

操作系统实验一：可变分区存储管理

姓名: 吴嘉锐

班级: F1803601

学号: 518021910082

实验题目

编写一个C程序，用char *malloc(unsigned size)函数向系统申请一次内存空间（如size=1000，单位为字节），用**循环首次适应法** addr = (char *)malloc(unsigned size)和free(unsigned size, char * addr)模拟UNIX可变分区内存管理，实现对该内存区的分配和释放管理。

实验目的

- 加深对可变分区的存储管理的理解；
- 提高用C语言编制大系统程序的能力，特别是掌握C语言编程的难点：指针和指针作为函数参数；
- 掌握用指针实现链表和在链表上的基本操作。

算法思想

可变分区存储

可变分区存储管理法并不预先将内存划分成分区，而是等到作业运行需要内存时就向系统申请，从空闲的内存区中“挖”一块出来，其大小等于作业所需内存大小，这样就不会产生“内存头”。

空闲表.png
Figure1: 空闲表
管理空闲内存区的数据结构可采用链接法和连续性表格法。本实验采用循环双链表。

循环首次适应法

- 把空闲表设计成顺序结构或双向链接结构的循环队列，各空闲区仍按地址从低到高的次序登记在空闲区的管理队列中。需要设置一个起始查找指针，指向循环队列中的一个空闲区开始查找。
- 循环首次适应法分配时总是从起始查找指针所指的头项开始查找。第一次找到满足要求的空闲区时，就分配所需大小的空闲区，修改表项，并调整起始查找指针，使其指向队列中被分配的后面的那块空闲区。下次分配时就从新指向的那块空闲区开始查找。
- 循环首次适应法的实质是起始查找指针所指的空闲区和其后的空闲区群需为较长时间未被分割过的空闲区，它们已合并成较大的空闲区可能性较大。比起首次适应法来，在没有增加多少代价的情况下却明显地提高了分配查找的速度。
- 释放算法基本同首次适应法一样。
- 首次适应法和循环首次适应法不仅可用于内存的分配和释放，也可用于进程图像交换的辅存空间的分配和释放，其管理空闲区的数据结构也相同。

设计概要

在本实验中，程序会首先初始化分配一定的内存地址空间，接着通过终端接收用户的输入信息进行地址分配并显示处理结果，直到用户输入退出程序指令后程序会回收分配的所有空间并结束程序。下为程序的总体设计图，一些细节没有显示（如错误处理等）。

设计概要.png
Figure 2: 程序设计摘要

数据结构及重要变量

本实验采用**双向循环链表**实现存储管理中空闲表的功能：
typedef struct map
{
 unsigned m_size; // 空闲区大小
 char *m_addr; // 空闲区起始地址
 struct map *next, *prior; // 下一个&上一个空闲区
} map_m;
...

由于采用首次适应法的思想，设置了一个**指向当前空闲区的全局变量**cur_map;

当然，**初始分配内存空间**也需要保存为一个变量buf;

除此之外，程序还设计了一个宏SIZE以保存**初始化空间的大小**

模块设计说明

此部分将自顶向下逐一阐述实验中设计的不同模块的功能、接口以及实现方法。

程序初始化部分

空闲表初始化

该模块通过函数init_map实现，会初始分配大小为SIZE的内存空间，生成第一个空闲表并返回其地址。该空闲表表的起始地址m_addr为分配内存空间的起始地址buf，大小size为SIZE，前序指针与后续指针均指向自己。

```
...
// initialize the addr space and the current map
map_m * init_map() {
    buf = malloc(sizeof(char)*SIZE); // 分配空间
    map_m * object = (map_m *) malloc(sizeof(map_m));
    // 初始化第一个空闲表块
    object->m_addr = buf;
    object->m_size = SIZE;
    object->next = object;
    object->prior = object;
    printf("map initialized, m_addr = %lu, size = %u\n",
(unsigned long) object->m_addr - buf, object->m_size);
// 显示初始化地址、大小
return object; // 返回空闲表指针给主函数，主函数会将其存于 cur_map 中
}
...
```

输入输出部分

一切输入输出的地址均当作偏移地址处理。

用户输入输出交互

该模块直接内置于主函数中。用户输入的指令会被 fgets 获取并将指令字符串保存在 cin 中，然后将其作为参数传递给判断函数 judge_str 进行指令类型的判断，分为辅助、分配、释放、退出和非法指令五种，然后根据指令类型作进一步处理。

```
...c
int Judge; // 用于判断用户输入的指令类型

大小
unsigned size; // 保存用户输入分配或释放内存空间的大小
unsigned long addr; // 保存用户输入释放内存的起始地址

char cin[100]; // 保存用户输入的指令字符串
char command[6]; // 保存用户输入的命令 (malloc、free)，仅用于占位
do
{
    printf("\n\n // 提示用户开始输入\n\nPlease enter a command(You can try\n\n");
    type ^help^");
    fgets(cin, 100, stdin);
    printf("\n");
    Judge = judge_str(cin);

    // 根据不同输入类型进行处理
    switch (Judge)
    {
        case 0: // help
            printf("malloc: m[alloc] size;\nfree: f[ree] size addr;\nexit: quit the program\n");
            break;
        case 1: // quit
            scanf(cin, "%s %u", command, &size);
            Lmalloc(size);
            break;
        case 2: // free
            scanf(cin, "%s %u %lu", command, &size, &addr);
            if (IfFreeIllegal(size, (char *) (addr + (unsigned long) buf))) == 0 {
                printf("illegal free command!\n");
                ShowMapState();
            }
            else Lfree(size, (char *) (addr + (unsigned long) buf));
            break;
        case 3: // illegal command
            printf("illegal command\n");
            break;
        default:
            break;
    }
    while (Judge != 4);
}
...
```

指令类型判断

该模块通过函数 judge_str 完成，功能在于通过返回一个 int 类型判断用户输入指令的类型。

```
0:help      1:malloc      2:free
            3:illegal     4:exit

...c
/*
0:help
1:malloc
2:free
3:illegal
4:exit

*/
int Judge_str (char *str) {
    if ( (strcmp(str, "help", 4) == 0) && (str[4] == '\n')) return 0;
    if (strcmp(str, "m", 2) == 0) {
        int i = 2;
        if (str[i] == '\n') return 3;
        while (str[i] != '\n')
        {
            if (!isdigit(str[i])) return 3;
            i++;
        }
        return 1;
    }
    if (strcmp(str, "malloc", 7) == 0) {
        int i = 7;
        if (str[i] == '\n') return 3;
        while (str[i] != '\n')
        {
            if (!isdigit(str[i])) return 3;
            i++;
        }
        return 1;
    }
    if (strcmp(str, "f", 2) == 0) {
        int i = 2;
        if (str[i] == '\n') return 3;
        while (str[i] != ' ')
        {
            if (!isdigit(str[i])) return 3;
            i++;
        }
        i++;
        if (str[i] == '\n') return 3;
        while (str[i] != '\n')
        {
            if (!isdigit(str[i])) return 3;
            i++;
        }
        return 2;
    }
    if (strcmp(str, "free", 5) == 0) {
        int i = 5;
        if (str[i] == '\n') return 3;
        while (str[i] != ' ')
        {
            if (!isdigit(str[i])) return 3;
            i++;
        }
        if (str[i] == '\n') return 3;
        while (str[i] != '\n')
        {
            if (!isdigit(str[i])) return 3;
            i++;
        }
        return 2;
    }
    if ((strcmp(str, "exit", 4) == 0) && (str[4] == '\n')) return 4;
    return 3;
}
...
```

显示空闲表状态

该模块用于从起始查找指针开始打印当前空闲表的信息。

```
...c
// show map
void ShowMapState() {
    printf("\ncurrent maps:\n");
    if (cur_map == NULL) {
        printf("All adrs have been malloced!\n");
        return;
    }
    map_m * ser_map = cur_map;
    do
    {
        printf("map: m_size = %u, m_addr = %lu\n",
            ser_map->m_size, (unsigned long) (ser_map->m_addr - buf));
        ser_map = ser_map->next;
    } while (ser_map != cur_map);
}
```

```
return;
}
...

// 空闲表处理部分

地址空间分配模块

该模块用函数 Lmalloc 实现，参数为所需分配空间大小 size，函数从起始查找指针开始寻找大小大于等于 size 的空闲表块进行地址分配。若没有符合条件的地址块则输出分配失败信息。

...c
// malloc
void Lmalloc(unsigned size) {
    map_m * ser_map = cur_map;
    if (cur_map == NULL) {
        printf("malloc failed!\n");
        ShowMapState();
        return;
    }
    do
    {
        if (ser_map->m_size > size)
        {
            cur_map = ser_map->next;
            ser_map->m_addr += sizeof(char) * size;
            ser_map->m_size -= size;
            printf("applied for size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                size, (unsigned long) (ser_map->m_addr - sizeof(char) * size - buf), (unsigned long) (ser_map->m_addr - 1 - buf));
            ShowMapState();
            return;
        }
        else if (ser_map->m_size == size)
        {
            printf("applied for size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                size, (unsigned long) (ser_map->m_addr - buf), (unsigned long) (ser_map->m_addr + sizeof(char) * size - 1 - buf));
            if (ser_map->next == ser_map) cur_map = NULL;
            else cur_map = ser_map->next;
            ser_map->prior->next = ser_map->next;
            ser_map->next->prior = ser_map->prior;
            free(ser_map);
            ShowMapState();
            return;
        }
        else {
            ser_map = ser_map->next;
        }
    } while (ser_map != cur_map);
    printf("malloc failed!\n");
    ShowMapState();
    return;
}
...
```

地址空间释放模块

释放模块由函数 Lfree 实现，参数为释放空间的大小、释放空间的起始地址。由于采用了双向链表，除了教材的四种分配情况，这里实现还增加了释放地址小于空闲表最小地址与大于空闲表最大地址、空闲表为空共五种情况讨论。

```
...c
// free
void Lfree(unsigned size, char * addr) {
    map_m * ser_map = cur_map;
    map_m * tmp_map;
    map_m * new_map;
    map_m * to be freed;

    // 空闲表为空
    if (cur_map == NULL)
    {
        cur_map = (map_m*) malloc(sizeof(map_m));
        cur_map->m_addr = addr;
        cur_map->m_size = size;
        cur_map->next = cur_map;
        cur_map->prior = cur_map;
        printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
            size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
        ShowMapState();
        return;
    }
    if (addr < (tmp_map = min_addr()->m_addr)
    {
        if (addr + size - 1 == tmp_map->m_addr - 1)
        {
            tmp_map->m_size = addr;
            printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
            ShowMapState();
            return;
        }
        else
        {
            new_map = (map_m*) malloc(sizeof(map_m));
            new_map->m_addr = addr;
            new_map->m_size = size;
            new_map->next = tmp_map;
            new_map->prior = tmp_map->prior;
            new_map->prior->next = new_map;
            tmp_map->prior = new_map;
            printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
            ShowMapState();
            return;
        }
    }
    // 分配地址大于最大地址
    if (addr > (tmp_map = max_addr()->m_addr)
    {
        if (tmp_map->m_addr + tmp_map->m_size - 1 == addr - 1)
        {
            tmp_map->m_size += size;
            printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
            ShowMapState();
            return;
        }
        else
        {
            new_map = (map_m*) malloc(sizeof(map_m));
            new_map->m_addr = addr;
            new_map->m_size = size;
            new_map->prior = tmp_map;
            new_map->next = tmp_map->next;
            tmp_map->next->prior = new_map;
            printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
            ShowMapState();
            return;
        }
    }
    // 寻找释放地址处于哪两个空闲表块之间
    for (tmp_map = cur_map; ! (tmp_map->m_addr < addr && addr < tmp_map->next->m_addr); tmp_map = tmp_map->next);
}
```

```
if (tmp_map->m_addr + tmp_map->m_size - 1 == addr - 1)
// case a, b
{
    if (addr + size - 1 < tmp_map->next->m_addr - 1)
    // case a
    {
        tmp_map->m_size += size;
        printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
            size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
        ShowMapState();
        return;
    }
    else // case b
    {
        tmp_map->m_size = tmp_map->m_size + size + tmp_map->next->m_size;
        if (cur_map = tmp_map->next)
        {
            if (cur_map = tmp_map->next)
            {
                to be freed = tmp_map->next;
                tmp_map->next = tmp_map->next->next;
                free(to be freed);
                printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                    size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
                ShowMapState();
                return;
            }
            else // case c, d
            {
                if (addr + size - 1 == tmp_map->next->m_addr - 1)
                // case c
                {
                    tmp_map->next->m_addr = addr;
                    printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                        size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
                    ShowMapState();
                    return;
                }
                else // case d
                {
                    new_map = (map_m*) malloc(sizeof(map_m));
                    new_map->m_addr = addr;
                    new_map->m_size = size;
                    new_map->prior = tmp_map;
                    new_map->next = tmp_map->next;
                    tmp_map->next->prior = new_map;
                    printf("freed size: %u, \nfrom addr: %lu, \n to addr: %lu\n",
                        size, (unsigned long) (addr - buf), (unsigned long) (addr + size - 1 - buf));
                    ShowMapState();
                    return;
                }
            }
        }
    }
    printf("free address failed");
    ShowMapState();
    return;
}
...
```

释放地址合法性判断

在执行地址释放之前，需要判断地址释放的合法性，即释放的空间必须处于初始分配的空间内，并且不能涵盖空闲区。参数为释放空间大小以及释放的起始地址。

```
...c
// 0: illegal 1: legal
int IfFreeIllegal(unsigned size, char * addr) {
    if (size <= 0) return 0;

    if (addr < buf || addr + size - 1 > buf + SIZE - 1) return 0;

    // map is empty
    if (cur_map == NULL) return 1;

    map_m * ser_map = cur_map;
    // addr cannot be in a map's addr area
    do
    {
        if (addr > ser_map->m_addr && addr <= ser_map->m_addr + ser_map->m_size - 1) return 0;
        ser_map = ser_map->next;
    } while (ser_map != cur_map);

    // a map's m_addr cannot be in addr and addr + size do
    if (ser_map->m_addr >= addr && ser_map->m_addr <= addr + size - 1) return 0;
    ser_map = ser_map->next;
    } while (ser_map != cur_map);

    return 1;
}
...
```

退出程序部分

当检测到用户输入 exit 指令时，程序将进入退出部分。

释放空间

主函数中会输出程序关闭提示，接着释放地址空间以及全部空闲表。

```
...c
printf("program closed\n");
free(buf);
return 0; free_all();

// 释放全部空闲表实现:
...c
void free_all() {
    map_m * ser_map = cur_map;
    map_m * to be freed;
    int num = getmapnum();
    for (int i = 0; i < num; i++)
    {
        to be freed = ser_map;
        ser_map = ser_map->next;
        free(to be freed);
    }
    return;
}
...
```

测试

测试方法

由于用户输入输出均采用了**偏移地址**（实验中**分配了1000单位地址，偏移范围为0~999**），测试相对简单。

测试会使用“./a.out </test_in/xxx.txt >/test_out/xxx.txt”命令直接生成输出文件，再对比是否符合预期。

测试将按照“循环首次”特性释放的9种情况以及部分错误处理分别进行。

为了便于管理输入和输出的测试文件，使用一个**linux shell 程序**通过读取所有测试输入自动生成输出文件，shell 程序如下：

