

Digitisation of Weather Records of Seungjeongwon Ilgi: A Historical Weather Dynamics Dataset of the Korean Peninsula (1623-1910)

Introduction

This study has exploited the daily weather records of *Seungjeongwon Ilgi* from the NIKH database. *Seungjeongwon Ilgi* is a daily record of the *Seungjeongwon*, the *Royal Secretariat* of the *Joseon Dynasty of Korea*. These diaries span from 1623 to 1910 and generally involve daily weather records in the entry header. Their observational site would be located in Seoul (N37°35', E126°59'). We have exploited the weather records from the NIKH database and classified the daily weather using text mining method. We have also converted the report dates from the traditional lunisolar calendar to the Gregorian calendar, to better contextualise our data into the contemporary daily measurements.

Data

We provide different formats (csv, xlsx, json) to facilitate the usage of data. The main contents of data are listed as below.

- **ID:** The unique identifier of a specific record in the metadata, which can also serve as the identifier to merge with external data in the NIKH digital database.
- **Traditional calendar:** The original lunar dates in the NIKH digital database, which are listed in data format "YYYY-MM-DD". More specifically, "L0" implies the leap year and "L1" implies the common year.
- **Leap:** The identifier of a leap year.
- **Gregorian calendar:** The Gregorian calendar date that converted by the traditional calendar date.
- **Weather Text:** The text that describe the weather conditions. Specifically, multiple weather descriptions of the same day have been put together.
- **Flag:** The computed value that indicates different combinations of weather conditions.
- **Volume:** The volume of text in the original record.
- **Herbal Volume:** The volume of text in the herbal record.
- **Sunny:** A dummy variable that represents whether the weather description contains the expression of sunny.
- **Cloudy:** A dummy variable that represents whether the weather description contains the expression of cloudy.
- **Rainy:** A dummy variable that represents whether the weather description contains the expression of rainy.
- **Snow:** A dummy variable that represents whether the weather description contains the expression of snow.
- **Wind:** A dummy variable that represents whether the weather description contains the expression of wind.

Import Data

```
# Python
# CSV file
import pandas as pd
data=pd.read_csv('~\SJWilgi_Seoul_Weather_YR1623_1910.csv',encoding="utf-8")
# JSON file
data=pd.read_json('~\SJWilgi_Seoul_Weather_YR1623_1910.json',encoding="utf-8")
# Excel file
data=pd.read_excel('~\SJWilgi_Seoul_Weather_YR1623_1910.xlsx') # Excel file
```

```
# R
# CSV file
library(readr)
data<- read_csv("Documents/SJWilgi_Seoul_Weather_YR1623_1910.csv")
# Excel file
library(readxl)
data <- read_excel("Documents/SJWilgi_Seoul_Weather_YR1623_1910.xlsx")
```