# 1、概述

本文档主要描述搜索程序——search和前台页面——portal的交互流程

二者的通信方式为套接字通信，降低二者的耦合度。降低程序的复杂度。

search端口设置为：19999

# 2、交互协议

交互的报文由报头和报文数据组成。

~~报头由5个int(4个字节)组成，含义依次为：~~

~~报文命令、报文数据长度、参数1、参数2、参数3~~

~~参数置0表示参数并未被赋予含义。~~

起始的时候，没有考虑到php中内存操作的API，刚才我查了一下，发现php内存操作可能有点困难。所以，放弃该报文格式，改为下面：

报头由5个字段组成，含义依次为：

报文命令、报文数据长度、参数1、参数2、参数3

各个字段之间由字符’\n’隔开，空字段表示该字段没有被赋予含义。

最后说明一下报文数据的格式，每个命令可以定义自己的报文数据格式。

# 3、报文命令定义

PORTAL\_SEARCH\_QUERY 1

//这是PORTAL 向 SEARCH 发出的 QUERY 命令

报头定义为：PORTAL\_SEARCH\_QUERY | 报文数据长度 |

参数1 | 参数2 | 参数3 （所有参数都为0）

后接报文数据

PORTAL\_SEARCH\_QUERY\_OK 2

//这是SEARCH 回复 PORTAL的 命令，表示查询结束，给出查询结果

报头定义为：PORTAL\_SEARCH\_QUERY\_OK | 报文数据长度 |

参数1（表示一共有多少个查询结果） | 参数2（本报文是第几个查询结果）

| 参数3 （）

后接报文数据

查询结果数据的组织方式：

title\nurl\n网页摘要\n

由此可以看到此命令中的报文数据长度字段可以忽略了，没有什么意义。因为报文数据是以\n来分隔的。

PORTAL\_SEARCH\_QUERY\_ARGUMENT\_ILLEGAL 3

//这是SEARCH 回复 PORTAL的 命令，表示查询结束，查询的关键词非法

//这里的非法包括：目前搜索引擎仅仅支持纯中文的搜索（由于UTF-8是变长编码，

所以为了简化分词程序设计，仅仅搜索中文）所以如果要查询纯英文，那么

关键词分词之后就是一个空串。不能搜索。

报头定义为：PORTAL\_SEARCH\_QUERY\_ARGUMENT\_ILLEGAL | 0 |

参数1 | 参数2 | 参数3 （所有参数都为0）

PORTAL\_SEARCH\_QUERY\_OTHER\_ERROR 4

//这是SEARCH 回复 PORTAL的 命令，表示查询结束，表示其他的未知的错误

报头定义为：PORTAL\_SEARCH\_QUERY\_OTHER\_ERROR | 0 |

参数1 | 参数2 | 参数3 （所有参数都为0）

# 4、常用const常量定义

CAN\_NOT\_CONNECT\_SEARCH 1

SEARCH\_LISTEN\_PORT 19999

# 5、Notes

①：网络交互过程中记得字节序的问题，发送前转换为网络字节序，接收到

后转为主机字节序。由于改为字节串传输方式，所以这个Note可以删除了。