システムディスクリプション

|  |
| --- |
| 記入日：2012年8月11日  対象トラック：　動詞トラック  システム名：　ハンドメイドルールシステム  チーム名：　ototake  性能：　ドライラン　R=0.675　 P=0.406　　F=0.507  　　　　フォーマルラン　R=　P=　F= |

1. 概要

本システムは，シンプルな人手による規則を用いた主語と動詞の一致に関する誤りを検出するシステムである．

1. 実行方法

実行環境・使用言語：Windows .NET Framework 4.0 または Linux Mono 2.10，C#

実行方法：（詳しくはREADME.txtを参照）

1. OpenNLPモデルファイル，WordNetデータベースファイルを適当なディレクトリにに置く（以下，~/resとする）
2. KJ Corpusデータを適当なディレクトリに置く（以下，~/dataとする）
3. 誤り検出  
   Console.exe d --target v\_agr --kjdir ~/data/corpus\_data --formal --res ~/res --out output\_dir
4. 使用データ，ツール，辞書

ツール名：IKVM.NET

　　URL：http://www.ikvm.net/

種別：仮想マシン（フリーソフトウェア）

用途：JavaプログラムであるApache OpenNLPの.NETアセンブリへの変換

ツール名：Apache OpenNLP

　　URL：http://opennlp.apache.org/

種別：解析器（フリーソフトウェア）

用途：トークン分割，POSタグ付け，Chunking

ツール名：SGMLReader

URL：https://github.com/MindTouch/SGMLReader

種別：SGML解析ライブラリ（フリーソフトウェア）

用途：KJ CorpusのXML解析

ツール名：DotNetZip

URL：http://dotnetzip.codeplex.com/

種別：ZIPライブラリ（フリーソフトウェア）

用途：結果出力用のZIP圧縮

ツール名：ConsoleLib

URL：https://bitbucket.org/shokai/consolelibnet

種別：コマンドパラメータ解析ライブラリ（フリーソフトウェア）

用途：コマンドラインオプション解析

辞書名：WordNet

　　URL：http://wordnet.princeton.edu/wordnet/download/current-version/

　　用途：Lemmatize

1. システムの詳細（システムの詳細を記入ください）

処理の流れは以下の通り．

1. 動詞句を抽出し，各動詞句の主語と人称変化動詞を推定する
2. 主語の人称と動詞の対応をチェック

主語の推定方法は以下の通り．

1. 動詞句の前の句が名詞句であれば，その名詞句の主名詞を主語とする
2. 1.の推定主語が関係代名詞の場合は，推定主語のひとつ前の名詞を主語とする
3. 1.の推定主語がthereの場合は，動詞句の後ろの名詞句の主名詞を主語とする

人称変化動詞の推定方法は以下の通り

1. 動詞句の先頭からPOSがVB\*となっている単語を探していく
2. 1.で最初に見つかった単語を人称変化動詞とする
3. セールスポイント

* トレーニングデータを必要としない