

Если плотность воздуха уменьшается с высотой неравномерно, то искажения формы солнечного диска могут быть самыми разнообразными, например и такими, какие показаны на рисунке.



Рисунок – Искажения формы солнечного диска

Заключение. Оптические явления являются неотъемлемой частью природы и техники, они выполняют важную роль в понимании окружающего мира и разработке различных устройств и систем. Оптика – это наука, изучающая свет и его взаимодействие с веществом, и она имеет широкий спектр применений в природных процессах и технологических приложениях. Одним из наиболее известных оптических явлений является отражение и преломление света. Это явление широко используется в оптических зеркалах, линзах и других устройствах. Оптические явления также оказывают влияние на понимание фундаментальных физических законов и явлений. Исследования в области оптики способствуют развитию квантовой оптики и новых методов управления светом на атомарном уровне [4].

1. Галузо, И.В. Оптические явления в атмосфере при астрономических наблюдениях // И.В. Галузо, В.А. Голубев // Фізика: проблеми викладання. – 2003. – № 6(35). – С. 76–92.
2. Мак-Картни, Э. Оптика атмосферы. – М.: Мир, 1979. – 408 с.
3. Миннарт, М. Свет и цвет в природе. – М.: Наука, 1969. – 344 с.
4. Редкие и необычные явления на небе // Энциклопедия для детей: Астрономия (Том 8). – М.: Аванта+, 2001. – С. 271–279.

ВОЗМОЖНОСТИ РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ДЛЯ КЛАССИФИКАЦИИ ТЕКСТОВ ПО ТЕМАТИКАМ

Щетина П.Д.,

студентка 4 курса ВГУ имени П.М. Машерова, г. Витебск, Республика Беларусь
Научный руководитель – Корчевская Е.А., канд. физ.-мат. наук, доцент

С развитием технологий стремительно меняются и интернет-технологии, что, в свою очередь, приводит к появлению разнообразных способов распространения вредоносных постов в сети. Специализированные службы, включая правоохранительные органы и независимые организации, активно отслеживают социальные сети и интернет-страницы для выявления и блокировки ложных или манипулятивных постов, подрывающих общественное доверие и провоцирующих насилие. Этот процесс требует значительных временных и человеческих ресурсов, что приводит к созданию небольших подразделений, специализирующихся на анализе контента по определённым темам. Для увеличения скорости и классификации постов могут использоваться нейронные сети.

В целом разработка интеллектуальной системы для классификации текстов по темам обеспечивает повышенную точность, гибкость и эффективность при анализе больших объёмов информации по сравнению с традиционными методами. Это позволяет пользователям быстро получать релевантные результаты, что значительно упрощает процесс поиска и извлечения нужных данных.

Материал и методы. Для классификации текстовых постов по тематикам важно понимать основные аспекты работы с текстовыми данными натурального языка.

Токенизация – это фундаментальный процесс, включающий разделение текстовых данных на более мелкие единицы, называемые лексемами. Эти лексемы обычно представляют собой отдельные слова или значимые фрагменты текста. Разбивая входной текст на лексемы, система создаёт более подробное представление, которое позволяет эффективно индексировать и извлекать информацию. Токенизация также включает в себя обработку пунктуации, удаление специальных символов и учёт специфики языка.

Стоп-слова – это часто встречающиеся в языке слова, которые не несут существенного значения для целей поиска, например, «это» или «является». В классификации с помощью нейронных сетей удаление подобных «стоп-слов» может значительно повысить эффективность за счёт уменьшения размера индекса и сосредоточения внимания на более информативных терминах.

Стемминг и лемматизация – это методы, которые помогают сопоставить различные формы слова в процессе поиска, сводя их к базовой или корневой форме. Стемминг подразумевает сокращение слов до их основы путём удаления префиксов, суффиксов и склонений, а лемматизация приводит слова к их канонической форме на основе их словарной статьи. Эти методы обеспечивают равенство различных грамматических форм слова [1].

Результаты и их обсуждение. Модели машинного обучения, в том числе нейронные сети, принимают на вход числовые данные. Тексты, состоящие из слов, не могут быть напрямую обработаны математическими моделями. Преобразование текстов в числовые последовательности позволяет использовать их в качестве входных данных для обучения.

Лучше всего с задачей классификации текстов справляются сети-трансформеры. Эти модели используют механизм внимания, который позволяет сосредоточиться на наиболее значимых частях входных данных, что особенно полезно при работе с текстом, где контекст играет ключевую роль. В отличие от традиционных рекуррентных нейронных сетей, которые обрабатывают данные последовательно, трансформеры обрабатывают все слова в предложении одновременно, что значительно ускоряет процесс.

Трансформеры работают с последовательностями, представляя каждое слово в виде векторного представления (эмбединга). Затем они применяют механизм внимания, который позволяет модели оценивать важность каждого слова относительно других слов в контексте. Это означает, что при анализе предложения модель может «обращать внимание» на слова, которые имеют большее значение для понимания смысла всего текста.

Заключение. Таким образом, понимая ключевые аспекты разработки интеллектуальной системы для тематической классификации текстов, специалисты могут оптимизировать процесс обработки информации, повысить точность классификации и предоставить пользователям возможность эффективно извлекать наиболее релевантные данные из больших массивов текстов. Это позволит создавать более эффективные инструменты для автоматической категоризации текстов, что значительно упростит поиск нужной информации.

1. Кришень, Д.А. Раскрытие возможностей эффективного текстового поиска / Кришень Д.А.; науч. рук. Ермоченко С.А. // Молодость. Интеллект. Инициатива: материалы XII Международной научно-практической конференции студентов и магистрантов, Витебск, 26 апреля 2024 года: в 2 т. – Витебск: ВГУ имени П.М. Машерова, 2024. – Т. 1. – С. 39–40. – Библиогр.: с. 40 (5 назв.) – URL: <https://rep.vsu.by/handle/123456789/43192> (дата обращения: 02.02.2025)