

Отзыв о презентации Яндекс–Королев: хорошо, но не очень. – Нестеров А. В.

Презентация новой поисковой системы Яндекса «Королев», несомненно, обладала драматургией, но недостаточный артистизм исполнителей и метафоричность контента, оставила неприятный осадок. Конечно, разработчики и их руководители не артисты, но когда они говорят о качестве контента, то необходимо пояснять, что они имеют в виду и их контент должен соответствовать этим требованиям.

Судя по всему, фишкой системы Королев является наличие тензорного поиска, по словам всего анализируемого контента, который выражается в виде вектора из трёхсот слов поискового образа документа. При этом с этим вектором работает «искусственный интеллект» в виде многослойной нейронной сетки, которая должна быть «обучена» многочисленными ассесорами с помощью «глубинного обучения».

Действительно, то, что называют нейросетью после «глубинного обучения», может узнать огурец и отличить его от помидора, но при этом она не понимает, что она делает и не может объяснить, как она это делает. Поэтому лучше такую дрессировку системы «черного ящика» называть «темным «обучением»», по аналогии с «темным веществом и материей». Наверное, метафоричность позволяет разработчика лучше уговаривать инвесторов и маскировать алгоритмы поиска от конкурентов.

Однако для пользователей не важно, как называется поисковая система и ее алгоритмы. Несомненно, что Королев будет выдавать более точное соответствие поискового образа запроса и поискового образа документа (медиа), что было показано в публикации «Нестеров А. В. Методы информационного поиска: тензорный подход. – Новосибирск: Препр. / ГПНТБ СО АН СССР, 91-4. - 1991. – 36 с.».

На презентации также отмечалось, что в системе Королев релевантность результата поиска будет выше, чем ранее. Однако проблема повышения релевантности возникла не вчера. С ней боролись давно, до

появления интернета, с помощью библиографов. Естественно, тогда поток документов был значительно меньше, но уже тогда было известно, что человек может прочитать из релевантных документов примерно 10%. Поэтому ему нужно найти эти десять процентов пертинентных документов. Многим людям нужны не только пертинентные, но детонационные документы, которые детонируют их на продуцирование креативных и/или когнитивных продуктов. Поэтому у команды Яндекса впереди еще много работы.

Практика показывает, что ни библиограф, ни поисковая система принципиально не может заменить способность пользователя анализировать результаты поиска. Формализация поискового образа запроса должна идти не только по формализации совокупности знаков, но и содержания документов (медиа). Содержание документа зависит не только от пользовательских запросов, но и от ссылок (связей), как между документами, так и между частями контента. Поэтому система Королев должна уметь настраиваться на индивидуальный поиск людей, позволяющей находить пертинентные и/или детонационные документы, как это было описано в публикации Нестеров А. В. Компьютерные методы и средства глубокой обработки, анализа и синтеза общедоступных документов. Новосибирск: Изд-во ГПНТБ Сибирского отделения АН СССР, 1991. - 214 с.

Конечно, публикаторы конкурируют за возможность попасть в первые пять страниц выдачи, а поисковые системы конкурируют между собой за точность поиска, и это стимул для прогресса в продуцировании новых возможностей по поиску документов в интернете. Однако тестирование некоторых запросов показывает, что первыми выходят рефераты студентов и школьников, т.е. неважно качества контента, а важно количество запросов. Про рекламу на первом месте разговор не идет. Это святое.

Если нейросетка Яндекса-Королев «поймет» мой текст и обратит на него внимание ребят из Яндекса, то мне будет интересно получить от них обратную связь.