

# Cyber\_Poem

Chinese Poems  
Encoding with TEI

Changxu Duan, Zhe Fan,  
Tianxiang Wang

# Structure

---



**Motivation**



**TEI Introduction**



**Workflow:**

Data Collect  
Data Prepare  
Data Access

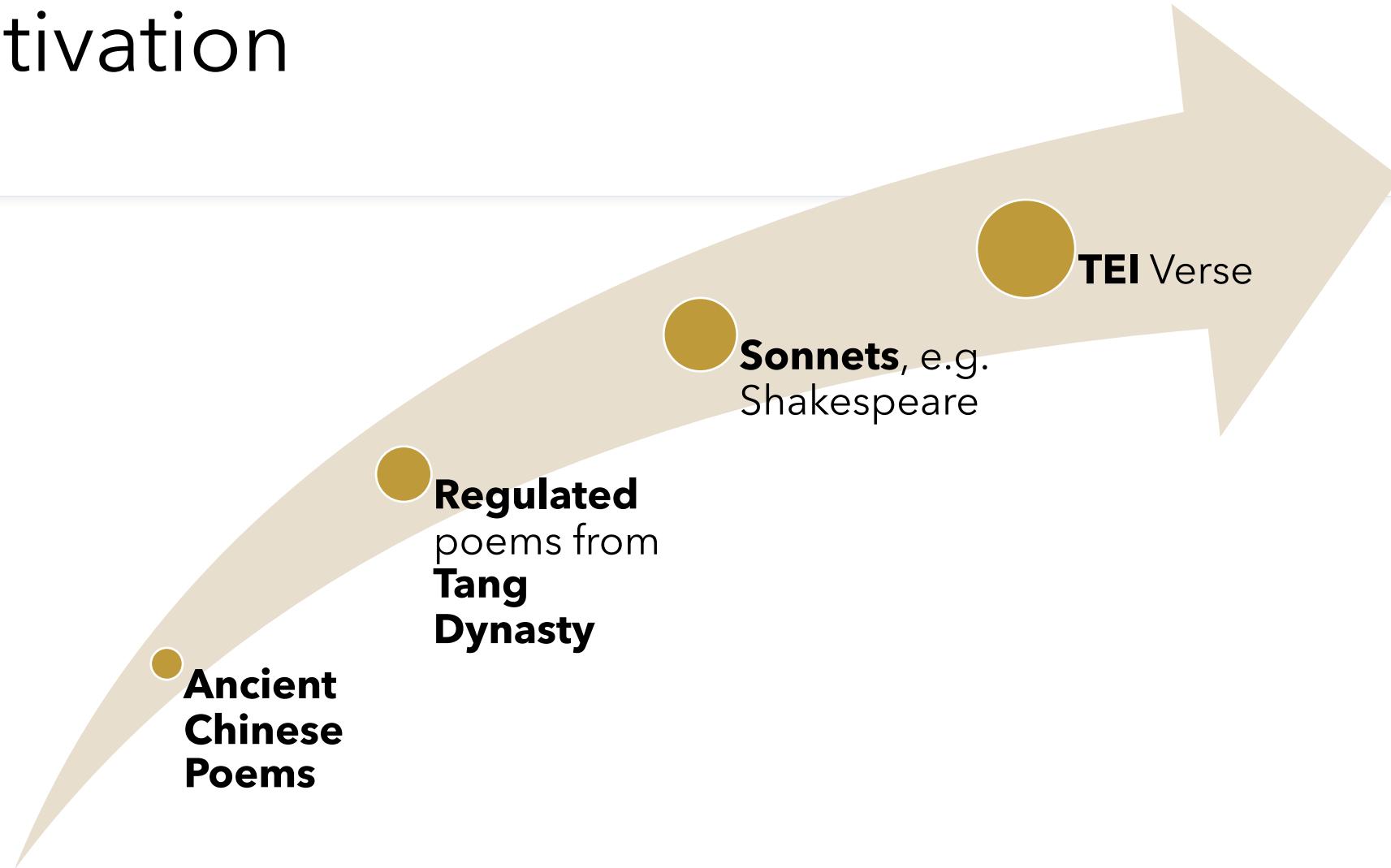


**Demo**



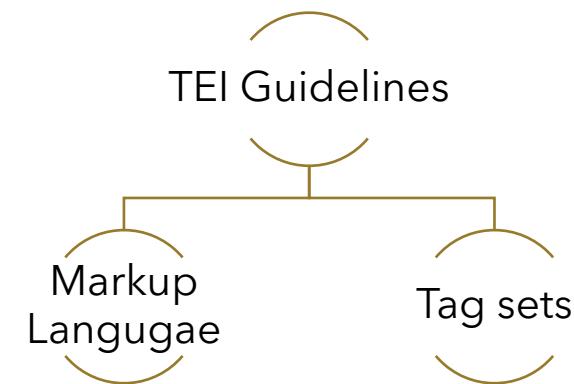
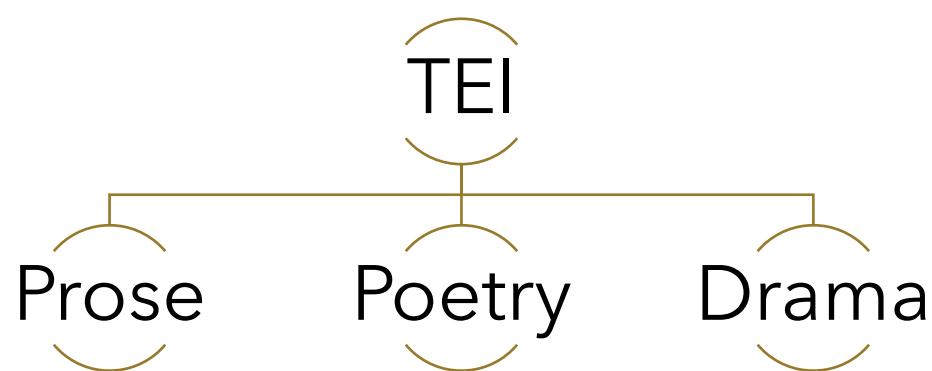
**Take-away [Issues encountered]**

# Motivation



# TEI (Text Encoding Initiative)

The Text Encoding Initiative (TEI) is a standard for the representation of textual material in digital form through the means of text encoding.



# Properties of “Chinese regulated poems”

- Five-character Quatrains (4/8 lines)
- Seven-character Quatrains (4/8 lines)
- The Metrical schemas (e.g. 4 for 5-char Quatrains)
- The rhyme rules



# Applying Text Technology

- Data Collect
- Data Prepare
- Data Access

# Data Collect

- Gathered 60 poems from blog



月落乌啼霜满天，江枫渔火对愁眠。姑苏城外寒山寺，夜半钟声到客船。

## 三四、江南曲（储光义）

日暮长江里，相邀归渡头。落花如有意，来去逐轻舟。

## 三五、逢雪宿芙蓉山主人（刘长卿）

日暮苍山远，天寒白屋贫。柴门闻犬吠，风雪夜归人。

## 三六、听弹琴（刘长卿）

泠泠七弦上，静听松风寒。古调虽自爱，今人多不弹。

## 三七、送灵澈上人（刘长卿）

苍苍竹林寺，杳杳钟声晚。荷笠带夕阳，青山独归远。

## 三八、绝句二首其一（杜甫）

迟日江山丽，春风花草香。泥融飞燕子，沙暖睡鸳鸯。

## 三九、绝句二首其二（杜甫）

江碧鸟逾白，山青花欲燃。今春看又过，何日是归年？

## 四十、八阵图（杜甫）

功盖三分国，名成八阵图。江流石不转，遗恨失吞吴。

## 四一、华子岗（裴迪）

日落松风起，还家草露晞。云光侵履迹，山翠拂人衣。

## 四二、题红叶（韩氏）

流水何太急，深宫尽日闲。殷勤谢红叶，好去到人间。

## 四三、塞下曲六首其二（卢纶）

林暗草惊风，将军夜引弓。平明寻白羽，没在石棱中。

## 四四、塞下曲六首其三（卢纶）

月黑雁飞高，单于夜遁逃。欲将轻骑逐，大雪满弓刀。

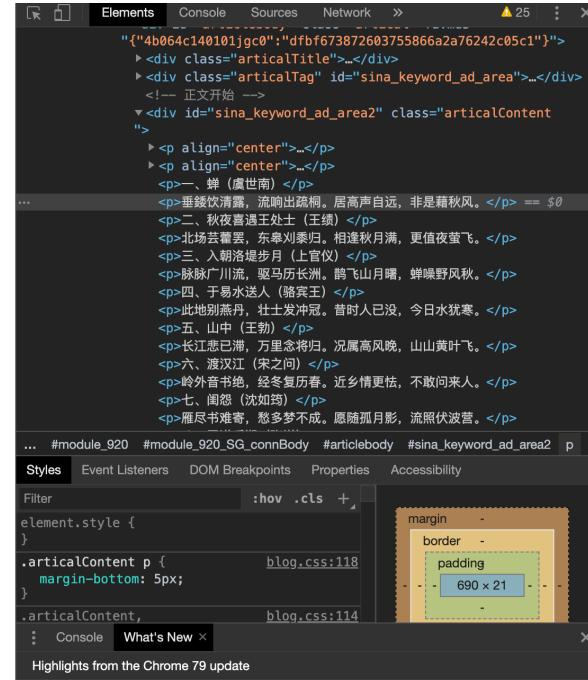
## 四五、登鹳雀楼（畅当）

迥临飞鸟上，高出世尘间。天势围平野，河流入断山。

# Data Collect

- Check the html structure of the website in developer mode.

大卜伤心处，穷穷送客亭。春风知别苦，个道柳杀青。  
三三、小长干曲（崔国辅）  
月暗送湖风，相寻路不通。菱歌唱不彻，知在此塘中。  
三四、江南曲（储光羲）  
日暮长江里，相邀归渡头。落花如有意，来去逐轻舟。  
三五、逢雪宿芙蓉山主人（刘长卿）  
日暮苍山远，天寒白屋贫。柴门闻犬吠，风雪夜归人。  
三六、听弹琴（刘长卿）  
泠泠七弦上，静听松风寒。古调虽自爱，今人多不弹。  
三七、送灵澈上人（刘长卿）  
苍苍竹林寺，杳杳钟声晚。荷笠带夕阳，青山独归远。  
三八、绝句二首其一（杜甫）  
迟日江山丽，春风花草香。泥融飞燕子，沙暖睡鸳鸯。  
三九、绝句二首其二（杜甫）  
江碧鸟逾白，山青花欲燃。今春看又过，何日是归年？  
四十、八阵图（杜甫）  
功盖三分国，名成八阵图。江流石不转，遗恨失吞吴。  
四一、华子岗（裴迪）  
日落松风起，还家草露晞。云光侵履迹，山翠拂人衣。  
四二、题红叶（韩氏）  
流水何太急，深宫尽日闲。殷勤谢红叶，好去到人间。  
四三、塞下曲六首其二（卢纶）  
林暗草惊风，将军夜引弓。平明寻白羽，没在石棱中。  
四四、塞下曲六首其三（卢纶）  
月黑雁飞高，单于夜遁逃。欲将轻骑逐，大雪满弓刀。



# Data Collect



# Data Collect



- **Get html elements**
- **Read contents**

# Data Collect



- **Get html elements**
- **Read contents**
- **Clean data**
- **Structure and classify data**

# Data Collect



- **Get html elements**
- **Read contents**
- **Clean data**
- **Structure and classify data**
- **Save the structured data as a json file**

# Data prepare - Encoding

- Load collected.json => [[title],[author],[content]]...
- **Prominence** (+/-) Analysis
- Automatically choose the meter schema
- Rhyme Evaluation
- Populate the xml file with data

## Encoded verse : teiHeader

- <titleStmt>: title statement
- <publicationStmt>: publication statement
- <sourceDesc>: source description
- <metDecl>: metrical notation declaration
  - pattern: a regular expression defining any value that is legal for this notation
- <metSym>: metrical notation symbol

```
<?xml version="1.0" ?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>格律诗</title>
      </titleStmt>
      <publicationStmt>
        <p>Text Technology</p>
      </publicationStmt>
      <sourceDesc>
        <p>-DFW-</p>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <metDecl pattern="(\+|\-)+">
        <metSym value="+>metrical prominence</metSym>
        <metSym value="->metrical non-prominence</metSym>
      </metDecl>
    </encodingDesc>
  </teiHeader>
```

```
<lg type="poem">
  <head>
    <title>入朝洛堤步月</title>
  </head>
  <lg rhyme="aaba" type="五绝仄起首句入韵">
    <l met="---++" rhyme="a">
      脉脉广川
      <rhyme>流</rhyme>
    </l>
    <l met="+----" real="+----" rhyme="a">
      驱马历长
      <rhyme>洲</rhyme>
    </l>
    <l met="+---+" real="---++" rhyme="b">
      鹊飞山月
      <rhyme>曙</rhyme>
    </l>
    <l met="---++" real="---++" rhyme="a">
      蝉噪野风
      <rhyme>秋</rhyme>
    </l>
  </lg>
  <signed>上官仪</signed>
</lg>
```

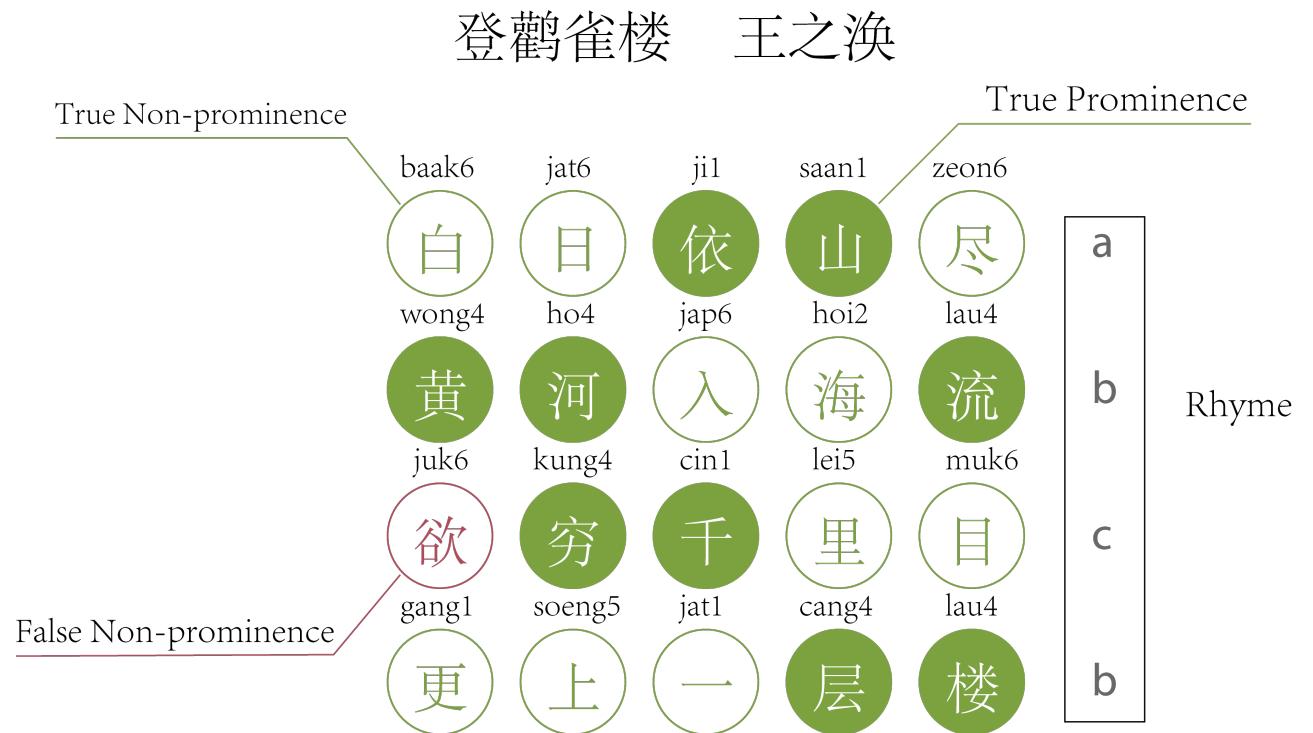
## Encoded verse : body

---

- <title>: title
- <lg>: line group
  - rhyme: applicable rhyme scheme
  - type: applicable metrical scheme
- <l>: line
  - met: metrical structure of the line
  - real: actual realization of the metrical structure
  - rhyme: rhyme of this line
- <rhyme>: rhymed token
- <signed>: author

# Data Access

- JavaScript front-end
- Binding and rendering HTML element with Vue.js
- Load XML object with DOMParser.parseFromString()



烟



Thank you