CSS/SCIS 論文に基づく 日本のセキュリティ系研究の推移調査

金岡 晃[1,a)]

概要：日本国内のセキュリティとプライバシーに関連する研究は、情報処理学会のコンピュータセキュリティシンポジウム（CSS）と電子情報通信学会の暗号と情報セキュリティシンポジウム（SCIS）が最も大きな 2 つの学術イベントと言ってよい。そこでは毎年 100 を超える研究が発表され、500 人を超える参加者が集まり議論をしている。技術の高度化から産業界への一般展開のスピードが速まるなか、新しい技術に関連するセキュリティとプライバシーの問題も増加し、深くなってきている。学術研究の成果も展開のスピードアップが望まれるところである。本稿では、CSS と SCIS で発表された研究論文の用語をテキスト分析することによりそれぞれのイベントで研究発表されるテーマの特徴や 10 年間での変化を考察し、日本におけるセキュリティとプライバシー関連の研究がどういったキーワードを中心に行われているかを見る。

キーワード：CSS、SCIS、 研究、トレンド

A survey on the transition of security-related research in Japan based on CSS/SCIS papers

Akira Kanaoka[1,a)]

Abstract: The Computer Security Symposium (CSS) of the Information Processing Society of Japan (IPSJ) and the Symposium on Cryptography and Information Security (SCIS) of the Institute of Electronics, Information and Communication Engineers (IEICE) are arguably the two largest academic events for research related to security and privacy in Japan. More than 100 research papers are presented there every year, and more than 500 participants gather to discuss them. With the increasing speed of technology advancement and its general deployment to industry, the security and privacy issues associated with new technologies are also increasing and becoming deeper. It is desirable to speed up the deployment of academic research results. In this paper, we analyze the terminology used in the research papers presented at CSS and SCIS to identify the characteristics of the research topics presented at each event and how they have changed over the past 10 years, as well as the keywords used in security and privacy-related research in Japan.

Keywords: CSS 2021, LaTeX, style files

1. はじめに

技術の高度化から産業界への一般展開のスピードが速まるなか、新しい技術に関連するセキュリティとプライバシーの問題も増加し、深くなってきている。学術的な研究も広く深く行われている。社会におけるセキュリティとプライバシーの重要性はますます高まっており、学術研究の成果も展開のスピードアップが望まれるところである。

新しい技術が急激に産業界を通じて一般に浸透する中で、セキュリティやプライバシーの懸念がなされ、それらについてのセキュリティやプライバシー研究も加速していく。その結果、セキュリティとプライバシーの研究分野全体で活発化と加速化がされている様子がうかがえる。

たとえば 4 大会議と呼ばれる IEEE S&P、USENIX Se-

1 東邦大学
Toho University, Funabashi, Chiba 274–8510, Japan
a) akira.kanaoka@is.sci.toho-u.ac.jp

© 2021 Information Processing Society of Japan

curity Symposium、ACM CCS、NDSS といった難関国際会議では投稿数・採録数の増加にともない投稿と査読プロセスが変わった。学術的なセキュリティとプライバシーの研究は日本のみならず世界的にも加速化していると言えよう。

日本国内では学術研究はさまざまな学会で行われている。中心的なのは情報処理学会と電子情報通信学会であろう。それぞれの学会でセキュリティやプライバシーに関連する研究会がありコミュニティが形成され、国内での学術研究が進んでいる。国内での議論を経て国際会議で発表され、世界のトップと比肩する研究が数多く生まれている。

それぞれの学会や研究会では、定期的に開催される研究会における論文発表が行われるが、それとともに年に 1 回などのペースで比較的大きな学術集会が研究会を母体にして開催される。日本のセキュリティ・プライバシーの学術研究では毎年 10 月頃に開催される情報処理学会 CSEC 研究会主催のコンピュータセキュリティシンポジウム（CSS）と、電子情報通信学会 ISEC 研究会主催の暗号と情報セキュリティシンポジウム（SCIS）が代表的である。それぞれ 100 件を優に超える論文が毎年投稿され、500 人を超える参加者が集まる大規模な学術集会となっている。

そこで発表される研究内容は、時事をつかんだものが多く、今後の動向や現在の研究のトレンドがどこにあるかなどが伺える。

著者は、これまでの CSS/SCIS での発表を縦断的に俯瞰することで、これまで行われてきた研究分野の概観やトレンドの推移ができるのではないかと考えた。そして概観や推移を知り、それを整理して発表することにより、学術以外の分野に向けて学術成果を伝える良い機会になるのではと考えた。

そこで CSS と SCIS の過去 10 年の論文をテキスト分析し、頻出する語の特徴やクラスター、共起関係を見ることで CSS と SCIS において発表される研究の特徴の違いや、過去 10 年での研究トレンドの変遷を見ることとした。調査では KH Coder を用いて計 5057 件の論文をテキスト分析を行った。

2. CSS と SCIS

2.1 CSS:コンピュータセキュリティシンポジウム

コンピュータセキュリティシンポジウム（Computer Security Simposium, CSS）は情報処理学会コンピュータセキュリティ（CSEC）研究会が主催するシンポジウムである。例年 10 月後半ごろに開催される。併催のワークショップが複数あり、2021 年のシンポジウム CSS2021 ではマルウェア対策研究人材育成ワークショップ 2021（MWS2021）、プライバシーワークショップ 2021（PWS2021）、ユーザブルセキュリティワークショップ 2021（UWS2021）、OSS セキュリティ技術ワークショップ 2021（OWS2021）開催される。

2.2 SCIS: 暗号と情報セキュリティシンポジウム

暗号と情報セキュリティシンポジウム（Symposium on Cryptography and Information Security, SCIS）は電子情報通信学会情報セキュリティ研究専門委員会（ISEC）が主催するシンポジウムである。例年 1 月後半ごろに開催される。

3. 研究の推移調査

日本における大規模なセキュリティ・プライバシー学術研究のイベントにおける論文を対象に、研究の推移調査を行う。

3.1 調査対象

調査対象としたのは、CSS2011 から CSS2020 までの 10 年分の CSS 発表論文と、SCIS2012 から SCIS2021 までの 10 年分の SCIS 発表論文とした。

3.2 調査方法

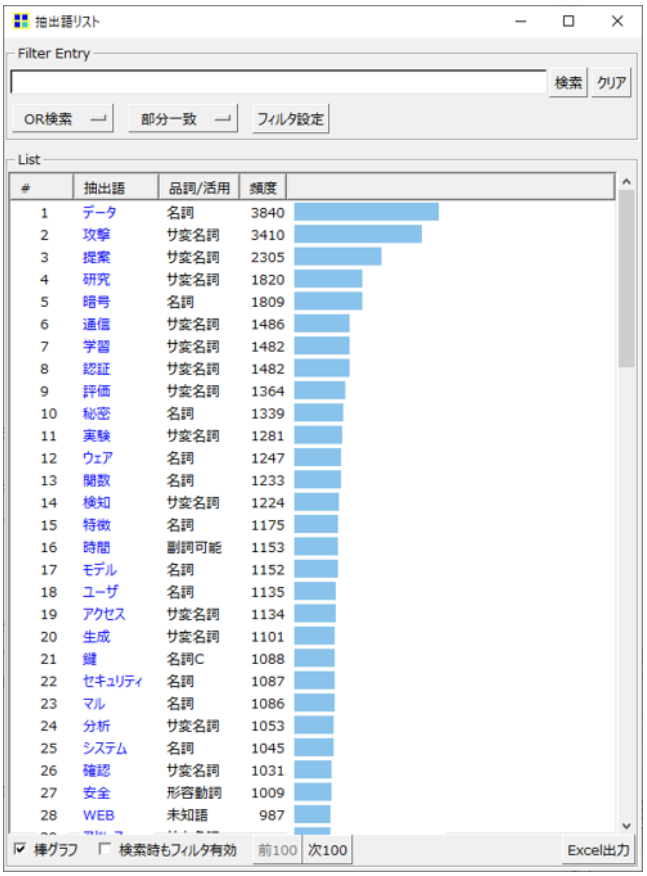
CSS と SCIS の 10 年分の論文 PDF をテキストに変換し、テキストを分析することで用語の頻出度合いやクラスタリング、共起関係を見ることで研究テーマの変遷や CSS や SCIS の違いを見に行く。

PDF のテキスト変換には pdftotext を利用した。pdftotext の出力結果のテキストファイルは UTF-8 文字エンコーディングで出力されるが、論文中の画像部分等は UTF-8 の文字コード以外のデータとして出力されるために出力結果のテキストファイルを直接分析ツールに採用することはできない。そこで、出力結果のテキストファイルから文字情報以外のデータを取り除き、そのデータを分析する。過去 10 年間に CSS と SCIS で発表された合計で 5,057 件の論文をテキスト化した。

分析 KH Coder[1], [2] を用いた。KH Coder はテキストデータから語を抽出し、それらの語を処理し抽出語のリスト作成や頻出語の表の出力、さらには語と語の結びつきを複数の手法により分析可能なツールである。テキスト分析において多くの採用事例がある [3]。

具体的な分析手順を解説する。まず上記方法で PDF からテキストを抽出する。CSS と SCIS の各年の論文テキストデータを 1 つのテキストファイルにし、それを KHCoder で分析をすることで、各イベントの年度ごとの特徴分析を行う。テキストファイルを KH Coder に登録した後、前処理メニューからテキストのチェックを実施し、分析に問題が生じるデータをあらかじめ取り除く。その後前処理を実行して、語の抽出等を実施する。

前処理が終了したのち、ツールメニューにより抽出語のリストを表示する。この時、抽出語として一般の名詞や動



#	抽出語	品詞/活用	頻度
1	データ	名詞	3840
2	攻撃	サ変名詞	3410
3	提案	サ変名詞	2305
4	研究	サ変名詞	1820
5	暗号	名詞	1809
6	通信	サ変名詞	1486
7	学習	サ変名詞	1482
8	認証	サ変名詞	1482
9	評価	サ変名詞	1364
10	秘密	名詞	1339
11	実験	サ変名詞	1281
12	ウェア	名詞	1247
13	関数	名詞	1233
14	検知	サ変名詞	1224
15	特徴	名詞	1175
16	時間	副詞可能	1153
17	モデル	名詞	1152
18	ユーザ	名詞	1135
19	アクセス	サ変名詞	1134
20	生成	サ変名詞	1101
21	鍵	名詞C	1088
22	セキュリティ	名詞	1087
23	マル	名詞	1086
24	分析	サ変名詞	1053
25	システム	名詞	1045
26	確認	サ変名詞	1031
27	安全	形容動詞	1009
28	WEB	未知語	987

図 1　KH Coder による CSS2020 論文データの分析結果（抽出語）

詞なども語としてリストアップされ、それらが上位に来るために本研究の本来目的には関係ない用語が上位を占めることになるので、それらの語を除外して再度抽出語のリストを出す。これらブロックする語は著者が判定した。例としては「THE」「OF」「AND」といった一般的な英単語と思しきものや、「行う」「用いる」「利用」「情報」「場合」といった年度やイベントにかかわらず出現すると考えられる語がある。詳細のリストは付録 A.1 より参照できる。

抽出語は KH Coder による品詞と活用の判定情報と頻度とともに表示される。図 1 は CSS2020 の分析結果である。語「データ」が 3840 回検出されて最も多い語となっていることがわかる。

この抽出語のうち、再頻出の 150 語を年度とイベントごとに抽出し、その推移を見る。

続いて、ツールメニューより多次元尺度構成法を実施する。多次元尺度構成法では、各語の間の距離を計算し、その距離をもとに座標を計算する。そしてその座標に基づいて図が作成されて図示される。KH Coder では距離計算手法は複数選択可能であるが本研究では Jaccard を選択した。座標計算についても複数手法が選択可能であるが、本研究では Kruskal を選択した。また座標系さんにおける次元は 2 とし、クラスター数は 8 とした。図 2 は CSS2020 の多次元尺度構成法の分析結果である。円の大きさは出現頻度を示し、各クラスターは同じ色の円で示される。この結果から

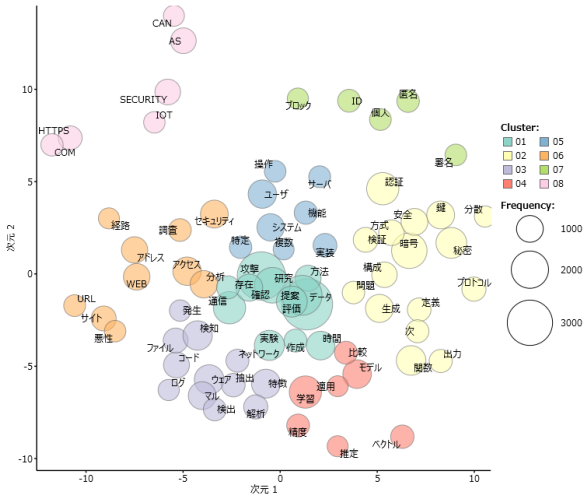


図 2　KH Coder による CSS2020 論文データの分析結果（多次元尺度構成法）

らは、攻撃や研究、データといった比較的一般的と思える用語が高い頻度とともにクラスタを形成していることがわかる。また「学習」「適用」「精度」といった近年盛んに研究されている機械学習のキーワードがクラスターを形成していることなどもわかる。調査ではさらに多次元尺度構成法による図のデータをイベント間、年度間で定性的に比較する。

同じくツールメニューより階層的クラスター分析を実施する。階層型クラスター分析では、クラスターを形成する過程と結果を視覚的に把握することができる。KH Coder では距離計さん手法が複数選択可能であるが、本研究では Jaccard を選択した。また同じくクラスター分析手法も複数手法が選択可能であるが、Ward 法を選択した距離に多次元尺度構成法で選択した Jaccard を採用したことにより形成されるクラスターが多次元尺度構成法で出力されたものと同様のものとなり、それぞれの用語間のクラスター形成がわかる。結果の図は大きなものになるために本稿での図示は控えるが、付録 A.1 にそれぞれのイベント、年度の結果を掲載する。

最後にツールメニューにより共起ネットワークを実施する。共起ネットワークではそれぞれの語の共起関係が分析され図示される。それぞれの語がどう共起するかで研究の方向性を視覚的に把握可能であると考えた。KH Coder では共起関係の係数計算にあたって手法を複数選択できるが、本研究では Jaccard を選択した。図 3 は CSS2020 の共起ネットワークの分析結果である。それぞれの語が頻度に従った円の大きさで表され、共起関係が強い語が直線により結ばれる。多次元尺度構成法と似たような図に見えるが、それぞれの語の結びつきが異なり、より研究の内容が想起できる内容になっていることがわかる。たとえば「検知」は「悪性」「サイト」「Web」等と結びつき、さらにそれらの語は「URL」「アクセス」と結びつくなど、悪性 Web

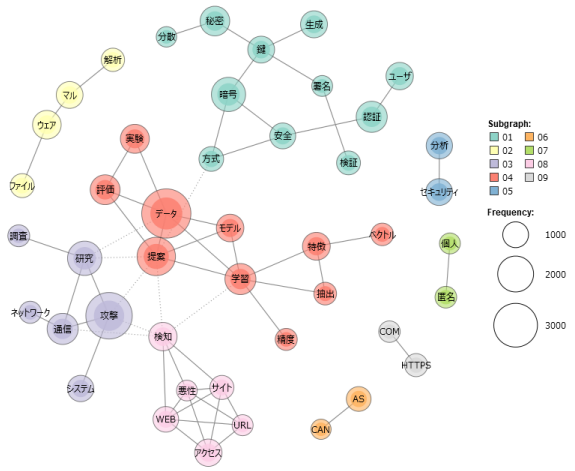


図 3　KH Coder による CSS2020 論文データの分析結果（共起ネットワーク）

表 1　CSS/SCIS の論文投稿数推移

年度	CSS	SCIS
2011	141	
2012	115	322
2013	140	308
2014	175	343
2015	184	322
2016	192	327
2017	221	367
2018	184	361
2019	223	338
2020	177	340
2021		277

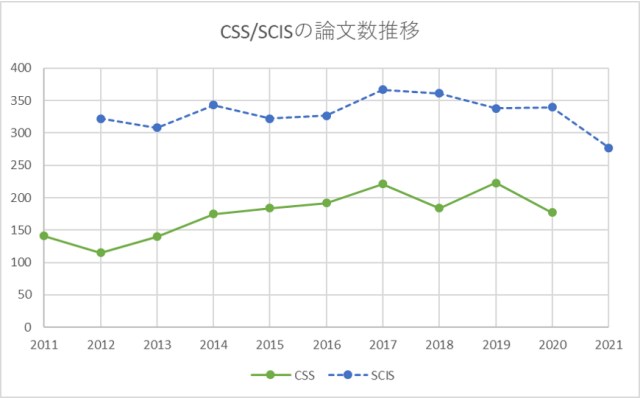


図 4　CSS/SCIS の論文投稿数推移

サイトの検知研究が盛んであることが伺える。

4.　調査結果

4.1　投稿論文数の推移

CSS/SCIS の投稿論文の推移を表 1 と図 4 で示す。

4.2　頻出語の推移と CSS/SCIS 間の差異

KH Coder の前処理結果で抽出された頻出語のイベント、年度ごとの違いや推移を示す。

CSS では、頻出語の上位はほとんどが変わっていない。語「攻撃」は CSS2012 から CSS2019 にわたって最も多く検出された。2 位は「データ」であり、CSS2020 では「データ」が 1 位になり「攻撃」が 2 位になった。「提案」は 3 位となることが多く、CSS2014 で 4 位になった以外は 3 位であった。その次に 10 年間を通して多く出る語は「通信」「暗号」と続く。「暗号」が 6 番目に位置づけされることは CSS の特徴と言えよう。

SCIS も頻出語の上位に大きな変化は見られない。語「暗号」と「攻撃」が 1 位を分け合っている様子が見て取れる。語「鍵」も継続して上位にあることと合わせて SCIS と CSS の違いが見て取れる。変遷としてはこの 2 年間で「データ」が上位に挙がってきていることが特徴的と言えよう。

その他の特徴的な語の変遷としては、機械学習関連の語の流行が見て取れる。CSS において語「学習」は 2016 年以前は上位 100 位に入っていなかったが、2017 年に 27 位、2018 年に 13 位、2019 年に 10 位、2020 年に 7 位と順位を上げている。同じく「モデル」も 2017 以降に上位に入ってくる語となっていた。「学習」は SCIS においても 2020 年に 29 位、2021 年に 13 位と上がってきている。

SCIS においては 2016 年に 9 位に「CAN」が出現するなど、自動車セキュリティ関連の用語が 2016 年を中心に上位に上がってきていた。

4.3　多次元尺度構成法によるキーワードのマッピングとクラスタリング

図 5 に CSS2011、図 6 に SCIS2012、図 7 に SCIS2021、の多次元尺度構成法による分析結果をそれぞれ示す。図 2 の CSS2020 の図と比較することでそれぞれのイベントにおける用語やクラスタの差異や、10 年間での用語の変化が見て取れる。

たとえば CSS では IP アドレスの分析に関連する用語が 2011 年ではクラスタを形成していたが 2020 年では関連するクラスタは形成されていない。マルウェアに関連するクラスタは CSS2011 では大きな塊となっていたが、2020 年ではクラスタの形成はされているが 1 つ 1 つの語は小さくなっておりその距離も広がっており、研究としてはより細分化されてきている様子がうかがえる。

SCIS では CSS と異なり暗号理論研究に関連する用語が多くプロットされており、それぞれがクラスタを形成している。CSS との性格の違いがよく表れていると言えよう。10 年間での変化は目立ったものは多くないが、機械学習関連のクラスタが形成されていることが見て取れる。

それぞれの年度・、イベントの結果の図は A.1 付録にて示すこととする。

表 2　CSS の頻出語推移 (2011-2020)

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1	攻撃	データ	攻撃	攻撃	攻撃	攻撃	攻撃	攻撃	攻撃	データ
2	データ	ユーザ	データ	データ	データ	データ	データ	データ	データ	攻撃
3	提案	提案	提案	通信	提案	提案	提案	提案	提案	提案
4	鍵	攻撃	通信	提案	暗号	認証	システム	研究	研究	研究
5	暗号	通信	システム	ウェア	通信	システム	評価	検知	暗号	暗号

表 3　SCIS の頻出語推移 (2012-2021)

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
1	暗号	鍵	暗号	暗号	暗号	攻撃	攻撃	攻撃	攻撃	攻撃
2	鍵	暗号	攻撃	攻撃	攻撃	暗号	暗号	鍵	データ	データ
3	攻撃	攻撃	鍵	鍵	鍵	鍵	鍵	暗号	暗号	鍵
4	方式	方式	データ	提案	データ	データ	データ	データ	提案	暗号
5	提案	提案	提案	方式	提案	提案	提案	提案	鍵	提案

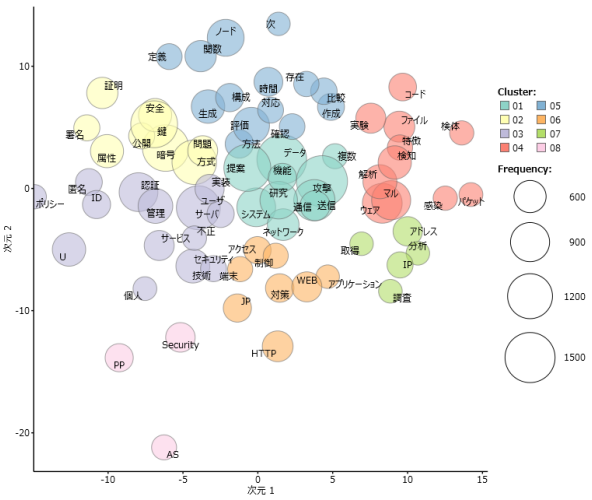


図 5　KH Coder による CSS2011 論文データの分析結果（多次元尺度構成法）

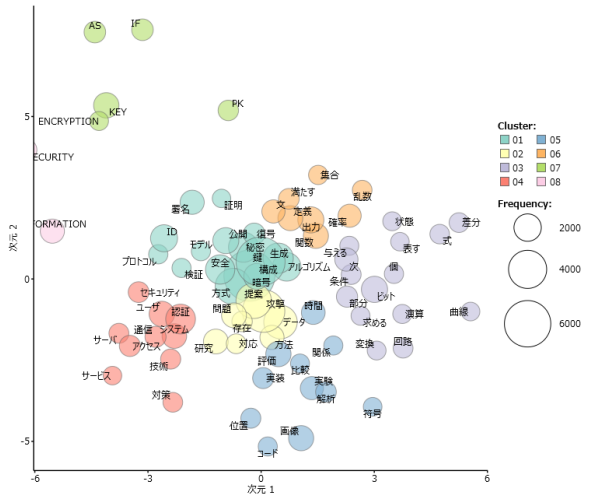


図 6　KH Coder による SCIS2012 論文データの分析結果（多次元尺度構成法）

4.4　階層的クラスター分析による用語のクラスタリング

階層型クラスター分析結果を比較することで前節で考察した結果をより深く議論可能であるが、それぞれの図を掲

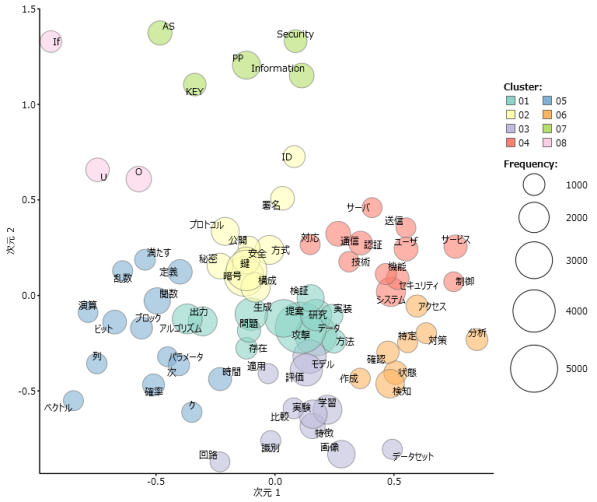


図 7　KH Coder による SCIS2021 論文データの分析結果（多次元尺度構成法）

示するスペースが足りないため、ここでは議論は避ける。それぞれの詳細の図は A.1 付録にて示すこととする。

4.5　共起ネットワークによる用語同士の関係性

図 8 に CSS2011、図 9 に SCIS2012、図 10 に SCIS2021、の共起ネットワークによる分析結果をそれぞれ示す。図 3 の CSS2020 の図と比較することでそれぞれのイベントにおける用語の共起関係の違いや、10 年間での用語の変化が見て取れる。

たとえば CSS では「攻撃」は「通信」「アドレス」「マル」「ウェア」との共起関係が示されているが、SCIS では「攻撃」は「モデル」「評価」「暗号」「安全」「鍵」との共起関係が示されている。同じ語でも共起する語が全く異なるものとなっており、CSS と SCIS の正確の差が最も強く表れている図と言って良いだろう。その他にも CSS はこの 10 年でデータを用いた研究がより盛んになっている様子がさまざまな語との共起関係から見て取れる。攻撃と共起関係にある用語とクラスタは CSS2011 ではマルウェアに関す

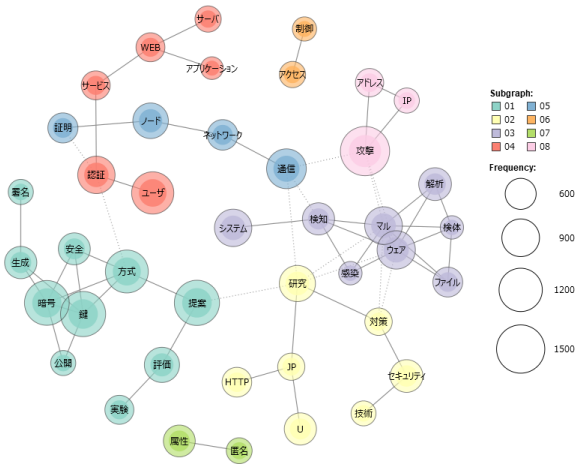


図 8 KH Coder による CSS2011 論文データの分析結果（共起ネットワーク）

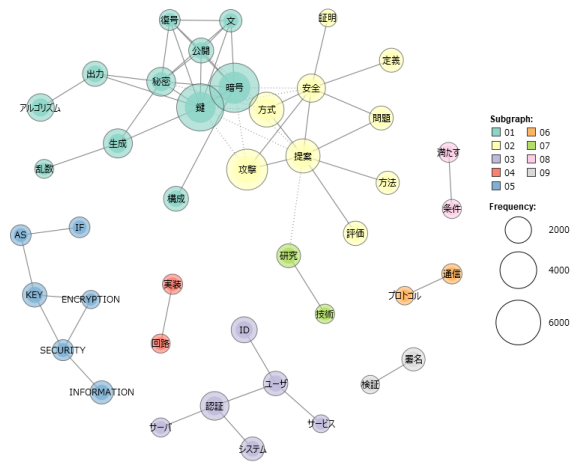


図 9 KH Coder による SCIS2012 論文データの分析結果（共起ネットワーク）

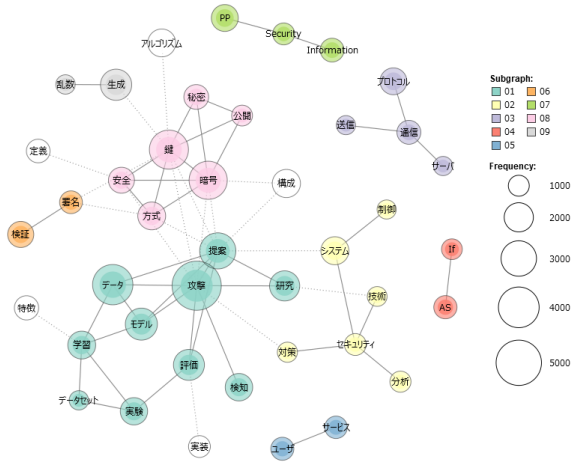


図 10 KH Coder による SCIS2021 論文データの分析結果（共起ネットワーク）

るクラスタであったが、CSS2020 になりマルウェアに関するクラスタは距離が離れ、悪性 Web サイトの検知に関するクラスタと近くなった点などが特徴的と言えよう。

5. まとめ

本研究では、過去 10 年間の CSS と SCIS の論文に焦点を当て、テキスト分析による語の頻出や繋がり、共起を調べることで CSS と SCIS での研究トレンドの差やこの 10 年間での変遷を考察していった。いずれの分析結果も CSS と SCIS では研究トレンドに大きな違いがあることが示されており、10 年での変化も大きくはないものの見られるものとなった。それぞれの分析の詳細情報は紙面の都合上掲載がかなわなかったが、それぞれの年度の結果を詳細に見ることでさらに見えるものもあろう。それぞれの図やデータは付録を通じて Web に掲載するのでそちらを参考いただきたい。

本稿をきっかけに、学術のコミュニティ自身が研究の変遷や違いを認識するだけにとどまらず、学術研究の成果をより広く産業界などに伝えることがかなえば幸いである。

参考文献

[1] 樋口耕一 2020 『社会調査のための計量テキスト分析 —内容分析の継承と発展を目指して— 第 2 版』 ナカニシヤ出版 サポートページ Amazon

[2] 樋口耕一 2004 「テキスト型データの計量的分析 —2 つのアプローチの峻別と統合—」 『理論と方法』 (数理社会学会) 19(1): 101-115 PDF FileIEEE/ACM 13th Working Conference on Mining Software Repositories (MSR), 2016, pp. 468-471.

[3] “ KH Coder を用いた研究事例” 、`https://khcoder.net/bib.html?year=2021&auth=all&key=`, （2021 年 8 月 23 日閲覧）

付　　録

A.1 詳細データ

詳細なデータを Github で公開している。`https://github.com/kanaoka-laboratory/IPaddrHardCodingSurvey`