А. С. Козицын, **И. М. Конев**, **Е. А. Степанов** (Москва, МГУ)**. Индексация коллекций документов в поисковой системе АСТАИ.**

Основной задачей любой поисковой системы является быстрое выполнение запросов по подбору документов согласно введенным пользователем параметрам. В качестве параметров запроса в системе АСТАИ [1] могут использоваться ключевые слова, термина, даты, рубрики, ресурсы и именные группы. Для осуществления быстрого поиска в системе АСТАИ, также как и в аналогичных системах, используются обратные индексы. Обратный индекс по каждому параметру состоит из набора файлов определенной структуры, содержащей информацию о документах, релевантных данному значению параметра. Каждому значению параметра соответствует ровно один файл.

В АСТАИ используются различные обратные индексы для дискретных и недискретных параметров.

Файл обратного индекса для каждого значения дискретного параметра содержит отсортированный список ключей документов с указанием значимости параметра для документа и позиций, в которых встречался данный параметр с выбранным значением.

Обратный индекс для недискретных параметров, состоит из файлов, соответствующих конкретному значению хэш-функции от значения параметра. В каждом файле этого индекса хранится отсортированный список ключей документов, с указанием значимости параметра для документа и списка принимаемых значений.

Для каждого параметра используется собственная хэш-функция, сохраняющая порядок и, по возможности, равномерно распределяющая множество документов на область возможных значений.

Дополнительно для каждого документа может запрашиваться информация об авторитетности документа. Эти данные хранятся в виде индекса на отдельных серверах и за счет небольшого размера полностью умещаются в оперативную память.

Язык запросов системы АСТАИ позволяет задавать любые логические выражения с использованием операций AND, OR, NOT. Использование индексов, описанных выше, позволяет вычислить значение логической формулы для документа и провести ранжирование за один проход по файлам индексов.

СПИСОК ЛИТЕРАТУРЫ

1. *Афонин С. А.*, *Козицын А. С.*, *Титов А. С.* Автоматизированная система обработки информации в интересах обеспечения безопасности критически важных объектов. — Проблемы безопасности и противодействия терроризму. Материалы конференции МГУ 25–26 сентября 2006 г. МЦНМО, 2007, с. 383–400