In [19]:
```
import pandas as pd
from wordcloud import WordCloud
import matplotlib.pyplot as plt
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize


df1 = pd.read_csv('考研weibo.csv', encoding='utf-8')
# df1 = pd.read_csv('计算机专业考研weibo.csv', encoding='utf-8')
df1 = df1.fillna('') # 将NaN值替换成空字符串


df2 = pd.read_csv('考研zhihu.csv',encoding='utf-8')
# df2 = pd.read_csv('计算机专业考研zhihu.csv',encoding='utf-8')
df2 = df2.fillna('') # 将NaN值替换成空字符串
#
text = ' '.join(df1['content'].tolist() + df1['topic'].tolist()+df2['title'].tolist()+d
# wc = WordCloud(font_path = r'./MSYH.TTC')
# wc.generate(text)

# plt.imshow(wc)
# plt.axis("off") # 不显示坐标轴
# plt.show()



# 分词并去除停用词
nltk.download('stopwords')
nltk.download('punkt')
stop_words = set(stopwords.words('chinese'))
tokens = word_tokenize(text.lower())
keywords = [word for word in tokens if word.isalpha() and word not in stop_words]

# 生成词云
wordcloud = WordCloud(font_path = r'MSYH.TTC',width=800, height=800, background_color='

# 显示词云
plt.figure(figsize=(8, 8), facecolor=None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad=0)
plt.show()
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```


法硕
everyday
日精进
mti初试 mti
众合法硕教育
热词背诵 翻硕热词
考虫萌酱
英语
孔玮vicky
翻硕热词 mti初试
我在
觅游带你考新传 新传
宝宝的少女心
米奇比心
胖鱼成长记
我免费围观了
扇贝
教育学 司南
高途
成功
和拖鞋哥一起学新闻传播 新传
上海对外经贸大学
新传 新闻传播学
kelly雯的日记
studyaccount
教师资格证
旭东翻硕
米奇大哭 米奇大哭
六一儿童节
回乡见闻 回乡见闻
谭剑波英语全程陪伴
觅游新传教育
华图法硕
政治徐涛
新东方王江涛 刘文涛老师和
刺猬大学成长计划
胖鱼成长记 有道考神
英语大王思思老师 一笑而过
小仙女 小仙女
今天我在
倒计时
决战
studyonline
碎碎念
百日行动派周思成
新传小小班
教育
二十遍之约
北京建筑大学
第一站
四六级
考虫萌酱 石雷鹏老师
考虫萌酱 王琢leo
越努力越优秀
就要语众不同
闲鱼也要有梦
翻译硕士 mti
汉硕 mti 水清木华翻硕
mti 桃子姐翻译硕士mti
新东方王江涛
刑法手越
就这一刻
竞赛凯哥
宇哥
托福备考干货
果汁成长日记
长孙博
冒个泡
复试
奋斗
成绩
规划
一万词汇二十遍之约
十年对比挑战
周思成 英语大王思思老师
带着边框一起拍
朱一龙齐衡
择校
今日贴纸打卡
新浪看点
焦虑是个好东西
民法杨烁
教师资格证网课
每日打卡
北大新传
数学周洋鑫
水清木华翻硕
k的打卡日常
华东政法大学
studyaccount分享平台
文运法硕 法律硕士
民法岳业鹏
新标日
觅游新传 觅游带你考新传
琦妈每日一句
励志语录
有道考神培训
墨墨背
上午
交互设计
法学理论杜洪波
法硕超话
法硕 文运法硕
米奇比心
上岸
恋上冬日
追星要趁早
基金从业资格考试
和鹿晗过新年
studyaccount梦想家
翻译硕士网
政治
cpa
studywithme
一笑而过
有道考神
人挺住
光学工程
宋智鸣
刑法车润海
法硕 华图法硕
司南 教育学
考研人数直逼
微博抽奖平台
百日行动派周思成
一笑而过数学周洋鑫
one字幕组
缇伊的学习打卡
人
二战
studyaccount分享平台 studyonline

In [20]:
```
import pandas as pd
from wordcloud import WordCloud
import matplotlib.pyplot as plt
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize


# df1 = pd.read_csv('考研weibo.csv', encoding='utf-8')
df1 = pd.read_csv('计算机专业考研weibo.csv', encoding='utf-8')
df1 = df1.fillna('') # 将NaN值替换成空字符串


# df2 = pd.read_csv('考研zhihu.csv',encoding='utf-8')
df2 = pd.read_csv('计算机专业考研zhihu.csv',encoding='utf-8')
df2 = df2.fillna('') # 将NaN值替换成空字符串
#
text = ' '.join(df1['content'].tolist() + df1['topic'].tolist()+df2['title'].tolist()+d
# wc = WordCloud(font_path = r'./MSYH.TTC')
# wc.generate(text)

# plt.imshow(wc)
# plt.axis("off") # 不显示坐标轴
# plt.show()



# 分词并去除停用词
nltk.download('stopwords')
nltk.download('punkt')
stop_words = set(stopwords.words('chinese'))
tokens = word_tokenize(text.lower())
keywords = [word for word in tokens if word.isalpha() and word not in stop_words]

# 生成词云
wordcloud = WordCloud(font_path = r'MSYH.TTC',width=800, height=800, background_color='

# 显示词云
plt.figure(figsize=(8, 8), facecolor=None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad=0)
plt.show()
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```


考研
考研政治英语
管理类联考数学
管理类联考数学 考研政治英语
计算机考研
抓码计算机考研
米奇比心
考研英语 考研法硕
考研英语
新年心愿 新年心愿
计算机二级
初级会计
教师资格证面试
教师资格证
山东师范大学 计算机技术
注册会计师考试
考研择校 山东师范大学
初级会计 考虫专四专八寒假班
有道考神考研 逢考必过
天道考研
隐藏消息的bug修好了
熙宝儿认真的生活
微信公众号
susie的实习日记
计算机
教师资格考试
计算机网络
河南城建学院团委
山东建筑大学
艺术升崩溃
dikw大讲堂
胜利
中级会计
工科到底该选什么专业呢
我的近期目标
操作系统
英语四六级
大学生
初级会计职称
软考
考研资料鱼
建筑
计算机复试
无明天安排
点对点教育
普通话考试
南师大考研
药 考试季
专业硕士 专业硕士
晒发际线大赛
春晚
普通话
专四专八
税务师考试
点对点计算机学院
考研结束下届学生狂奔占座
初级会计成绩 注册会计师考试
何凯文刘一男
省考
省考笔试
哔哩哔哩动画
十年
论文发表推荐期刊 双向暗恋太甜了
四六级
王道计算机考研
如何扛过最缺钱的日子
湖南大学计算机
噢耶
计算机组成原理
控制科学与工程
everyday日精进
新年愿望单
大学生英语竞赛
中医执业医师
何凯文考研英语每日一句
北大cat
一级学科
雅思
税务师
微博联考
专四
我的专业日常
新春旺好运
考研专业课
考虫专四专八寒假班 新东方专四专八
浙点对点
评职称
抓码每日一题 计算机专业考研见上
it研究所
计算机考研复试项目经历
考研择校
考研政治
考研复试英语资历优化
计算机专业综合
考研专业
教师资格证成绩
你说我说海院
教育学
点对点考研
考研上岸
初级会计教材
期刊
保研推免
宝宝的少女心
抓码每日一题
疤痕效应
考研数学
亚投行 亚投行
公安
一级建造师
考研政治经验分享
非法学
数据结构
决战考研
课辞 课辞
吉祥三保
专八
二级建造师
考研复试
何凯文
论文
考研资料酱
公务员
河北工程大学
c语言版
硬核导师催你交论文
刘恺威杨幂离婚
王道课程讲解
证券从业资格考试
小菲菲的碎碎念
考研励志语录
北大计算机辅助翻译
大厂最尬男孩
初级会计成绩
考研公共课
汤家凤 何凯文
三月de日常
mpacc
西南大学考研复试流程
考研网课
点对点计算机
我在
北大cat考研
愉快
国考
年终总结
二战mpacc
农
医学生的日常
我免费围观了
图片 图片
haek
微博问答
考研打卡
科技潮生活

In [ ]: