**과목명: 데이터베이스시스템**

**분반 1**

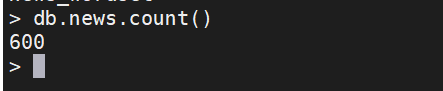
**<<Project #3>>**

**서강대학교 [컴퓨터공학]**

**[20161598]**

**[손동현]**

1. MongoDB 기본 질의
   1. 전체 기사의 개수를 구하시오



* 1. 기술/IT면에 실린 기사들의 개수를 구하시오



* 1. 스포츠 기사 중 ‘우승’ 단어를 포함하는 기사들의 개수를 구하시오.



1. RDB vs NoSQL DB에 대한 비교

* 텍스트 마이닝을 수행할 때에는 RDB보다 NoSQL DB를 사용하는 것이 좋은가?

RDB, 즉 관계형 데이터베이스 모델은 행과 열(튜플과 어트리뷰트)로 이루어진 테이블 형태의 데이터 구조와 그 데이터들 간에 연결을 정의하고 있다. 관계를 기반으로 한 데이터베이스 개체의 내부 관계가 아닌 외부의 관계를 적용한 것이어서 우리 세상에 존재하는 거의 모든 형태의 자료를 표현할 수 있는 꽤 완벽한 모델이다. RDB는 배우기 쉽고 효율적이며 기능적으로 완전하다는 장점이 있으며 독립성이 매우 강하다 특징을 갖고 있다. 하지만 최근 들어 SNS의 등장에 따라 어마어마한 자료가 매 초마다 쏟아지게 되면서, 이러한 정보의 양은 기존의 데이터 모델로 분석, 처리할 수 있는 양을 넘어서게 되었다. 이것을 보고 빅데이터라고 하고, 이 빅데이터를 RDB로 처리하기에는 데이터 형태의 일관성이 떨어진다는 단점이 있다. 따라서 이를 개선하기 위해 NOSQL이 나왔는데 Not only SQL로써 SQL 계열 쿼리 언어를 사용할 수 있으며 RDB와는 다르게 table이나 column과 같은 스키마 개념이 없고 트랜잭션을 지원하지 않는다. 즉 전통적인 관계형 데이터베이스 보다 덜 제한적인 일관성 모델을 이용하는 데이터의 저장 및 검색을 위한 매커니즘을 제공하고, 정해진 규격이 없으므로 비정형 데이터를 다루는데 유리하다. NOSQL은 단순 검색 및 추가 작업에 최적화 되어 있으며 처리 속도나 효율이 매우 뛰어나다는 장점이 있다.

텍스트 마이닝에서는 테이블 간의 조인을 수행할 필요가 없고, 대용량 데이터를 처리해야 한다. 또한 비정형 데이터를 다루기도 해야 한다. 이런 점을 미루어 보았을 때, 텍스트 마이닝을 수행할 때는 RDB보다 테이블 schema가 유동적이고, 대용량 데이터를 저장 가능한 NOSQL을 사용하는 것이 더욱 효율적이다.