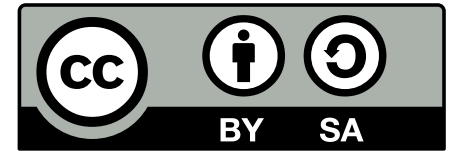
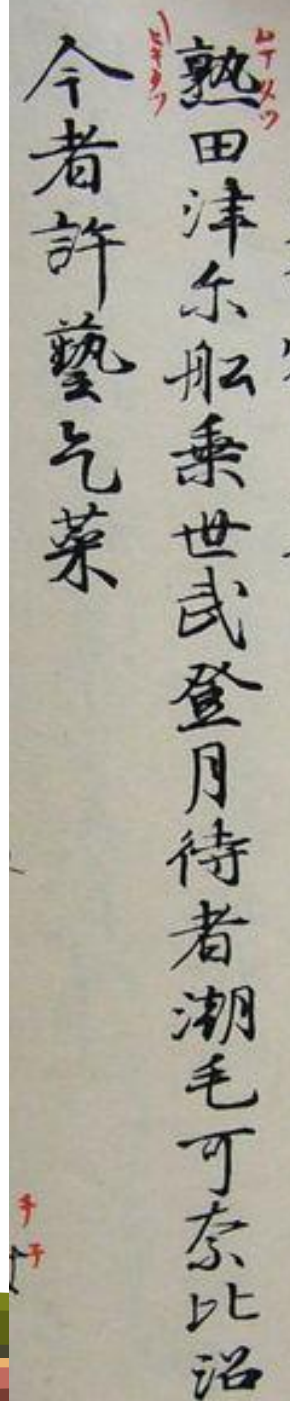


# Manyoushuu (万葉集)

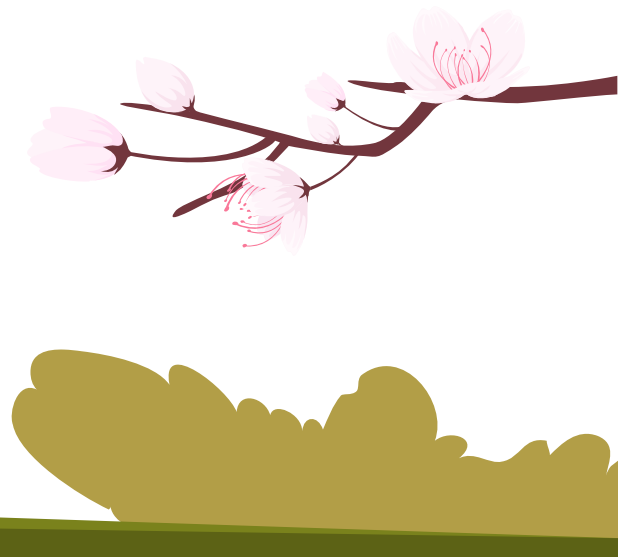
*Jakub Levý*

*[jakub.levy15@gmail.com](mailto:jakub.levy15@gmail.com)*

*Faculty of Mathematics and Physics  
Charles University*



unless otherwise stated



# Content



- Japanese 101
  - Writing System
  - Mora
  - Waka
  - Haiku
- Manyoushuu
  - About
  - Example in detail

- Project
  - Visualization Goals
  - Overview
  - Technologies
  - Look at the Code
  - (Live) Examples
  - Conclusion
- References





# Japanese 101



# Japanese Writing System (1)



- No script existed until ~ AD 285
- Chinese symbols not perfect due to language differences
  - Sino-Japanese (音読) texts unintelligible
- Japanese reading of Chinese characters invented (訓読)
- How to capture the exact sound (of Japanese names)?
  - Use an arbitrary Chinese symbol with the correct 音読 reading
- Ex. *no* could be written as 能, 乃, 農, 濃, 迺, ...
- The list was then reduced and partially defined
- *Manyougana* (万葉仮名)





# Japanese Writing System (2)



- *Manyoushuu* is written in *Manyougana*
- Grammatical forms are captured by certain *Manyougana* characters
- Characters are complicated, but often used
  - Simplification
- In 10<sup>th</sup> century, shortcuts defined as *Katakana* (片仮名)
  - First Japanese syllabic alphabet
  - Considered less aesthetic than *Manyougana*



| n   | w-   | r-   | y-   | m-  | h-   | n-   | t-   | s-   | k-   |   |    |
|---|--|--|--|---|--|--|--|--|--|---|----|
| <br><i>N</i> | <br><i>WA</i>   | <br><i>RA</i>   | <br><i>YA</i>   | <br><i>MA</i>   | <br><i>HA</i>   | <br><i>NA</i>   | <br><i>TA</i>   | <br><i>SA</i>   | <br><i>KA</i>   | <br><i>A</i>   | -a |
|   | <br><i>WI</i>   | <br><i>RI</i>   |  | <br><i>MI</i>   | <br><i>HI</i>   | <br><i>NI</i>   | <br><i>CHI</i>  | <br><i>SHI</i>  | <br><i>KI</i>   | <br><i>I</i>   | -i |
|   |  | <br><i>RU</i>   | <br><i>YU</i>   | <br><i>MU</i>   | <br><i>FU</i>   | <br><i>NU</i>   | <br><i>TSU</i>  | <br><i>SU</i>   | <br><i>KU</i>   | <br><i>U</i>   | -u |
|   | <br><i>WE</i>  | <br><i>RE</i>  |  | <br><i>ME</i>  | <br><i>HE</i>  | <br><i>NE</i>  | <br><i>TE</i>  | <br><i>SE</i>  | <br><i>KE</i>  | <br><i>E</i>  | -e |
|   | <br><i>WO</i> | <br><i>RO</i> | <br><i>YO</i> | <br><i>MO</i> | <br><i>HO</i> | <br><i>NO</i> | <br><i>TO</i> | <br><i>SO</i> | <br><i>KO</i> | <br><i>O</i> | -o |

# Japanese Writing System (3)



- In parallel, *sougana* (草仮名) was used in private correspondences
  - Few selected *Manyougana* characters written in highly cursive style (草書体)
- Generally for woman
  - Not expected to know Chinese characters
- *Sougana* was called *onnade* (女手)
- Becomes popular in 10<sup>th</sup> century
- Contains “*woman gracefulness*”
- Slightly transformed and called *Hiragana* (平仮名) nowadays



|         |         |        |        |        |        |        |         |        |         |        |
|---------|---------|--------|--------|--------|--------|--------|---------|--------|---------|--------|
| 无<br>えん | 和<br>わ  | 良<br>ら | 也<br>や | 末<br>ま | 波<br>は | 奈<br>な | 太<br>た  | 左<br>さ | 加<br>か  | 安<br>あ |
|         | 爲<br>ゐる | 利<br>り |        | 美<br>み | 比<br>ひ | 仁<br>に | 知<br>ち  | 之<br>し | 機<br>き  | 以<br>い |
|         |         | 留<br>る | 由<br>ゆ | 武<br>む | 不<br>ふ | 奴<br>ぬ | 川<br>かわ | 寸<br>す | 久<br>く  | 宇<br>う |
|         | 恵<br>ゑ  | 礼<br>れ |        | 女<br>め | 部<br>ぶ | 祢<br>ね | 天<br>てん | 世<br>せ | 計<br>けい | 衣<br>い |
|         | 遠<br>を  | 呂<br>ろ | 与<br>よ | 毛<br>も | 保<br>ほ | 乃<br>の | 止<br>と  | 曾<br>そ | 己<br>こ  | 於<br>お |





| n   | w-   | r-   | y-   | m-  | h-   | n-   | t-   | s-   | k-   |   |    |
|---|--|--|--|---|--|--|--|--|--|---|----|
| <br><i>N</i> | <br><i>WA</i>   | <br><i>RA</i>   | <br><i>YA</i>   | <br><i>MA</i>   | <br><i>HA</i>   | <br><i>NA</i>   | <br><i>TA</i>   | <br><i>SA</i>   | <br><i>KA</i>   | <br><i>A</i>   | -a |
|   | <br><i>WI</i>   | <br><i>RI</i>   |  | <br><i>MI</i>   | <br><i>HI</i>   | <br><i>NI</i>   | <br><i>CHI</i>  | <br><i>SHI</i>  | <br><i>KI</i>   | <br><i>I</i>   | -i |
|   |  | <br><i>RU</i>   | <br><i>YU</i>   | <br><i>MU</i>   | <br><i>FU</i>   | <br><i>NU</i>   | <br><i>TSU</i>  | <br><i>SU</i>   | <br><i>KU</i>   | <br><i>U</i>   | -u |
|   | <br><i>WE</i>  | <br><i>RE</i>  |  | <br><i>ME</i>  | <br><i>HE</i>  | <br><i>NE</i>  | <br><i>TE</i>  | <br><i>SE</i>  | <br><i>KE</i>  | <br><i>E</i>  | -e |
|   | <br><i>WO</i> | <br><i>RO</i> | <br><i>YO</i> | <br><i>MO</i> | <br><i>HO</i> | <br><i>NO</i> | <br><i>TO</i> | <br><i>SO</i> | <br><i>KO</i> | <br><i>O</i> | -o |

# Japanese Writing System (4)



- Modern Japanese use
  - Hiragana, Katakana, Kanji (Chinese characters), Latin, Arabic numerals

SHII DII ICHI MAI TI SHA TSU NI MAI RI N GO SAN KO O KA I TA I TO OMOI MA SU  
CD 1 枚、Tシャツ 2 枚、リンゴ 3 個を買いたいと思います。

I want to buy 1 CD, 2 T-shirts, and 3 apples so much.

- Notice SOV (subject-object-verb) word order



# Mora



- One mora is the length of time it takes to pronounce one *kana* (*hiragana* or *katakana*) character

馬鹿 → <sup>BA KA</sup>  
ばか

- 2 K, 2 M, 2 S

大学 → <sup>DA I GA KU</sup>  
だいがく

- 4 K, 4 M, 3 S

結婚 → <sup>KE K KO N</sup>  
けっこん

- 4 K, 4 M, 2 S

東京 → <sup>TO U KYO U</sup>  
とうきょう

- 5 K, 4 M, 2 S

- Generally  $\#moraes \neq \#kana \neq \#syllables$



# Waka (和歌 or 倭歌)



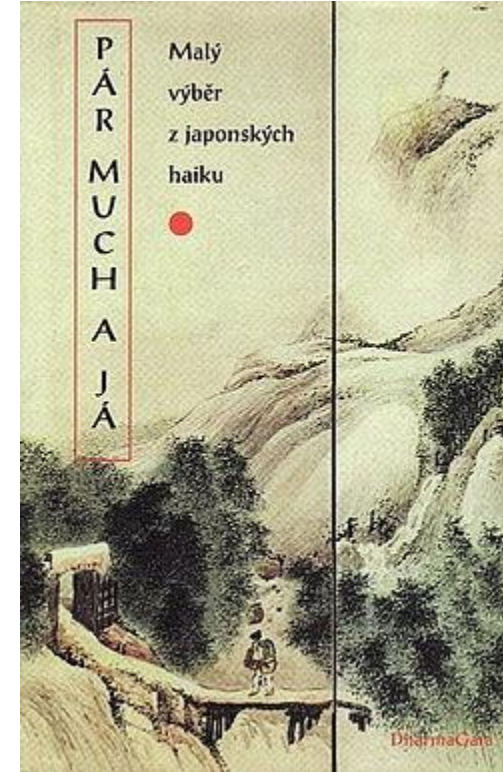
- Poetry in classical Japanese literature
- Many forms
  - Differ by the number of morae
- *Tanka* (短歌)
  - Major genre
  - 5-7-5-7-7
- *Katauta* (片歌)
  - Shortest form of *Waka*
  - 5-7-7
- *Sedouka* (旋頭歌)
  - *Katauta* twice
  - 5-7-7-5-7-7
- *Chouka* (長歌)
  - 5-7 repeated at least twice
  - Concluded with 5-7-7 ending
- *Bussokusekika* (仏足石歌)
  - *Tanka* with an extra 7 phrase at the end
  - 5-7-5-7-7-7



# Why is Haiku (俳句) missing?



- First emerged in 17<sup>th</sup> century
- Much younger than all forms of *Waka*
- 5-7-5
- Not relevant for us

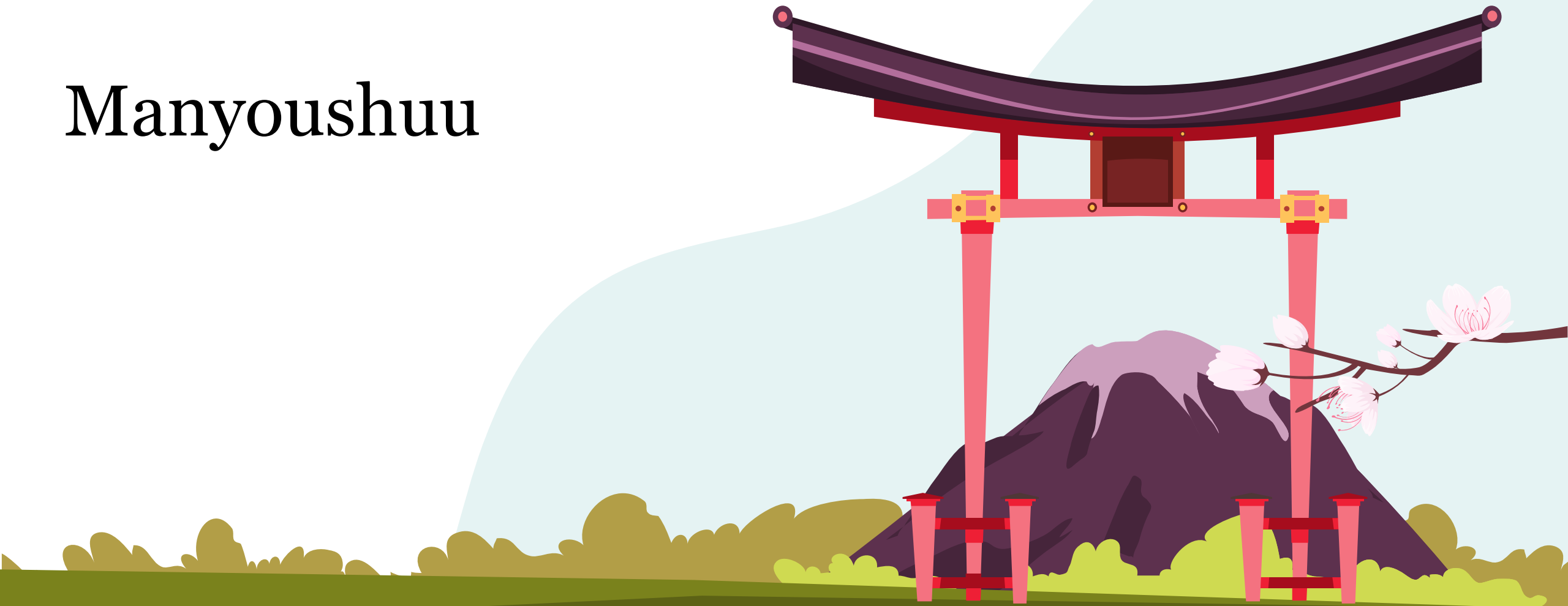


古池や  
蛙飛び込む  
水の音





# Manyoushuu



# Manyoushū (万葉集) (1)



- 万葉集 ~ “collection of ten thousand leaves”
- Oldest collection of *Waka* poetry
- Written in *Manyōgana*
- More than 4000 *Waka* poems
- 561 authors
  - 70 woman
  - People of various statuses: emperors, aristocrats, peasants, street performers, ...
- Over 2100 poems by unknown authors



# Manyoushū (万葉集) (2)



- First poem by emperor *Nintoku* (313-399)
- Last poem from AD 759
- Most poems probably from 7<sup>th</sup> and 8<sup>th</sup> century
- Probably compiled by *Ootomo no Yakamochi* in 8<sup>th</sup> century

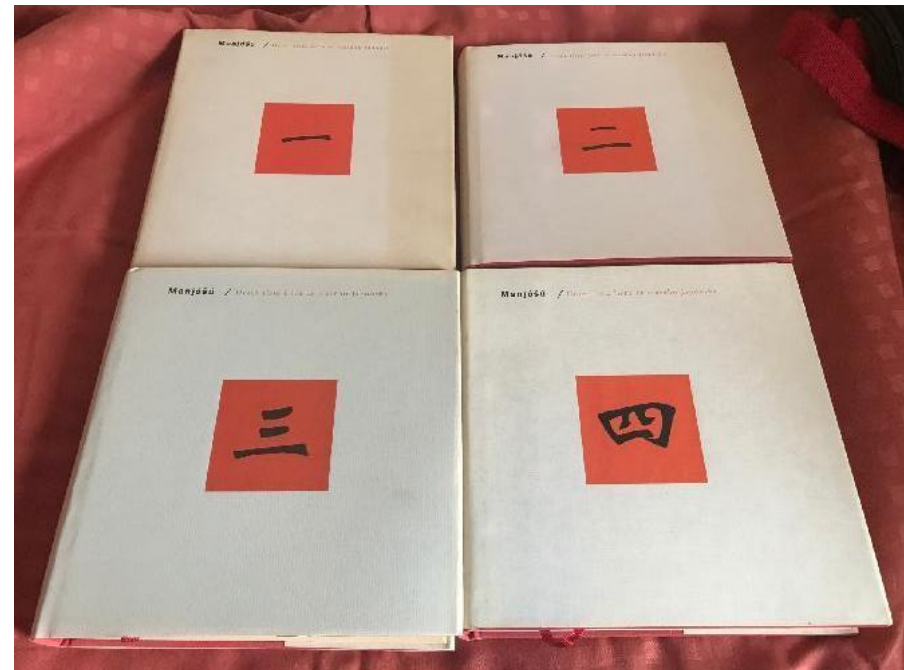




# Manyoushū (万葉集) (3)



- Complete translation of the whole anthology available only in
  - French
  - **Czech** (by Antonín Líman)



# Uta 8 (Princess Nutaka's Poem)



|   | Original | Modern                          | Hiragana only                     |   |
|---|----------|---------------------------------|-----------------------------------|---|
| 5 | 熟田津尔     | NIGI TA TSU NI<br>熟田津に          | NI GI TATSU NI<br>にぎたつに           | 5 |
| 7 | 船乗世武登    | FUNA NO RI SE MU TO<br>船乗りせむと   | FU NA NO RI SE MU TO<br>ふなのりせむと   | 7 |
| 5 | 月待者      | TSUKI MA TE BA<br>月待てば          | TSU KI MA TE BA<br>つきまてば          | 5 |
| 7 | 潮毛可奈比沼   | SHIHO MO KA NA HI NU<br>潮もかなひぬ  | SHI HO MO KA NA HI NU<br>しほもかなひぬ  | 7 |
| 7 | 今者許藝乞菜   | IMA WA KO GI I DE NA<br>今は漕ぎ出でな | I MA WA KO GI I DE NA<br>いまはこぎいでな | 8 |

U lod'ky v zálivu Nikita

čekáme na měsíc...

Příliv nám přeje,

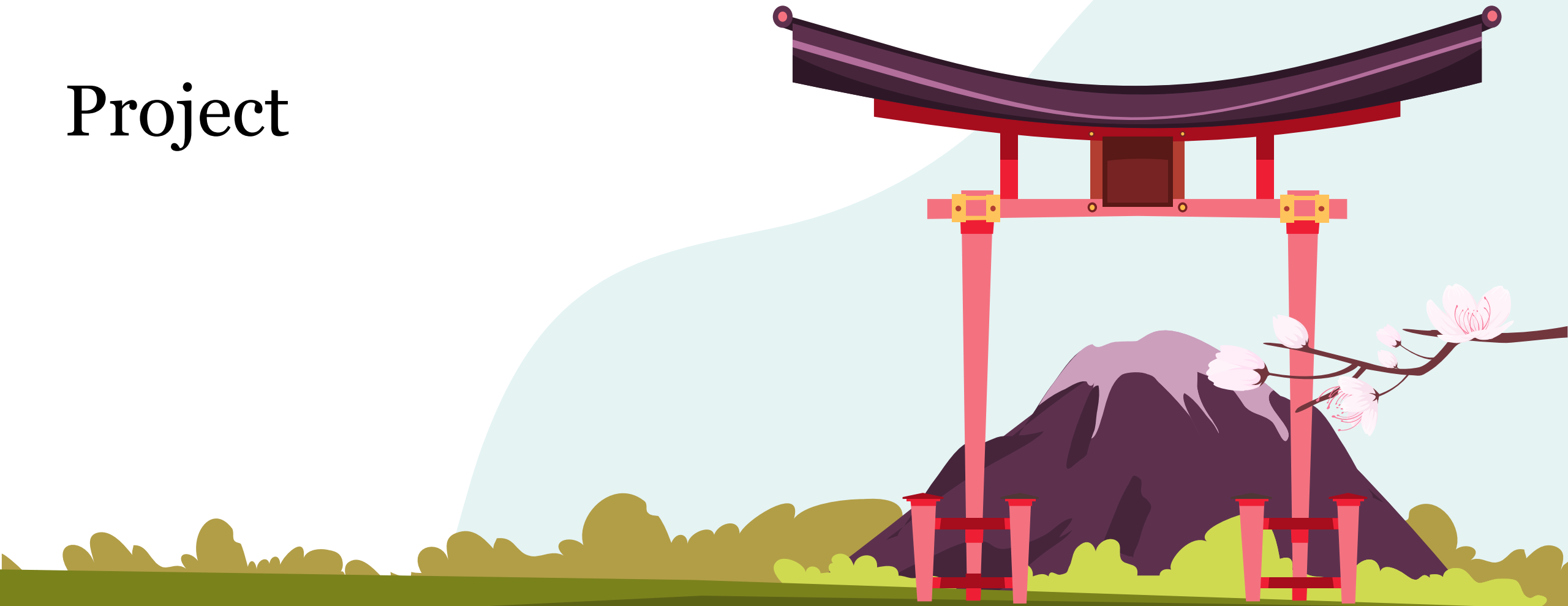
nuže, k veslům, k veslům!

(Translated by Antonín Líman)





# Project



# Visualization Goals

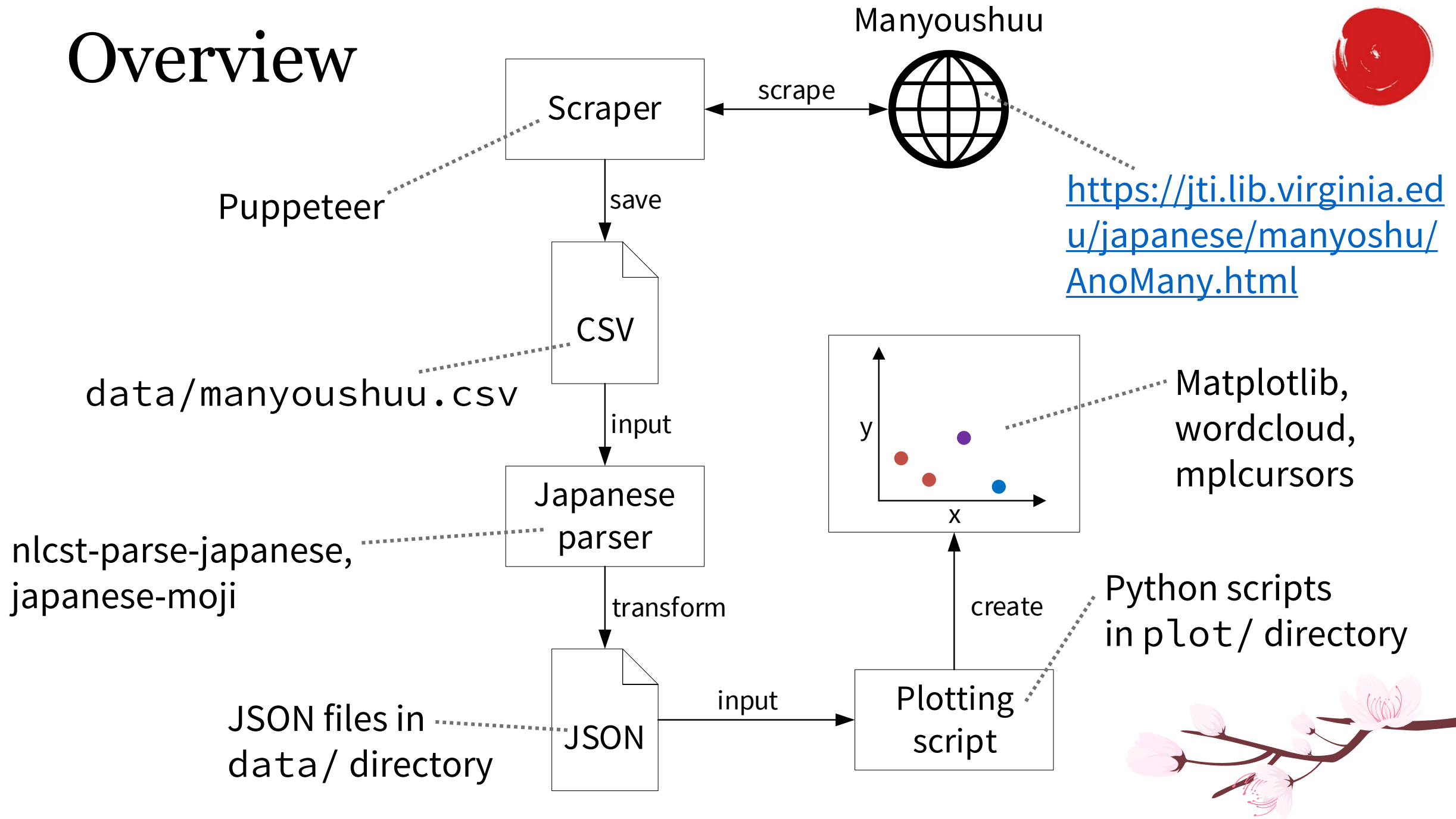


- Word cloud of the most used words
- Frequency of a *kanji* vs the number of different words the *kanji* is used in
  - And be able to somehow quickly see JLPT level of the *kanji*
- Check the compliance of the form of a poem in modern Japanese
- Calculate the number of poems for every *Waka* form
- Show the occurrence of a *kanji*/word across all poems

Different forms of a word must be handled correctly  
持たず and 持ち are a form of 持つ



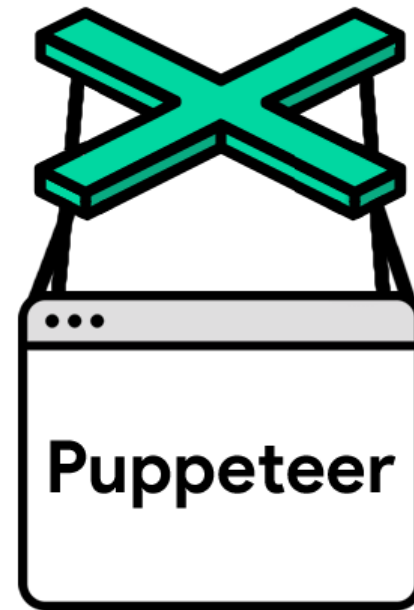
# Overview



# Puppeteer



- Node.js library providing API to control Chrome/Chromium
- Uses DevTools Protocol



# nlcst-parse-Japanese (1)



- Parser for Japanese text
- Output format is *NLCST*
  - **N**atural **L**anguage **C**oncrete **S**yntax **T**ree format

nlcst



# nlcst-parse-Japanese (2)



Parser output for 持ち (boilerplate nodes omitted)

```
{
  "type": "TextNode",
  "value": "持ち",
  "position": {
    "start": {
      "line": 1,
      "column": 8,
      "offset": 7
    },
    "end": {
      "line": 1,
      "column": 10,
      "offset": 9
    }
  },
  "data": {
    "word_id": 3144310,
    "word_type": "KNOWN",
    "surface_form": "持ち",
    "pos": "動詞",
    "pos_detail_1": "自立",
    "pos_detail_2": "*",
    "pos_detail_3": "*",
    "conjugated_type": "五段・タ行",
    "conjugated_form": "連用形",
    "basic_form": "持つ",
    "reading": "モチ",
    "pronunciation": "モチ"
  }
},
```

A decorative illustration of a cherry blossom branch with pink flowers and buds, located in the bottom right corner of the slide.



# japanese-moji



- Handy functions for lazy people
  - isValidKanji
- Supports creation of custom validators

```
kanaValidatorOptions = {  
  characterSets : [  
    CharacterSet.Katakana,  
    CharacterSet.KatakanaPhoneticExtension,  
    CharacterSet.HalfWidthKatakana,  
    CharacterSet.Hiragana  
  ]  
}  
isValidKana = createStrictValidator(kanaValidatorOptions)
```

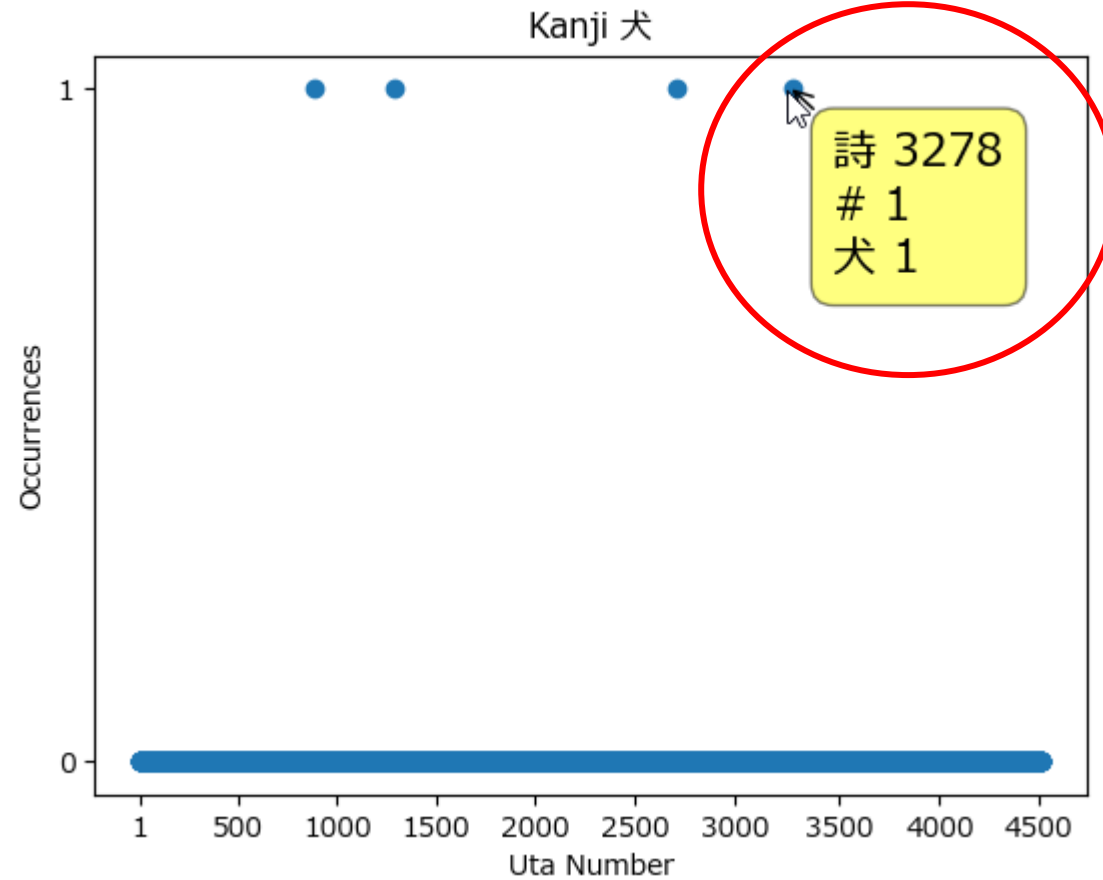
文字  
**Japanese Moji**

A toolkit to validate Japanese characters



# mplcursors

- Interactive data selection tooltips for Matplotlib



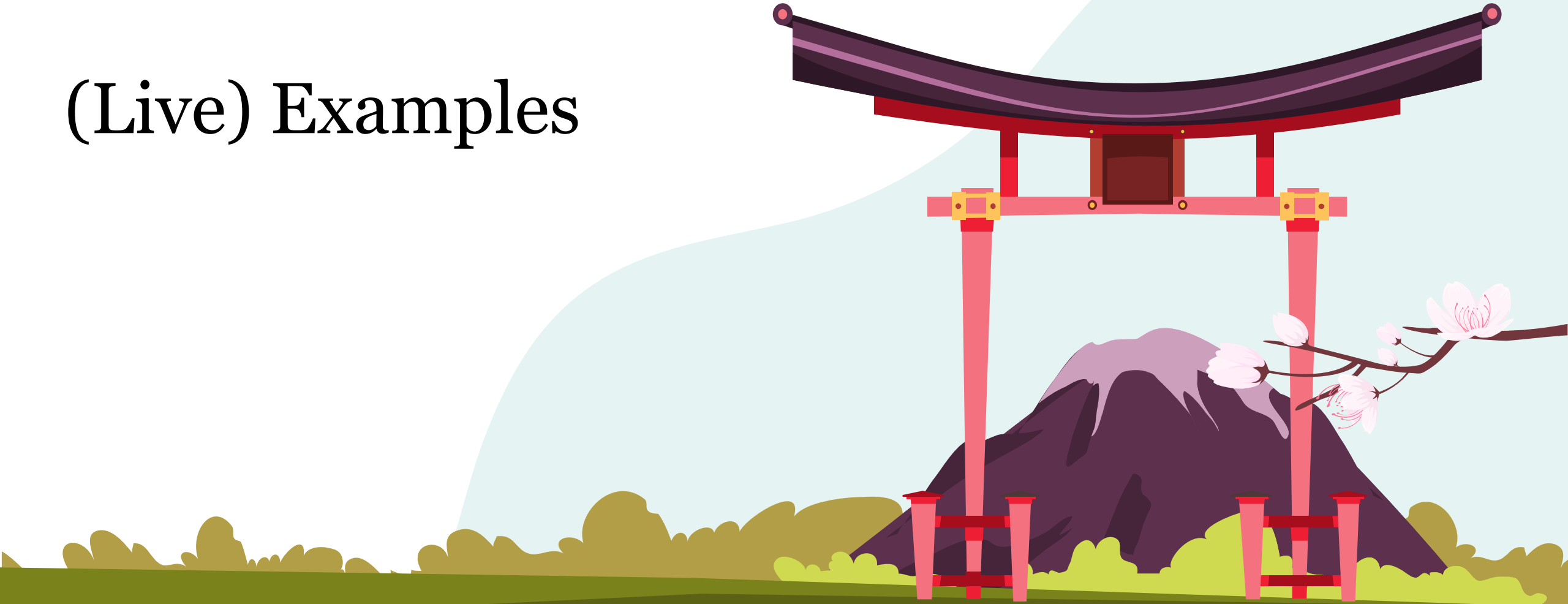


# Look at the Code





# (Live) Examples



行く 聞く 旅 波 田 衣 鹿 奈良 立つ 雁 原 海 人 越 間  
汝 思 へ 朝 磯 道 向 船 母 廻 照 御 欲 妹 鳴 来 楫  
夕 玉 藻 重 斎 八 荒 今日 置 漕 ぐ 降 上 知 行 霧 声 来 楫  
生 夢 寒 持 天 何 月 命 問 梅 雪 嶺 秋 高 着 行 寄 出 言 夜 浦  
隠 出 家 今夜 誰 天国 公 泣 繁 惜 大 和  
風 娘 大 離 吉 野 年 門 麻 呼 渡 名 海 水 十 刈 繼 去 辺 鏡 千 玉 瀬 萩  
待 居 島 袖 鳴 経 天地 宿 寝 時 春 鳥 心 玉 萩  
悲 しい

- Other
- JLPT N1
- JLPT N2
- JLPT N3
- JLPT N4
- JLPT N5

# Order of Examples



- TODO: zadefinovat order v kteremu budou veci ukazovany
- Případně mít vždycky slide k exemplu a něco napsáno ohledně toho co má být na slidu vidět – co je výsledek



# Conclusion





# Code & Slides

<https://gitlab.mff.cuni.cz/levyjak/visualization-project>





# References (1)



- Japonsko-český studijní znakový slovník, druhé vydání
- Manjóšú – Deset tisíc listů ze starého Japonska, díl první
- Manjóšú – Deset tisíc listů ze starého Japonska, díl druhý
- Pár much a já: malý výběr z japonských haiku
- <https://www.japanesewithanime.com/2019/11/mora.html>
- <https://www.japanesewithanime.com/2017/12/compound-kana.html>
- <https://ja.wikipedia.org/wiki/%E5%92%8C%E6%AD%8C>
- <https://www.britannica.com/art/haiku>



# References (2)



- <https://ja.wikipedia.org/wiki/%E4%B8%87%E8%91%89%E9%9B%86>
- <https://en.wikipedia.org/wiki/Man%27y%C5%8Dsh%C5%AB>
- <https://cs.wikipedia.org/wiki/Manj%C3%B3%C5%A1%C3%BA>
- <https://github.com/puppeteer/puppeteer>
- <https://github.com/azu/nlp-pattern-match/tree/master/packages/nlcst-parse-japanese>
- <https://github.com/arjunvegda/japanese-moji>



# References (Pictures) (1)



- [https://en.wikipedia.org/wiki/Katakana#/media/File:Table\\_katakan\\_a.svg](https://en.wikipedia.org/wiki/Katakana#/media/File:Table_katakan_a.svg)
- <https://en.wikipedia.org/wiki/S%C5%8Dgana#/media/File:Sogana.png>
- [https://en.wikipedia.org/wiki/Hiragana#/media/File:Table\\_hiragan\\_a.svg](https://en.wikipedia.org/wiki/Hiragana#/media/File:Table_hiragan_a.svg)
- <https://www.japanpowered.com/japan-culture/the-life-and-impact-of-matsuo-basho>



# References (Pictures) (2)



- [https://www.databazeknih.cz/img/books/13\\_/13545/big\\_par-much-a-ja-yD0-13545.jpg](https://www.databazeknih.cz/img/books/13_/13545/big_par-much-a-ja-yD0-13545.jpg)
- <https://www.istockphoto.com/cs/vektor/%C5%BEab%C3%AD-sk%C3%A1k%C3%A1n%C3%AD-izolovan%C3%A1-%C5%BE%C3%A1ba-sk%C3%A1%C4%8De-nab%C3%AD%C3%A9m-pozad%C3%AD-gm1130502836-299008045>
- <https://www.stockio.com/free-icon/nature-icons-blue-fish>
- [https://www.historyofcreativity.com/content/6/s\\_creator.001\\_1589061380.jpg](https://www.historyofcreativity.com/content/6/s_creator.001_1589061380.jpg)



# References (Pictures) (3)



- <http://www.c-d-f.cz/media/images/authors/antonin-liman.jpg>
- <https://cdn.aukro.cz/images/sk1622082133201/730x548/manjosu-dil-i-az-iv-komplet-101033023.jpeg>
- <https://user-images.githubusercontent.com/10379601/29446482-04f7036a-841f-11e7-9872-91d1fc2ea683.png>



# References (Miscellaneous)



- Template
  - <https://www.slideegg.com/japan-theme-presentation>
  - <https://www.shutterstock.com/cs/image-vector/set-handdrawn-traditional-japanese-symbols-red-418482400>
- Speech bubbles
  - <https://www.elearningdesigner.co.uk/freebies>

