

РЕФЕРАТ

УДК 002.001

Д. О. АВЕТИСЯН, Г. С. МОВСЕСЯН

ПСИХОЛОГИЧЕСКИЕ И СТАТИСТИЧЕСКИЕ АСПЕКТЫ ПОНЯТИЯ РЕЛЕВАНТНОСТИ

Понятие релевантности является ключевым в теории и практике информатики. При эксплуатации как документальных, так и фактографических информационно-поисковых и информационно-логических систем весьма важное значение имеют вопросы, связанные с оценкой степени соответствия найденной информации информационной потребности.

Естественно, чрезвычайно сложное психологическое явление информационной потребности не всегда удается точно, однозначно и исчерпывающе сформулировать в виде информационного запроса. Под одним и тем же информационным запросом различные специалисты в принципе могут подразумевать различные информационные потребности. Элементы субъективизма, неминуемо сопровождающие процедуру языковой формулировки различных семантических категорий, стали причиной многочисленных научных споров вокруг понятия релевантности (пертиненности), призванного служить для оценки степени близости найденной информации и информационного запроса (информационной потребности).

В отличие от понятия формальной вероятностной меры понятие релевантности содержит также прагматическую меру оценки потребителем найденной информации, что в значительной степени зависит от психологических факторов потребителя и не поддается исчерпывающей формализации.

Путем предварительной статистической обработки данных эксперимента, проведенного сотрудниками Абовянского завода радиодеталей и Государственного Союзного конструкторского бюро приборов и элементов автоматики, была разработана информационно-статистическая модель оценки релевантности контрольной группой потребителей. В качестве экспериментального массива документов служили 880 рефератов из реферативных журналов «Автоматика, телемеханика и вычислительная техника» (№№11, 12, 1969 г.). В качестве экспериментального массива запросов служили 10 запросов, сформулированных потребителями. В контрольную группу входили по 5 специалистов из вышеупомянутых организаций.

Путем статистического подытоживания результатов оценки семантической близости очередных пар документов-запросов различными

членами контрольной группы потребителей, удалось определить статистическую вероятность ($P=0,85$) того, что статистическое среднеквадратическое отклонение различных оценок при числе потребителей, равном десяти, при пятибальной шкале оценки не превышает величину 1,24. Удалось построить семейство кривых определения необходимого числа потребителей контрольной группы для нахождения с заданной точностью и достоверностью объективной меры оценки семантической близости.

Эксперимент показал, что при прочих равных условиях чем выше квалификация специалистов, входящих в контрольную группу, тем меньше дисперсия оценок семантической близости и, следовательно, больше точность и достоверность определения математического ожидания с одинаковым числом членов контрольной группы.

Иллюстраций 3. Библиографий 12.

Лаборатория нейробионики АН АрмССР

Поступило 17.III 1971 г.

Полный текст статьи депонирован
в ВИНТИ