

# 自動要約文生成システムの作成

C0119321 山岡 尚広

## 1. はじめに

近年ニュースアプリの出現やSNSなどの流行により非構造データの量が急激に増えている。人間自体の処理性能の上限はあまり変化しないが、データ量自体は増えていっている。自分の欲しい情報を探し、一読するだけでも多大な時間を費やすことも珍しくなくなってきている。その中で自分の欲しい情報をなるべく素早く詳細に集めるために、単語ごとにフォーカスを当てる要約文が役に立つのではないかと思った。

## 2. 関連研究

### 2.1 情報検索のクエリに基づく自動要約

桜井・内海 (2004) は、情報検索のためのクエリに基づく文書自動要約について報告している。彼らはユーザが検索したクエリに基づき、クエリ検索で得た文書の集合から文書 D と文書 D' を用意した。その後文書 D の要約と文書 D' から文書 D に特徴的な内容に基づく要約を作成し、その二つで選択された文を文書 D に出現した順で並べることで通常的重要文抽出方法より高い評価を得ることでできたということ報告している。しかし要約に求める条件が細かいほど評価が落ちるといった問題点がある。

### 2.2 自動要約における文重要度に関する研究

内山・井佐原 (2000) は、重要文に選ばれる特徴的要素が自動要約にどの程度寄与するかの報告をしている。彼らは文の重要度はタイトルとの類似度で決めるという仮定の元、単語間の共起関係を利用した類似度を参考にした方が高精度な要約ができたという結果が出されている。

### 2.3 本研究の立場

本研究では情報検索のための時間浪費についての問題について、文単位的重要度計算と注目単語に対しての位置関係情報や頻出頻度情報を利用した重要度計算を行うことで要約に必要な文や単語を算出していく。

## 3. 一つの単語に注目した自動要約生成システム

本システムでは、まず原文に含まれる単語の品詞を半別するため原文に形態素解析を行う。その後文単位的重要度計算を行い得られた結果の要約と注目単語に対して位置関係情報や各単語の頻出頻度を利用した単語単位的重要度計算を行った結果の要約を比較し、共通する部分が多い箇所を優先し

て、最終的な要約文として生成する。

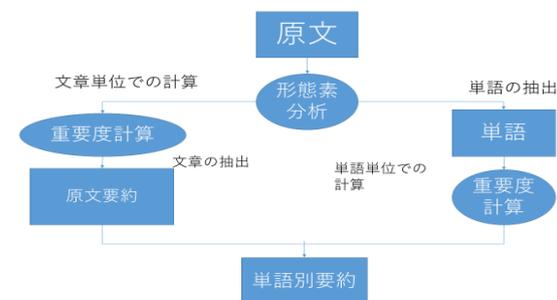


図1. 本研究で構築するシステム処理の流れ

## 4. 研究計画

	月						
内容	Aug-22	2022年9月	2022年10月	2022年11月	2022年12月	2023年1月	2023年2月
類似研究の検索							
要約システムの体験							
重要単語の定義付け							
重要文・語の計算式							
システムの開発							
卒業研究準備・発表							

## 5. 進捗状況

現在インターネットなどで公開されている要約できるソースコードを数個試した。また自分の研究と類似している論文を参考にし、単語や文の重要度を定義付ける計算式など今回のシステムに参考できるものを探している。

## 6. おわりに

テキストデータの急増によるユーザの情報処理負担増加を軽減する解決策として一つの単語に注目し、その単語に注目して要約文を生成するシステムの概要、およびその構築について述べた。

## 参考文献

[1] 桜井俊彦,内海彰 “情報検索のためのクエリに基づく文書自動要約” 言語処理学会(2004)  
 [2] 内山将夫,井佐原均, ”自動要約のための文重要度の比較“ J-STAGE 7 巻 4 号 p-261-270 (2000)