

言語処理学会第24回年次大会ワークショップ

形態素解析の今とこれから

開催趣旨

単語分割・形態素解析技術は、日本語を対象にした自然言語処理研究・開発において避けて通れない処理として認識されており、これまで、数多くの辞書・システムが開発されてきました。ユーザの選択肢が増えた利点はあるものの、既存の形態素解析器・辞書・品詞体系から、応用に則したシステムを選択することは、言語処理研究者でも難しく、なんとなく新しそうだから、研究室の方針だから、先輩が使っていたからといった、場当たりの基準で選択されていることが少なくありません。本ワークショップでは、単語分割・形態素解システム、辞書の開発者にお集まりいただき、各システムの特徴、開発方針、想定ユーザ、他のシステムとの違いを明らかにすると共に、応用処理と形態素解析システムとのベストマッチングを支援することを目的としています。また、ニューラルネットワーク言語処理におけるサブワード化のように、単語分割そのものの必要性が揺らぎつつある現状を踏まえ、再度、単語分割・形態素解析の研究・応用における立ち位置をみなさんと議論したいと思えます。

スケジュール

発表申込受付開始：2018/01/11

発表申込受付締切：2018/2/28

プログラム公開：2018/3/5

ワークショップ：2018/3/16（金）9:30-17:00

招待発表

- 「形態素解析システム JUMAN++」 河原大輔, Arseny Tolmachev (京都大学 大学院情報学研究科)
- 「汎用形態素解析システムMeCab」 工藤拓 (グーグル合同会社)
- 「日本語の単語分割・品詞推定」 森 信介 (京都大学学術情報メディアセンター教授)
- 「製品利用可能な形態素解析器 Sudachi」 高岡一馬 (株式会社ワークスアプリケーションズ)
- 「電子化辞書『UniDic』を中心に見たリレーショナル・データベースによる統合的言語資源管理環境」 岡 照晃・中村 壮範 (国立国語研究所 コーパス開発センター)
- 「単語分かち書き用辞書生成システム NEologd の現状と今後」 佐藤 敏紀 (LINE株式会社)

招待講演

タイトル: 形態素解析辞書で曖昧性解消に挑む

講演者: 坂本美保氏

所属: 株式会社ワークスアプリケーションズ ワークス徳島人工知能NLP研究所

概要:

形態素解析の基本機能は、文を形態素に分割し各々の形態素の品詞を判別することだが、自然言語処理を必要とするアプリケーションに応用するとなると、さらに詳細な語句認識が求められることが多い。例えば、異なる表記でも同じ語（表記揺れ）であると認識したり、あるいは同じ表記でも異なる語（同表記異義語）であると認識できること等が期待される。本発表では、これまで様々なアプリケーションの辞書開発に携わってきた筆者の経験を元に、アプリケーションへの応用を前提とした形態素解析辞書の構築ノウハウについて紹介する。基本的な語彙から応用先の機能に関連性の高い語彙

に至るまで、どのように採否を決定し情報付けを行うか、具体例とともに紹介する。また最後に、当社で開発した最新の形態素解析器Sudachiの辞書データについて、その特色を簡単に紹介する。

講演者略歴:

1990年～2015年 株式会社ジャストシステム 勤務

2017年～ 株式会社ワークスアプリケーションズ ワークス徳島人工知能NLP研究所 勤務

一般発表

一般発表は 15分の通常の発表と5分程度の Lightning talk の2種類があります。

詳細は <https://sites.google.com/view/nlp2018ws> をご覧ください。