

## 全自動言語処理システムを用いたツール作成学習テンプレートの網羅的整備と共有

メタデータ	言語: ja 出版者: 静岡大学 公開日: 2019-05-13 キーワード (Ja): キーワード (En): 作成者: 狩野, 芳伸 メールアドレス: 所属:
URL	<a href="http://hdl.handle.net/10297/00026569">http://hdl.handle.net/10297/00026569</a>

**科学研究費助成事業 研究成果報告書**

平成 30 年 6 月 18 日現在

機関番号：13801

研究種目：若手研究(A)

研究期間：2014～2017

課題番号：26700019

研究課題名(和文)全自動言語処理システムを用いたツール作成学習テンプレートの網羅的整備と共有

研究課題名(英文)Application workflow templates with automated natural language processing system

## 研究代表者

狩野 芳伸 (Kano, Yoshinobu)

静岡大学・情報学部・准教授

研究者番号：20506729

交付決定額(研究期間全体)：(直接経費) 18,200,000円

研究成果の概要(和文)：本研究の目的は、自然言語処理のアプリケーション構築を容易にするプラットフォームの実装である。同時に、機械学習の学習索性テンプレートを網羅的に実装・配信できるようにするのが目的である。そのために自然言語(テキスト・音声)について、機械学習統合の資源を「作る」ための全自動システムを構築する。本研究はおおきく、プラットフォームの構築、プラットフォーム上で動作するコンポーネントの構築、それらを利用した実証的な応用研究にわかれる。構築したシステムを、法律分野、医療分野、対話システム、論文処理などいくつかの異なるテーマにおける研究の基盤として利用し、その有効性を検証した。

研究成果の概要(英文)：This research aims to build a platform that allows users to construct natural language processing applications in an easier way. Another aim is to provide machine learning workflow templates. I built an automated system that can create natural language processing resources with machine learning integration. My research consists of my platform, my components that are runnable on the platform, and application research using these platform and components. I confirmed usefulness of my system by applying to a couple of different fields including legal document processing, medical document processing, dialog system and scientific literature mining.

研究分野：自然言語処理

キーワード：自然言語処理

## 1. 研究開始当初の背景

自然言語(テキスト・音声)の大規模処理は日々需要が高まっており、それに応えるには再利用性の高い資源(ツール・データ)の整備と機械学習の利用が必須である。これまで既存資源を「使う」ための自然言語処理プラットフォームを構築してきたが、自然言語処理ではドメイン依存性が高く、データに応じたツールの再学習が必要であることが多かった。

## 2. 研究の目的

本研究の目的は、自然言語処理のアプリケーション構築を容易にするプラットフォームの実装である。同時に、機械学習の学習素性テンプレートを網羅的に実装・配信できるようにするのが目的である。そのために自然言語(テキスト・音声)について、機械学習統合の資源を「作る」ための全自動システムを構築する。

## 3. 研究の方法

本研究はおおきく、プラットフォームの構築、プラットフォーム上で動作するコンポーネントの構築、それらを利用した実証的な応用研究にわかれる。最後の応用研究については、研究成果の項で述べる。

プラットフォームには、教師つき機械学習手法を用いた訓練と、その結果の評価について、汎用性・互換性・可搬性を保ちつつ、容易に利用できるように、かつこれまでのプラットフォーム実装を極力生かし統合的なシステムとして利用可能になるように設計を検討し、実装した。

また、いくつかの実用的・複雑なシステムについて、コンポーネント化と学習ワークフローの構築を行った。具体的には試験問題の自動解答、検索、質問応答、対話システムなどである。

## 4. 研究成果

このようにして構築したプラットフォームやコンポーネントを応用し、さまざまなタスクについてアプリケーションを実行する実験を行った。利用は下記に述べるいくつかの異なる研究テーマにまたがり、具体的には、大学入試の自動解答、医療言語情報処理、法律文書の処理などである。

法律文書の処理については、国際コンテスト型ワークショップであるCOLIEE(Competition for Legal Information Retrieval and Entailment)のオーガナイザーとして、法情報学の国際ワークショップ

JURISIN(Juris Informatics)併設のワークショップとして、またロンドンで開催された法と人工知能の国際会議 ICAIL(International Conference on Artificial Intelligence and Law)内にてワークショップをオーガナイズした。COLIEEは日英両言語で司法試験(民法短答式)の自動解答を試みるものであり、コンテストの運営と同時に参加者として自動解答システムを構築した。システムは日本語を対象に法律文書の処理を行うものである。

医療分野においては、電子カルテの処理や問診データの自動分析を行った。いずれも、固有表現など単語レベルの処理から、係り受けや述語項構造などまでさまざまなレベルでの自然言語処理技術が必要となり、プラットフォームの有用性を確認する応用としても適切な課題であった。

対話システム分野においては、会話ゲーム「人狼」のプレイヤー自動化を目標とする人狼知能プロジェクトなど話し言葉の処理と文生成を対象にプラットフォームを利用してシステム構築を行った。

学術論文データの自動処理においては、日本語が対象の他の研究とは異なり英語を対象とし、多言語への対応を確認した。こうした幅広い応用から、プラットフォームの有用性と実用性を検証した。

さらに、利用者の利便性を考慮し、従来コンポーネント化の際に準拠していた国際標準よりも、開発者にとってより容易に利用可能なAPIを定義しアクセスを可能にする方法を提案し、プロトタイプの実装を行った。これは開発者が標準を学習するハードルを下げるとともに、標準の縛りを緩くすることでソフトウェア設計の柔軟性を高め、一方で極力従来の標準準拠の実装からの変更を小さくするAPI仕様によりコンポーネント実装の共通化を図ったものである。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計5件)

1. 狩野 芳伸. 「ロボットは東大に入れるか」という企て: 番外編 さまざまな正誤問題の自動解答 -世界史, 日本史から医師国家試験, 司法試験まで-. 情報処理, 58(7), pp.616-617. 2017 査読無
2. 狩野 芳伸. コンピューターに話が通じるか: 対話システムの現在. 情報管理 59(10), pp.658-665. 2017. 査読無

3. 伊藤 詩乃, 田中 佑岳, 狩野 芳伸, 榊原康文. 医師国家試験を自動解答するプログラムの構築. 人工知能学会論文誌 31(6), 2016. 査読有 DOI: <https://doi.org/10.1527/tjsai.AI30-G>
  4. 松崎拓也, 横野光, 宮尾祐介, 川添愛, 狩野 芳伸, 加納隼人, 佐藤理史, 東中竜一郎, 杉山弘晃, 磯崎秀樹, 菊井玄一郎, 堂坂浩二, 平博順, 南泰浩, 新井紀子. 「ロボットは東大に入れるか」プロジェクト: 代ゼミセンター模試タスクにおけるエラーの分析. 自然言語処理, 23(1) pp.119-159. 2016. 査読有
  5. 石岡恒憲, 狩野 芳伸, 大津起夫. 全文検索による試験解答システム - 新規作成問題の類似文書検索を中心として -. 大学入試研究ジャーナル, 25, pp.129-135, 2015. 査読有
- [学会発表](計 59 件)
1. 渋木英潔, 阪本浩太郎, 石下円香, 狩野 芳伸, 三田村照子, 森辰則, 神門典子. NTCIR-13 QA Lab-3 における大論述問題の事例分析. 言語処理学会第 24 回年次大会(NLP2018). 2018.
  2. 狩野 芳伸, 佐藤 健. 司法試験自動解答の国際コンテスト型ワークショップ COLIEE と法律文書処理. 第 4 回自然言語処理シンポジウム(情報処理学会 NL-NLC 研究会合同). 2017.
  3. Kotaro Sakamoto, Hideyuki Shibuki, Madoka Ishioroshi, Akira Fujita, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori and Noriko Kando. Automatic Evaluation of World History Essay Using Chronological and Geographical Measures. The 8th International Workshop on Evaluating Information Access (EVIA 2017), NTCIR-13 conference. 2017.
  4. Masahito Sakishita and Yoshinobu Kano. Classification of Tweet Posters for Diseases by combined Rule-Based and Machine Learning Method in NTCIR-13 MedWeb. NTCIR-13 conference. 2017.
  5. Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishioroshi, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori and Noriko Kando. Overview of the NTCIR-13 QA Lab-3 Task. NTCIR-13 conference. 2017.
  6. Shoko Wakamiya, Mizuki Morita, Yoshinobu Kano, Tomoko Ohkuma and Eiji Aramaki. Overview of the NTCIR-13 MedWeb Task. NTCIR-13 conference. 2017.
  7. Yoshinobu Kano. Towards Text Mining for Meta-analysis of Neuroscience Papers. Second International Workshop on SCientific DOcument Analysis (SCIDOCA 2017). 2017.
  8. 箕輪 峻, 滝波 秋穂, 小川 ちひろ, 三原 直樹, 真木 裕子, 柴 淳, 狩野 芳伸. プロトコルを用いた意味理解による人狼知能の自然言語エージェント. 第 8 回対話システムシンポジウム, 人工知能学会 音声・言語理解と対話処理研究会 (SLUD) 第 81 回研究会. 2017.
  9. 阿部 元樹, 梅井 良太, 狩野 芳伸, 網川 隆司, 西田 昌史, 西村 雅史. 音響情報を利用した音声対話システムにおける破綻検出. 第 8 回対話システムシンポジウム, 人工知能学会 音声・言語理解と対話処理研究会 (SLUD) 第 81 回研究会. 2017.
  10. 狩野 芳伸, 佐藤 健, Mi-Young Kim, Randy Gooble. 司法試験自動解答の国際コンペティション COLIEE と法律文書の自然言語処理. 法と言語学会研究例会. 2017.
  11. 狩野 芳伸. 「ロボットは東大に入れるか」大学入試自動解答の成果、技術的課題と今後. 代表企画人・全体司会. 第 16 回情報科学技術フォーラム (FIT 2017). 2017.
  12. 中村 洋文, 片上 大輔, 鳥海 不二夫, 大澤 博隆, 稲葉 通将, 篠田 孝祐, 狩野 芳伸. 人間同士の議論の疑似体験を実現する人狼知能対戦システムの開発. 第 33 回ファジィシステムシンポジウム (FSS2017). 2017.
  13. 谷口 諒輔, 狩野 芳伸. 述語項構造を用いた単純な司法試験自動回答器の構築. 選奨セッション, 第 16 回情報科学技術フォーラム (FIT 2017). 2017.
  14. 狩野 芳伸, 稲葉 道将, 篠田 孝祐, 大澤 博隆, 片上 大輔, 大槻 恭士. しゃべる人狼知能 ~ 人工知能による自然言語人狼対戦. CEDEC CHALLENGE セッション, コンピュータエンターテインメントデベロッパーズカンファレンス 2017 (CEDEC 2017). 2017
  15. Hideyuki Shibuki, Kotaro Sakamoto,

- Madoka Ishioroshi, Akira Fujita, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori and Noriko Kando. Chronological and Geographical Measures for Evaluation of World History Essay QA in University Entrance Exams. Open Knowledge Base and Question Answering Workshop (OKBQA 2017) at the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017). 2017.
16. Hirofumi Nakamura, Daisuke Katagami, Fujio Toriumi, Hirotaka Osawa, Michimasa Inaba, Kosuke Shinoda and Yoshinobu Kano. Generating human-like discussion by paraphrasing a translation by the AIWolf Protocol using Werewolf BBS Logs. FUZZ-IEEE 2017. 2017.
17. Ryosuke Taniguchi, Yoshinobu Kano. Legal Yes/No Question Answering System Using Case-Role Analysis. In: Kurahashi S., Ohta Y., Arai S., Satoh K., Bekki D. (eds) New Frontiers in Artificial Intelligence. JSAI-isAI 2016 (JSAI-isAI 2016 Workshops, LENLS, HAT-MASH, AI-Biz, JURISIN and SKL. 2017
18. Yoshinobu Kano, Mi-Young Kim, Randy Goebel, Ken Satoh, Overview of COLIEE 2017 In 4th Competition on Legal Information Extraction and Entailment (COLIEE 2017), 16th International Conference on Artificial Intelligence and Law (ICAIL 2017). 2017.
19. Yoshinobu Kano, Reina Hoshino, Ryosuke Taniguchi. Analyzable Legal Yes/No Question Answering System using Linguistic Structures. In 4th Competition on Legal Information Extraction and Entailment (COLIEE 2017), 16th International Conference on Artificial Intelligence and Law (ICAIL 2017). 2017.
20. 崎下 雅仁, 狩野 芳伸. 日本語電子カルテから ICD コードを自動的に出力する. 情報処理学会第 79 回全国大会. 2017.
21. 洪水英潔, 藤田彬, 阪本浩太郎, 石下円香, 狩野芳伸, 三田村照子, 森辰則, 神門典子. 世界史論述問題の自動評価のための時間的・地理的情報の利用. 言語処理学会第 23 回年次大会. 2017.
22. 狩野芳伸, 稲葉通将. 人狼知能大会における自然言語処理部門の構想. 言語処理学会第 23 回年次大会. 2017.
23. 箕輪峻, 狩野芳伸, 柴淳, 黒田和也. 係り受け関係を利用したルールベースによる自然言語人狼エージェントの構築. 第 2 回 GAT2017 (2nd Game AI Tournament 2017). 2017.
24. 中村洋文, 片上大輔, 鳥海不二夫, 大澤博隆, 稲葉通将, 篠田孝祐, 狩野芳伸. 人狼 BBS のログデータを利用した人狼知能の発話プロトコル文の言い換えによる人間らしさの表出. 第 4 回人間共生システムデザインコンテスト (Human Symbiotic Systems - Design Contest (HSS-DC), 人間共生システム研究部会 第 22 回 HSS 研究会. 2017
25. Yoshinobu Kano. Between Platform and APIs: Kachako API for Developers. Third International Workshop on Worldwide Language Service Infrastructure and Second Workshop on Open Infrastructures and Analysis Frameworks for Human Language Technologies (WLSI3n0IAF2) at the 26th International Conference on Computational Linguistics (COLING 2016). 2016.
26. Masahito Sakishita, Yoshinobu Kano. Inference of ICD Codes from Japanese Medical Records by Searching Disease Names. Clinical Natural Language Processing Workshop at the 26th International Conference on Computational Linguistics (COLING 2016). 2016.
27. Eiji Aramaki, Yoshinobu Kano, Tomoko Ohkuma, Mizuki Morita. MedNLPDoc: Japanese Shared Task for Clinical NLP. Clinical Natural Language Processing Workshop at the 26th International Conference on Computational Linguistics (COLING 2016). 2016.
28. Yoshinobu Kano. Answering Yes-No Questions by Penalty Scoring in History Subjects of University Entrance Examinations. Open Knowledge Base and Question Answering Workshop (OKBQA2016) at the 26th International Conference on Computational Linguistics (COLING 2016). 2016.
29. 中村 洋文, 片上 大輔, 鳥海 不二夫, 大澤 博隆, 稲葉 通将, 篠田 孝祐, 狩野 芳伸. 人狼 BBS のログデータを利用

- した人狼知能の発話プロトコル文の言い換え生成. HCG シンポジウム 2016. 2016.
30. Yoshinobu Kano. Machine solves examinations: benchmarking natural language processing technology. Seminar in Institute of Information Science. 2016.
31. Yoshinobu Kano. Challenges in Natural Language Processing: Question Answering and Dialog System. Keynote Speech at IR Workshop 2016. 2016.
32. Mi-Young Kim, Randy Goebel, Yoshinobu Kano, and Ken Satoh. COLIEE-2016: Evaluation of the Competition on Legal Information Extraction/Entailment, Tenth International Workshop on Juris-informatics (JURISIN 2016). 2016.
33. Ryosuke Taniguchi, Yoshinobu Kano. Legal Yes/No Question Answering System using Case-Role Analysis. Competition on Legal Information Extraction/Entailment (COLIEE), Tenth International Workshop on Juris-informatics (JURISIN 2016). 2016.
34. Yoshinobu Kano. Table Data Extraction for Text Mining in Neuroscience Papers. First International Workshop on Scientific Document Analysis (SCIDOCA 2016). 2016.
35. 柴淳, 狩野芳伸. 単語の意味の距離から検出する会話破綻. 対話破綻検出チャレンジ, 人工知能学会 言語・音声理解と対話処理研究会 (SLUD) 第 78 回研究会第 7 回対話システムシンポジウム. 2016.
36. 狩野 芳伸. コンピュータは言葉を操れるか? ~自然言語処理からみた人工知能の現在と応用, CEDEC 2016. 2016.
37. 鳥海 不二夫, 片上 大輔, 狩野 芳伸, 丹野 宏昭. [人工知能学会×CEDEC コラボセッション] 人狼知能大会技術セッション, CEDEC 2016. 2016.
38. 渋谷 英潔, 石下 円香, 阪本 浩太郎, 藤田 彬, 狩野 芳伸, 三田村 照子, 森 辰則, 神門 典子. NTCIR-12 QA Lab-2 における QA システムの課題-センター試験の結果を中心として. 情報処理学会第 227 回自然言語処理研究会 (IPJSJ-NL227). 2016.
39. Masahito Sakishita, Yoshinobu Kano. KIS: Inference of ICD Codes by Rule-Based Method from Feature of Medical Record in NTCIR-12 MedNLPDoc. In Proceedings of NTCIR-12 workshop. 2016.
40. Eiji Aramaki, Mizuki Morita, Yoshinobu Kano, Tomoko Ohkuma. Overview of the NTCIR-12 MedNLPDoc Task. In Proceedings of NTCIR-12 workshop. 2016.
41. Hideyuki Shibuki, Kotaro Samamoto, Madoka Ishioroshi, Akira Fujita, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori, Noriko Kando. Overview of the NTCIR-12 QALab2 Task. In Proceedings of NTCIR-12 workshop. 2016.
42. 梶原健吾, 鳥海不二夫, 稲葉通将, 大澤博隆, 片上大輔, 篠田孝祐, 松原仁, 狩野芳伸. 人狼知能大会における統計分析と SVM を用いた人狼推定を行うエージェントの設計. 人工知能学会第 30 回全国大会 (JSAI2016). 北九州国際会議場. 2016.
43. Yoshinobu Kano, Text Mining for Neuroscience Papers, 4th International Workshop of Advances in Neuroinformatics (AINI 2016) and 14th INCF Nodes Workshop. 2016.
44. 渋谷英潔, 阪本浩太郎, 石下円香, 藤田彬, 狩野芳伸, 三田村照子, 森辰則, 神門典子. 大学入試世界史論述問題における専門知識の有無による評価の一致に関する考察. 言語処理学会第 22 回年次大会 (NLP2016). 2016.
45. 箕輪峻, 狩野芳伸. 複数のルールによる評価を総合的に判断する人狼エージェントの実装. Game AI Tournaments 第一回大会 (GAT 2016). 2016.
46. 谷口諒輔, 狩野芳伸. 単語間共起及びキーワード抽出を用いたルールに基づく対話破綻自動検出器の構築と評価手法の検討. 言語処理学会第 22 回年次大会 (NLP2016). 2016.
47. Yoshinobu Kano. Keyword and Snippet Based Yes/No Question Answering System for COLIEE 2015. Competition on Legal Information Extraction/Entailment (COLIEE), Ninth International Workshop on Juris-informatics (JURISIN 2015). 2015.

48. 大澤博隆, 狩野芳伸, 稲葉通将, 片上大輔. パネルトーク: どのようなプロトコルを設計するべきか? 自然言語処理, 対話エージェント, 対話分析の観点から. 「人狼知能コンテスト」キックオフ. 人工知能学会合同研究会. 2015.
49. 谷口諒輔, 狩野芳伸. キーワード抽出を用いたルールに基づく対話破綻自動検出器の構築. 第 75 回 人工知能学会 言語・音声理解と対話処理研究会 (SIG-SLUD) 第 6 回対話システムシンポジウム: 特別セッション: 対話破綻検出チャレンジ. 2015.
50. 榎原康文, 伊藤詩乃, 田中佑岳, 佐藤健吾, 洪繁, 狩野芳伸. 医師国家試験を自動解答するプログラムの開発. 第 1 回日本医療情報学会「医用知能情報学研究会」人工知能学会「医用人工知能研究会」(SIG-AIMED) 合同研究会 (JAMI & JSAI AIM 合同研究会). 2015.
51. 阪本 浩太郎, 石下 円香, 藤田 彬, 渋谷 英潔, 狩野 芳伸, 三田村 照子, 森辰則, 神門 典子. 大学入試の世界史論述問題における質問応答システムの自動評価に関する一考察. 第 222 回情報処理学会自然言語処理研究会 (SIG-NL), 2015.
52. 松崎拓也, 横野光, 宮尾祐介, 川添愛, 狩野芳伸, 加納隼人, 佐藤理史, 東中竜一郎, 杉山弘晃, 磯崎秀樹, 菊井玄一郎, 堂坂浩二, 平博順, 南泰浩. 『ロボットは東大に入れるか』プロジェクト 代ゼミセンター模試タスクにおけるエラーの分析 ワークショップ 自然言語処理におけるエラー分析 (兼: Project Next NLP 報告会), 言語処理学会第 21 回年次大会. 2015.
53. 石下円香, 阪本浩太郎, 渋谷英潔, 狩野芳伸, 三田村照子, Wang Di, 森辰則, 神門典子. 大学入試問題を解くための質問応答システムにおける現状と課題. 言語処理学会第 21 回年次大会. 2015.
54. Yoshinobu Kano. Kachako: all-in-one automated NLP platform and multiple domain UIMA toolkit. The Second International Workshop on Worldwide Language Service Infrastructure (WLSI). 2015.
55. Eiji Aramaki, Mizuki Morita, Yoshinobu Kano and Tomoko Ohkuma. Overview of the NTCIR-11 MedNLP-2 Task. NTCIR-11 conference. 2014.
56. 狩野 芳伸. 社会問題のエラー分析. 「ロボットは東大に入れるか」エラー分析会議. 2014.
57. Yoshinobu Kano. Answering Yes-No Questions by Keyword Distribution: KJP System at NTCIR-11 RITEVal Task. NTCIR-11 conference. 2014.
58. Yoshinobu Kano. Solving History Exam by Keyword Distribution: KJP System at NTCIR-11 QALab Task. NTCIR-11 conference. 2014.
59. Hideyuki Shibuki, Kotaro Sakamoto, Yoshinobu Kano, Teruko Mitamura, Madoka Ishioroshi, Kelly Y. Itakura, Di Wang, Tatsunori Mori and Noriko Kando. Overview of the NTCIR-11 QA-Lab Task. NTCIR-11 conference. 2014.
- 〔図書〕(計 3 件)
1. 宮崎 真, 阿部 匡樹, 山田 祐樹, 井隼 経子, 小野 史典, 門田 宏, 高橋 康介, 羽倉 信宏, 平島 雅也, 荒牧 勇, 有賀 敦紀, 池上 剛, 大泉 匡史, 狩野 芳伸, 黒田 剛士, 鈴木迪諒, 田中 章浩, 西村 幸男, 宮脇 陽一, 吉江 路子, 渡邊 克巳, 門田 浩二, 河内山 隆紀, 戸松 彩花, 吉田 真一. 日常と非日常からみるこころと脳の科学. コロナ社. 198p. 2017
  2. 狩野 芳伸, 大槻 恭士, 園田 亜斗夢, 中田 洋平, 箕輪 峻, 鳥海 不二夫. 人狼知能で学ぶ AI プログラミング~ 欺瞞・推理・会話で不完全情報ゲームを戦う人工知能の作り方~. マイナビ出版. 320p. 2017
  3. 鳥海 不二夫, 片上 大輔, 大澤 博隆, 稲葉 通将, 篠田 孝祐, 狩野 芳伸. 人狼知能 だます・見破る・説得する人工知能. 森北出版. 168p. 2016
6. 研究組織
- (1) 研究代表者  
 狩野 芳伸 (KANO, Yoshinobu)  
 静岡大学・情報学部・准教授  
 研究者番号: 20506729