**DB project#3**

**Text mining with MongoDB**

학번: 20171639

학과: 컴퓨터공학

학번: 박소희

1. 데이터베이스 서버에 접속해 MongoDB 질의를 수행하고 결과 화면을 screenshot으로 첨부할 것.

1. 전체 기사의 개수를 구하시오.



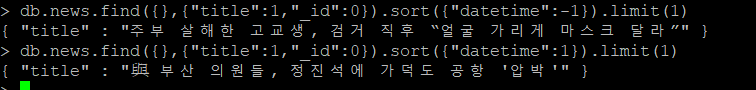
1. ‘연합뉴스’ 사에 실린 기사들의 개수를 구하시오



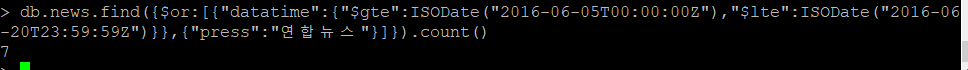
1. 2016년 6월 1일부터 2016년 6월 12일까지 실린 기사 개수를 구하시오



1. 전체 기사를 날짜 별로 정렬해서 가장 빠른 날짜와 늦은 날짜의 기사 제목을 출력하시오.



1. ‘연합뉴스’ 사에 실린 기사 혹은 2016년 6월 5일부터 2016년 6월 20일까지 실린 기사의 개수를 구하시오.



2. 텍스트 마이닝을 수행할 때에는 RDB보다 NoSQL DB를 사용하는 것이 좋은가? 이에 대한 답을 하고 이유를 기술하시오.

: 텍스트 마이닝이란 비정형 데이터 마이닝의 유형 중 하나로 자연어 처리 기술과 문서 처리기술을 적용하여 유용한 정보를 추출, 가공하는 것을 목적으로 하는 기술이다. 실생활에서 만들어지는 대부분의 문서는 형태로 보관되며 제목, 저자, 출판날짜 등과 같은 구조적인 특징들과 문서의 요약, 내용과 같은 크기가 일정하지 않은 비구조적 요소들을 포함하기에 반구조적 데이터로 분류된다.

RDB란 관계형 데이터 모델에 기초를 둔 데이터베이스이다. 관계형 데이터 모델이란 데이터를 구성하는데 필요한 방법 중 하나로 모든 데이터를 2차원의 테이블 형태로 표현해 준다. RDB는 데이터의 독립성이 강하다.

NoSQL은 Not Only SQL로서 기존의 관계형 DBMS가 가지고 있는 특성 뿐 만 아니라 특성들을 부가적으로 지원한다는 것을 의미한다. NoSQL 데이터베이스는 기존의 관계형 데이터베이스보다 더 융통성있는 데이터 모델을 사용하고 데이터의 저장 및 검색을 위한 특화된 메커니즘을 제공한다. 이를 통해 NoSQL 데이터베이스는 단순 검색 및 추가작업에 있어서 매우 최적화된 키 값 저장 기법을 사용하여 응답속도나 처리효율 등에 있어서 매우 뛰어난 성능을 나타낸다.

따라서 비구조적인 요소들을 포함하는 텍스트 마이닝의 경우 NoSQL에 유리하다. 왜냐하면 NosQL은 RDB와 달리 데이터 관의 관계를 정의 하지 않고 ,RDB보다 훨씬 더 대용량의 데이터를 저장 가능하고, 테이블의 스키마가 유동적이어서 이번 프로젝트와 같은 텍스트 마이닝의 경우 NoSQL이 RDB보다 적합하다.