

英字ニュース記事を用いた英文多読支援 Android アプリの開発

中野 明^{1, a)}

概要: 国際化に対応できる技術者を育成するために、英語の指導は重要といえる。しかしながら、理科系の学生の英語への興味・関心は必ずしも高いとはいえない。本研究は Android タブレットの持つ携帯性と、ネット上に豊富にあるある英語のニュース記事を利用した英字ニュース記事の多読支援 Android アプリの開発を目指す。英字ニュースを利用することで、リアルタイムな内容や、興味のある内容の英文で学習することができる。また、ネット上からニュース記事を取得する際に、その記事の難易度を計算することで、アプリは、使用者の英語レベルに応じた英文記事の提示ができる。

Development of An Android Application for Extensive Reading with English News Articles

NAKANO AKIRA^{1, a)}

Abstract: It is important to study English, in order to become engineers working abroad. However, in most cases, students in science course are not interested in English. In this research, we develop an android application for extensive reading with English news articles. We think that it is easy for students to read real time news and some news category which they are interested in. In preprocessing, we evaluate of difficulty of articles. Thereby, this application can propose articles which are appropriate for each user.

1. はじめに

今日、国際化に対応できる技術者を育成するために、英語の指導は重要といえる。しかしながら、理科系学生の英語への興味・関心は必ずしも高いとはいえない。また、英語の授業以外において英語に触れる機会もほとんどない。本研究では Web 上に豊富にある英字ニュースサイトと、辞書サイトを資源として活用することで、学習者の英語の能力に応じた記事を提示できるアプリケーションの開発を目指す。対象とするニュース記事は、日本の新聞社が海外へ発信している日本のニュースの英訳版とする。

多読は英語運用能力の向上が指摘される学習方法であり、精読学習とは異なり辞書を引かずに読み進める学習方法である。そのため、本アプリケーションでは、収集した英字ニュース記事の評価を行い、学習者が多読できる範囲の記事を評価する機能を実装し、データベースを構築する。これまでに windows OS 上で動作するプロトタイプシステムの開発を行っている[1]。本稿では、携帯性の高いデバイスである Android 端末への移植実装の報告を行う。

2. システムの概要

本システムの構成図を図 1 に示す。本システムは主に二つの機能を持つ。一方はニュース記事を Web 上から取得し、その際に単語の難易度情報を抽出し計算する英字ニュース記事収集機能。もう一方は、学習者の能力に応じたニュー

ス記事の一覧を表示し、本文を表示する英文多読支援機能である。

英字ニュース記事収集機能は JDK により開発しており、パソコン上にて稼動させ、英文多読支援機能は Android SDK により開発し、Android タブレットやスマートフォンなどで利用する。また、使用するデータベースシステムは Android に標準搭載されている SQLite を用いている。SQLite で作成されるデータは、ひとつの独立したファイルにまとめられる特徴がある。そのため、PC でデータを作成し、そのファイルを用いて Android デバイスでのデータベース検索を実現している。

3. 英字ニュース記事収集機能

本機能は、多読の対象である英文を指定した URL から収集し、難しさに関する評価分析を行う機能である。現在、この機能は、自動で定期巡回が行っている。本機能では、記事見出し解析、単語レベル解析、記事本文解析の 3 つの処理がある。

本機能には予め各新聞会社の「英字ニュースサイトにおける見出し一覧ページの URL」が指定されている。まず始めに、この URL から各記事の URL を抽出し、さらにその記事の URL を辿って得られる記事のソースコード (HTML ファイル) の解析を行い、記事の見出し情報の解析を行う。見出し情報には、記事見出し、発信日、ジャンル、URL などを調べる。ジャンルもソースコードの解析から取得する。そのためニュースサイトによって分類の細かさが異なる。

次に、記事本文から使用している単語を抜き出して、そ

1 久留米工業高等専門学校
Kurume National College of Technology
a) nakano@kurume-nct.ac.jp

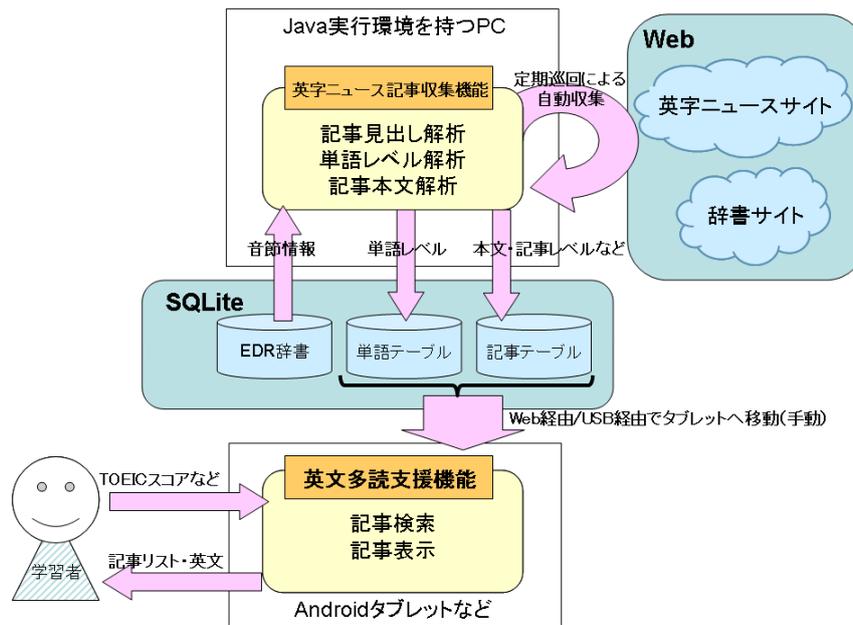


図 1 システム構成図
 Figure 1 a framework of the system

の単語の難易度を調べる（以下、この「単語の難易度」のことを単語レベルと呼ぶ）。単語レベルを調べる際には、英語辞書サイトを用いる。

最後に、記事本文の解析を行う。この解析では、記事中に含まれる単語の延べ数、単語の実数、各単語レベルの頻度、記事の難易度などを調べる（以下、この「記事の難易度」のことを記事レベルと呼ぶ）。

単語・記事本文の解析で得られた情報はそれぞれデータベースに追記し（図1の単語テーブル、記事テーブル）、英文多読支援機能に使用する。

3.1 単語レベル解析

本機能では、単語レベルを調べるために、英和辞典サイトである Weblio を用いている。Weblio では、単語を表1に示すように評価している。ただし、単語レベルには13以上も存在しており、現在確認されている最も高いレベルは30である。レベル13以上の単語に関しては、使用頻度の低いもの、専門用語などが中心となっている。

この解析は、処理に時間のかかるため、この解析で一度調べた単語レベルは、単語テーブルに追記し、辞書サイト検索の前に単語テーブルを検索することで処理の軽減を行っている。

3.2 記事本文解析

記事本文解析では、本文中に使用している単語の延べ数や単語レベルの頻度といった値の他に、記事レベルの計算を行う。記事レベルは、現在2種類を計算している。一方は、単語レベルの頻度の情報から「使用されている単語の

表 1 Weblio における単語の評価表

Table 1 An evaluation table of words in Weblio.

レベル	英検	TOEIC	学校
1	3級	220	中学校
2	準2級	350	高校1年
3	準2級	350	高校2年
4	2級	470	高校3年
5	2級	470	大学
6	準1級	600	最難関大学
7	準1級	730	最難関大学
8	準1級	730	最難関大学
9	1級	860	大学院
10	1級	950	大学院
11	1級		
12	1級		

平均単語レベル」を計算し、その値を記事レベルとする。他方は、記事本文における1文あたりの平均単語数と1単語あたりの平均音節数から「英文の読みやすさ」を計算し、その値を記事レベルとする。算出の式は、Flesch-Kincaid Grade Level (FKG) を用いた[2]。各単語の音節数は、EDR電子化辞書を用いて行っている[3]。

4. 英文多読支援機能

本機能では、学習者に応じたニュース記事群を提供する記事検索と、記事表示の2機能を持つ。記事検索では、学習者に TOEIC スコア、ジャンル、記事のワード数の3項目から絞り込み検索ができるインターフェースを提供する（図2）。現在の Android アプリでは、「使用されている単語の

平均レベル」を記事レベルとし、入力された TOEIC スコアを、表 1 に基づいてレベルに変換し、その値の付近の記事レベルの記事を検索する。

また、記事表示では、学習者に英文を提示する際に、入力 TOEIC スコア以上の単語にリンクを貼った HTML を表示する(図 3)。このリンクは、辞書サイトに連結されており、リンクをタップすることで意味の検索ができる。また、入力 TOEIC スコア以上の単語については、本文の下にリストとして表示する。



図 2 記事検索画面

Figure 2 an interface for searching articles



図 3 記事表示画面 (画像の色は反転)

Figure 3 an interface for showing an article

5. 予備調査

多読に適した記事を収集できるかどうかを調べるため予備調査を実施した。調査は、ニュースサイトとジャンルの 2 軸から、記事の難易度の状況を調べた。調査の対象とな

る記事は、本システムが収集した 2012 年 11 月 1 日から 30 日までの 2060 件とする。結果を表 2 に示す。記事の難易度評価には、レベル 2 ~ 1 2 までの単語を対象とし、それ以外のレベル 1 は一般語、レベル 1 3 以上は特殊語として除外している。ニュースサイトごとの各ジャンルの記事レベルの集計結果を表 3 に示す。

表 2 より、記事収集機能の記事見出し解析が機能していること、ジャンルの分類にはニュースサイトごとに差があることが確認できた。

表 3 の平均値 (AVE) と標準偏差 (STDEV) より、収集した記事の多くは、高校 2 年から大学までの難易度であることが分かった。この結果より、英字ニュース記事は、英語を苦手とする TOEIC スコア 300 点台の英文多読の教材として利用できることがわかった。

表 2 ニュースサイトとジャンル別の記事件数

Table 2 Number of articles in each news site and genre.

News site	national	world	sports
NHK	211	271	0
JapanTime	493	0	182
Asahi	154	0	21
Yomiuri	657	11	60

表 3 ニュースサイトとジャンル別の記事レベル

Table 3 difficulty of articles in each news site and genre.

News site	national	world	sports	AVE	STDEV
NHK	3.92	4.11		4.02	0.50
JapanTime	4.11		4.33	4.17	0.48
Asahi	4.42		3.97	4.37	1.03
Yomiuri	3.98	3.96	3.96	3.98	0.40
AVE	4.06	4.10	4.22		
STDEV	0.53	0.53	0.58		

6. おわりに

今後の課題として、単語レベルと記事レベルの検討が必要と考えている。単語レベルに関しては、各単語に 1 つ設定された値であるため、品詞、意味、使い方による差が表現されていない。記事レベルに関しては、意味・文脈・文法・利用者の既習知識などが表現されていない。現状の単純な評価において、どの程度有効であるか検証をした後に、各レベル解析の改良を行いたいと考える。

参考文献

- 1) 中野明: 英字ニュース記事を用いた英文多読支援システムの開発, 日本産業技術教育学会第 28 回情報分科会, pp.41-42(2013)
- 2) 染谷泰正: Word Level Checker(Online Ver. 2.0), <http://someya-net.com/wlc/readability.html>
- 3) EDR 電子化辞書 http://www2.nict.go.jp/out-promotion/techtransfer/EDR/J_index.html