**«Сеть»: анализ контекстов. На пути к моделированию предметной области «Нейронные сети»**

Веренева Юлиана Юрьева

Аспирант Московского государственного университета имени Михаила Васильевича Ломоносова, Россия, Москва

Нейронные сети — тема, которая актуальна для многих направлений. В нашем исследовании данная предметная область рассматривается с точки зрения тезаурусного представления информации. Цель исследования – разработать лингвистическую онтологию предметной области «Нейронные сети». Данная онтология базируется на тесной связи значения языковых единиц и терминов предметной области [[Loukachevitch N.](https://istina.ipmnet.ru/workers/1252816/), [Dobrov B.](https://istina.ipmnet.ru/workers/1265723/):1]

На первом этапе мы сформировали корпус текстов для дальнейшего извлечения терминов. Нашей задачей было сформировать репрезентативный корпус для проведения исследования. Общий объем корпуса составил 4 235 страниц, 1 025 581 слов. Корпус может быть подразделен на два раздела: научно-популярная литература и научная литература. В соответствие с идеей, что корпус представляет собой уменьшенную модель языка или подъязыка, мы посчитали необходимым разделить корпус на учебную и научно-популярную литературу. Для нас важно изучить употребление кандидатов в термины в разных контекстах, чтобы определить контексты функционирования, а также использовать наш анализ для отделения профессионализмов и терминов, так как мы вслед за Н.Н. Болдиной считаем, что профессионализмы и термины не взаимозаменяемы, так как профессионализмы хоть и семантически близки к терминам, но не образуют системы знаний, эмоциональны и образны [Болдина 1996:11].

Научно-популярные тексты представляют собой такие научно-популярные издания, как «Создаем нейронную сеть» (Тарик Рашид), «Введение в нейронные сети» (Евгений Черняк), «Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow» (Олерьен Жерон), «Нейронные сети и глубокое обучение: учебный курс»( Аггарвал Чару), материалы таких курсов, как «Переходим на следующий уровень» , «Программирование на питоне», «Строим первые модели» (Дмитрий Макаров), Обучение нейросетей. Стэнфордский курс. 2019 год.( Fei-Fei Li, Justin Johnson, Serena Yeung.), а также все статьи, представленные в корпусе). Большую долю научной литературы составляют монографии и учебные пособия («Нейронные сети: полный курс» (Хайкинг Саймон), «Глубокое обучение: погружение в мир нейронных сетей» (Николенко С., Надурин А., Архангельская Е.), «Искусственные нейронные сети и их приложения» (Гафанов Ф.М, Галимянов А.Ф.).) и статьи из специализированных журналов («Применение нейронных сетей для анализа текстов»( (Титов И.В., Третьяков В.С. )), «Использование нейронных технологий для повышения квалификации специалистов налоговых служб» (Фомина Е.А.), «Концепт предобученных языковых моделей в контексте инженерии знаний» (Понкин Д.И. ). Всего корпус включает 59 текстовых единиц на русском языке (3 научных статьи, 4 книги для специалистов технических областей, 4 научно-популярных книги, одна из которых написана российским автором, 4 курса и 44 статьи с таких сайтов, как «Хабр»). При работе с изданиями по нейронным сетям мы намеренно исключили предметные указатели, которые представляют собой готовый список терминов.

Второй этап работы представлял собой статистический анализ корпуса и выявление списка слов с указанием их частотности в тексте. Мы выделили 10 слов, которые являются терминами-основами предметной области: сеть (8505), нейрон (6797), обучение (6461), функция (5174), модель (4704), значение (4346), слой (4191), данные (3009), вектор (2999), вес (2698).

Третий этап исследования заключается в анализе семантических отношений слов для моделирования предметной области. Данная работа предполагает анализ всех (8505) контекстов корпуса с исследуемой леммой (в рамках данного доклада рассматривается термин «сеть»); выделение семантических отношений в соответствии с классификацией Л.В. Найхановой [Найханова 2008: 42-50] (квантитативные отношения (отношения синонимии, отношения корреляции), квалитативные отношения (отношение «Объект-свойство», отношение «Часть -Целое», отношение «Род-Вид», отношение «Объект действия – действие – субъект действия»)).

Результатом данного этапа является список конструкций для каждого семантического отношения, наглядное соотношение семантических связей, характерных для предметной области «Нейронные сети».

**Список литературы:**

1. Болдина, Надежда Николаевна. Профессиональная лексика в современном русском языке Учеб. пособие к спецкурсу / Н. Н. Болдина; Пенз. гос. пед. ун-т им. В. Г. Белинского. 1996.
2. Найханова Л.В. Технология создания методов автоматического построения энтологий с применением генетического И автоматного программирования: Монография. Улан-Удэ. 2008. С. 42-50.
3. [Loukachevitch N.](https://istina.ipmnet.ru/workers/1252816/), [Dobrov B.](https://istina.ipmnet.ru/workers/1265723/) [Proceedings of Computational Linguistics in Bulgaria Conference](https://istina.ipmnet.ru/collections/119961851/).]: <https://istina.msu.ru/publications/article/3161637/>