシン：

メインメモリ 4096MB(割り当て)
プロセッサ 2(割り当て)

OS：
Ubuntu 14.04.5 LTS 64bit

主な使用ソフトウェア：
CoreNLP 3.4.1
openjdk version 1.8.0_111
Python 2.7.6

(2) **検証に使用したデータ**

セキュリティベンダが提供するブログから無作為に選択した 58 件の英語記事を使用した．このデータはあらかじめタグ付けされているためこのタグをもとに評価する．

(3) **検証方法**

CoreNLP のアノテーターの固有表現抽出を主に使用し検証を行った．

5. 評価

4 章での提案方式を評価する方法として，公開されているすでに専門家によってタグ付けされている情報との比較が可能である．ここでは，非構造化テキスト情報(記事)に対して既にタグ付けされている情報と，提案方式で付与したタグとの一致度を比較することによって評価を行う．評価対象とする記事は，非構造化テキスト情報を主として構成されるデータを使用した．評価はタグとの完全一致を確認し，比較を行う上で大文字と小文字の違いについては問わないこととした．

5.1 CoreNLP の固有表現抽出を使用した結果

CoreNLP の固有抽出表現を使用し，比較対象のタグと比較した結果は以下の表 5.1 のとおり．

表 5.1 CoreNLP での自動タグ生成した結果(n=58)

	比較対象のタグ	自動タグ生成した数	一致した数
総タグ数	216	2425	45
平均タグ数	3.724138	41.81034	0.77586

比較対象のタグで評価したところ，CoreNLP で自動タグ生成したものとの一致率は **20.83%**であった．

5.2 CoreNLP 以外の手法との比較による評価

上記の結果を受けて，考察を行う上で異なる手法でのタグ付けを試みることにした．今回は，一般的なサービスとして提供されているキーフレーズ抽出システム A と，キーワード抽出システム B を使用し比較した．なお前提条件として，キーフレーズ抽出システム A は API 経由で行い，その際 1 記事あたりの送信制限があったため，CoreNLP で使用したテキストデータと同等ではなく，量を減らす加工を行ったものを使用した．また，キーワード抽出システム B については，記事を URL ベースで解析し得られるキーワードを利用した．

(1) **キーフレーズ抽出システム A の結果**

表 5.2 キーフレーズ抽出システム A の結果(n=58)

	比較対象のタグ	システム A	一致した数	一致率 (%)
総タグ数	216	1595	47	21.7593
平均タグ数	3.724138	27.5	0.81034	-

(2) **キーワード抽出システム B の結果**

表 5.3 キーワード抽出システム B の結果(n=58)

	比較対象のタグ	システム B	一致した数	一致率 (%)
総タグ数	216	2751	79	36.5741
平均タグ数	3.724138	47.43103	1.36207	-

5.3 評価に対する考察

(1) **CoreNLP を使用した一致した際の一致率と課題**

今回の CoreNLP の固有表現抽出の手法では，ワードのみが抽出された．その結果，比較対象のタグに含まれる 1 語以上のワードで構成されるフレーズについて抽出することができなかった(例：exploit kit など)．この問題については，2 語以上を組み合わせて表現する必要があるため，現状の固有表現抽出だけでは解決ができない．残された課題として，CoreNLP は語間の依存関係などの情報も抽出可能なため，固有表現抽出だけでなく依存関係などを組み合わせることでより理想に近い形でタグを抽出することが期待できる．また今回は固有表現抽出で抽出できるもののみ抽出された，CoreNLP では名詞や動詞も個別にタグ付けされるためこれらの情報を MISP のために用意されている分類情報や同義語情報などを利用してフィルタすることでサイバーセキュリティの用語で特化した自然言語処理が可能になると期待できる．これは専門家がタグ付けを行うような，タグ生成の精度向上も期待できる．

(2) **提案手法の良い点と残された課題**

今回提案した自動タグ生成システムの評価により明らかになった点として，あらかじめ提案した意図(期間を示す固有表現の抽出を期待するもの)とは異なるが，この固有表現によってテキスト情報の中から CVE 番号がすべて正しく抽出できた点があげられる．ただしこの影響により，CoreNLP で抽出したタグ数が増加している．原因は，対象となったデータの中に Windows Update を通知するような

情報がありその中に CVE 番号が大量に含まれていることがある．対照的に，比較対象のタグについては CVE 番号が記載されていたとしてもすべての CVE 番号がタグとして付与されていない場合が多かった．これは記事を書いた専門家が重要なもののみタグ付けした可能性が考えられる．また比較対象のタグにはなかったが，テキスト情報の中に国名(Taiwan，Japan など)や組織名(INTERPOL など,この情報は比較対象のタグにも存在), 人物名が抽出できていたことがあげられる．これらは，意図した固有表現が抽出できたことを示している．

(3) **システム A，B を使用した際の一致率と良い点**

システム A と B の結果を比較すると，CoreNLP よりも一致率がよかったといえる．その要因として 2 語以上のフレーズに対応しており，センスメイキングに利用可能な証拠となる情報であるコンテキスト情報が高まっていることが分かった．システム A，B 双方とも，専門家が付与した比較対象としたタグよりもより良いタグとなるキーフレーズやキーワード抽出を行っているといえるものもあった．また，一致率という点では完全一致を見たため，上記の追加されたワードなどによって一致率が下がっているものがある．全体としては，比較対象となるタグと同じ意味合いのものも抽出されていた．一方で比較対象となるタグと比較すると，タグとなるフレーズやワードが多く抽出されすぎているため絞り込みといった点で改善の余地があるといえる．また，これらの結果から提案した CoreNLP での自動タグ生成システムも改善の余地が残されている．

6. 考察

6.1 SOC のセキュリティオペレーションに対する適応

本稿で提案したタグの自動生成は，あくまでアナリストを支援するための一パーツにしか過ぎない．アナリスト中心のオペレーションでは，関連情報を分析する際に SIEM を使用している．そのため，本論で提案した自動生成されたタグを, SIEM と連携することができれば, アナリストが行う分析を効率化し，分析を強化することができる．

また，SOC のオペレーションでは，世界各国にアナリストが配置され複数人の共同作業の中で分析を行っている．タグという短いワードにより，言語の壁を越えて迅速な情報共有を実現できる可能性がある．さらにアナリスト間で情報の共有を行うにあたり，SOC アナリストの人数が増えるほど情報共有の問題が生じやすいため関連付けを行うタグの効力が発揮できるといえる．

6.2 SOC 以外の立場における適応の検討

SOC 以外の立場の組織においてもサイバー状況認識を向上させるためにタグを用いることは有益であるといえる．例えば，CSIRT なども事前にインシデントハンドリングに備えるために常日頃から情報収集を行うことが求められる．またインシデント発生時には，タグというキーワードで必要となる情報を認識できれば迅速なインシデントハンドリングが可能になる可能性がある．特に今回の CoreNLP の結果は，脆弱性情報である CVE 番号の抽出が可能なことを示した．このような脆弱性情報は CSIRT においても非常に役立つ情報といえる．

6.3 AI と自然言語処理のさらなる活用

今回 OSINT で活用可能な非構造化テキスト情報に対して，自動タグ生成を行った．これにより，人の手によるタグ付けに近づけることが可能であることが示された．しかし，まだ完全な理想的な人の手によるタグ付けには至っていない．タグ付けの問題点として，組織や属人的なタグ付けルールがある．一度付与したタグ付けのルールを人の手によって変更しようとした場合，情報量が多ければ莫大な手間がかかってしまうという点がある．この点について，今後 AI(機械学習)と自然言語処理のさらなる応用できれば, タグ付けを使用する組織がある時点で, AI にタグ付けの指示を出すことですべてのタグを自動的に付与しなおしてくれるような仕組みが実現可能となる．そのためには AI がワードの意味や文脈を認知する仕組みが必要となる．

また，今回は提案で用いなかったが，CoreNLP やシステム B には記事中のワードやフレーズからポジティブ，ニュートラル，ネガティブといった分類，感情情報の高い割合などを示すことが可能である．こうした情報を利用することによって，付与するタグの色を変化させアナリストの状況認識を向上させ，注目すべき情報を抽出・提供するといったことも検討の余地がある. また自然言語の関係性から，今後は AI による文書要約も期待される．最後に OSINT で得られる情報はその特性から発信者の意図が含まれる場合があるため，立場によらずこのような技術によりニュートラルな状況判断が行うことにも期待できる．

7. おわりに

サイバー空間で発生する事象を，アナリストが状況認識し情勢判断・意思決定に役立てることが重要である．しかし，サイバー攻撃が増加しそれに伴いアナリストが取り入れる情報量が増加している．またそうした情報をアナリスト間で共有するという問題もあった．本稿では，非構造情報のテキスト情報に対して自動タグ生成を試みた. そして，人が情報をセンスメイキングしやすい形での自動タグ付けを実行可能なことを明らかにした．今後も多層防御の観点から，機械学習の手法や AI などを取り入れアナリストの分析強化につながる状況認識の方法について検討していきたい．

謝辞 ご協力頂いた皆様に，謹んで感謝の意を表する．

参考文献

[1] サイバーセキュリティ戦略本部, ”日本年金機構における個人情報流出事案に関する 原因究明調査結果”, http://www.nisc.go.jp/active/kihon/pdf/incident_report.pdf (2016/06/18 アクセス).
[2] YOMIURI ONLINE, ”ＪＴＢ個人情報７９３万件流出か？…標的型攻撃の巧妙な手口”, http://www.yomiuri.co.jp/science/goshinjyutsu/20160615-OYT8T50004.html (2016/12/17 アクセス).
[3] 経団連事務局, ”経団連事務局コンピュータのマルウェア感染”, http://www.keidanren.or.jp/announce/2016/1115.html (2016/12/17 アクセス).
[4] 日本経済新聞, ”経団連の情報、漏洩の可能性　サイバー攻撃か”, http://www.nikkei.com/article/DGXLASFS10H10_Q6A111C1EAF000/ (2016/12/17 アクセス).
[5] JPCERT/CC, ”高度サイバー攻撃への対処におけるログの活用と分析方法”, https://www.jpcert.or.jp/research/apt-loganalysis.html (2016/06/18 アクセス).
[6] マクニカネットワークス, ”標的型攻撃の実態と対策アプローチ”, http://www.macnica.net/file/security_report_20160613.pdf (2016/09/15 アクセス).
[7] 産経デジタル, ”狙われる中小企業、従業員１０人なのに「標的」に…甘いセキュリティー突き大企業の機密情報も！？”, http://www.sankei.com/west/news/160309/wst1603090004-n2.html (2016/7/17 アクセス).
[8] Mandiant, ”M-trends Asia pacific”, https://www.fireeye.com/blog/threat-research/2016/08/m-trends_asia_pacifi.html (2016/09/15 アクセス).
[9] NTT 技術ジャーナル, ”新たなサイバー攻撃の出現と 今後のセキュリティ研究開発の方向性”, http://www.ntt.co.jp/journal/1208/files/jn201208008.pdf (2016/07/17 アクセス).
[10] 日立ソリューションズ情報セキュリティブログ, ”脅威インテリジェンス(Threat Intelligence)とは”, http://securityblog.jp/words/threat_intelligence.html (2016/12/26 アクセス)
[11] Webroot, ”Threat Intelligence:What is it, and How Can it Protect You from Today's Advanced Cyber-Attacks?”, https://www.gartner.com/imagesrv/media-products/pdf/webroot/issue1_webroot.pdf (2016/12/26 アクセス)
[12] Wikipedia, ”諜報活動”, https://ja.wikipedia.org/wiki/%E8%AB%9C%E5%A0%B1%E6%B4%BB%E5%8B%95 (2016/12/26 アクセス)
[13] Wikipedia, ”List of intelligence gathering disciplines”, https://en.wikipedia.org/wiki/List_of_intelligence_gathering_disciplines (2016/12/26 アクセス)
[14] Wikipedia, ”インテリジェンス・サイクル”, https://ja.wikipedia.org/wiki/%E3%82%A4%E3%83%B3%E3%83%86%E3%83%AA%E3%82%B8%E3%82%A7%E3%83%B3%E3%82%B9%E3%83%BB%E3%82%B5%E3%82%A4%E3%82%AF%E3%83%AB (2016/12/26 アクセス)
[15] NTTGroup, ”2015 グローバル脅威インテリジェンス・レポート”, pp.58
[16] CIRCL, “MISP - Malware Information Sharing Platform and Threat Sharing”, http://www.misp-project.org/
[17] 中島庸介ほか, ”オープンソース・インテリジェンスの競争分析への活用の戦略的枠組み　テキスト・マイニングによる日本の製薬業界の 2010 年問題における M&A 情報分析を事例として”, INTELLIGENCE MANAGEMENT vol.3, No.1 / 2011, http://lab.sdm.keio.ac.jp/idc/yasui/papers/j7_2011intelligence%20management_nakajima_yasui.pdf (2017/1/10)
[18] NATO OSINT Handbook v1.2 - Jan 2002, http://www.oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf (2017/1/3 アクセス)
[19] Diane Staheli ほか, ”Collaborative Data Analysis and Discovery for Cyber Security”, https://www.usenix.org/system/files/conference/soups2016/wsiw16_paper_staheli.pdf (2016/8/31 アクセス)
[20] Wikipedia, ”OODA ループ”, https://ja.wikipedia.org/wiki/OODA%E3%83%AB%E3%83%BC%E3%83%97 (2016/12/26 アクセス)
[21] “軍隊式人を動かすマネジメント書評”, http://www.sinkan.jp/special/us_military_formula/ (2016/12/26 アクセス)
[22] Peter Pirolli, PARC, “The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis”
[23] NISC, ”サイバーセキュリティ戦略”, http://www.nisc.go.jp/active/kihon/pdf/cyber-security-senryaku-set.pdf (2016/12/17 アクセス)
[24] 情報セキュリティアドバイザリボード，総務省における情報セキュリティ政策の推進に関する提言，http://www.soumu.go.jp/main_content/000217000.pdf (2016/12/26 アクセス)
[25] 名和利男，ASERT, ”求められるサイバー攻撃対処能力”, http://jp.arbornetworks.com/%e6%b1%82%e3%82%81%e3%82%89%e3%82%8c%e3%82%8b%e3%82%b5%e3%82%a4%e3%83%90%e3%83%bc%e6%94%bb%e6%92%83%e5%af%be%e5%87%a6%e8%83%bd%e5%8a%9b/ (2016/12/26 アクセス)
[26] 小松ほか，”サイバーセキュリティ分析における状況認識の研究”，Computer Security Symposium 2016
[27] Ulrik Franke ほか，”Cyber situational awareness e A systematic review of the literature”，COMPUTERS & SECURITY 46 (2014) 18-31
[28] Edwin Tump, ”Machine Learning for Cyber Security Intelligence”, 27th FIRST Conference，https://www.first.org/resources/papers/conf2015/first_2015_-_tump-_edwin_-_machine_learning_for_cyber_security_intelligence_20150611_fw.pdf (2016/01/16 アクセス)
[29] Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60., http://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf (2016/12/31 アクセス)
[30] Jiakai Yut ほか，"Automated Framework for Scalable Collection and Intelligent Analytics of Hacker IRC Information"，2016 International Conference on Cloud and Autonomic Computing