데이터베이스시스템 PROJECT3

20161595 배성현

**1. MongoDB 질의문 및 결과**

1)전체 기사의 개수를 구하시오.



2) 기술/IT면에 실린 기사들의 개수를 구하시오.



3) 스포츠 기사 중 ‘우승’ 단어를 포함하는 기사들의 개수를 구하시오.



**2. RDB vs NoSQL에 대한 비교**

RDB는 Relational Database를 의미하며, 관계형 데이터 모델이 기반을 둔 데이터베이스이다. 관계형 데이터 모델이란 데이터를 구성할 때에 모든 데이터를 2차원 테이블 형태로 표현해준다. RDB에서는 데이터가 스키마를 통하여 명확하게 정의되에 데이터베이스 테이블에 저장되었기 때문에, 데이터의 무결성을 보장하고 각 데이터를 중복 없이 저장할 수 있으며 Transaction을 통해 작업의 원자성을 보장한다. 또한 RDB에서의 데이터는 데이터 사이의 관계를 통하여 연결된 여러 개의 테이블에 분산되어 저장되게 된다.

NOSQL은 Not Only SQL은 RDB 데이터베이스와 반대되는 방식을 사용하며, 데이터 사이의 관계를 정의하지 않는다. 또한 RDBMS에서는 스키마에 따라서 데이터를 관리해야 하지만, NoSQL의 경우에는 스키마가 없기 때문에 RDB보다는 더 유연하게 비정형의 데이터를 다룰 수 있고, 대용량의 데이터를 관리할 수 있다. 또한 데이터가 어플리케이션이 필요로 하는 형식으로 저장되기 때문에 데이터를 읽어오는 속도가 빠르다. 하지만 동시에 스키마가 없기 때문에 데이터에 대해 규격화된 결과값을 얻기 어려우며, 데이터의 중복이 나타날 수 있다. 그리고 NoSQL에서는 RDB의 테이블과 같은 개념으로 Collection이라는 형태로 데이터를 관리하게 된다.

텍스트 마이닝에서는 RDB보다 NoSQL을 사용하는 것이 더 좋다. 텍스트 마이닝에서는 조인연산을 할 일이 없고, 위의 프로젝트에서와 같이 방대한 비정형의 텍스트(뉴스들)에서 의미가 있는 정보(프로젝트 내의 wordset, frequent\_itemset 등)를 추출해내어 다른 정보와의 연계성(프로젝트 내의 strong 연관규칙)을 파악하는 등 일을 해야 하기 때문에 비정형 데이터, 대용량의 데이터를 더 유연하게 처리 가능한 NoSql이 RDB보다 더 효율적이다.