

УДК 662.642: 621.926.7

М.Г. Киселева, А.Н. Садовникова (4курс, каф. САиУ), А.А. Денисов, д.т.н., проф.

СОЗДАНИЕ ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ ПО ТЕМЕ “МЕРЫ ИНФОРМАЦИИ”

Существуют различные способы измерения информации. В 1928 г. Р.Хартли определил информацию сообщения как $H = m \log n$, где m – количество символов в сообщении, а n – количество доступных для использования различных символов.

С точки зрения современной теории информации эта мера применима только тогда, когда символы выбираются независимо один от другого, и выбор любого из символов равновероятен.

Наиболее полное развитие теории информации применительно к проблемам ее передачи было осуществлено в сороковые годы К.Шенноном. Шеннон связал понятие информации с понятием энтропии и предложил следующую меру количества информации:

$$H = \sum_{i=1}^n p_i \log p_i ,$$

где p_i – вероятность того, что будет выбран i -й символ из полного набора в n символов, которые вырабатывает источник сообщений.

Величина H (энтропия) измеряется в битах (от англ. binary digit – двоичная система). А.А. Харкевич связал ценность информации с целью деятельности, предложив рассматривать энтропийную меру Шеннона как меру вероятности недостижения цели.

Е.К. Войшвилло предложил различать информацию восприятия (знаки) и информацию-значение, смысловое содержание для потребителя информации (означаемое знака).

А.А. Денисов показал, что информацией для потребителя является пересечение чувственной (информации восприятия) и логической (потенциала) информацией, в результате чего формируется единое понятие – информационный смысл. Кроме того, в концепции А.А. Денисова учитывается, что далеко не всегда может быть строго определено вероятностное пространство. Поэтому предложена информационная мера, учитывающая степень влияния ресурсов на реализацию целей. Эта мера позволяет учесть не только вероятность достижения цели p_i , но и вероятность q_i того, что данная информация будет использована лицом, принимающим решение, при реализации цели:

$$H = -\sum_i q_i \log_2(1 - p_i)$$

Разработана информационно-поисковая система, которая позволяет получить данные о методах измерения информации.

В работе ИПС определена следующим образом:

$$IPS = \langle D, Q, TS, RL, IND, KSS, D', N \rangle ,$$

где D – множество документов;

Q – множество информационных запросов;

TS – технические средства (устройства, необходимые для записи и хранения поисковых образов, для хранения документов и осуществления процесса сопоставления поисковых образов документов (ПОД) с поисковым образом запроса (ПОЗ));

RL – информационно-поисковый язык (ИПЯ);

IND – правило индексирования;

KSS – критерий смыслового соответствия;

D' – ответ на информационный запрос;

N – люди, взаимодействующие с системой (те, кто пользуются данной ИПС, осуществляют индексирование документов и информационных запросов, выбирают стратегию поиска).

В системе информационный поиск осуществлялся по автору, названию, году, издательству, количеству страниц, ключевым словам, следующим образом: каждому документу, вводимому в поисковый массив, ставится в соответствие ПОД, в качестве которого, в данной системе, является заглавие документа. В виде такой же краткой характеристики создан информационный запрос (ПОЗ). Процедура поиска осуществляется при сопоставлении ПОД и ПОЗ. Оценка качества поиска проводилась по критериям: релевантность и пертинентность.

Для сопоставления ПОЗ и ПОД разработан ИПЯ. Для перевода с естественного языка на ИПЯ введена система индексирования. Если обозначить d_i за отдельную книгу в поисковом массиве D , то ее можно проиндексировать для того, чтобы найти из общего массива по разным признакам:

$d_i = (A, NAZV, IZD, G, STR, CS)$,

где A -автор;

NAZV -название;

IZD -издательство;

G -год издания;

STR -количество страниц;

CS -ключевые слова.

Система может быть использована для поиска информации о различных мерах, для выбора меры.