

Разработка системы семантического поиска файлов на основе их метаданных

Автор: Я.С. Рогов, студент, Университет ИТМО, Санкт-Петербург
Научный руководитель: к.т.н., доцент, С.В. Быковский, Университет ИТМО

С момента появления и распространения компьютеров иерархическая файловая система (ИФС) была стандартом де факто. Относительная простота реализации, высокая производительность и понятность используемых структур делали такие файловые системы естественным выбором для использования в компьютерах.

Однако с развитием компьютерных систем более важным стала возможность категоризировать и быстро находить необходимую информацию: иначе говоря, иметь возможность находить информацию не по ее физическому расположению на носителе, а по значению и назначению, т.е. по ее семантике.

Задачу простой классификации файлов можно решить при помощи средств ИФС, объединяя директории на одном уровне иерархии по тематике (например, исполнитель/год/альбом для музыкальных файлов). Однако при организации более сложной системы возникают следующие проблемы:

1. невозможность категоризации данных по нескольким темам (ключам);
2. неиспользование данных, хранимых как часть формата отдельных файлов;
3. повторное представление метаинформации в структуре каталогов (избыточность).

Для решения данных проблем необходима система, которая использует уже существующую информацию о файлах для организации файлов в специальную структуру и которая предоставляет единый интерфейс поиска необходимых файлов по их смысловой информации, то есть система семантического поиска.

Система основывается на метаинформации - информации о файлах, которая описывает сам файл и его содержимое, и использует эти метаданные для создания внутренней базы данных, при обращении к которой пользователю предоставляется возможность находить необходимые им файлы по их описанию.

В данной работе предлагается архитектура системы семантического поиска, описываются её возможности, а также создан прототип такой системы и произведён анализ её характеристики производительности (например, скорость обработки файлов).

Руководитель ОП: _____ Алиев Т.И. _____ (подпись)

Научный руководитель: _____ Быковский С.В. _____ (подпись)

Автор: _____ Рогов Я.С. _____ (подпись)