

**Е. А. Гладкая**  
(УО «БГТУ», Минск)

212

## ОРГАНИЗАЦИЯ ПОЛНОТЕКСТОВОГО ПОИСКА ПРИ ХРАНЕНИИ НАУЧНЫХ СТАТЕЙ

Поисковая система (ПС) – программно-аппаратный комплекс с веб-интерфейсом, предоставляющий поиск информации. Программной частью ПС является поисковая машина (поисковый движок) – комплекс программ, обеспечивающий функциональность ПС и обычно является коммерческой тайной компании-разработчика ПС.

MS SQL Server содержит компонент FTS или Full-Text Search, который позволяет выполнять поиск данных в текстовых документах и дает возможность разработчику создать свою ПС.

Исходные данные необходимо проиндексировать, для этого определяются границы лексем с помощью делителей текста.

Сразу, после заполнения индекса, пользователи могут отправлять полнотекстовые запросы, базовая единица такого запроса называется лексемой (token).

Также создаются списки стоп-слов – слова, которые имеют низкую значимость и не используются при полнотекстовом поиске.

Каждый индексированный термин имеет свой коэффициент релевантности, который определяет важность и значимость концепта данного термина. Значимость измеряется количеством вхождений термина в текст.

Заполнение полнотекстового индекса FTS происходит автоматически или в результате изменения данных в соответствующих таблицах. Оно может происходить сразу или по расписанию, в обоих случаях может быть полным или инкрементальным.

Создать полнотекстовый индекс можно с помощью T-SQL, а также в самой среде MS SQL Server. В результате получаем полнотекстовый индекс, к которому выполняются полнотекстовые запросы. Результатами запросов могут быть конкретные слова, фразы, синонимические формы конкретного слова.

Столбцы, которые хранят файлы с поддерживаемым расширением, также можно использовать при полнотекстовом поиске. Интерпретация таких данных происходит с помощью фильтра, который реализует интерфейс IFilter, извлекает текстовую информацию и представляет для индексирования. Данный подход полнотекстового поиска был использован при разработке веб-приложения, в котором осуществляется хранение научных статей.