

理科課題研究 研究計画書(詳細版)

高校Ⅱ年	組	番	氏名 Genbu Hase
------	---	---	---------------

※ 提出締切：6月30日(土)
※ 提出先：指導教員に直接提出する

研究テーマ	Shiina ―― 形態素解析を用いた AI の会話の学習		
目的	<p>そのテーマを選んだきっかけ、研究の動機、研究の背景、研究のねらい（何を解き明かすのか）を詳しく記述する。</p> <p>AI がどのようにして言語を認識していくかを研究する。</p>		
先行調査	<p>研究を始めるにあたって、事前に学んだことや、調べてわかった事柄を詳しく記述する。</p> <ul style="list-style-type: none">形態素解析という手法が利用できる日本語の文法構造は複雑化しているAI を作るのは容易ではない		
仮説	<p>実験する前の自分の考えや、実験結果予測を、今までの知識・経験をふまえて記述する。その仮説に至った理由も記しておく、予想と違う結果になったときに考察しやすい。</p> <p>世界には沢山の言語があるが、その文法構造は数種に分類できる。 すなわち、文法構造が解析できれば、容易に言語の組成が可能になるということである。 そこで日本語をベースとした形態素解析によってデータを収集し、日本語の文法構造を記憶させ、AI と会話したとしても、比較的違和感が少ない程度の意思疎通を図る。 また、過去の事例との比較のために、言語合成処理内で最も有効と思われる理論（統計学・確率論）を組み込み、それによる正確さの変化を調査する。</p>		
準備	<p>実験に使用する器具・薬品・設備・生体等を箇条書きする。使用予定分量・回数も詳細に記述する。</p> <ul style="list-style-type: none">開発用 PC（Windows 81 / x64+）<ul style="list-style-type: none">➢ Visual Studio Code（x64+）➢ Mastodon<ul style="list-style-type: none">✧ 板橋井 (itabashi0i0.jp)✧ ニコフレ (friends.nico)やる気（←一番大事）元気（←二番目に大事）	分量・回数	分量・回数
実験計画	<p>研究日程と場所をできるだけ具体的に記述する。とくに自宅外で実施する場合は、相手施設側の都合も考慮して計画を立てる。また、夏休み中に実験が完結できるように、無理なく計画を立てること。</p> <p>言語解析 AI「Shiina」（以下しいちゃん）は以下の動作環境にて製作を進める。 詳しくは https://github.com/Itabashi-don/Shiina を参照のこと。</p> <ul style="list-style-type: none">Node.js（JavaScript）		

方法	<p>実験手順を箇条書きで記述する。文章だけでなく図も示すことが望ましい。できるだけ、客観的な数値を求める方法を考えること。対象実験をする場合は、どの条件を変えて、どの条件をそろえるのかを明確にして表にまとめる。精度を上げるために何回同じ実験をするのかも書く。枠内に収まらない場合は、別紙に記述して添付する。</p>
	<ol style="list-style-type: none"> 1. <u>Mastodon 連携処理の製作</u> 既に製作済み。 2. <u>形態素解析データの収集</u> 認識精度を上げるために、複数のデータを収集し、判断材料とする。 ニコフレの連合タイムラインにてしいちゃんを学習モードで起動し、数時間程度放置する。 3. <u>文章の合成</u> 形態素解析したデータを基に、合成結果をテストする。 テスト基準の詳細情報は Shiina レポジトリを参照のこと。 4. <u>修正・改善</u> テスト結果を基に、改善点を見出し修正・改善を行う。
まとめ方	<p>どのようなデータを測定・記録して、どのようにレポートにまとめるのかを記述する。表やグラフの下書きを示すことが望ましい。</p> <p>文章の合成に違和感を覚えなくなった所でログデータを退避させ、再学習を行う。 解析データ100件ごとに文章を10回生成し、その中で最も違和感がなかったものを抽出し、DB の総数との相関性を測る。</p> <p>流れてくるツイートによって結果が変わる可能性があるが、そこは誤差として考慮する。</p>

※ 指導教員処理欄（生徒は記入しない）

提出日	2018 年 月 日	評価	合格 ・ 再提出
指導教員			
面談記録			