

**С. Я. Вишнеvский, А. А. Шимоволос**

*(ГрГУ им. Я. Купалы, Гродно)*

## **Выделение базовых понятий для построения**

### **онтологии на множестве текстов отзывов туристов**

Построение онтологии в заданной предметной области требует выделения ее базиса – словаря понятий и последующей его систематизации и классификации. Оптимальным источником для наполнения такого словаря могут быть текстовые описания и отзывы туристов.

Для достижения поставленной цели, было принято решение основной акцент сделать на поиск существительных в тексте и последующей их классификации. Для этого необходимо: провести лексический разбор текстов, на полученных данных выполнить морфологический анализ, выбрать требуемый набор понятий.

Лексический разбор выполняем на основе регулярных выражений, выделяя слова в тексте. Морфологический анализ проводим с использованием словаря словоформ[1]. Это необходимо по нескольким причинам: во-первых, такой подход позволяет получить наиболее точное описание слова и его нормальную форму; во-вторых, в зависимости от используемого словаря, позволяет исправлять орфографические ошибки и восстанавливать общепринятые сокращения. Базовую классификацию слов проводим на основе морфологических признаков, в частности определяем по признаку одушевленности существительные типов объекты и субъекты.

Подробно рассмотрим предложенный алгоритм. При определении очередного слова выполняем его морфологический анализ, получаем набор морфо-характеристик и нормальную форму слова, если слово является именем существительным, добавляем его в словарь согласно признаку одушевленности. Если для одного слова найдено несколько морфологических наборов, лексическую омонимию не разрешаем, а учитываем все слова при дальнейшем рассмотрении.

Полученные словари существительных обладают свойством полноты, так как выделены все существительные, а также каждое добавленное в словарь понятие присутствует в нем в единственном экземпляре, так как добавляется в словарь только его нормальная форма.

Полученные наборы можно использовать в качестве базиса для построения онтологии туристических предпочтений, доопределив эти наборы связями и правилами вывода. Заметим также, что описанный алгоритм может быть обобщен и использоваться в других задачах выделения понятий из естественно-языковых текстов.

## **Литература**

1 Система ДИАЛИНГ [Электронный ресурс] – Режим доступа: <http://www.aot.ru> – Дата доступа: 17.01.2014.