

構造を用いた英語エッセイ検索機能の実現

Implementation of an Example Essay Retrieval Function Using the Structure of an Essay

坂本 洵一朗^{*1}, 萩原 大貴^{*2}, 國近 秀信^{*3}

Junichiro SAKAMOTO^{*1}, Daiki HAGIHARA^{*2}, Hidenobu KUNICHIKA^{*3}

^{*1}九州工業大学大学院情報工学府

^{*1}Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology

^{*2}九州工業大学情報工学部

^{*2}Computer Science and Systems Engineering, Kyushu Institute of Technology

^{*3}九州工業大学大学院情報工学研究院

^{*3}Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology

Email: j.sakamoto@minnie.ai.kyutech.ac.jp

あらまし：英語の初学者がエッセイを書く場合、英語の論理展開法に則った正しいエッセイを参考にすることは有用な方法の一つである。本研究の目的は、英語の論理展開法に則り、かつユーザの目的に合致したエッセイを検索するための英語エッセイ検索機能の実現である。本機能は、エッセイの種類ごとの典型的な構造を表現したエッセイ展開スキーマを基にして、適切な構造のエッセイを検索する。

キーワード：英語学習支援, エッセイライティング, 例文検索

1. はじめに

英語の初学者がエッセイを書く場合、英語の論理展開法に則った正しいエッセイを参考にすることは有用な方法の一つである。書籍に載っているエッセイは有用であるが数が限られているため、WWW上のエッセイを検索する方法が考えられるが、その構造が適切か否かを判断することは難しい。これまでに構造に着目した英文の検索システム⁽¹⁾は実現されているが、文章レベルの検索は一般的ではない。そこで我々は、文章の構造を用いた英語エッセイ検索システムの実現を目指している。本研究では、WWWより自動収集したエッセイの構造が同定された後を想定し、エッセイの構造を用いて検索を行う英語エッセイ検索機能の実現を目的とする。

2. エッセイ検索に必要な機能

本研究では、英語の論理展開法の理解が不十分なユーザを想定している。よって、英語の論理展開法に則った正しいエッセイであり、かつユーザが書きたい内容に近いエッセイを検索できるようにする必要がある。一般にエッセイの構造が英語の論理展開法と完全に一致しているとは限らないため、本システムではエッセイの構造と英語の論理展開法との一致度を採点し、エッセイとともにユーザへ提示する。また、ユーザが書きたい内容に近いエッセイを検索できるようにするため、キーワードを用いた絞り込み機能を提供する。

3. エッセイの構造に関する知識

エッセイの採点を行うためには、英語の論理展開法に関する知識が必要となる。エッセイやエッセイに含まれるパラグラフの論理展開法にはいくつかの種類があり、それらの構造は異なる。我々はこれまで

に、エッセイやパラグラフの種類ごとに典型的な構造を表現した11種類のエッセイ展開スキーマ⁽²⁾、および、11種類のパラグラフ展開スキーマ⁽³⁾を定義した。例として、図1に手順を説明するProcess Essayのエッセイ展開スキーマの構造情報を示す。

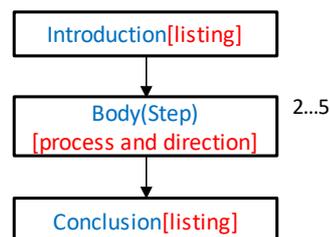


図1 Process Essayのエッセイ展開スキーマ

図中の四角は一つのパラグラフに対応し、その内部ではパラグラフの役割を示している。また、[]内は対応するパラグラフの種類を表す。四角の右上の数字はそのパラグラフの個数の範囲を表す。

4. 英語エッセイ検索システム

図2に本システムの概要を示す。エッセイ収集部では、WWW上からエッセイを収集し、それらの構造分析を行う。構造分析を行ったエッセイはエッセイデータベースに保存する。エッセイ検索部では、まず、ユーザが目的のエッセイの種類等の条件を指定する。その後エッセイデータベース内に存在するエッセイをエッセイ検索機能の下位機能である採点機能で採点し、得点順にソートして出力する。

5. 英語エッセイ検索機能の実現

エッセイ展開スキーマの構造は典型的な構造であるため、エッセイの構造と完全に一致するとは限ら

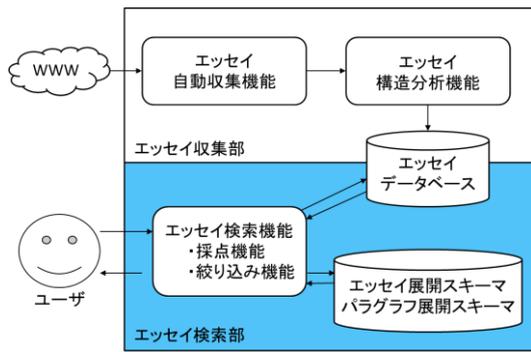


図2 システムの概要

ない。つまり、構成要素の種類や数、並びが異なる場合がある。よって本研究では、両者の構造の類似度を数値で表す採点機能を実現した。本研究で扱う構造はエッセイの構成要素とその並びから成ることから、構成要素の種類を用いた採点および構成要素の並びを用いた採点を行い、それらの平均点を最終的な得点とする。また内容や構造を基にした検索結果の絞り込み機能を実現した。以下、各機能について述べる。

5.1 構成要素の種類を用いた採点機能

エッセイ展開スキーマおよびエッセイデータベースでは XML 形式で構造が表現されており、それらの比較により採点を行う。具体的には、まず採点を行うエッセイの構造を分析し、構成要素のうち数に制約のある要素の個数を同定した後に、エッセイ展開スキーマを変形して理想的な構造を生成する。理想的な構造における個数は、採点を行うエッセイで、ある構成要素の個数が不足している場合はその要素の個数の範囲の最小値とし、個数が超過している場合は最大値とする。その後、採点を行うエッセイと生成した理想的な構造とを比較し、要素の過不足から100点満点に換算しエッセイレベルの得点とする。続いて、採点を行うエッセイに含まれる各パラグラフを同様に採点し、その平均点をパラグラフレベルの得点とする。最後に、エッセイレベルとパラグラフレベルの得点の平均を求め最終的な得点とする。

5.2 構成要素の並びを用いた採点機能

本機能では、構成要素の前後関係に着目して採点を行う。具体的には、まず採点を行うエッセイの連続する構成要素の組を取り出し、その合計数に応じて各組に按分した値を各組の得点として与える。次に要素の組がエッセイ展開スキーマに存在するか否かを調べる。もし存在する場合は持ち点をそのまま与え、2つの要素の間に異なる要素が存在する場合や組が存在しない場合は減点する。これらの得点の合計をエッセイレベルの得点とする。続いて、採点を行うエッセイに含まれる各パラグラフを同様に採点し、その平均点をパラグラフレベルの得点とする。最後に、エッセイレベルとパラグラフレベルの得点の平均を求め最終的な得点とする。

5.3 絞り込み機能

多くの検索結果から必要なエッセイを探すための機能である。ユーザが指定できる条件には、エッセイ中の文数の指定、特定の構成要素を参照する際に用いる構成要素の指定、書きたい内容に近いエッセイに絞り込むための単語の指定がある。

6. 評価

本稿で述べた英語エッセイ検索機能の有用性の確認のため、指定した種類のエッセイが検索可能か否かを調査した。準備として、書籍で種類が明示されているエッセイの構造を手で分析して構成要素を表すタグを付与し、エッセイの種類ごとに4個のエッセイをエッセイデータベースに保存した。その後各エッセイを11種類すべてのエッセイ展開スキーマを用いて採点を行い、採点法の妥当性を評価した。具体的には、エッセイと同じ種類のエッセイ展開スキーマを用いて採点したときのみ得点が高くなるか否かをDunnettの多重比較を用いて有意水準5%にて調査した。

調査の結果、採点を行うエッセイとエッセイ展開スキーマが対応しているときの得点の平均が、対応していないときの得点の平均と比較して有意差があることがすべてのエッセイの種類で確認できた。よって、英語エッセイ検索機能が英語の論理展開法に則ったエッセイの検索に対して有用であると考えられる。しかし、今回評価に使用したエッセイは1種類につき4個と限られていたため、今後エッセイを増やして更なる調査を行う必要がある。

7. おわりに

本論文では、自動収集したエッセイの構造が同定された後を想定し、エッセイの構造を用いて必要なエッセイの検索を行う英語エッセイ検索機能の実現と評価について述べた。

今後は、エッセイを増やした上で採点法の妥当性の再調査を行うとともに、検索時間を短縮するよう改良することが必要である。また、エッセイ自動収集機能およびエッセイ構造分析機能の実現を予定している。

参考文献

- (1) 三好康夫, 越智洋司, 金西計英, 岡本竜, 矢野米雄: “英作文支援における句構造情報を利用した用例検索ツール”, 日本教育工学会論文誌, Vol.23, No.3, pp.283-294 (2003)
- (2) Kunichika, H. Ikeda, Y. and Takeuchi, A.: “Essay Development Schemata to Support English Composition”, Proc. of ICCE 2014, pp.815-817 (2014)
- (3) Kunichika, H. and Takeuchi, A.: “Paragraph Development Schemata to Support English Composition”, Proc. of AIED 2007, pp.602-604 (2007)