*Confidential（绝密）*

携程旅行网

**纠错词词库搭建及零少结果词处理**

产品需求文档

版本号: 0.1

旅游事业部：王晓英

2017年7月10日

All Rights Reserved

# **目录**

[目录 2](#_Toc489984207)

[0. 版本历史 3](#_Toc489984208)

[1. 需求背景 3](#_Toc489984209)

[2. ROI预估 3](#_Toc489984210)

[4. 适用范围 3](#_Toc489984211)

[4需求描述 4](#_Toc489984212)

[4.1 纠错词 4](#_Toc489984213)

[4.2 零少结果词的丢词分析 5](#_Toc489984214)

[4.3用户搜索关键词处理 5](#_Toc489984215)

[5 需求负责人 6](#_Toc489984216)

# 版本历史

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **修订日期** | **修订内容** | **修订人** | **修订版本号** |
| 1 | 7月10日 | PRD草稿 | 王晓英 | 0.1 |

# 需求背景

丰富纠错词库数据，降低零少结果率，提升用户体验

# ROI预估

|  |  |  |
| --- | --- | --- |
|  | | 跟团游 |
| 关键词 | TotalUV | 1241605 |
| 订单占比 | 8.47% |
| 搜索词到订单转化率 | 0.08%（预估提升10%） |

1. ROI = log（81\*792\*12+250\*52）=5.89

# 适用范围

**页面范围：**聚合Tab、出发地参团Tab、目的地参团Tab、自由行Tab、自由行列表页、跟团游列表页，周边游跟团游，周边游自由行

**产品范围：**

新产品形态=‘跟团游’、‘半自助游’、‘私家团’

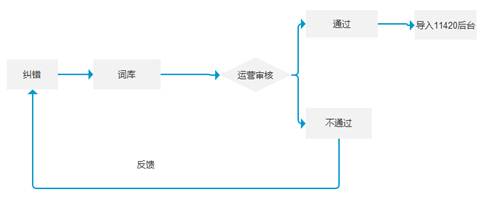
新产品类型=‘出境旅游’、‘国内旅游’、‘境外N日游’、‘境内N日游’、‘邮轮’、‘国内游学’、‘出境游学’、‘国内微游学’、‘出境微游学’。及产品类型=‘签证’的上线有效产品。

新产品形态=自由行，新产品类型=出境旅游、国内旅游、境内N日游、境外N日游

**渠道范围**：APP 、H5、online

# 4需求描述

## 4.1 纠错词



1. 根据提供的人工整理的错词和正确词对应关系列表，人工整理出错词模式。
2. 拉取POI和目的地的数据，根据错词模式生成对应的错词库，该错词库包含正确的词（POI、目的地）及相应的错词之间的映射关系。
3. 拉取过去1年半的历史搜索词，排除可以匹配到POI和目的地的搜索词（即正确的搜索词，这部分不是我们的处理目标），剩下的就是错误的搜索词（可能仍有干扰项，需人工筛选），也就是我们的目标处理范围。
4. 将错词库与上一步得到的历史错误搜索词进行匹配，计算错词库对历史搜索词的覆盖度。

如果覆盖度较小，则对于没有覆盖到的历史错误搜索词，人工再次整理错误模式，再次生成错词库并进行收缩，这可能是一个迭代的过程。

1. 对于历史错误搜索词（包含可匹配到错词库的、以及经人工审核确认是错词的），建立Ngram语言模型，然后计算错词库中每一个错词的转移概率（即该错词可能发生的概率），排除掉转移概率小的错词，从而对错词库进行收缩。
2. 对收缩后的错词进行进一步筛选（比如有些生成的错词恰好是某个poi或目的地，要排除掉），并由业务人工抽样检查准确率

目前整理出的错词模式如下：

1. 前后颠倒：比如马代尔夫
2. 形同字：比如九塞沟
3. 拼音相同：比如西林国勒
4. 拼音相似：比如爱情海

* 平舌音和翘舌音互换
* 前鼻音和后鼻音互换
* 例如： Z换成ZH，SH换成S，IN换成ING，ENG换成EN

具体细则如下：

'z':'zh','c':'ch','s':'sh',

'zh':'z','ch':'c','sh':'s',

'an':'ang','en':'eng','in':'ing','on':'ong',

'ang':'an','enh':'en','ing':'in','ong':'on'

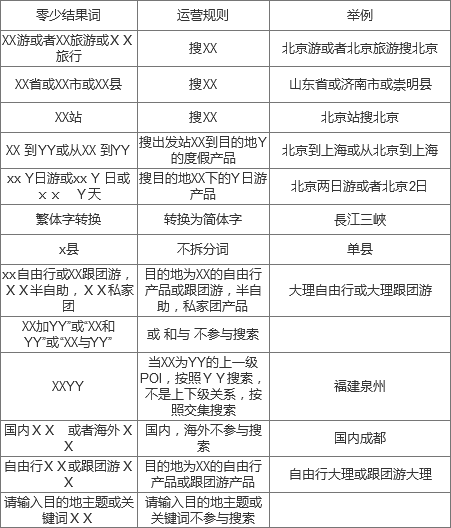
1. 中文转数字：比如5日游

## 4.2 零少结果词的丢词分析

根据用户原始输入词及转化之后的词，对零少结果词分析，得到搜索上可以丢掉的词 ，增加丢词词库数据

## 4.3用户搜索关键词处理

通过运营总结的搜索词 ，制定运营规则，通过运营规则降低用户搜索词的零少结果率



# 5 需求负责人

|  |  |
| --- | --- |
| 需求list | 负责人 |
| 4.1 | 于天池 |
| 4.2 | 于天池 |
| 4.3 | 周燕军 |