

仕 様 書

1 件名

チャット会話データ(テキスト)に含まれる単語に対する興味の有無のラベル付け作業
業務

2 業務目的

人間同士のチャット会話データ(テキスト)に含まれる単語(名詞)に対し、「興味あり」、
「判断不能・興味なし」を二者択一で付与した電子データを収集する。

3 業務内容

2名の話者によるチャット会話データとラベル付けする単語(名詞)の電子データ(CSV
形式又はtxt形式)を本学から提供する。

会話データ(テキスト)と、その対話中で使用された単語から、それぞれの単語に対す
る2名の話者の「興味あり」、「判断不能 or 興味なし」のいずれかのラベルを付与する。

実施にあたっては、各単語に対し、10名がそれぞれ個別にラベル付けを行い、その
結果をすべて提出するものとする。

作業はEXCELなどの表計算ソフト上で実施することが可能である。また、希望により、
本学教員が作成した作業専用のソフト(Windows、MacOS、Linux対応)を提供することも可
能である。

【付与基準】

当該単語を含む発話をした話者に、「あなたは「〇〇(当該単語)」に興味があります
か？」もしくは「あなたは「〇〇(当該単語)」に関する話題に興味がありますか？」
という質問をした際のその話者の回答を予想し、「興味あり」、「判断不能 or 興味なし」
のいずれかのラベルを付与する。

質問の「〇〇」に単語を入れた際、日本語として意味が通らない場合は「判断不能・
興味なし」のラベルを付与する。

【例】

①対話

話者1：趣味は何ですか？

話者2：趣味はピアノを弾くことですね。

話者1：：どんなのを弾くんですか？

話者2：下手なのですが、J-POPを好んで弾いています。

②名詞とラベル付け

趣味 話者 1[判断不能 or 興味なし] 話者 2[判断不能 or 興味なし]

ピアノ 話者 1[判断不能 or 興味なし] 話者 2[興味あり]

J-POP 話者 1[判断不能 or 興味なし] 話者 2[興味あり]

【参考】

- ・概ね 1 対話につきラベル付け対象の単語は 50～200 個程度（2 名分なので 100～400 単語程度）、300 対話である。作業時間は 1 対話につき 20 分程度であり、20 分×300 対話×10 名 = 60000 分 = 1000 時間 ほどの延べ作業時間が必要と想定される。すなわち、作業者が 10 名同時に作業すれば 100 時間、20 名であれば 50 時間で作業が完了する計算である。

4 成果物

ラベル付与済み電子データ一式

5 情報管理

契約を履行するうえで知り得た個人情報に関しては、別紙「個人情報取扱特記事項」に従い、適正に業務を履行すること。

また、情報の取扱いにあたっては、漏洩の防止等適正な管理のために必要な措置を講じること。

6 履行期限

平成 29 年 3 月 27 日（月）とする。

ただし、作業完了分は逐次納入すること。

7 その他

- (1) 落札者は、契約締結後、本学担当教員に連絡を行い、詳細な調整を行うものとする。
- (2) 成果物に対する著作権、特許権その他一切の権利は本学に帰属するものとする。
- (3) 受託者は、本業務の履行について、特許権その他第三者の権利の対象となっている処理方法等を使用する場合、これらの権利に関する一切の責任を負うものとし、これらに要する費用は受託者の負担とする。
- (4) この仕様書に疑義があるとき、又は定めのない事項について、甲・乙協議して定めるものとする。