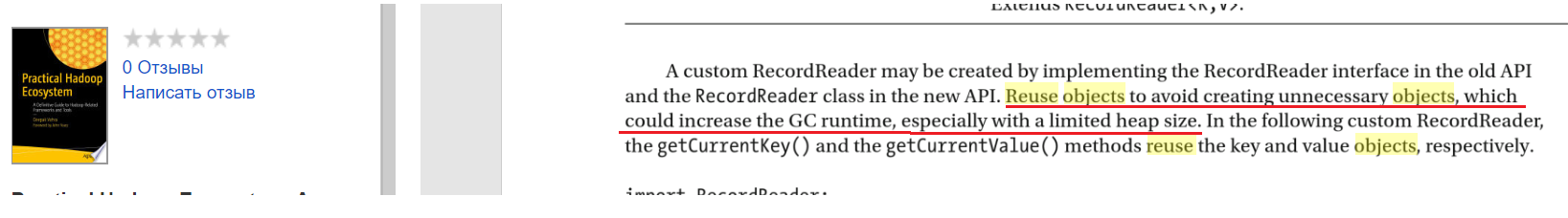
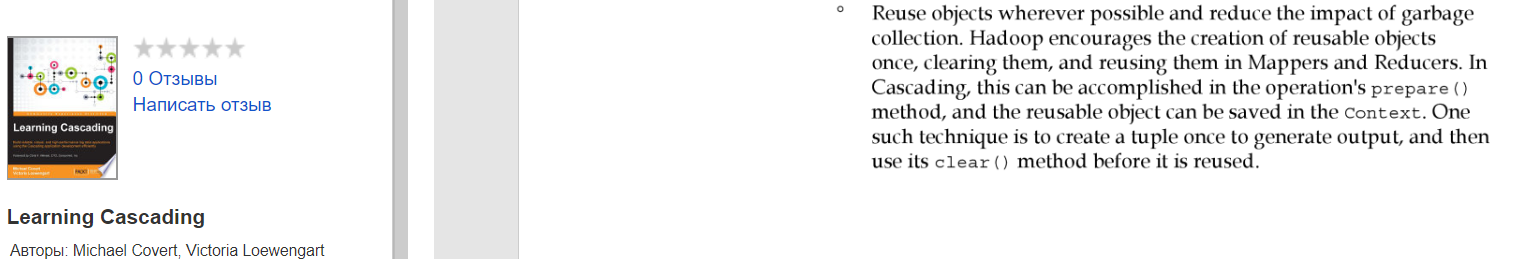
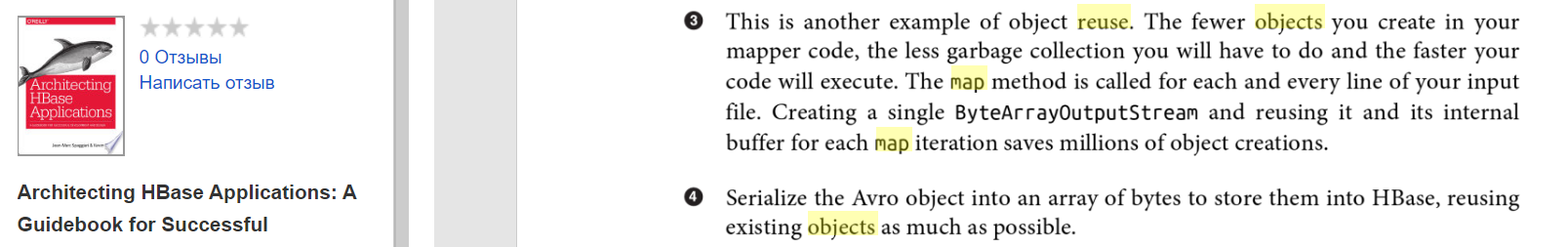
1. Зачем в технологии Map\_Reduce переиспользуются одни и те же объекты с подменой их значений?

Ответ:

Дело в работе сборщика мусора, который будет замедлять обработку данных из-за постоянного запуска при превышении определенного объема памяти, так как будет создаваться огромное количество новых объектов. Если создавать по объекту на каждую строку - то больше времени будет тратится на сборку мусора, чем логику обработки. Сериализация/Десериализация тоже уже как следствие оптимальнее работает, не создавая лишних объектов, а просто копируя байты из потока в заготовленный буфер. Источники:







<https://softwareengineering.stackexchange.com/questions/228571/hadoop-and-object-reuse-why> (Creating a million new strings is always going to be slower than using a StringBuilder to manipulate a single string)

<https://m.facebook.com/notes/cloudera/7-tips-for-improving-mapreduce-performance/207887162002/> (пункт 6, о переиспользовании writable)

таск на введение переиспользования объектов - <https://issues.apache.org/jira/browse/HADOOP-2399>

public class MapRunner#run