

Programming Assignment #2

Due: October 2, 11:59:59 PM

1. Introduction

- 이번 과제를 통해 프로세스 생성 프로그래밍에 익숙해 지도록 한다.

2. Overview

- PA #1에서 작성한 `parse_url()` 함수를 발전시켜 입력 URL 형식이 `http://` 로 시작하면, 자식 프로세스(child process)로 `wget` 응용 프로그램(`/usr/bin/wget`)을 실행하여 주어진 URL 주소의 파일을 받아온 다음 표준 출력으로 내보내거나, 주어진 경로에 저장하도록 구현한다.

3. Specification

- `parse_url()` 함수 내부에서 `input_url`이 올바른 형식일 경우, `get_file()` 또는 `get_http()` 함수를 호출한다.

```
int get_file(char *input_url, char *output_path); // PA #1 재사용
int get_http(char *input_url, char *output_path); // 신규 구현
```

- 입력된 URL이 PA #1에서 정의한 **파일 프로토콜**을 따르면, `get_file()` 함수를 호출한다. `get_file()`의 기능은 PA #1에서 구현했던 `parse_url()` 함수의 기능과 정확히 일치한다.
 - ✓ PA #1의 내용 중 필요한 부분을 `get_file()` 함수로 분리하여 모듈화한다.
 - ✓ PA #1에서 주어졌던 제약 사항은(파일 포맷 정의, 에러 처리 방법 등)은 이번 과제에서도 동일하게 적용된다.
- 입력된 URL이 `http://` 로 시작하면 `get_http()` 함수를 호출한다.
- Arguments - `get_http()`
 - ✓ 첫 번째 인자 `input_url`은 `http://` 로 시작하는 **http 프로토콜**로 이루어진 일반적인 웹 주소이다. `input_url`에 위치한 파일을 `wget`을 이용해 받아온다.

- ◆ http 프로토콜만 처리한다. (예로 https:// 형태이면 제외)
- ✓ 두 번째 인자 `output_path`는 `input_url`에서 `wget`을 이용해 받은 데이터를 저장할 파일의 경로명이고, `NULL`이 들어올 수 있다.
 - ◆ `output_path`가 `NULL`로 주어지는 경우는 `${HOME_DIR}/temp.out` 파일로 저장하고, 해당 파일을 다시 읽어 표준 출력으로 내보낸다.
 - ◆ `output_path`가 `NULL`이 아닌 값으로 주어진 경우, 해당 위치에 기존 파일이 존재하지 않는다. 따라서 이미 존재하는 파일을 다루는 방법을 고려할 필요가 없다. 주어진 그대로 `wget`으로 넘긴다.
- 자식 프로세스의 생성
 - ✓ `get_http()` 함수 내에서 `fork()` 함수를 호출하여 자식 프로세스를 생성하고, 자식 프로세스 안에서 `exec*()` 계열의 함수를 이용하여 `/usr/bin/wget`을 실행하여 `input_url`에 위치한 파일을 `output_path`에 저장한다.
- 결과의 반환
 - ✓ `get_http()` 함수를 통해 파일을 저장한 경우 해당 파일의 크기를 반환한다.
 - ✓ 파일 크기를 정확하게 반환하기 위해서는 자식 프로세스 `wget` 실행이 끝난 뒤에 측정하여야 한다.
 - ◆ 부모 프로세스에서 `wait*()` 계열 함수를 사용하여 자식 프로세스가 종료된 이후에 측정하도록 한다.
 - ✓ `parse_url()` 함수의 리턴 값은 `get_http()` / `get_file()` 함수의 리턴 값과 같도록 한다.
 - ✓ 에러가 발생한 경우, 아래 절에 명시한 방법으로 처리한다. 에러의 조건과 처리 방법은 다음 절에서 설명한다.
- 에러 처리
 - ✓ 프로그램 수행 중 예외 상황이 발생하면, `parse_url.h`에 정의된 다음의 `define` 값을 리턴한다.

```

#define ERR_NO_FILE      (-1)

#define ERR_NOT_ALLOWED  (-2)

#define ERR_INVALID_URL  (-3)

#define ERR_ACCESS_FAIL  (-4)
```

- ◆ 파일이 존재하지 않는 경우
 - ERR_NO_FILE
- ◆ 파일이 존재하나, 이용할 권한이 없을 때
 - ERR_NOT_ALLOWED
- ◆ 파일/http 프로토콜 형식이 틀렸거나, 다른 프로토콜로 URL을 요청할 경우
 - ERR_INVALID_URL
- ◆ wget이 주어진 URL로부터 데이터를 받아서 저장하지 못한 경우
 - ERR_ACCESS_FAIL

4. Background and Additional Information

- 리눅스는 Man-page (manual-page)를 통해 상세한 메뉴얼을 제공한다.
- 잘 모르는 사항은 TA에게 질문하거나, 구글을 통한 검색을 권장한다.
- wget 응용 프로그램
 - ✓ wget은 주어진 URL로부터 데이터를 받아 저장하는 프로그램이다. http, https, ftp 등의 프로토콜을 사용할 수 있다.
 - ✓ wget이 설치되어 있지 않은 경우에는 “# apt-get install wget” 명령어를 통해 wget을 설치할 수 있다.
 - ✓ 과제 수행에 필요한 중요 옵션
 - ◆ -O, --output-document : 문서를 파일로 저장한다. (대문자임에 주의)
 - ◆ -q, --quiet : wget 수행 중 나타나는 메시지를 출력하지 않는다. 채점 서버는 프로그램 수행시 표준 출력으로 나온 내용을 비교하여 평가하기 때문에, 과제 수행 시 이 옵션을 사용해야 한다.
 - ✓ 기타 자세한 사용 방법
 - ◆ “\$ wget --help”, “\$ man wget” 을 참고한다.
 - ◆ http://www.gnu.org/software/wget/manual/html_node/index.html
 - ✓ “\$ wget -O pa1.pdf http://cs1.skku.edu/uploads/SSE2030F13/1-PA1-130911.pdf” 명령을 내리면, cs1 서버에 있는 PA #1 문서를 현재 디렉토리에 pa1.pdf라는 이름으로 저장한다.
 - ✓ wget 프로그램은 정상적인 종료 시 0을, 에러 상황에선 그 외의 값을 리턴한다.

- 프로세스 관련 시스템 콜 함수의 사용 방법은 강의 교안을 참고한다.
 - ✓ 특히, 자식 프로세스의 리턴값을 구해오는 방법을 참고한다.

5. Restriction

- 과제는 본인이 직접 설치한 리눅스 환경에서 구현한다.
 - ✓ 테스트 서버에서 컴파일, 실행시킬 수 없는 코드는 과제 제출물로 인정하지 않는다.
 - ✓ 테스트 서버는 Linux kernel 3.0-32bit, gcc 4.6.1 를 이용한다.
- 구현을 완료하면 <http://sys.skku.edu>에 과제를 제출하고, 수행하여 결과를 확인한다.
- 과제 점수는 컴파일/실행 가능 여부, 작성한 함수의 완성도, 출력 결과 및 작성한 문서에 의하여 평가된다.
- 시스템 콜을 사용하여 프로그램을 작성하며, Standard C library의 malloc(), free(), getenv() 이외 함수는 사용하지 않는다.
- 파일 입출력의 경우 open(), read(), write(), close() 시스템 콜 함수 중 필요한 것을 선택하여 사용한다.
- 프로세스 생성의 경우 fork(), exec*() 계열, wait*() 계열 POSIX 함수와 관련된 매크로 함수를 선택하여 사용한다.
- 추가적으로 stat(), fstat() 시스템 콜 함수를 사용할 수 있다.
- 필요한 경우 함수를 직접 parse_url.c 안에 구현하여 추가할 수 있다.
- parse_url.c 이외의 파일은 수정할 수 없고, 설사 수정해서 제출하더라도 채점 서버에서 자동으로 제외하여 컴파일한다.

6. Skeleton Codes

- 본 과제 수행을 위하여 아래와 같이 5개의 파일이 주어진다.
 - ✓ Makefile GNU make utility를 위해 사용되는 파일.
 - ✓ main.c 구현한 함수를 호출하고 에러 발생시 문자열을 출력하는 main()을 포함한 파일
 - ✓ parse_url.h parse_url() 함수 구현을 위해 필요한 헤더 정보
 - ✓ parse_url.c parse_url(), get_file(), get_http() 함수를 구현할 파일

7. Hand in instruction

- 작성한 프로그램 코드 상단에 이름과 학번을 주석으로 표기한다.

- 작성한 과제 코드는 "학번.tar.gz" 형태로 압축해 <http://sys.skku.edu>에 제출한다.
 - ✓ 작업 디렉토리에서 "\$ make tar 학번" 명령을 이용해 파일을 "학번.tar.gz" 으
로 압축할 수 있다.
- 작성한 과제 코드의 디자인과 별도로 구현에 관한 내용을 담은 보고서를 "학번.pdf"
파일로 채점 서버에 제출한다.

8. Logistics

- 과제 제출 결과는 <http://csl.skku.edu/SSE2030F13/Assignments> 에서 확인할 수 있다.
- 과제 제출 기한은 <http://sys.skku.edu> 서버시간을 기준으로 하며, 기한 이후 24시간
내에 제출할 경우 30%, 48시간 내에 제출할 경우 60% 감점한다. 그 이후 제출이 불
가능하며, 0점 처리한다.
- 과제에 대한 의논은 함께 할 수 있으나, 프로그램 소스코드 작성은 스스로 해야 한
다.
 - ✓ 다른 사람의 과제를 copy 한 경우, 두 사람 모두 0점 처리하며 **학점상의 불이익
이 있다.** 인터넷 등에서 찾은 소스 코드를 그대로 copy 한 경우에도 0점처리한
다. **두 번 이상 적발되면 F 학점**을 받는다.