

Аннотация

При обработке естественного языка встаёт необходимость распознавать имеющиеся в тексте сущности, а также уметь выстраивать из них цепочку уровней общности. Для обеспечения обучения предлагается создать датасет вложенности понятий, опираясь на имеющиеся ресурсы, путём создания разметки на тексте. После создания датасета появится возможность проведения экспериментов по качеству возможного распознавания сущностей. Важность решения задачи состоит в его востребованности для автоматической обработки большого количества текста.