「ドメイン依存の固有表現抽出技術の現状」付録：ドメイン依存の固有表現抽出に関連する論文リスト

・本リストは、ドメイン依存の固有表現抽出技術の現状を調査するにあたり、以下3つの学会の論文を参照し、その中からドメイン依存の固有表現抽出をテーマとしたものを抽出したものです。
  ・言語処理学会年次大会（2019年3月～2022年3月）
  ・言語処理学会論文誌（2018年1月～2021年12月）
  ・情報処理学会NL研究会（2018年5 月～2021年9 月）
  ・電子情報通信学会テキストアナリティクス・シンポジウム（2011年7月～2021年11月）
・論文の本文内で紹介していない研究も含みます。
・論文内で紹介した研究はタイトル部分をハイライトしています。
・順序は、論文内でのドメインの出現順→論文内で紹介したものは本文内の出現順・紹介していないものは学会別・発表時期の新しい順となっています。
・本リストは論文執筆作業中の参考材料として、理解可能な範囲で整理したものです。

ドメイン	発表学会／掲載誌	論文タイトル（ハイライトは本文に記載したもの）	著者	概要	課題設定／着眼点	手法	手法の一部となっているデータ（学習データ、辞書など）	入力データ	出力	URL
化学	言語処理学会年次大会（2019年）	外部知識源を使用したWikipediaからの化合物情報抽出	邊土名朝飛, 野中尋史, 小林暁雄, 関根聡	日本語版Wikipediaの化合物記事から、原材料・製造方法を抽出する。化合物の属性6つのうち、原材料・製造方法は抽出が難しいとされるが、その抽出精度を高めるために深層学習（Bi-LSTM＋CRFモデル）と外部知識源から作成した化合物辞書を組み合わせる。	化合物の属性の内、抽出が難しい原材料・製造方法について精度高く抽出したい。化合物の情報抽出に深層学習モデルを適用する場合、化合物の種類が非常に多く、トレーニングデータ中での出現頻度が低い（または出現しない）化合物名が大量にあることで精度が低下しやすい。	ベースをBi-LSTM＋CRFモデルとし、外部知識源から作成した化合物辞書を組み合わせることで出現頻度の低い表現による精度低下を防ぐ。辞書はWikiData、PubChem、ChEBIから構築したものと、日化辞を併用。	・WikiData、PubChem、ChEBIから構築した化合物名辞書と日化辞 ・上記辞書で化合物名を置換した文 → Bi-LSTM+CRFモデルの入力とする ・学習データは、森羅プロジェクトで公開されている構造化データとそれに対応するWikipedia化合物記事	Wikipediaの化合物記事	化合物の原材料と製造方法	https://anlp.jp/proceedings/annual_meeting/2019/pdf_dir/B4-8.pdf
化学	言語処理学会年次大会（2021年）	構文情報とラベルなしデータを用いた化学分野の関係抽出	新城大希, 徳永健伸, 牧野拓哉, 岩倉友哉	化学論文から、化学物質間の相互作用等の関係を自動抽出する。BioBERTの関係抽出精度をさらに高めるために、OpenIEを補助タスクとしたマルチタスク学習を行う。またPubMedにラベルを付与し学習に利用する。	BioBERTは大量のラベル付きデータを必要とすること、構文情報を利用していないことに改善の余地がある。	BioBERTによる関係抽出と、OpenIEによる補助タスクを組み合わせる。またPubMedにラベルを付与し学習に利用する。	・CHEMPROTなどのラベル付きデータとPubMedにラベルを付与したデータで固有表現抽出器・関係抽出器を学習 ・PubMedのアブストラクトにラベルを付与したものを、既存のラベル付きデータとともに学習に利用	化学関係の論文や特許など。（実験はChemprot、GAD、EU-ADRで行った）	化学物質間の関係	https://www.anlp.jp/proceedings/annual_meeting/2021/pdf_dir/C5-4.pdf
化学	情報処理学会 NL研究発表会 第241回（2019年）	辞書を用いたコーパス拡張による化学ドメインのDistantly Supervised固有表現認識	辰巳 守祐, 後藤 啓介, 進藤 裕之, 松本 裕治	Distant Supervisionを使った自動アノテーションのノイズの除去手法と、既存の辞書を使ったコーパス拡張によるRecall 向上手法	専門分野の固有表現認識では学習データのアノテーションコストが高い。それを解決する1手法として、DSによる自動アノテーションがある。だがDS NERでは教師データのノイズが多く、またRecallが低いことが課題となっている。	ノイズに関しては、4 fold cross validationによるNERの予測を行い、自動アノテーションと予測が不一致の場合はノイズとみなす。Recallに関しては、DSで生成したデータ中の特定単語と辞書の単語を入れ替えることで辞書単語を含む疑似センテンスを作り、DSデータに加えてコーパスを拡張する。	化学物質名辞書（CTD、Mesh、PubChemからの収集）、論文アブストラクト（MedlineからPubMed経由収集）、人手アノテーションデータ（ChemdNERのTestデータ）	なんらかの化学系文書（実験では論文アブストラクト）	文書内の固有表現（化学物質名など）へのアノテーション	https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=198848
化学	言語処理学会年次大会（2020年）	無機材料文献からの合成プロセス抽出のための関係抽出	牧野晃平, 國吉房貴, 小澤順, 三輪誠	無機材料の合成プロセスは複数文にわたり記述されるため、文間の関係を対象とした関係抽出手法が必要となる。深層学習モデルとルールベースモデルの2つを提案。	材料開発の分野では、膨大な文献に記述された合成プロセスの解析により探索・開発時間短縮する技術へのニーズがある。合成プロセスは複数文にわたり記述されるため、文間の関係を対象とした関係抽出手法が必要となる。	深層学習モデルとルールベースモデルの2つを提案。前者はBRANのTransformer部をBERTに置き換えたモデル。後者はデータセットの観察により定義。	下記（学術文献からの無機材料合成プロセス抽出のためのグラフ表現）で作成した、文献にタグ付けされたコーパスを拡張。材料を示すタグについてサブラベルを追加した。	タグ付きの無機材料文献	文献中の用語間の関係	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/G4-2.pdf
化学	言語処理学会年次大会（2021年）	文献抄録中の主題材料に着目した超伝導材料に関する情報抽出	山口京佑, 旭良司, 佐々木裕	文献抄録から超電導材料に関する情報をスロット抽出する。	効率的な材料探索を目指すMaterials Informaticsにおいては材料・物質特性について大量のデータが必要だが、多くのデータが構造化されていない。よって、文献抄録から超電導材料に関する情報をスロット抽出するシステムの構築を目指す。	固有表現・関係・イベントを抽出するモデルと、主題材料分類モデル（それぞれニューラルネットワーク）と、これらを統合しスロット抽出するルールベースモジュール。	超電導材料に関する文献抄録1,000件に対し固有表現クラスなどを人手でタグ付けしたもの。	超電導材料に関する文献抄録	超電導材料のElement（元素名、化合物名など）、Doping（ドーピング操作）、Value（45%などの定量表現）、SC（超電導特性に関連する固有表現）など	https://www.anlp.jp/proceedings/annual_meeting/2021/pdf_dir/P9-10.pdf
化学	言語処理学会年次大会（2021年）	Relation Extraction Task for Inorganic Material Synthesis Procedure	Shanshan liu, 松本裕治	無機材料合成手順の手順表現抽出から関係抽出までをパイプラインとして（＝jointでなく）行う。	関係抽出タスクは現実的な課題における検証が不十分であるため、無機材料合成手順の抽出という課題を選択した。	Bi-LSTMとALTOPを比較する。言語表現モデルとしてSciBERTを利用。	熱電材料に関する論文241件に人手でエンティティと関係をタグ付けしたもの。	無機材料に関する論文など	無機材料に関する「Material」「Condition」「Method」「Process」	https://www.anlp.jp/proceedings/annual_meeting/2021/pdf_dir/C7-2.pdf
化学	言語処理学会年次大会（2020年）	複数の事前学習モデルを併用した化学分野の関係抽出	肥合智史, 嶋田和孝, 渡邊大貴, 三浦明波, 岩倉友哉	化学ドメインで固有表現間の関係抽出を行うためのより軽量なモデルの検討。	化学分野の関係抽出にはBERTモデルが成果を上げているが、サイズが大きくなりがちであり計算時間がかかる。そこでContextual String Embeddings（CSE）を使った手法を提案する。	まずBiLSTMを使った文字レベルの言語モデルを事前学習する。関係分類にはBiLSTM-Attentionモデルを用いる。その入力として、既存の二種類の分散表現（GloVe分散表現とElmo分散表現）に加えて、CSEを利用する。	PubMed、PMC、ChemRxivからの科学論文アブストラクトと本文。	化学論文などの文書	固有表現間の関係（遺伝子と疾患、タンパク質と化合物など）	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/G4-3.pdf

「ドメイン依存の固有表現抽出技術の現状」付録：ドメイン依存の固有表現抽出に関連する論文リスト
・本リストは、ドメイン依存の固有表現抽出技術の現状を調査するにあたり、以下3つの学会の論文を参照し、その中からドメイン依存の固有表現抽出をテーマとしたものを抽出したものです。
　・言語処理学会年次大会（2019年3月～2022年3月）
　・言語処理学会論文誌（2018年1月～2021年12月）
　・情報処理学会NL研究会（2018年5 月～2021年9 月）
　・電子情報通信学会テキストアナリティクス・シンポジウム（2011年7月～2021年11月）
・論文の本文内で紹介していない研究も含みます。
・論文内で紹介した研究はタイトル部分をハイライトしています。
・順序は、論文内でのドメインの出現順→論文内で紹介したものは本文内の出現順・紹介していないものは学会別・発表時期の新しい順となっています。
・本リストは論文執筆作業中の参考材料として、理解可能な範囲で整理したものです。

ドメイン	発表学会／掲載誌	論文タイトル （ハイライトは本文に記載したもの）	著者	概要	課題設定／着眼点	手法	手法の一部となっているデータ （学習データ、辞書など）	入力データ	出力	URL
化学	言語処理学会年次大会（2020年）	Contextual Subword Embeddingsを考慮した文書からの化合物名抽出実験	関根裕人, 浦澤合, 乾孝司, 岩倉友哉	化合物名をより細かいサブワードに分割したものを利用した化合物名抽出。	化合物名抽出では極端に長い単語や未知語の存在が課題となる。そこで「methyl」「amino」 などのサブワードを利用した抽出を行う。	単語を分割したサブワードをBiLSTMを使ったモデルで処理。それを単語系列のBiLSTM-CRFモデルの入力とする。	CHEMDNERコーパスを実験用データとする。（論文Abstract1万件に化合物を人手でタグ付けしたもの）	化学論文などの文書	化合物名 （サブワード情報を使うことにより未知語でも抽出可能に）	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B2-4.pdf
化学	言語処理学会年次大会（2020年）	自動生成した学習データを用いたマルチタスク学習によるタンパク質と化学物質間の関係抽出	新城大希, 西川仁, 徳永健伸, 牧野拓哉, 岩倉友哉	BioBERTの精度改善	BioBERTは ・ラベル付きデータの量が限定的（作成コストがかかる） ・関係抽出では構文情報が有用だが、BioBERTでは利用していない	・Open IEの手法（文内から「２つのエンティティとその関係」を抽出）で補助タスクの学習データ作成 ・主タスク（関係抽出）とOpen IEで抽出されたペアかどうかを分類する補助タスクを同時学習して主タスクの精度向上を図る	CHEMPROTのうち、タンパク質と化学物質のペアを含む文	化学論文などの文書	タンパク質と化学物質の関係	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B2-5.pdf
化学	言語処理学会年次大会（2020年）	Data Augmentation Technique for Process Extraction in Chemistry Publications	Yuni Susanti, Hikaru Yokono, Hiroaki Yoshida	化合物合成プロセス抽出のためのデータ拡張	化学ドメインでのラベル付きデータは少なく、作成コストが高い。既存のデータを最大限活用すべくデータ拡張を行う。	・入力文と類似する文を学習データから探す・ ・入力文中のエンティティを、類似文の中で同じエンティティタイプを持つエンティティに置き換えることにより、元の文に意味の近い文を新たに得る。 プロセス抽出はBiLSTM-CRFで行う。学習に拡張データを使う割合を変化させ、抽出性能を比較する。	合成プロセス235件に対し、化学ドメインの専門家が材料や環境条件などをタグ付けしたもの。	化合物合成プロセスを含む文書	化合物合成プロセス	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B2-6.pdf
化学	言語処理学会年次大会（2020年）	無機化合物を対象とした論文に対する化学物質名抽出システムの性能分析	町光二郎, 吉岡真治	生命医化学分野で学習した言語モデルを用いた化学物質名抽出	機械学習ベースのシステムでは化学物質名抽出の際、コーパスに存在しない無機化合物の再現率が低い。ニューラル言語モデルベースのシステムとサブワードによる単語分解の枠組みを利用することで再現率を高められるのではないか。	・BioBERTを抽出に利用 ・WordPieceをサブワードに利用	・BioBERTをCHEMDNERで学習 ・評価データには、ナノ結晶デバイス開発分野の論文5件に化学物質名や物質特性をタグ付けしたもの。	化学論文などの文書	化学物質名	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B2-3.pdf
化学	言語処理学会年次大会（2020年）	学術論文からのポリマー・溶媒の固有表現および溶解性の自動抽出	山口泰弘, 進藤裕之, 松本裕治	ポリマーと溶媒のスパン予測（固有表現抽出）と、ポリマーと溶媒の間の溶解性の関係抽出	物質化学論文において、ポリマーのデータは数値として表にまとめられている事が多いのに対し、溶解性に関する情報はテキストに記述されることが多い。これを機械学習モデルで自動的に抽出したい。	・固有表現抽出はBiLSTM-CRF ・関係抽出はBiLSTM ・固有表現抽出・関係抽出ともに、単語埋め込みにchar-CNNとBERT-Base、SciBERTを用いて性能比較	ポリマーと溶媒のスパンをタグ付けした599文	物質化学論文などの文書	ポリマー、溶媒、溶解性の関係	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B2-2.pdf
化学	言語処理学会年次大会（2020年）	Extraction of Inorganic Material Synthesis Procedure from Literature	Liu Shanshan, Fusataka Kuniyoshi, Jun Ozawa, Masaki Kiyono, Yuji Matsumoto	無機化合物からの情報抽出において、従来の抽出対象は固有表現およびアクショングラフ程度にとどまっていた。これを一歩進め、合成時の圧力など、素材合成工程全体の抽出に取り組む。	無機化合物合成は複数の工程を経て行われる。各工程は、何らかの素材に対し、何らかの条件下で、何らかの加工をすることである。したがって、素材、条件、加工を抽出した後、それらの関係を抽出することで、化合物合成過程抽出ができると考えた。	材料合成過程を記述したデータに、エンティティと関係をタグ付けしたもの（Kuniyoshi et al.2019）を使う。固有表現抽出部分と関係抽出部分で構成。前者はELMo、BiLSTM、CRFを組み合わせたもの。加工エンティティにはルールベースも併用、加工表現は17種類に分類した。関係抽出部分は複数手法を比較。	材料合成過程を記述したデータに、エンティティと関係をタグ付けしたもの（Kuniyoshi et al.2019）、素材エンティティを含むデータ（Kim et al.2018）	無機化合物合成過程を記述した文書	無機化合物、その材料、加工時の環境条件、加工内容（mix、meltなど）	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B2-1.pdf
化学	言語処理学会年次大会（2019年）	化合物の同義語辞書を用いた固有表現抽出	渡邉大貴, 田村晃裕, 二宮崇, 牧野拓哉, 岩倉友哉	化合物名抽出と化合物の言い換えをマルチタスク学習することで、化合物名抽出の性能を改善する。	化合物名には表記ゆれが多いため、表現の同一性を学習する必要がある。	アテンションに基づくニューラル機械翻訳（ANMT）で化合物を言い換える。PubChem名称辞書の同一IDの化合物ペアを学習して言い換えモデ	PubChem名称辞書（言い換えモデルの教師データ作成に利用）	化学ドメインの文書	化合物名	https://anlp.jp/proceedings/annual_meeting/2019/pdf_dir/B4-2.pdf
化学	言語処理学会年次大会（2019年）	学術論文からのポリマー溶解性データの自動抽出	岡博之, 吉澤篤志, 進藤裕之, 松本裕治, 石井真史	学術論文からのポリマーとその溶解性（どの溶媒に対し可溶性を持つか）の関係抽出	現在ポリマーデータを人手で抽出しDB化している。それを効率化したい。	・ポリマー名抽出はルールベース ・溶媒名抽出は辞書マッチング ・関係抽出はルールベース	・溶媒名辞書（135件）	化学ドメインの文書	ポリマー名ー良溶媒名の関係	https://anlp.jp/proceedings/annual_meeting/2019/pdf_dir/B4-4.pdf
化学	言語処理学会年次大会（2019年）	化学ドメインにおける教師無し固有表現抽出	辰巳守祐, 進藤裕之, 松本裕治	化学物質名の固有表現抽出において、以下を明らかにする。 ・分散表現抽出において、文／単語／サブワード／文字ベースのうち、どの処理単位が最も有効か ・Distant Supervisionで生成された擬似コーパスからどのようにノイズを取り除くか	化学物質名の固有表現抽出では （1）新しい化学物質が日々生み出される→文脈情報を用いた分散表現を使う手法が考えられてきた。では、分散表現抽出はどのような単位で行うべきか。 （2）辞書やコーパス作成に専門知識が必要　→　Distant Supervision　→　ノイズが生じるのでいかに取り除くか。	（1）Flairによる分散表現抽出を文字ベースとサブワードベースで行ったうえでBiLSTM-CRFでの固有表現抽出を行い、性能を比較。 （2）学習済みの固有表現抽出器に推論させてFalse Positiveとなる単語を除去する。	実験では以下を使用。 ・分散表現の事前学習データとしてMedline ・固有表現抽出器の学習データとしてChemdNERとMedline ・Distant Supervisionに使う化学物質辞書としてCTDとMeSH（計40万語）	化学ドメインの文書	化学物質名	https://anlp.jp/proceedings/annual_meeting/2019/pdf_dir/B4-3.pdf

「ドメイン依存の固有表現抽出技術の現状」付録：ドメイン依存の固有表現抽出に関連する論文リスト

・本リストは、ドメイン依存の固有表現抽出技術の現状を調査するにあたり、以下3つの学会の論文を参照し、その中からドメイン依存の固有表現抽出をテーマとしたものを抽出したものです。
- ・言語処理学会年次大会（2019年3月～2022年3月）
- ・言語処理学会論文誌（2018年1月～2021年12月）
- ・情報処理学会NL研究会（2018年5 月～2021年9 月）
- ・電子情報通信学会テキストアナリティクス・シンポジウム（2011年7月～2021年11月）

・論文の本文内で紹介していない研究も含みます。
・論文内で紹介した研究はタイトル部分をハイライトしています。
・順序は、論文内でのドメインの出現順→論文内で紹介したものは本文内の出現順・紹介していないものは学会別・発表時期の新しい順となっています。
・本リストは論文執筆作業中の参考材料として、理解可能な範囲で整理したものです。

ドメイン	発表学会／掲載誌	論文タイトル （ハイライトは本文に記載したもの）	著者	概要	課題設定／着眼点	手法	手法の一部となっているデータ （学習データ、辞書など）	入力データ	出力	URL
化学	自然言語処理	化学分野への言語処理の応用	岩倉友哉,吉川和	化学分野での自然言語処理技術の応用状況調査	-	-	-	-	-	https://www.jstage.jst.go.jp/article/jnlp/27/4/27_969/_pdf/-char/ja
化学	情報処理学会 NL研究発表会 第249回（2021年）	複数の補助教師データセットを用いた固有表現抽出の学習手法	市川 智也, 渡邉 大貴, 田村 晃裕, 岩倉 友哉, 馬 春鵬, 加藤 恒夫	複数（実験では７種類）の化学・科学技術分野の補助教師データを使うことで固有表現抽出の精度を向上する手法。	近年NERの精度向上のために行われている補助学習では、補助学習用の教師データが1種類のみだった。精度をさらに高めるために、複数の教師データを使う手法を提案する。	教師データを複数（7種類）用意し、（1）データごとの学習を順次行う、または（2）全種類のデータをひとつの補助学習に用いる。	メインの教師データ（CHEMDNER）＋7種類の教師データ（NCBI Disease、BC5CDR（DiseaseとDrug/Chem）、BC2GM、JNLPBA、LINNAEUS、	化学ドメインの文書	化合物名、化学式、化合物の商品名といった化学ドメインの固有表現	https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=212204&item_no=1&page_id=13&block_id=8
化学	人工知能学会全国大会									
化学	人工知能学会全国大会									
化学	人工知能学会全国大会									
化学	人工知能学会全国大会									
化学	人工知能学会全国大会									
化学	電子情報通信学会 テキスト・シンポジウム 第15回（2019年）	学術文献からの無機材料合成プロセス抽出のためのグラフ表現	國吉房貴, 小澤順, 藤井幹也, 森川幸治, 中田透, 井垣恵美子, 日比野純一, 清野正樹, 三輪誠	文献からの無機材料合成プロセス抽出を目的とした、合成プロセスのグラフ表現手法を検討。	有機化学分野と比較し、無機化学分野ではテキストにタグ付けされたコーパスがほとんど存在しない。そのためテキストからの合成プロセス抽出手法開発が停滞している。	合成プロセスを以下のようにグラフ表現する。 ・原料・処理・条件を3つのノードで表現 ・原料・条件間、処理・条件間、処理・処理間の順序関係をエッジで表	合成プロセスをグラフ表現でアノテーションしたデータ	化学ドメインの文書	無機材料合成プロセス	https://www.ieice.org/ken/paper/20190927V1OB/
化学	電子情報通信学会 テキスト・シンポジウム 第2回（2012年）	特許文書からの化学物質情報の抽出	田中一成・池田紀子	特許文書から、化学物質名と化学式の対をルールにより抽出する。	化学物質にはさまざまな異表記があるが、新規物質が多い、コストが高いなどの理由で、辞書による名寄せには限界がある。	特許文書から、「プロパン（CH3CH2CH3）」のように「片仮名、英数、『酸』などの一部の漢字、括弧が連続して並ぶ文字列」を抽出、括弧前の文字列と括弧内の文字列を化学物質名ー化学式と見なした。さらにゴミを排除する規則を作り、物質名と化学式の関係を抽出した。さらに、抽出した情報といくつかの命名規則を組み合わせることで、直接抽出できなかった物質名に対応する化学式を生成した。		特許文書	化学物質名と化学式の対	https://www.ieice.org/ken/paper/20120830W0WC/
医療・薬事	言語処理学会年次大会（2019年）	電子カルテ自由記述部分からの皮膚疾患における重症度抽出	加藤由美, 平川聡史, 梶山晃平, 堀口裕正, 狩野芳伸	電子カルテの自由記載部分から、抗がん薬によって起こる代表的な４つの皮膚障害の重症度を抽出	電子カルテの自由記載から疾患の重症度（有無でなく）を判定する例はまだ少ない。否定語の有無、疾患の状態を表すキーワードを使い、精度向上を図る。	SVMと多項式カーネルでの学習。比較として、ルールベースでの推測。	・模擬カルテに人手でタグ付けしたもの。（<karte>タグの属性として、診察日・医師の診断結果のICDコードとGrade）を付与） ・上記に対し、発疹の性状について9つの観点でキーワード付与 ・発疹の性状等を表す語550語の人手作成辞書	模擬電子カルテデータ	日付、疾患、重症度	https://anlp.jp/proceedings/annual_meeting/2019/pdf_dir/E3-2.pdf

「ドメイン依存の固有表現抽出技術の現状」付録：ドメイン依存の固有表現抽出に関連する論文リスト
・本リストは、ドメイン依存の固有表現抽出技術の現状を調査するにあたり、以下3つの学会の論文を参照し、その中からドメイン依存の固有表現抽出をテーマとしたものを抽出したものです。
　・言語処理学会年次大会（2019年3月〜2022年3月）
　・言語処理学会論文誌（2018年1月〜2021年12月）
　・情報処理学会NL研究会（2018年5 月〜2021年9 月）
　・電子情報通信学会テキストアナリティクス・シンポジウム（2011年7月〜2021年11月）
・論文の本文内で紹介していない研究も含みます。
・論文内で紹介した研究はタイトル部分をハイライトしています。
・順序は、論文内でのドメインの出現順→論文内で紹介したものは本文内の出現順・紹介していないものは学会別・発表時期の新しい順となっています。
・本リストは論文執筆作業中の参考材料として、理解可能な範囲で整理したものです。

ドメイン	発表学会／掲載誌	論文タイトル （ハイライトは本文に記載したもの）	著者	概要	課題設定／着眼点	手法	手法の一部となっているデータ （学習データ、辞書など）	入力データ	出力	URL
医療・薬事	言語処理学会年次大会（2020年）	医薬品添付文書からの薬剤情報抽出システム	小島諒介, 岩田浩明, 中津井雅彦, 奥野恭史	市販薬の添付文書のPDF群から、効能などの情報を抽出できるシステム。	市販薬の情報は新薬開発効率化に有用だが、企業横断の情報源が乏しい。医薬品添付文書のPDFはまとめてアクセスしやすいため、そこから情報抽出するシステムを構築する。	・PDFからのテキスト・レイアウト抽出 ・抽出対象の情報のアノテーション（添付文書をSGML化したデータを代替とした） ・「比較的規模の小さい」、「BiLSTMを含む」ネットワークでの深層学習 ・クエリに対し回答を返すインターフェース	・添付文書をSGML化したデータ	市販薬添付文書（または薬剤関連文書一般）、クエリ	クエリへの回答（薬剤の効能、禁忌など）	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/G4-5.pdf
医療・薬事	言語処理学会年次大会（2019年）	データベースの説明文を利用した薬物相互作用抽出	浅田真生, 三輪誠, 佐々木裕	薬物相互作用抽出に、薬物データベース（DrugBank）にある薬物説明文を利用する。	薬学論文等で報告される薬物相互作用を効率よくデータベースに登録するため、深層学習を使った自動抽出手法が検討されている。そこに既存の薬物データベースを活用したい。	DrugBankの薬物説明文と、相互作用抽出元となる文書それぞれをCNNで表現し、ふたつのCNNを同時に学習する。	DrugBankの薬物説明文（薬物名に対し説明文が対応）	薬学論文等	薬物間の相互作用（動態的作用、薬力学的作用、併用の際の推奨、相互作用有無）	https://anlp.jp/proceedings/annual_meeting/2019/pdf_dir/B4-9.pdf
医療・薬事	自然言語処理	病名アノテーションが付与された医療テキスト・コーパスの構築	荒牧 英治, 若宮 翔子, 矢野 憲, 永井 宥之, 岡久 太郎, 伊藤 薫	電子カルテに対し病名（および当該カルテの患者においてその症状が発生しているかどうか）をアノテーションした、45,000テキストの大規模コーパス構築。	大量の電子カルテをアノテーションしたコーパスが存在していない。	電子カルテへの病名アノテーションの詳細な仕様を作成し、フィージビリティを検討する。また、作成したコーパスを使った病名抽出器を構築しアノテーションを検証。	日本内科学会に報告された44,761件の症例報告。あらかじめ少数のデータを作成し、それを機械学習することで全体に仮のタグを付け、効率化を図った。	症例報告	病名、およびその症状が発生しているかどうか	https://www.jstage.jst.go.jp/article/jnlp/25/1/25_119/_pdf/-char/ja
医療・薬事	電子情報通信学会 テキスト・シンポジウム 第9回（2016年）	Twitterを用いた皮膚障害情報の抽出	阿部健一, 吉田博哉	消費者被害、なかでも皮膚障害被害の拡大を未然に防ぐべく、Twitterから皮膚障害情報を抽出したい。最終的には、被害を出している商品名や企業名を特定したいが、ここではその前段階として、消費者被害に関するツイートを特定するシステムを構築。	化粧品等による健康被害を早期に発見するため、Twitterから被害事例を抽出したい。ここでは被害事例を含むらしいツイートの特定を行う。	危険表現を含むツイートの中からノイズとなる文やTwitterユーザーを排除し、皮膚障害らしさ（信頼度）を付与してコーパスを構築。システムのユーザーが入力する危険表現を含むツイートを出力する。	・皮膚障害に関する危険表現を取得するためのキーワードとして、「かぶれた」等6種類の語を使用。	Twitterデータ、クエリ（危険表現）	危険表現を含むツイート	https://www.ieice.org/ken/paper/20160909zbks/
企業情報・金融	言語処理学会年次大会（2020年）	ニュース記事からの企業キーワード抽出	奥田裕樹, 髙橋寛治	Sansanでユーザー向けに配信しているニュース記事（新聞や通信社の記事＋企業プレスリリース）から、「企業活動の中で生まれたモノやサービスを表す名称」を抽出。	Sansanでは、ユーザーが名刺に記載された企業の情報を検索する際、「名刺に書かれている内容以上の情報での検索」を可能にしている。そのために必要となる企業キーワードの収集を自動化したい。	ニュース記事の中で鉤括弧で囲まれた部分を抽出し、それが企業キーワードであるかどうかの二値分類を行う。判定には、BiLSTM-CRFのうち、Contextual String Embeddings（CSE）による系列モデルを適用した。	2019年に配信されたニュース記事のうち3,978件からカギカッコ部分7,225件抽出、企業キーワードかどうかを人手で判定したもの。	Sansanでユーザー向けに配信しているニュース記事（新聞や通信社の記事＋企業プレスリリース）	企業キーワード	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/P2-13.pdf
金融・企業情報	言語処理学会年次大会（2020年）	StruAPを用いた金融分野の開示文書からの情報抽出	柳井孝介, 佐藤美沙, 十河泰弘 , 山脇功一, 渋谷淳	有価証券報告書などから、売上、利益、キャッシュフローなど投資家が投資判断に使える情報を抽出。	有価証券報告書や有価証券届出書は膨大に開示されており、投資家が必要とする情報は１文書あたり５０〜８０項目と、人手での理解にコストがかかる。これら文書はXBRLで公開されており、「〇〇をベンチマークとして」のような定型的表現も多いため、これらを活用できる。また提案手法であるStruAPを使うことで、表現そのものではなく、木構造のパターンを使った抽出が可能になっている。	抽出したいもの自体の辞書ではなく、木構造のパターンと関係を表す表現の辞書でマッチングする。	・木構造パターン474件 ・辞書80語	有価証券報告書などの金融文書	『売上高』「キャッシュフロー」等の経営指標名、および「8.0%」のような経営指標値。	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B1-4.pdf
金融・企業情報	自然言語処理（2020年）	金融・経済ドメインを対象とした言語処理	坂地 泰紀, 和泉 潔, 酒井 浩之	金融・経済ドメインでの自然言語処理の状況調査	-	-	-	-	-	https://www.jstage.jst.go.jp/article/jnlp/27/4/27_951/_pdf/-char/ja
金融・企業情報	自然言語処理（2021年）	クラウド名刺管理サービスに関連する自然言語処理の取り組み	髙橋 寛治, 真鍋 友則	名刺管理サービスに関連して行われる自然言語処理の活用事例紹介（本表では企業名抽出について述べる）	名刺交換した相手の社名とニュースを紐付け、ビジネス上の気付きを与えたい。	ルールベース（≒辞書マッチ）および機械学習（Transformerモデル）	企業名辞書、機械学習で間違いやすい項目のブラックリスト	ニュース記事	該当する企業に関連するニュース	https://www.jstage.jst.go.jp/article/jnlp/28/1/28_297/_pdf/-char/ja

「ドメイン依存の固有表現抽出技術の現状」付録：ドメイン依存の固有表現抽出に関連する論文リスト
・本リストは、ドメイン依存の固有表現抽出技術の現状を調査するにあたり、以下3つの学会の論文を参照し、その中からドメイン依存の固有表現抽出をテーマとしたものを抽出したものです。
　・言語処理学会年次大会（2019年3月～2022年3月）
　・言語処理学会論文誌（2018年1月～2021年12月）
　・情報処理学会NL研究会（2018年5 月～2021年9 月）
　・電子情報通信学会テキストアナリティクス・シンポジウム（2011年7月～2021年11月）
・論文の本文内で紹介していない研究も含みます。
・論文内で紹介した研究はタイトル部分をハイライトしています。
・順序は、論文内でのドメインの出現順→論文内で紹介したものは本文内の出現順・紹介していないものは学会別・発表時期の新しい順となっています。
・本リストは論文執筆作業中の参考材料として、理解可能な範囲で整理したものです。

ドメイン	発表学会／掲載誌	論文タイトル （ハイライトは本文に記載したもの）	著者	概要	課題設定／着眼点	手法	手法の一部となっているデータ （学習データ、辞書など）	入力データ	出力	URL
企業情報・金融	電子情報通信学会 テキスト・シンポジウム 第3回（2013年）	企業WEBページからの企業の事業に関連するキーワードの自動抽出	勝田研一郎・酒井浩之	企業の事業内容を素早く把握すべく、事業関係キーワードを抽出	企業に関する知識の浅い学生が就職活動をするにあたり、自身の関心や専門分野に近い企業を探すことは難しい。そこでたとえば「テキストマイニング」で検索した結果「野村総合研究所」や「マクロミル」といった関連性の高い企業がヒットするようなシステムを作りたい。	1）企業のWebページから名詞を抽出し頻度順に並べる 2）IDF値の低い名詞を排除することで一般的すぎる語を除く 3）ただしIDFが低い語を人手でチェックし、重要なものであれば残す 4）出現する企業数の少ない語（特殊すぎる用語）を除く 5）IDF値の低い名詞を含む語も排除する ※ たとえば「凸版印刷」において、「印刷」は１で抽出され、２で排除されるが３で復活し、５で「印刷ページ」「印刷画面」といった明らかに不適切なものを排除できる。		ニュース記事	企業に関連性の深いキーワード	https://www.jstage.jst.go.jp/article/pjsai/JSAI2013/0/JSAI2013_2H2OS2b1/_article/-char/ja/
機械加工	言語処理学会年次大会（2021年）	機械加工文書における用語入れ子構造とトリガワードを考慮した用語関係同時抽出	稲熊陸, 小島大, 東孝幸 , 三輪誠, 古谷克司, 佐々木裕	機械加工技術文書内の「切削速度が増加すると切削温度が増す」といった因子（切削速度・切削温度など）とそれらの関係（Aが増すとBも増す　など）を、トリガワード（「物理量を示す二用語間の変化を表す単語」）を考慮して抽出	ものづくり現場の高齢化により、技術継承が難しくなっている。熟練を要する業務に、工程策定業務があり、それを行うには機械加工因子間の関係の知見が必要となる。そのため、機械加工関連の文書から、機械加工因子と、因子間の関係を同時に抽出できる技術が求められている。	入れ子構造を考慮した用語抽出と、「トリガワード（物理量を示す二用語間の変化を表す単語）」を考慮した関係抽出を行う。	・切削加工に関する教科書文に対し、用語同士の関係ラベル（例：Aが大きくなるとBが大きくなる→Positive、など4パターン）を付与したもの ・各文のトークン分割のために、Sentencepieceを日本語版Wikipediaで事前学習する。	機械加工の切削加工に関する教科書文（文中に関係ラベルを付与）	機械加工用語、用語間の関係4種類（正の相関・負の相関・AはBの一種の関係・定性的な関係）	https://www.anlp.jp/proceedings/annual_meeting/2021/pdf_dir/B1-5.pdf
機械加工	言語処理学会年次大会（2020年）	入れ子構造を考慮した機械加工用語抽出	稲熊陸, 小島大, 東孝幸, 三輪誠, 古谷克司, 佐々木裕	機械加工分野の技術者の判断支援や知見の継承支援のために知識ベースを作りたい。その第一歩として、機械加工文書から機械加工用語とその関係を抽出する。	機械加工用語は「切削加工」が「切削」と「加工」を含むように入れ子構造になっている事が多い。	BERT→畳み込みニューラルネットワーク→トークン数ごとの用語抽出を行う。	機械加工分野の教科書2,881文に対し人手でアノテーションしたもの。	機械加工文書	機械加工用語	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B5-3.pdf
文学（小説）	言語処理学会年次大会（2021年）	小説あらすじを用いて学習した系列ラベリングモデルによる小説本文からの人物情報抽出の性能検証	岡裕二, 安藤一秋	あらすじのテキストで事前学習したモデルを使い、小説から人物の性別や年齢、職業などを抽出する。	ライトノベルなどの作品が増え、作品を探す労力が増大している。特に、小説の内容に踏み込んだ検索機能が実装されていない。小説内の人物相関図やあらすじの生成を目指すことでそれに資する。	小説のあらすじデータにタグ付けして4つの深層学習モデルで学習、固有表現抽出。	・小説のあらすじデータ（NIIのWebcat Plusから、Wikipediaの日本の小説家一覧の小説家名で検索したもののうち、同「日本のファンタジー作家一覧」にある作家の作品、または「ファンタジー」という単語を含むもの。 ・小説の本文データ（なろう小説API利用） ・人手によるタグ付け	小説の本文またはあらすじ	文中にある名前、性別、年齢、容姿、職業、所属、場所、人物関係など	https://www.anlp.jp/proceedings/annual_meeting/2021/pdf_dir/B1-1.pdf
文学（小説）	言語処理学会年次大会（2020年）	系列ラベリングによる小説のあらすじからの人物情報抽出の検討	岡裕二, 安藤一秋	小説のあらすじから人物の属性などを抽出	同上。本稿ではあらすじからの人物情報・人物関係表現抽出手法を検討。	小説あらすじテキスト（1008件、約5000文）に対し、名前・性別・年齢表現をタグ付け。CRFでラベリングを行う。素性として、表記等とともに文字uni-gram、bi-gram 等を加え、8パターンの素性組み合わせを作って性能比較。	タグ付けしたあらすじデータ	小説のあらすじ	登場人物の名前、性別表現、年齢表現、容姿や特性表現、職業や立場表現、組織・種族名、その他（異星人、神等）、地名や建物名、人物関係表現	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/B5-4.pdf
食	電子情報通信学会 テキスト・シンポジウム 第13回（2018年）	レストラン・レビューにおける食べ物・飲み物表現の抽出	新堂安孝, 友利涼, 富田紘平, 兼村厚範, 森信介	レストランレビューをマーケティングに利用するため、「そば」のような単純な表現ではなく「「香り高くのどごし抜群のおいしい十割そば」のような長く複雑なものを抽出したい。現状ではその難しさを定量的に表したデータもないため、その評価も含めて実験を行う。	レストランレビューをマーケティングに利用するため、固有表現抽出のRecall、Precisionともに改善したい。「そば」のような単純な表現ではなく「「香り高くのどごし抜群のおいしい十割そば」のような長く複雑なものを抽出したい。現状ではその難しさを定量的に表したデータもないため、その評価も含めて実験を行う。	飲み物・食べ物表現を人手でアノテートしてモデルの学習、開発、評価を行う。モデルはナイーブなCRFベースと、DNNを用いたCRFベースを比較。	・2016年の食べログレビューデータ230万件を元に50ジャンルから一定数のレビューを抽出したもの ・上記から飲み物・食べ物表現を人手でアノテーションしたもの	食べログのレストランレビュー	「香り高くのどごし抜群のおいしい十割そば」のように、食べ物・飲み物＋その性質を表す表現。	https://www.jstage.jst.go.jp/article/pjsai/JSAI2018/0/JSAI2018_1I2OS1_/_article/-char/ja/

「ドメイン依存の固有表現抽出技術の現状」付録：ドメイン依存の固有表現抽出に関連する論文リスト

・本リストは、ドメイン依存の固有表現抽出技術の現状を調査するにあたり、以下3つの学会の論文を参照し、その中からドメイン依存の固有表現抽出をテーマとしたものを抽出したものです。
　・言語処理学会年次大会（2019年3月～2022年3月）
　・言語処理学会論文誌（2018年1月～2021年12月）
　・情報処理学会NL研究会（2018年5 月～2021年9 月）
　・電子情報通信学会テキストアナリティクス・シンポジウム（2011年7月～2021年11月）
・論文の本文内で紹介していない研究も含みます。
・論文内で紹介した研究はタイトル部分をハイライトしています。
・順序は、論文内でのドメインの出現順→論文内で紹介したものは本文内の出現順・紹介していないものは学会別・発表時期の新しい順となっています。
・本リストは論文執筆作業中の参考材料として、理解可能な範囲で整理したものです。

ドメイン	発表学会／掲載誌	論文タイトル （ハイライトは本文に記載したもの）	著者	概要	課題設定／着眼点	手法	手法の一部となっているデータ （学習データ、辞書など）	入力データ	出力	URL
食	情報処理学会 NL研究発表会 第237回（2018年）	文字分散表現に基づく単語分類情報を用いたレシピ固有表現抽出	平松 淳 , 若林 啓 , 原島 純	ドメイン（この例では料理）に関連する言語資源を使った固有表現抽出	固有表現抽出の教師データをドメインごとに構築するのはコストが大きい。よって、文中の単語をカテゴリに分類し、分類情報を固有表現抽出器の入力として利用する。	LampleらのBiLSTM-CRFの処理に、単語分類器からの情報を追加する。文中の単語について、オントロジーでの属性ラベルを予測する分類器を学習し、固有表現抽出器の特徴量に組み込む。	・レシピNEコーパス（笹田ら） （クックパッドの手順データに固有表現を付与したもの） ・料理オントロジー（Nanbaら） （料理関係の用語に上位下位関係、同義語などを付与したもの）	料理ドメインのテキスト （長期的には、より一般的なドメインに適用したい）	料理関係の固有表現	https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=191443
食	人工知能									
みやげ品	情報処理学会 NL研究発表会 第243回（2019年）	固有表現抽出によるブログテキストからの品名・店名抽出	池田 流弥 , 安藤 一秋	「現地でしか購入できない土産」の情報をブログなどのUGCから抽出する手法	現地でしか購入できない土産情報をWebから収集するシステムを構築するにあたり、既存の固有表現抽出手法の多くはUGCや日本語テキストに対し未評価である。よって、日本語ブログから構築したデータを用い、CRFと深層学習による固有表現抽出の性能評価を行う。	CRFとBiLSTM-CRFの2モデル、BiLSTM-CNN-CRF、Char-BiLSTM-CRF。	・日本の著名な土産をまとめたOMIYA!による土産名をクエリとし、Yahoo!ブログの菓子・デザートカテゴリでヒットしたブログ記事のうち680エントリに人手で固有表現タグを付与したもの。	ブログ記事	土産名、土産／菓子店名	https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=200744&item_no=1&page_id=13&block_id=8
みやげ品	言語処理学会年次大会（2019年）	深層学習によるブログ記事からの土産の品名・店名抽出	池田流弥, 安藤一秋	「現地でしか購入できない土産」の情報をブログなどのUGCから抽出する手法	多くの人が旅行の際に「現地でしか購入できない土産」を求めているが、「現地でしか購入できない」という情報がほとんどない。そこでWebから収集した情報をもとに、土産品情報を抽出したい。 （本研究では「現地でしか購入できないかどうか」は判定せず、土産品情報全体を抽出する）	Yahoo!ブログから土産情報の書かれた記事を収集し、土産の品名・店名を人手でタグ付けする。BiLSTM-CRFでの抽出とCRFのみでの抽出を比較する。	・土産サイトOMIYA!の土産名リスト（ブログ記事収集時のクエリとして使用） ・ブログ記事（680件）への土産品名・店名にタグ付けしたデータ（学習用） ・Wikipedia本文全体（分散表現構築用）	ブログ記事	土産名、土産／菓子店名	https://anlp.jp/proceedings/annual_meeting/2019/pdf_dir/P3-4.pdf
交通	言語処理学会年次大会（2020年）	オントロジー形式アノテーションを対象とした交通用語・関係抽出と正誤問題の回答	鈴木直樹, Savong Bou, 三輪誠, 佐々木裕	交通関係の文書からの用語関係抽出にあたり、タグ付けデータの形式を変えることで精度を向上させた。	従来の交通関係用語抽出／関係抽出は精度が不十分だった。用語分類において似た意味のタグがあること、関係の分類が細かすぎたことが要因ではないか。	似た意味を持つタグを整理し、関係を用語としてタグ付け（オントロジー形式でアノテーション）する。関係の種類は47種類→5種類に削減。AlanらによるFlairで用語抽出。	オントロジー形式でタグ付けしたデータ。	交通法規に関する文書	交通用語と用語間の関係	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/P6-24.pdf
交通	言語処理学会年次大会（2019年）	CNNを用いた交通教則からの交通用語間関係抽出	八木智也, 三輪誠, 佐々木裕	自動運転での利用を目指した交通オントロジーのための、交通用語同士の関係抽出。	現在の自動運転プログラムには交通法規やマナーなどの知識がコード内に組み込まれており、更新が難しい。そのような知識を交通オントロジーとして整理し、自動運転プログラムから参照するという構造にすれば、法改正や異なる国での利用も容易になると考えられる。オントロジー構築には膨大な作業が必要となるため、それを補助するようなテキストからの情報自動抽出が求められる。	交通文書（約2000文）内の用語間に66種類のタグを付け、CNNを用いたモデルで用語ペア間の関係を学習する。	・交通文書へのタグ付けデータ	交通文書	文書内の用語同士の関係	https://anlp.jp/proceedings/annual_meeting/2019/pdf_dir/P2-1.pdf
その他	言語処理学会年次大会（2020年）	会議録に含まれる法律名を対象としたEnd-to-Endのエンティティリンキングの性能評価	桧森拓真, 木村泰知, 荒木健治	会議録から法律名（皮肉表現など多様な表記ゆれ含む）を抽出。	法律名には「働き方改革関連法」を「過労死促進法」と呼ぶなど、人手でも難しい表記ゆれがありうる。そのため、メンションを「法律名を表す語句」とし、法律名のメンション抽出および曖昧性解消の両方（End-to-End Approach）を行う。	"国会・地方議会会議録上の法律名部分にタグ付け（メンションとそれにひもづくWikipedia記事のアノテーション）を行う。 deeppavlov（BERT）で固有表現抽出モデルを作成してメンションを抽出。それをWikipedia記事と結びつけて曖昧性解消を行う。"	"・国会・地方議会会議録10日分（5つの検索語に対し各2日分）に対し、人手でタグ付けを行った。 ・タグ付けの中で、各メンションと対応するWikipedia記事を結びつけた。"	法律名を含む会議録	法律名と、それに対応するWikipedia記事。	https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/E1-2.pdf
その他	人工知能学会全国大会（2019年）	ECサイトにおける商品タイトルからの商品名抽出	張 培楠	ECサイトの商品タイトルから商品名だけを抽出。	商品タイトルはSEO対策のために商品名以外の要素が多く付随しておりわかりにくい。一般的な文章と違い名詞・名詞句の羅列で有ることが多く、このタスクに特化した手法が必要。	Term Weighting問題としてTF-IDFを使う手法と、系列ラベリング問題としてCRFを使う方法、BiLSTM-CRFを使う手法を比較した。	Yahoo!ショッピングから抽出した商品タイトル1万5000件にタグ付けしたもの。	商品タイトル （例：【純正品】HPインクカートリッジ【送料無料】）	商品名 （例：インクカートリッジ）	
その他	人工知能学会全国大会（2020年）	構文解析情報を用いたテキストからの数値情報の抽出	黒土 健三, 森本 康嗣, 佐藤 美沙, 柳井 孝介	論文から技術トレンドを把握するための数値情報抽出。ここでは応用物理分野の論文を対象とする。	科学論文中の数値情報からは、ムーアの法則のような技術トレンドを読み取れるのではないか。	StruAP（係り受け構造に基づくルールベースの抽出ツール）を用いる。		論文 （実験では半導体パッケージング技術に関する論文）	数値と項目名のペア	

「ドメイン依存の固有表現抽出技術の現状」付録：ドメイン依存の固有表現抽出に関連する論文リスト
・本リストは、ドメイン依存の固有表現抽出技術の現状を調査するにあたり、以下3つの学会の論文を参照し、その中からドメイン依存の固有表現抽出をテーマとしたものを抽出したものです。
　・言語処理学会年次大会（2019年3月～2022年3月）
　・言語処理学会論文誌（2018年1月～2021年12月）
　・情報処理学会NL研究会（2018年5 月～2021年9 月）
　・電子情報通信学会テキストアナリティクス・シンポジウム（2011年7月～2021年11月）
・論文の本文内で紹介していない研究も含みます。
・論文内で紹介した研究はタイトル部分をハイライトしています。
・順序は、論文内でのドメインの出現順→論文内で紹介したものは本文内の出現順・紹介していないものは学会別・発表時期の新しい順となっています。
・本リストは論文執筆作業中の参考材料として、理解可能な範囲で整理したものです。

ドメイン	発表学会／掲載誌	論文タイトル （ハイライトは本文に記載したもの）	著者	概要	課題設定／着眼点	手法	手法の一部となっているデータ （学習データ、辞書など）	入力データ	出力	URL
その他	自然言語処理（2021年）	将棋解説文への固有表現・モダリティ情報アノテーション	亀甲 博貴, 松吉 俊, John Richardson, 牛久 敦, 笹田 鉄郎, 村脇 有吾, 鶴岡 慶雅, 森 信介	実世界に紐づいた自然言語処理のテストベッドとして、将棋のゲーム局面に付随する解説文に着目。人手による単語分割・固有表現・モダリティ表現・事象の事実性をアノテーションして将棋解説文コーパスを作成した。	シンボルグラウンディング課題が注目され、画像や動画のような実世界と自然言語をひもづけたコーパスが数多く公開されている。将棋解説文は過去や未来への言及を含み、その多くがゲーム木に対応付けることができるため、固有表現やモダリティをタグ付けしたコーパスを作ることでシンボルグラウンディング研究に活用できるのではないか。	専門家による解説文が付与されたプロ棋士同士の対局棋譜6523件を配信サイトから収集。うち9棋譜の解説文に、将棋分野の固有表現とモダリティ表現を人手でタグ付けする。このタグ付けデータでPWNERを学習させ、残りの解説文に対し自動でタグを付与する。	9棋譜の解説文に将棋固有表現・モダリティを人手でタグ付けしたもの	将棋解説文	将棋固有表現（将棋特有の符号、定跡、局面に関する表現など）、モダリティ	https://www.istage.jst.go.jp/article/jnlp/28/3/28_847/_pdf/-char/ja
その他	電子情報通信学会 テキスト・シンポジウム 第8回（2015年）	陸上競技ブログからの活動記録抽出の試み	佐野正和, 福原知宏, 増田英孝, 山田剛一	スポーツ選手のブログから「バック懸垂 10回 x 2」のような練習メニューを抽出することで、他の選手のトレーニング状況を把握し、モチベーションを高めるシステム。	スポーツ選手のモチベーション向上のためにはライバルの存在が有効だが、必ずしも身近にライバルが得られるとは限らない。そこでブログを情報源として他の選手のトレーニング状況を知ることのできるシステムを構築したい。	Webから陸上競技選手のブログをクローリングし、活動内容・大会記録を正規表現により抽出。活動内容（メニュー名・量など）は隠れマルコフモデルを用いて活動量を抽出。	・陸上競技用語辞書（「坂ダッシュ」などのメニュー名） ・ブログ記事611件に対し、活動内容をタグ付けしたもの	陸上競技選手のブログ	メニュー名・内容・量・セット数（例：ベンチプレス・65キロ・10回・２）	https://www.jstage.jst.go.jp/article/tsgshinpo/2015/0/2015_0_1/_pdf/-char/ja
その他	電子情報通信学会 テキスト・シンポジウム 第4回（2013年）	Twitterを用いた特定エリアにおける注目話題の抽出とその可視化	六瀬聡宏, 清水真, 古橋慎之介, 高畑洋貴, 近藤直人, 佐藤智貴, 遠藤岳, 渡辺雅史, 内田理	特定地域のリアルタイム性の高い話題をTwitterから抽出する。	自治体などが観光・防災などの目的でポータルサイトで提供する情報では、非公式イベントや突発的状況に対応できない。そこで特定エリアにおけるリアルタイム性の高い注目話題を抽出するシステムを構築したい。	対象エリアを含むツイートの収集→注目話題の抽出（Yahoo!デベロッパーネットワークのキーフレーズ抽出APIによる特徴語抽出・重要度スコア付与）→注目話題へのスコア付与（Yahoo!によるスコア x 時間経過）→スコアに基づく注目話題の選定と提示		地名を含むツイート	地名に関連した注目話題（獅子舞・中華街など）	https://www.jstage.jst.go.jp/article/tsgshinpo/2013/0/2013_0_1/_pdf/-char/ja