

《大唐西域记》可视化期末报告

小组成员：陈凯瑞，王蕙钰，胡良

1. 项目概述

本课程设计是将《大唐西域记》和《大唐大慈恩寺三藏法师传》中玄奘西行的故事进行可视化，呈现玄奘西行的行程相关信息如路线、时间、地点，途径地区的地理人文信息如物产、文化、疆域范围，以及在当地的经历见闻等历史信息。同时加入远读，对文本结构信息如篇幅等进行直观呈现和横向对比。

2. 数据收集与分析

2.1 数据收集

通过对《大唐西域记》和《大唐大慈恩寺三藏法师传》的详细整理，收集到三种类型的数据，包括表格数据集、空间数据集和文本数据集。

表格数据中每个数据项代表《大唐西域记》中记载的一个地点，其属性包含该地点的各方面信息，具体如图 1 所示。

空间数据集是通过将《大唐西域记》原文与网上的玄奘西行路线图（如图 2 所示，网址：<https://www.ageeye.cn/map/53779/>）对照，获取相应地点的经纬度坐标得到的，是呈现玄奘西行轨迹路线的地图数据，以 geojson 的格式保存每个点的坐标及相应信息。

文本数据集通过对每个地点收集对应的文本得到，包含文本内容、文本结构信息（如字数、句长等），如图 3 所示。

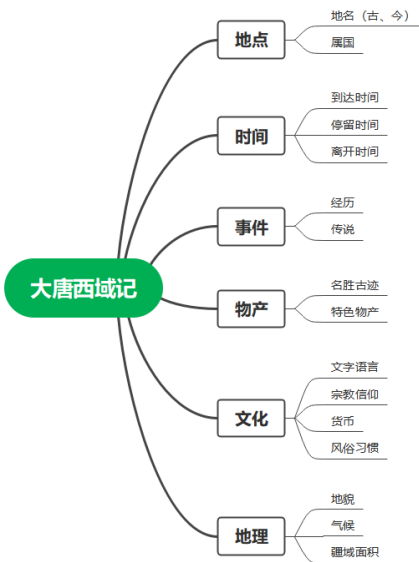


图 1 表格数据属性

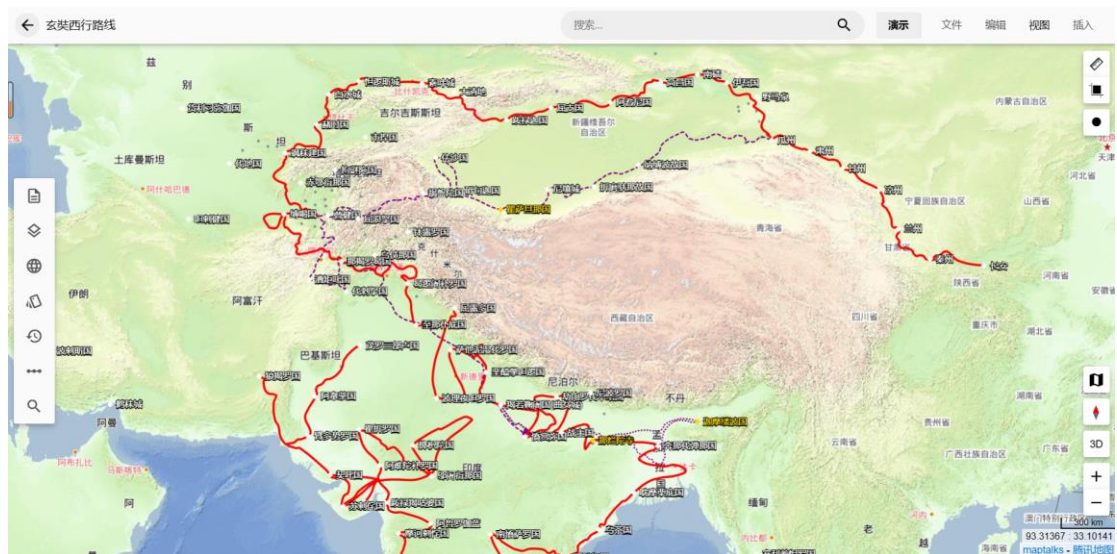


图 2 玄奘西行路线

国家	字数	句数	平均每句字数
阿普尼国	204	15	13.6
屈支国	911	66	13.80303
跋禄迦国	72	5	14.4
凌山	95	7	13.571429
大清池	78	4	19.5
素叶水城	67	5	13.4
千泉	101	7	14.428571
咄逻私城	34	3	11.333333
小孤城	54	2	27
白水城	31	2	15.5
恭御城	40	4	10

图 3 文本结构数据（部分）

2.2 数据分析

由于数据以非结构化的文本数据为主,对数据的初步探索主要是在表格数据的各个属性上做类别频数统计,观察数据集在各个属性上的分布情况。探索结果包括:

- 文字语言、特色物产、货币方面,存在大量缺失(未记载)且类别繁多,不适合进行分类,适合直接展示。
- 疆域面积、宗教信仰、气候方面,缺失较少且类别较少,适合作为分类依据。

3. 系统设计

3.1 任务分析

地图数据中的每一个数据点与表格数据中的每一个数据项相对应,因此将地图数据作为可视化的基础,以位置这一视觉通道呈现。

文本数据中既包含“地名”、“属国名”、“疆域”等可以在地图中显性表现出来的属性,也包含“人名”、“经历”、“文字”等不易从地图中呈现的属性。因此,根据是否可以在地图中直

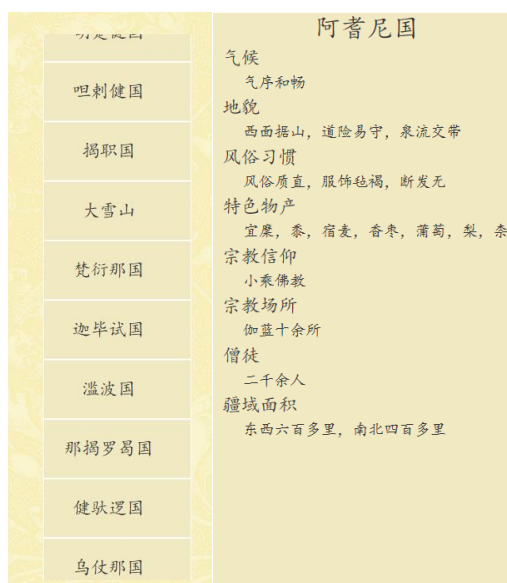


图 5 侧边栏与信息展示区



图 6 时间轴

3.4 文本远读

文本信息由一串矩形表示。每一个矩形代表一个地点，矩形长度代表地点对应文本字数，宽度代表该地点对应句数，颜色代表宗教、疆域、气候的类型（可供选择），鼠标悬停至矩形上会显示该地点的名称与内容（宗教、疆域、气候的类型）。通过远读，可以对文本结构信息如篇幅、内容信息如气候等进行直观呈现和横向对比。



图 7 文本远读

4. 系统展示

地图可视化界面可以通过拖拽缩放地图和时间轴进行范围选择、点选图层进行展示内容筛选以及点击/滑动左侧国家目录或点击地图上地点进行交互以展示国家相关信息。文本远读部分可以通过鼠标悬停展示国家名称和相关属性，通过点选下拉框进行展示属性筛选。



图 8 经行国家范围选择



图 9 侧边栏与地点交互和信息展示



图 10 地图与侧边栏交互



图 11 悬浮展示国家信息



图 12 选择文本远读展示属性

5. 探索发现

通过探索得到部分结果如下（例）：

- 从轨迹可视化来看，玄奘的活动路线呈现曲折性和反复性，这里可能说明玄奘当时对西域的了解确实不足，并没有很完善的路线规划，是以一种探索的态度完成这个旅行。
- 结合时间和经行国家，我们发现玄奘行进的速度是与相关国家地区的发达程度（通过疆域面积、宗教繁荣情况、记载文本量反映）成反比的。
- 今北印度、巴基斯坦、阿富汗等地区多信奉佛教，南印度地区多信奉印度教。临近的国家在宗教信仰相似度更高。

6. 不足与展望

- 轨迹可视化界面原定设计多个图层对各项内容进行筛选，受限于内容划分类别过多，数据处理难度较大，未能及时整理出所需数据。
- 远读部分原定再设计带滚动条界面，可与地图可视化界面联动，且更清晰地比对各个内容。视觉通道设计有待优化。

7. 分工

成员	分工
陈凯瑞	数据收集与整理、轨迹可视化、报告撰写
王蕙钰	数据收集与整理、PPT 制作、报告撰写
胡良	数据收集与整理、文本远读、报告撰写