# REPORT

# 보고서 작성 서약서

- 1. 나는 타학생의 보고서를 복사(Copy)하지 않았습니다.
- 나는 타학생의 보고서를 인터넷에서 다운로드 하여 대체하지 않았습니다.
- 나는 타인에게 보고서 제출 전에 보고서를 보여주지 않았습니다.
- 4. 보고서 제출 기한을 준수하였습니다.

나는 보고서 작성시 위법 행위를 하지 않고. 성 군 인으로서 나의 명예를 지킬 것을 약속합니다.

과 목:시스템SW 실습2

담당교수: 정진규

학 과: 반도체시스템공학과

학 번: 2018314788

이 름: 오해성

제 출일: 2020.11.20

## 서 론

본 리포트는 배쉬쉘의 일부 기능을 구현한 미니쉘 프로그램에 대해 다룹니다.

본론에서는 파일 설계도와 FLOW CHART 를 시작으로 구현 파일의 각 함수에 대해 ADT 를 설명하고 코드의 중요한 부분에 대해서만 간략히 설명합니다.

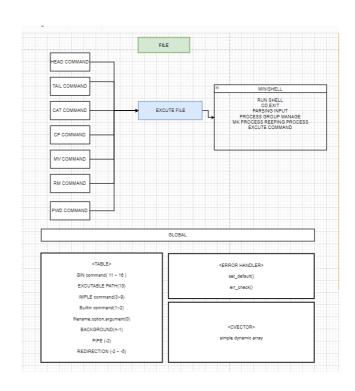
순서는 다음과 같습니다.

- 1. TABLE
- 2. CVECTOR
- 3. ERRHANDELR
- 4. HEAD
- 5. TAIL
- 6. CAT
- 7. CP
- 8. MV
- 9. RM
- 10. PWD
- 11. MINISHELL

#### 파일 설계도

우측 도면과 같이 구현한 소스파일은 총 11개 입니다. (헤더 파일 제외)

실제로 미니쉘은 구동하기 위한 미니쉘 파일과 binary 파일을 만들기 위한 커맨드 관련 파일 7개 위 파일의 동작을 돕기 위해 글로벌하게 쓰이는 파일 3개 입니다.



#### **FLOW CHART**

프로그램의 실행은 다음과 같이 간단합니다.

exit 커맨드가 들어오지 않으면 반복해서 명령어를 입력받고 유효한지 판단 후 유효하다면 foreground 인지 background 인지 판단하여

미니쉘의 생성된 자식(새로운 프로세스 그룹 리더)이 커맨드를 실행합니다.

이 리더 프로세스가 파이프라인이나 리다이렉션에 따라 자식을 생 성하여

실행 커맨드를 관리합니다.

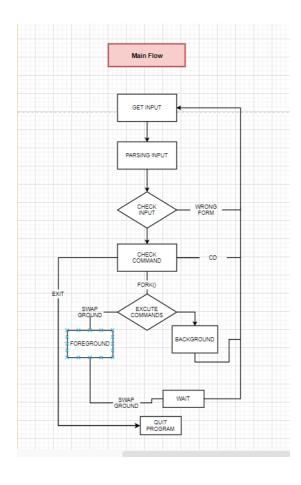
이때 foreground 라면 미니쉘의 자식 그룹은 세션의 foreground 그룹이 되고

미니쉘은 자식의 종료를 기다립니다.

background 라면 미니쉘은 자식의 종료를 기다리지 않고 다음 커맨드를 계속해서 받습니다.

자식의 실행이 종료되거나 시그널을 받아 stop 상태가 된 경우 미니쉘 프로세스에서 이를 reaping 합니다.

reaping 시 그룹 단위로 signal 을 보내 좀비 프로세스가 생성되지 않도록 합니다.



#### **TABLE**

실행 인자를 빠르게 파악하기 위해 문자열을 숫자에 매칭 시켰습니다.

비교연산자를 이용하여 빠르게 문자열의 타입을 파악할 수 있고

숫자를 조합하여 어떤 종류의 리다이렉션인지 빠르게 확인 할 수 있습니다.

C에서 기본적으로 제공되는 map 구조의 stion 없어 strcmp 로 간단히만

구현했습니다.

빈번히 사용되는 table 이므로 트라이나 해싱을 이용하여 더 최적화된 map 으로 개선이 필요합니다.

#### **CVECTOR**

객체 지향 언어가 아닌 C의 특성 상

정형화된 동적배열이 존재하지 않아 매번 동적 할당 후 해제해야 되는 불편함이 있어

자주 쓰이는 int 형으로만 동적배열 자료구조를 만들었습니다, 동작은 C++ vector 자료구조와 유사하며 push 연산 시 bucket의 끝에 다다르면 bucket 을 2배로 늘립니다.

메모리의 해제는 clear 함수를 이용하여 해제합니다.

직접적인 구현은 cvector.c 을 보면 알 수 있으나 크게 복잡한 부분이 없어 이 문서에서는 생략합니다.

#### **ERRORHANDLER**

에러를 매번 소스코드에서 핸들하면 코드가 지저분해지기도 하고 비효율적이라 생각해 따로 에러를 처리해주는 파일을 만들었습니다.

이 파일을 이용하는 소스코드에서는 set\_default 함수로 errno 를 기본값으로 설정 후

동작이 끝난 후 err\_check 함수로 error 가 있다면 그에 해당하는 error 를 형식에 맞게

출력하도록 하였습니다.

```
1 #include "definition.h"
2 #include "errhandler.h"
3
4 void set_default(void)
5 {
6    errno = 0;
7    return;
8 }
9
10 void err_check(char * str)
11 {
12    if(!errno) return;
13    switch (errno)
15    {
16    case EACCES:
17         fprintf(stderr, "%s: Permission denied\n",str);
18         break;
19
20    case EISDIR:
21         fprintf(stderr, "%s: Is a directory\n",str);
22         break;
23
24    case ENOENT:
25         fprintf(stderr, "%s: No such file or directory\n",str);
26         break;
27
28    case ENOTDIR:
29         fprintf(stderr, "%s: Not a directory\n",str);
30         break;
31    case EPERM:
32         fprintf(stderr, "%s: Not a directory\n",str);
33         break;
34         break;
35    default:
36    default:
37         printf("%s: Error occurred: <\s\s\n",str,errno);
38         break;
39    }
40
41    return;
42 }</pre>
```

#### **HEAD**

head 커맨드의 옵션이 사용된 경우와 사용되지 않은 경우 파일명이 인자로 들어온 경우와 그렇지 않은 경우를

라인수를 나타내는 변수 otk 와 파일 포인터인 inf 함수로 구분하여 29번 라인에서 otk 만큼의 라인을 읽어들이고 표준 출력으로 출력합니다.

```
1 | Finclude "definition.h"
2 | finclude "definition.h"
2 | finclude "errhandler.h"
3 | int main(int argc_char * argv[])
5 | (
6 | set_default();
9 | int otk = 0;
9 | FILE * inf = NULL;
10 | if(argc > 2 && istrcmp(argv[1],"-n")) otk = atoi(argv[2]);
11 | if(ctk && argc == 4) || (totk && argc==2))
12 | if((otk && argc == 4) || (totk && argc==2))
13 | if(int) | if(int) | if(int) |
14 | (
15 | inf = fopen(argv[3-(t!!otk)*2],"r");
16 | iv(int) |
17 | (
18 | err_check("head");
19 | iv(int) |
19 | iv(int) | iv(int) |
19 | iv(int) |
10 | iv(int) |
11 | iv(int) | iv(int) |
12 | iv(int) | iv(int) |
13 | iv(int) | iv(int) |
14 | iv(int) | iv(int) |
15 | iv(int) | iv(int) |
16 | iv(int) | iv(int) |
17 | iv(int) | iv(int) |
18 | iv(int) | iv(int) |
19 | iv(int) | iv(int) |
10 | iv(int) | iv(int) |
11 | iv(int) | iv(int) |
12 | iv(int) | iv(int) |
13 | iv(int) | iv(int) |
14 | iv(int) | iv(int) |
15 | iv(int) | iv(int) |
16 | iv(int) | iv(int) |
17 | iv(int) | iv(int) |
18 | iv(int) | iv(int) |
19 | iv(int) | iv(int) |
10 | iv(int) | iv(int) |
11 | iv(int) | iv(int) |
12 | iv(int) | iv(int) |
13 | iv(int) | iv(int) |
14 | iv(int) | iv(int) |
15 | iv(int) | iv(int) |
16 | iv(int) | iv(int) |
17 | iv(int) | iv(int) |
18 | iv(int) | iv(int) |
19 | iv(int) | iv(int) |
10 | iv(int) | iv(int) |
11 | iv(int) | iv(int) |
12 | iv(int) | iv(int) |
13 | iv(int) | iv(int) |
14 | iv(int) | iv(int) |
15 | iv(int) | iv(int) |
16 | iv(int) | iv(int) |
17 | iv(int) | iv(int) |
18 | iv(int) | iv(int) |
19 | iv(int) | iv(int) |
10 | iv(int) | iv(int) |
10 | iv(int) | iv(int) |
11 | iv(int) | iv(int) |
12 | iv(int) | iv(int) |
13 | iv(int) | iv(int) |
14 | iv(int) | iv(int) |
15 | iv(int) | iv(int) |
16 | iv(int) | iv(int) |
17 | iv(int) | iv(int) |
18 | iv(int) | iv(int) |
19 | iv(int) | iv(int) |
10 | iv(int) | iv(int) |
11 | iv(int) | iv(int) | iv(int) |
12 | iv(int) | iv(int) | iv(int) |
13 | iv(int) | iv(int) | iv(int) |
14 | iv(int) | iv(int) | iv(int) |
15 | iv(int) | iv(int) | iv(int) | iv(int) |
16 | iv(int) | iv(int) | iv(int) |
```

#### **TAIL**

파일의 앞 부분은 head 파일과 동일합니다.

head는 앞에서 읽어들이는 대로 바로 출력하면 되는 반면 tail은 그렇게 간단히 구현되지 않았습니다,

가장 단순하게 tail 을 구현하는 방법은 파일을 라인대로 읽어 전부 메모리에 올린 후 (전체 라인 수 - k) 라인부터 전체 라인 수 까지 출력하는 방식입니다,

이 방식은 메모리 상 효율적이지 않으므로 출력할 파일의 라인 크기 K 에 해당하는 배열에 가장 끝 K 라인의 offset만 저장하는 식으로 최적화 하였습니다.

배열은 원형 큐 형식으로 가장 뒤의 k 라인의 offset만 항상 저장되고

마지막 포인팅하는 위치부터 배열을 전부 한번씩 출력하면 됩니다.

이때 fseek 함수를 이용하여 해당하는 offset으로 이동한 후 다시 라인을 읽어

출력합니다.

STDIN 의 경우 offset 을 사용할 수 없기 때문에 char \*\* 배열을 K만큼 동적할당하여 위와 유사한 방식으로 처리하였습니다.

#### CAT

CAT 의 경우 head 와 구현이 크게 다르지 않습니다. 파일인지 STDIN 인지 판단 후 22 라인에서 해당하는 입력을 받으면서 입력의 끝까지 계속해서 출력합니다.

#### CP

CP 커맨드의 경우 인자 형식이 맞지 않으면 우선 오류를 처리합니다.

파일의 복사는 입력과 출력을 위한 두개의 파일 디스크립터를 열어서

한 라인씩 읽고 출력하는 식으로 합니다.

파일 디스크립터를 열 때 파일이 존재하지 않거나 권한이 없을 수 있으므로 23번 라인과 같이 에러가 발생한 경우 에러 메세지를 출력하고 프로그램을 종료 합니다.

30번 라인부터 56번 라인까지는 cp 의 dest 부분이 파일의 경로일수도 있고 바꿔야 할 파일일 수도 있으므로 이를 처리하기 위해 dest 의 마지막 문자가 '/' 인 경우 src의 파일명을 붙여주고

마지막 문자가 '/' 가 아닌 경우 해당 문자열이 파일인지 디렉토리 인지 판단하기 위해 opendir 함수의 반환값을 확인하여 디렉토리라면 dest 에 '/' 와 파일이름을 붙여줍니다.

#### MV

mv 커맨드의 경우 파일 디스크립터를 지정하지 않고 rename 함수를 이용하여 cp 와 유사한 전처리 후 구현하였습니다.

#### RM

RM 의 경우 별다른 옵션이 없고 unlink 함수를 이용하면 파일을 삭제할 수 있어 간단히 구현하였습니다.

```
#include "definition.h"
2  #include "errhandler.h"
3
4  int main(int argc,char * argv[])
5  {
6    set_default();
7    if(argc == 2) remove(argv[1]);
8    err_check("rm");
9    return 0;
10 }
```

#### **PWD**

pwd 커맨드 또한 별다른 옵션 없이 getcwd 함수를 이용하여 배열에 현재 경로를 담은 후 출력하는 식으로 구현하였습니다,

```
1 #include "definition.h"
2 #include "errhandler.h"
3
4 char pwd[1024];
5
6 int main(void)
7 {
8    set_default();
9    getcwd(pwd,|sizeof(pwd));
10    printf("%s\n",pwd);
11    err_check("pwd");
12    return 0;
13 }
```

#### **MINISHELL**

설명에 앞서 C 언어의 특성 상 객체 단위로 사고하기가 힘들어 많은 함수들이 MINISHELL 에 치중된 경향이 있습니다.

지금 생각해보면 파일을 파싱하는 함수들은 따로 분리하는 것이 맞지 않았나 싶습니다, 그렇기에 MINISHELL 의 소스코드는 흐름대로 보기에 복잡한 측면이 있어

최대한 실행 흐름대로 따라가면서 주 흐름 외 사용되는 함수는 ADT만 간단히 설명하겠습니다.

#### **MAIN**

MINISHELL 함수를 구동하는 몸통입니다.

프로그램의 명령을 계속해서 받는 while 루프로 들어가기전 초기 세팅을 하게 됩니다.

- getcwd 함수를 이용하여 구현한 커맨드의 실행파일 위치 저 장
- getpgid 함수를 이용하여 미니쉘의 그룹id 를 mpgid 변수에 저장
- 미니쉘 함수는 SIGINT,SIGTSP를 무시하도록 설정
- SIGCHLD 시그널에 대해 reaping 하는 함수로 설정

이러한 초기 설정 후 while 루프를 돌면서 터미널로 부터 입력을 받고 eval 함수를 사용하여 명령어를 분석 , 실행하게 됩니다.

후에 서술하겠지만 eval 함수에서 명령어를 실행하게 된다면 미니 쉘의 프로세스가

명령어를 실행하는 자식프로세스를 생성하고 그룹을 분리한 뒤 그 뒤로 생성되는

자식프로세스는 모두 명령실행프로세스에 의해 생성되고 같은 그 룹으로 관리되기 때문에

killpg 함수를 이용하여 명령실행프로세스 그룹의 모든 구성원에게 SIGKILL 을 날려

놓치는 좀비프로세스가 발생하지 않도록 합니다.

#### **EVAL**

명령어를 평가하고 실행하는 함수입니다.

이 함수에 너무 많은 로직을 담은게 프로그램의 단점입니다.

우선 사용되는 변수들을 선언,초기화 한 뒤 90번 라인에서 기본적인 파싱을 거칩니다.

95번 라인에서 단축평가로

invalid\_checker 와 parsecommand 함수를 사용합니다.

기본적으로 invaild\_check 에서 커맨드가 유효한 커맨드 인지 검 사하고

유효하다면 0을 반환하여 parsecommand 함수가 실행되게 됩니다,

parsecommand 함수는 parseline 함수에서 1차적으로 파싱된 배열을

파이프 단위로 다시 토큰화 합니다.

이때 86번 라인 cvector vsz,vtp 배열에

토큰화된 커맨드의 인자 개수와 리다이렉션 종류가 담기게 됩니다. parsecommand 단계에서도 2차적으로 커맨드의 유효성을 검증하여

결과적으로 유효하지 않은 커맨드에 대해서는 97번 라인으로 유효한 커맨드의 경우 102 번 라인으로 가서 명령어를 수행하게 됩 니다.

102번 라인에 도달하는 시점에 커맨드는 vsz,vtp 배열에 담긴 정 보로

토큰화 되있다고 할 수 있습니다.

108라인에 백그라운드가 아니면서 cd,exit인 경우 실행 후 eval 함수를 종료하게 됩니다.

115번 라인부터 133번 라인까지 프로세스 관리의 핵심을 담고 있습니다.

미니쉘이 자식을 생성한 후 foreground 라면 명령이 전부 실행되고 종료되기 까지 기다리고

background 라면 main 함수 반복문으로 혼자 돌아가게 됩니다.

미니쉘에 생성된 자식프로세스를 편의상 명령프로세스라고 칭하겠습니다.

123라인에서 명령 프로세스는 곧바로 새로운 그룹을 자신의 id로 설정하게 됩니다.

이 때 백그라운드 옵션이 아니라면 tcsetpgrp 함수를 이용하여 세 션의 터미널과 통하는

그룹을 자신으로 만듭니다.(자신의 그룹을 포어그라운드 그룹으로 설정합니다,)

```
inf(int 1=8;3<0n;++1,++c)
{
    int saz = vsz.arr[1] - (vtp.arr[1]/3) * 2;
    int pc = c;

    char ** command = malloc(sizenf(char*) * (saz+1));

    char ** f1 = NULL;
    char * f2 = NULL;

    is char * f2 = NULL;

    is is char * f2 = NULL;

    is is char * f2 = NULL;

    is is char * f3 = argv[c++];

    is is char f4 = argv[c++];

    is char f4
```

SIGTSTP,SIGINT를 무시하는 미니쉘 프로세스와는 다르게 명령프로세스는 이 시그널들을 받아야합니다.

다만 SIGINT 의 경우 명령프로세스가 SIGINT 를 받아 종료되는 경 우 명령프로세스의 자식프로세스들은

좀비 프로세스가 되므로 이를 맊기 위해 또다른 handler를 SIGINT 에 설정합니다.

134번 라인은 background 이면서 built-in command 경우 아무런 동작도 하지않으므로 이를 맞춰주기 위해서

넣었습니다. 조금 억지스럽게 구현한 느낌이 있습니다.

152번 라인부터 251 라인까지 명령프로세스가 명령어를 실행하는 부분입니다.

for 문으로 파이프로 분리된 인자개수만큼 돌면서

144번 라인부터 165번 라인까지 parsecommand 함수에서 담은 정보를 가지고

실제적인 토큰화를 하게 됩니다.

이때 리다이렉션 뒤에 오는 파일이름은 차례대로 f1,f2 에 할당되게 되고

char \*\* command 에는 실행하는 커맨드 옵션 인자만 담기게 됩니다.

그 뒤로는 파이프라인과 리다이렉션에 대한 처리가 이루어 지게 되는데

명령프로세스가 커맨드마다 fork 하여서 자식은 실제적인 실행을 하고

명령프로세스는 자식의 종료를 기다리게 됩니다.

마지막 커맨드가 아니면 다음 파이프라인으로 dup2 를 사용하여 출력을 보내고

첫 커맨드가 아니면 dup2 로 전 파이프라인에서 입력을 받게 됩니다.

vtp 에 저장되어있는 커맨드의 리다이렉션 타입에 따라 파일 디스크립터를 생성 후 dup2 함수로 적절히 연결한 후

213 번 라인부터 238 번 라인까지 execv 함수를 이용해 명령을 실행 후 종료하게 됩니다.

마지막 명령 프로세스의 종료 시 명령프로세스가 포어그라운드 였다면 다시 미니쉘 프로세스를 포어그라운드 그룹으로 만들고 종료하게 됩니다.

### 결 론

이 리포트에서는 직접 배쉬쉘을 모방한 미니쉘 프로그램을 구현하고 전체적인 흐름과 특정한 문제들에 대해서 간단히 다루었습니다.

내부 구현에 대한 정확한 이해는 소스코드를 보고도 이해할 수 있게 끔 단순하게 짰다고 생각합니다.

과제를 구현하면서 시그널에 대한 이해와 프로세스에 대한 이해가 많이 향상된 것 같습니다. 다만 위에서 계속 언급하였듯이 모듈화 측면에서 아쉬움이 많이 남습니다. 객체지향언어를 사용한다면 더욱 더 깔끔하게 쉘을 구현할 수 있었지 않았을까 싶습니다.