

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 3 区分
 【発行日】平成 19 年 11 月 1 日 (2007.11.1)

【公開番号】特開 2005-92883 (P2005-92883A)
 【公開日】平成 17 年 4 月 7 日 (2005.4.7)
 【年通号数】公開・登録公報 2005-014
 【出願番号】特願 2004-269036 (P2004-269036)
 【国際特許分類】

G 0 6 F 17/28 (2006.01)

G 0 6 F 17/27 (2006.01)

G 1 0 L 13/06 (2006.01)

【F I】

G 0 6 F 17/28 U

G 0 6 F 17/27 E

G 1 0 L 5/04 E

【手続補正書】
 【提出日】平成 19 年 9 月 14 日 (2007.9.14)
 【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

言語モデルの学習を行うためのコーパスを作成する方法であって、
 定義済み単語および規則と一致するテキストの第 1 のコーパスから単語候補のリストを
 抽出すること、

前記第 1 のコーパスのテキストについての単語主型の指示を供給する前記第 1 のコーパスの注釈にアクセスすること、

関連付けられている文字列の形態主型および形態副型を形成する部分の組合せを、前記注釈の関数として示す、前記第 1 のコーパス内の複数の形態タグを供給すること、および

前記第 1 のコーパスより大きいテキストの第 2 のコーパスに、前記第 1 のコーパスに基づいて前記第 2 のコーパス内のテキストについての単語主型の指示の注釈を付けること
 を備えたことを特徴とする方法。

【請求項 2】

前記注釈は、前記単語が語彙であるか、形態的派生単語であるか、ファクトイドであるか、名前付きエンティティであるかを示す指示を備えたことを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記形態主型は、接辞添加、反復、分割、マージ、および見出し不変変化詞のうちの 1 つであることを特徴とする請求項 1 または 2 に記載の方法。

【請求項 4】

前記形態タグを供給することは、前記部分の組合せに対する発語の一部を示すことをさらに含むことを特徴とする請求項 1、2 または 3 に記載の方法。

【請求項 5】

前記形態タグを供給することは、前記部分の組合せに対する文字パターンを示すことをさらに含むことを特徴とする請求項 1 ないし 4 のいずれかに記載の方法。

【請求項 6】

前記第 2 のコーパスを使用して、言語モデルを学習すること、および、

前記言語モデルを使用してテキストのテストコーパスを処理して、前記テストコーパスのテキスト内の単語を示す出力と前記テストコーパスのテキストの分かち書きの指示とを供給すること

をさらに備えたことを特徴とする請求項 1 に記載の方法。

【請求項 7】

前記言語モデルの出力を、前記テストコーパスの分かち書きを定義する所与の注釈と比較すること、および、

前記比較に基づいて前記言語モデルの出力を評価すること

をさらに備えたことを特徴とする請求項 6 に記載の方法。

【請求項 8】

前記比較することは、前記出力と前記所与の注釈の双方に現れる一致した単語を特定することを含むことを特徴とする請求項 7 に記載の方法。

【請求項 9】

前記評価は、一致する単語に基づいた有効性の値を供給することを特徴とする請求項 8 に記載の方法。

【請求項 10】

分かち書き言語モデルを評価する方法であって、

注釈付きのコーパスに基づいて分かち書き言語モデルを構成すること、

前記分かち書き言語モデルを、前記注釈付きのコーパスとは異なる、分かち書きされていないテキストのテストコーパスに適用すること、

前記分かち書き言語モデルの出力を、前記テストコーパスの単語の所与の分かち書きと比較すること、および、

前記出力と前記所与の分かち書きとの比較に基づいて、前記分かち書き言語モデルを評価し、前記分かち書き言語モデルの有効性の指示を、個々の単語の型の関数として供給すること

を備えたことを特徴とする方法。

【請求項 11】

前記評価は、前記所与の分かち書き内の単語と一致する、前記出力内の単語を特定することをさらに含むことを特徴とする請求項 10 に記載の方法。

【請求項 12】

前記比較は、前記出力および前記所与の分かち書き内の人名、地名、組織名、オーバーラップするあいまいな文字列、およびカバーするあいまいな文字列を比較することを含むことを特徴とする請求項 10 に記載の方法。

【請求項 13】

前記有効性の指示は、人名、地名、組織名、オーバーラップするあいまいな文字列、およびカバーするあいまいな文字列の前記比較に基づいて供給することを特徴とする請求項 12 に記載の方法。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】発明の名称

【補正方法】変更

【補正の内容】

【発明の名称】言語モデルの学習を行うためのコーパスを作成する方法