



8/3 ~ 8/7 5주차 - 3

게시글 버전 되돌리기에 대한 고민...

버전 되돌리기

1. 게시판에 게시글이 등록된다.
2. 사용자가 수정한 내용을 저장하고 깃에서 처럼 이전에 작성한 내용을 그대로 보고자 한다면 어떻게 해야할까.
3. Version 테이블을 따로 만들어 그대로 저장한다.
4. content의 제한이 70만자 이므로 대략 3MB이다. (280만 바이트)
5. 사용자가 70만자의 내용중 1글자만 바뀌서 30개를 저장한다면 이 버전을 관리하는데 DB에서 60MB가 필요하고 기타 다른 버전 정보(시간, 작성자, 제목 등)들을 저장한다면 이 보다 약간 더 저장 공간을 필요로 한다.
6. github는 파일의 변환된 내용을 기록하여 파싱하는 내용이지 않을 까 라는 생각을 하였다.
7. 텍스트의 변화내용을 기록하는 것을 찾아보았다.
8. 리눅스에서 diff나 patch 같은 파일 변경 내용을 알 수 있는 명령어가 있다.

diff

두 파일 또는 프로젝트의 diff를 줄별로 비교 및 분석해주는 명령어입니다.

patch

패치파일을 이용해 프로젝트나 파일에 패치를 적용하는 명령어입니다.

diff를 이용해 patch 파일 만들기가 가능하다

두 파일간의 diff를 이용하여 patch파일을 만들고 파일에 patch파일을 적용하면

A - 원본

B- 변화된 파일

diff를 이용해 A와 B사이의 patch파일을 구한다.

A에 patch파일을 적용하면 언제든지 B파일을 얻을 수 있다.

```
index 1394ce8..2ce6036 100644
--- a/main.c
+++ b/main.c
@@ -1,5 +1,6 @@
#include <stdio.h>
```

어떤 라인에 어떤 문장이 변경되었는지 알 수 있다.

diff에 대한 라이브러리를 검색하였고

kpdecker/jsdiff


A javascript text differencing implementation. Contribute to kpdecker/jsdiff development by creating an account on GitHub.

 <https://github.com/kpdecker/jsdiff>



google/diff-match-patch

The Diff Match and Patch libraries offer robust algorithms to perform the operations required for synchronizing plain text. Diff: Compare two blocks of plain text and efficiently return a list of

 <https://github.com/google/diff-match-patch>



이 둘을 알 수 있게 되었다.

jsdiff는 javascript만 지원하고 뷰가 익숙하지않아 이를 사용하기 어려웠다.

반면에, google/diif-match-patch는 java,python, c++등 다양한 언어가 지원되고 백엔드에서 에러처리를 하기위해 자바언어를 사용할수 있는 이 라이브러리를 택했다.

버전 저장 알고리즘은 이렇다.

1. 최초 오리지날 게시글 내용이 저장된다.
2. 게시글 수정시 diff를 이용하여 patch파일을 구하고 10개의 새로운 내용의 저장은 patch파일만을 기록한다.
3. 이 patch파일의 크기는 원본 게시글보다 보통 적고 게시글 전체가 전혀 다른 내용으로 바뀐다면 더 길 수도 있다.
4. mysql context를 저장한 데이터의 바이트는 이렇다.

TYPE	사용되는 바이트	속성
TINYTEXT(n)	최대 n (<=255)	문자열로 취급
TEXT(n)	최대 n (<=65535)	문자열로 취급
MEDIUMTEXT(n)	최대 n (<=16777215)	문자열로 취급
LONGTEXT(n)	최대 n (<=4294967295)	문자열로 취급

mediumtext를 사용하였고

70만자 * 4바이트 - > 280만 바이트 이다. 한글일 경우

70만자가 새로운 내용으로 바뀐다면 (게시글을 다 지우고 새로운 내용을 썼다 가정)

280만 * 2 이다 560만 바이트가 필요하다.

mediumtext는 1600만 바이트를 저장할 수 있다.

즉 저장공간이 충분하다.

5. patch파일이 70만자보다 커질 수 있는데 이렇게 저장하는 이유는 ?

- 앞서 적었지만 70만자의 버전을 100개를 저장한다하면 300MB정도 필요하다.
- 70만자를 한번에 다 지우고 글을 다시 쓰는 사용자 보단 다른 글을 새로 작성할 가능성이 더 높고.
- 이러한 큰 수정보단 작은 수정을 거치는 일이 보통 사용측면에서 많다고 생각해서 이고 변화내용만 기록한다면 분명 전부 저장하는 것보다 저장공간에서 이득이 생길 것이라 생각한다.

6. 10개의 버전마다 patch파일이 아닌 내용을 저장하려고 한다. (full content 저장)

7. 버전 36의 게시글을 보다가 사용자가 25번의 게시글을 보려고 한다면

8. 20번 버전의 context에서 21 ~ 25의 patch파일을 적용하여 25번 버전의 게시글을 볼 수 있게 구현하려고 한다.