

目次

刊行のことば …2
第6巻について …3

第1章 ウェブ検索概論 小野正弘 7

1. はじめに…………… 7
2. ウェブにおける日本語…………… 8
3. 検索エンジン…………… 10
4. Google と Yahoo!を中心とした検索エンジンの動作…………… 13
検索結果の大前提…13/「一個車両」の検索…14/オノマトペ
の検索…22/検索エンジンの特徴…31
5. Google による類義語の研究…………… 35
「走る」と「駆ける」…35/内省結果の確認①…38/内省結果の
確認②…43/ワイルドカード検索…47
6. その他の検索エンジン…………… 50
Baidu…50/Mooter…51
7. おわりに…………… 53

第2章 ウェブ検索の応用 岡島昭浩 59

1. はじめに…………… 59
2. フレーズ検索など…………… 61
プラス演算子 (+) …61/フレーズ検索 (" ") …62/自然文
検索…64
3. AND検索, OR検索, NOT検索…………… 64
AND検索…65/OR検索…65/NOT検索…66/OR検索の
応用…67
4. ワイルドカード検索…………… 68
grep 型検索との違い…68/フレーズ検索との組み合わせ…69
5. 時間的変化の検出…………… 69
日付指定検索…69/Google デスクトップサーチ…71

6. 大量な用例採集のために…………… 72
検索結果を AND 検索と NOT 検索で分ける…72/検索結果を時
系列によって分ける…73
7. サイト指定検索…………… 73
特定のサイト内で検索する…73/サイト指定によるさまざまな
用例の検索…74/データのダウンロード…76
8. ブログ・Twitter 検索…………… 77
goo ブログ検索ほか…77/ Twitter の検索…78/リアルタイム
検索…80
9. 用例の探せるウェブ上のデータベース…………… 80
10. 検索の容易化のために…………… 81
アドオン機能…81/検索プラグイン…83/さまざまな検索プラ
グイン…85
11. おわりに…………… 85

第3章 ウェブを利用した研究例 服部匡 89

1. これまでの状況…………… 89
2. ウェブを利用する利点と問題点…………… 90
3. ウェブからのデータ利用の方法…………… 92
検索エンジンを利用してデータを取得する方法…93/分析・加
工された形で提供されるデータを利用する方法…95/ウェブ
コーパスを作成し利用する方法…101
4. ウェブを利用した研究の例…………… 102
共起関係(コロケーション)の抽出・分析…103/周辺的な用例
の発見・用例の広がり探索…112/辞書にない語形の発見
…116/語形や表記の変異の分析…118/文法的変異の分析…120
/話者の属性による言語的特徴の相違の分析…123/ネット社会
に特有の語や表記…126/文体の分析…127/使用機器による言
語特徴の相違の分析…129
5. おわりに…………… 129

ウェブ検索概論

1. はじめに

本章では、検索エンジンを利用して行う、ウェブにおける日本語に関する研究について概説する。

ウェブで用いられている日本語は、日本語研究のためのコーパスとして十分に魅力的なものであるということは、荻野(2008)、黒田他(2008)等に述べられているとおりであるが、その最大の魅力は、なんといっても、データの規模が極めて大きく(量)、しかも多様であること(質)であろう。

一般に、扱ったデータの量が大きければ大きいほど、そこから得られる情報に基づいた結論の信頼性は高くなる。例えば、ある日の新聞の1ページだけを見て出した結論と、すべてのページを見て出した結論のどちらが信頼できるかを考えてみれば、答えは明白であろう。さらに、それが、1カ月分、1年分、10年分…と増えていったら、信頼性はますます高くなっていくに違いない。とはいえ、10年分の新聞を独力で読み通すには、個人の力では絶望的な労力と時間を要する。それを一気に調査し終えるような手段があったら、と思う。検索エンジンとは、まさに、その夢をかなえるものなのである。

一方、データの量が仮に十分なものであっても、それがまったく均一的なものであったら、データとしての価値は低下してしまう。

まったく同じ例が10回あるよりも、10種類の例があつて10回になっているほうが、いや、それよりも、a-3回・b-2回・c-2回・d-1回・e-1回のように、aからeの5種類あつて全部で10回になっているほうがあ

第4章 自分のプログラムでウェブにアクセスできると 小木曾智信 137

1. はじめに…………… 137
2. ウェブ上のデータをダウンロードしてミニコーパスを作る…………… 138
ウェブ上のデータを利用するには…138/データのダウンロード…140/ダウンロードに際しての注意点…148/ダウンロードしたファイルを加工して利用する…151/ミニコーパスを「ひまわり」で利用する…157/形態素解析の必要性…168
3. ウェブのデータと形態素解析・データベースを利用した研究例…………… 169
調査の対象…169/ UniDic を使った形態素解析…172/ データベースを使った集計…173/ 分析の例…173
4. ウェブ API ……………… 177
ウェブ API とは…177/ ウェブ API を使ったプログラミング…180
5. まとめ…………… 182

第5章 ウェブと他のコーパスとの比較 前田広幸 185

1. ウェブと他のコーパスとの比較にあたって…………… 185
2. コーパスのタイプ分類の諸観点…………… 187
3. ウェブコーパスが一般的傾向として有する特性…………… 189
4. ウェブコーパスと他のコーパスの利用について…………… 190
大規模コーパス利用で収集されるようになった用例について…191/使用度数の時系列推移・書き手の属性とのかかわりの分析…203/複数のコーパスからえられる検索結果をもとにした比較…208
5. ウェブと他のコーパスとの比較のまとめ…………… 210