

■募集情報

募集番号	2022-3
部署名	ユニバーサルコミュニケーション研究所 先進的音声翻訳研究開発推進センター
インターンシップ課題名	Efficient tree-structured decoding methods for language processing
インターンシップ内容の概要	<p>Natural language processing on low resource languages has been long suffering from insufficient data. Multilingual models have rapidly gained popularity since they can support various languages with a single model.</p> <p>To automatically mine the universal representations across languages, we introduce Huffman code to character representation as a similarity metric. We also introduce Hierarchical Softmax (H-Softmax), which is an approximation inspired by binary trees. By replacing the flat Softmax layer with words as leaves. H-Softmax allows us to decompose calculating a sequence of probabilities at different time-steps into a single time-step and reducing the calculation on the normalization over all words.</p>
課題に関する問い合わせ先	先進的翻訳技術研究室 研究統括 田中英輝 hideki.tanaka@nict.go.jp
応募条件	<ul style="list-style-type: none"> ・ 学年：大学院以上 ・ 専攻学科等：計算機科学 ・ 自然言語処理、機械翻訳に関する知見があること。
実施場所	ユニバーサルコミュニケーション研究所（けいはんな）
実施期間	2022年4月1日～2023年3月31日の間で60日間
受入予定人数	1名
選考課題	なし
備考	