



ЗАДАЧА АВТОМАТИЧНОЇ ГЕНЕРАЦІЇ ТЕХНІЧНОЇ ДОКУМЕНТАЦІЇ

Авдєєва Ю. О.

Національний технічний університет

"Харківський політехнічний інститут"

м. Харків, вул. Пушкінська, 79/2, тел. (057) 707-64-60, 707-63-60.

e-mail: avdeevajuli@gmail.com

На даний час у зв'язку з швидким ростом кількості текстової інформації, в тому числі в мережі Інтернет, виникає необхідність у швидкій розробці прикладних програмних систем для автоматичної або автоматизованої обробки текстів природною мовою (ПМ – текстів). Прикладами такої обробки є збір і фільтрація даних з різних джерел, витяг знань, реферування, анотування та т.і. Розробка систем для вирішення такого роду завдань має ряд складностей, в першу чергу – необхідність інтеграції великого числа програмних компонентів, що реалізують алгоритми обробки ПМ – тексту, що працюють на різних його рівнях (обробка слів, речень, абзаців тощо).

Генерація текстів на природній мові (ГПМ) – це область досліджень в рамках напряму Автоматична Обробка Текстів (АОТ), що вивчає питання автоматичного породження текстів на природних мовах (ПМ) з деякого формального представлення інформації. Традиційно генерацію поділяють на три етапи:

- Макропланування – побудова структури тексту;
- Мікропланування – побудова планів речень;
- Мовне оформлення – реалізація побудованих планів речень відповідними граматичними структурами [1].

У прикладних системах генерації виділяється п'ять стійких проміжних уявлень даних:

- Об'єктна структура змісту тексту – концептуальна структура змісту тексту, побудована у відповідності зі структурами предметної області (ПО) з об'єктів і відносин ПО.
- Структура тексту – структура дискурсивних відносин, організуюча повідомлення в текст (зокрема, риторична структура).
- Повідомлення – елементарні предикативні фрейми з об'єктами та з відношенням ПО.
- Семантичні подання речень тексту (зокрема, Standard PHP Library – стандартна бібліотека для мови програмування PHP (SPL)).
- Лексикалізація граматичної конструкції речень тексту.

Генерація текстів – дозволяє отримати на основі одного вхідного тексту безліч унікальних текстів, а добре зроблений макет для генерації і подальша перевірка згенерованих текстів гарантує їх читабельність. Для того, щоб розробити систему автоматичної генерації технічної документації, було взято один з методів – шаблонний. Цей метод спирається на вже існуючий шаблон,



який відповідає тексту [2]. Спочатку користувач вибирає певний шаблон, який потрібен для генерації тексту, потім заповнює необхідні поля ключовими словами за потрібною тематикою.

Шаблон являє собою текст з різного роду спеціальними синтаксичними вставками. Ці самі синтаксичні вставки задають алгоритм за яким генеруються різні варіанти тексту.

Три варіанта підтримуваних шаблонів :

- {a | b | c} – програма вибирає одне з поданих значень, наприклад, «b»;
- [a | b | c] – комбінація значень, вибирає наприклад варіант – «bac»;
- [+ I + a | b | c] – вибирає комбінацію значень використовуючи роздільник, в даному випадку результат буде – «b i a i c» [4].

На основі проаналізованої літератури, були виділені шаблони для анотації та змісту інструкції для користувача (рис. 1).

{Программа | Программный продукт}{Picasa} версии {3.9} бесплатная, разработана компанией { Google }. Выпуск последней версии{11 сентября | 2012 | 11 сентября 2012}, язык интерфейса[++ Русский | Украинский | English | Deutsch]. Платформа{Microsoft Office | Linux | Mac OS}. Скачать последнюю версию можно на официальном сайте{ <http://picasa.google.com/> }.

Рисунок 1. Приклад шаблона для анотації інструкції

На виході автоматичної генерації технічної інструкції користувача буде виведено згенерований текст. Приклад такого тексту приведено нижче.

*«Программный продукт Picasa (бесплатный) Google
Год выпуска: 11 сентября 2012
Версия: 3.9.0 Build 136.07
Платформа: Windows
Язык интерфейса: Многоязычный + Русский
Посетите официальный сайт и скачайте последнюю версию
<http://picasa.google.com/> ».*

Список літератури

1. Болдасов М.В., Соколова Е.Г. Генерация содержания запроса пользователя системой InBASE в виде естественно-языкового высказывания // Труды Международного семинара Диалог'2002 по компьютерной лингвистике, Том 2, Протвино, Июнь 2002, С. 49-60.

2. Соколова Е.Г., Шаров С.А. К многоязыковой генерации руководств пользователя: начальный этап проекта AGILE. // Труды Международного семинара Диалог'98 по компьютерной лингвистике и ее приложениям (ред. А.С. Нариньяни), Казань. С. 848-859.

3. Котов Р. Г. Лингвистика и информационная технология / Р. Г. Котов // Лингвистические вопросы алгоритмической обработки сообщений. М.: Наука, 1983, С. 84-96.

4. Paiva, D.S. A survey of applied natural language generation systems. Technical report ITRI-98-03, Informationechnology Research Institute, University of Brighton, Brighton, UK.