

1. Тема НДР: Розробка алгоритмічного забезпечення інтелектуальних систем обробки текстової інформації на природній мові.

2. Номер державної реєстрації теми: 0109U005736

3. Назва бюджетної програми: "Прикладні дослідження і розробки за напрямками науково-технічної діяльності вищих навчальних закладів та наукових установ" (КПКВ 2201040).

4. Терміни виконання: початок – 01.2009, закінчення – 12.2010.

5. Керівник НДР: к.т.н., професор Вороной Сергій Михайлович

6. Короткий зміст викладу запиту: *Метою наукового дослідження є розробка теоретичних засад та алгоритмічного забезпечення інтелектуальних систем обробки текстової інформації на природній мові.*

Предметом дослідження виступають семантичні закономірності в різнотипних процесах утворення слів російської мови.

Об'єкт дослідження – способи, засоби і моделі словотворення російської мови.

Основне завдання роботи – пошук шляхів підвищення ефективності експертної системи навчання словотворенню шляхом скорочення кількості похідних слів у результаті словотворчого синтезу за рахунок формалізації особливостей словотворчої семантики.

Наукові задачі, що вирішуються в роботі задля досягнення основної мети:

1. Формалізація семантики словотворчих формантів та розробка моделей семантичних одиниць словотворення.

2. Розробка словотворчої бази знань та алгоритмів її обробки.

3. Розробка алгоритму морфологічного аналізу, що використовує для отримання граматичних характеристик словотворчу базу знань.

4. Розробка методу словотворчого синтезу, що враховує особливості семантики словотворчих формантів.

5. Практична реалізація запропонованих моделей і методів у вигляді експертної системи навчання словотворенню російської мови.

6. Дослідження розроблених алгоритмів з метою визначення ефективності їх використання.

Заплановані очікувані наукові результати: алгоритмічне забезпечення функціонування інтелектуальних навчальних, пошукових та аналітичних систем при обробці текстової інформації на природній мові.

7. Отримані наукові і (або) науково-технічні результати:

1. Винайдено нові базисні лексичні функції, які дозволили розробити формальну модель семантики похідного слова у вигляді суперпозиції формальних представлень семантики формантів слова, що скорочує обсяг даних, які зберігаються.

2. Розроблено математичну модель вузла дерева, що містить словотворчі форманти, правила словотворення і чергування, яка дозволяє скоротити обсяг похідних слів при словотворчому синтезі.

3. Вдосконалений метод морфологічного аналізу Волкової І.А., завдяки використанню запропонованих дерев морфем, які складають словоформу, що дозволяє використовувати для отримання граматичних характеристик словоформи словотворчу базу знань.

4. Запропоновано поняття міри семантичної близькості слів, використання якої в якості оціночної функції в розробленому евристичному алгоритмі словотворення дозволяє знизити часові витрати на побудову слова по заданій семантиці.

8. Отримана науково-методична і (або) науково-технічна продукція

В ході досліджень виконано розробку бази даних афіксів і виконано програмну реалізацію компонента морфологічного аналізу, проведено адаптацію структури мережевої словотворчої

бази знань і алгоритму побудови слова із заданою семантикою під конкретну СКБД та інші компоненти системи.

Експертна навчальна система взаємодіє з учнем у двох основних режимах: в першому режимі підсистема демонструє і пояснює тому, хто навчається, хід рішення задачі побудови слова по заданій семантиці; другий режим полягає у вирішенні аналогічної задачі учнями під постійним контролем з боку системи, яка аналізує отриману від учнів інформацію на кожному кроці рішення задачі, коректує його дії в разі невдалих кроків і може перейти до першого режиму навчання за проханням того, хто навчається, тобто продовжити рішення задачі за нього. Важливим засобом, за допомогою якого учень набуває практичних навичок у словотворенні російської мови, є можливість інтерактивного словотворення.

Сторінка інтерактивного словотворення складається з машини переходів по словотворчому гнізду (МПСГ), яка завантажується на сервері і спеціальної програми на клієнті. МПСГ також відповідає за контроль над діями того, хто навчається. Кожна вхідна подія, що виникає внаслідок дій того, хто навчається, порівнюється з допустимою дією. Коли досягнуто останній стан, результати передаються планувальнику навчання для прийняття рішення, чи досяг учень мети вивчення, пов'язаної зі сторінкою інтерактивного словотворення.

9. Відмінні риси і перевага отриманих результатів (продукції) над вітчизняними або зарубіжними аналогами чи прототипами

Для оцінки ефективності та надійності результатів роботи алгоритму морфологічного аналізу було проведено серію експериментів з залученням експертів, які визначали такі граматичні характеристики словоформи, як частина мови, рід, число і відмінок. Для цих же словоформ були отримані аналогічні граматичні характеристики за допомогою запропонованого алгоритму. Результати експериментів показали, що у двох з двадцяти проведених експериментів відповіді експерта та системи не співпали. В обох випадках у дереві словоформ (закінчень) виникала омонімія. Так як аналізовані словоформи не мали суфіксу, то подальший аналіз дерева суфіксів, як запропоновано в алгоритмі для усунення омонімії, був неможливий. Однак, найбільш важливою з морфологічних ознак для подальшої коректної роботи алгоритму семантично-орієнтованого словотворчого синтезу є частина мови, яка була правильно ідентифікована у всіх експериментах. На підставі аналізу отриманих результатів можна зробити висновок, що алгоритм є ефективним і надійним.

Ефективність алгоритму пошуку вузла дерева, який відповідає ключовому слову, і алгоритму пошуку вузла із заданою семантикою була оцінена за допомогою функції часової складності, що залежить від кількості морфем у слові. За результатами значної кількості проведених експериментів встановлено, що розроблений модифікований алгоритм перевершує раніше відомий еквівалентний алгоритм за часом обчислень на простих деревах у 1,2 рази, на складних – у 2 рази, а розроблений евристичний алгоритм – у 1,6 і 2,3 рази відповідно.

Було проведено понад 1000 серій експериментів програмної реалізації узагальненого алгоритму словотворчого синтезу. У кожній серії синтезувалися слова різних частин мови і з різною кількістю морфем, з яких складається ключове слово. Кількість морфем у найбільш поширених словах варіювалася від одної до п'яти, причому, найменше число морфем, з яких складаються прикметники, становить два форманта. Аналізуючи отримані результати, можна відзначити, що точність словотворчого синтезу в усіх випадках становить не менше 93%, що перевищує правильність використання словотворчих моделей російськомовною людиною на 10-15%.

Було визначено семантичну близькість між заданою користувачем семантикою похідного слова (у формальному представленні) і семантикою слова, отриманого в результаті словотворчого синтезу. На підставі отриманих результатів зроблені наступні висновки: міра близькості між заданою і отриманою словотворчою семантикою похідного слова варіюється від 93,5 до 100 відсотків; найбільш часто вживані слова російської мови (слова, які складаються з двох або трьох морфем) утворюються у відповідності із заданою семантикою (міра семантичної близькості не опускається нижче 0,96).

З допомогою розробленої експертної системи навчання проведена діагностика знань учнів шляхом використання підходу, запропонованого в роботах Г.О. Атанова та І.Н. Пустиннікової. Ймовірності підтвердження гіпотез відповідями учнів складають в середньому 0,95 відсотків. На підставі отриманої статистики планувальник навчання експертної системи перетворює форму вправ у відповідності зі знаннями та вміннями учнів.

10. Практична цінність результатів та продукції (галузі економіки та суспільства, де можливе їх використання, конкурентоспроможність та інвестиційна привабливість, ступінь впровадження)

Результати проведеної наукової роботи мають наступну практичну цінність:

1. Розроблені формалізації та алгоритми можуть використовуватися як в експертних системах і навчальних середовищах, так і в інформаційно-пошукових системах, модулях перевірки правопису і системах машинного перекладу.

2. Розроблена експертна система навчання словотворенню сприяє більш швидкому накопиченню словникового запасу учнів і розвиває навички використання різних словотворчих моделей при вивченні російської мови.

3. Розроблений діалоговий компонент і Web-орієнтований інтерфейс експертної навчальної системи спрощують завдання семантики похідного слова та дозволяють використовувати систему в дистанційному навчанні.

4. Результати наукових досліджень використані в науково-дослідних роботах відділу комп'ютерно-інформаційних технологій Інституту проблем штучного інтелекту НАН України.

11. Використання результатів та продукції у навчальному процесі

Результати проведених досліджень використовуються в навчальному процесі кафедри систем штучного інтелекту Державного університету інформатики і штучного інтелекту при підготовці фахівців зі спеціальностей «Інтелектуальні системи прийняття рішень» та «Програмне забезпечення автоматизованих систем», а саме: в тематиці дипломних і магістерських робіт, при підготовці лекційних курсів «Технології комп'ютерного навчання», «Системи штучного інтелекту», «Експертні системи», при проведенні лабораторних та практичних занять.