

004.421

• •

« »

-
-
-

-
-

.

-

[4,5].

-

[1].

-

-

.

-

:

-

[1-3].

-

-

-

.

-

[6-8]

-

-

-

,

-

-

-

,

-

-

(1) $WX \Leftarrow Wdd -$

(2), FA -

$Wx \quad Wdd$

$dist \leq \max dist$
 $|context| \ll |WX|,$

; $< 5 l,$

G ,

G() .

I

(context ,

Wx).

1 2,

$dist = 1$

(I context *)

$dist = 2 - (|context| * 2)$;

$dist = \max dist$

(I context I* y nTMdst)

[11].

$|context| * \wedge^{maxdis/}$
 $=$

context Wx

$\{context\} -$
 $\{context\},$

Wx .

$Icontext * |WX|^{maxdisf}$
 $=$

w Wdd

Icontext\.

$|WX| * \sum_{l=1}^{[maxdistl]} I + Icontext * \sum_{l=1}^{[max / 2]} X$

Порівняємо спочатку кількості слів, які вибираються із словника для аналізу.

$$|context| * \sum_{i=1}^{\max dist} y^i \leq |W_x| * \sum_{i=1}^{\lceil \max dist / 2 \rceil} y^i + |context| * \sum_{i=1}^{\lceil \max dist + 1 / 2 \rceil} y^i \quad (3)$$

Для випадку $\max dist = 1$ $|context| * y \leq |context| * y$ - це коректний запис. Для випадку $\max dist = 2$ маємо: $|context| * (y + y^2) \leq \leq |W_x| * y + |context| * y \Rightarrow |context| * y^2 \leq \leq |W_x| * y \Rightarrow |context| * y \leq |W_x|$. Можна стверджувати, що остання нерівність є справедливою за умови, що потужності відповідних множин $|context|$ та $|W_x|$ співвідносяться як $|context| \ll |W_x|$.

Розглянемо тепер кількість операцій попарної перевірки слів для алгоритму, що передбачає аналіз вершин графу в обох напрямках. Якщо $\max dist = 1$, повинно відбутися $|context| * y * |W_x|$ перевірок; у випадку $\max dist = 2$ - відповідно $(|context| * y * |W_x| + |context| * y * |W_x| * y)$ перевірок. У загальному вигляді кількість попарних перевірок можна визначити як $|context| * |W_x| * \sum_{i=1}^{\max dist} y^i$, що дорівнює кількості перевірок, передбаченій алгоритмом пошуку слів в одному напрямку.

Вершини, розташовані одна від одної на відстані, більшій за $\max dist = 2$, розглядати недоцільно, адже семантичний зв'язок між ними майже не простежується [4, 5, 10]. Якщо ж існує необхідність у перевірці таких вершин, необхідно для конкретного випадку провести аналіз потужностей множин слів $|context|$ та $|W_x|$ та, відповідно до (3), обрати найефективніший принцип побудови шляхів між їх елементами.

Таким чином, у процесі орфокоорекції (за виконання умов $\max dist \leq 2$ та $|context| \ll |W_x|$) при формуванні набору вершин із заданою мірою семантичної близькості до певного

Припустимо, що виконується така нерівність:

контекстного оточення доцільно рухатись графом від вершин, які відповідають словам контексту.

Висновки

Для забезпечення алгоритмів контекстноорієнтованого виправлення помилок необхідними даними про семантичні зв'язки між словами природної мови запропоновано використовувати лексико-семантичний словниковий ресурс. Сформульований новий підхід до визначення міри семантичної близькості заданого варіанта виправлення до контекстного оточення спотвореного слова та обґрунтовано ефективність його реалізації. Показано, що у межах задачі пошуку слів, семантично пов'язаних із заданим контекстом, рух лексико-семантичним словником доцільно здійснювати у напрямку від вершин, які відповідають словам контексту.

Перспективним напрямком подальшого вивчення питання побудови орфокооректорів на основі контекстноорієнтованих методів визначення варіантів виправлення є введення ранжування семантичних відношень лексико-семантичного словникового ресурсу за певними критеріями та врахування ваги відповідних дуг графу G під час вибору його наступних вершин для аналізу.

Список літератури

1. Kukich K. Techniques for Automatically Correcting Words in Text // ACM Computing Surveys. - 1992. - Vol. 24, № 4. - P. 377-439.
2. Johannes Schaback, Fang Li Multi-Level Feature Extraction for Spelling Correction In Proceedings of the IJCAI-2007 // Workshop on Analytics for Noisy Unstructured Text Data. - Hyderabad, India, 8 January 2007. - P. 79-86.
3. Пецак М. М. Нариси з комп'ютерної лінгвістики. - Ужгород: Закарпаття, 1999. - 199 с.

4. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2000. – 384 с.

5. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. – М.: Издательский центр "Академия", 2006. – 304 стр.

6. Белецкий Е.В. Методы, модели и алгоритмы контекстного обмена данными и их использование при проектировании информационных систем: Дис. канд. техн. наук: 05.13.06/ХНУРЕ. – Х., 2000. – 143с.

7. Минский М. Фреймы для представления знаний: Пер. с англ. – М.: Энергия, 1979. – 151с.

8. Моделирование языковой деятельности в интеллектуальных системах. – М.: Наука, 1987. – 279с.

9. Плескач В.Л., Рогошина Ю.В. Агентні технології: Монографія. – К.:Київ.нац.торг.-екон.ун-т, 2005. – 338с.

10. Марченко О.О. Алгоритми семантичного аналізу природномовних текстів: Дис. канд. фіз.-мат. наук:01.05.01/КНУ ім. Тараса Шевченка. – К., 2005. – 150с.

11. Ландэ Д.В. Поиск знаний в Internet.– М.: Диалектика, 2005. – 271с.