

# An Attempt at Creating Integrated Retrieval for Chinese Excavated Materials: An Implementation of a Search Function across Interpretations of Ancient Characters

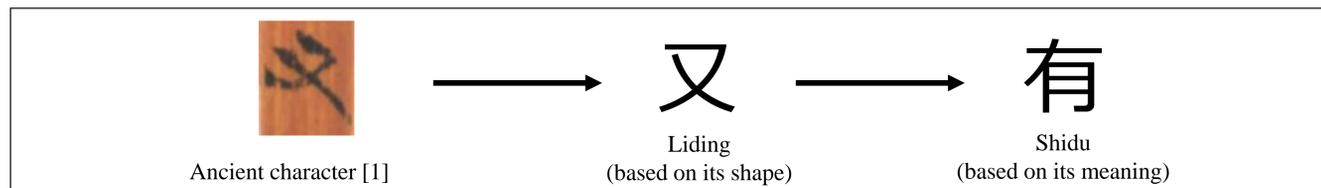
Shumpei Katakura (The University of Tokyo)

## 1. Background

- Many important excavated materials from the Warring States period (5th to 3rd centuries B.C.) have been rediscovered in mainland China since the 1950s.
- To interpret sentences and characters on new materials, we refer to other excavated materials. However, as of now, there are **not enough digitalized text data** about them.
- We should gather and accumulate many interpretation data of excavated materials in order to **create sufficient text data**. In addition, we should **create search function which can find characters and sentences easily through various interpretations** through this text data.

## 2. Interpretations of Excavated Materials

- There are mainly two types of interpretations: “Liding” (隸定) and “Shidu” (稊讀).
  - “Liding” is the interpretations **pertaining the shape of characters**: here, components of ancient characters are modified to those of Hanzi (漢字). (It is like “transcription”.)
  - “Shidu” is the interpretations **on the meaning of characters**: the use of an ancient character is examined as corresponds to that of Hanzi. (It is like “identification”.)



An example of “Liding” and “Shidu”.

- Generally, we can agree how to interpret as “Liding”. On the other hand, sometimes there are various “Shidu” interpretations on an ancient character by many researchers.

## 5. Future prospect

- We plan to incorporate this search system into our digital archive on development.

## 3. Original text data

- We have been creating original text data with metadata (e.g., split number, IDS information) for search convenience.
- For now, we prepare a dataset (including about 13,000 records) of an excavated material with multilayered interpretations. When we succeed in creating our search function, to begin with, we are going to demonstrate on the bases of this dataset.

An example of original dataset (Baoshang Chu Bamboo-slips 包山楚簡).

1	A	B	C	D	E	F	G	H
type	title	slip number	orders of appearance	Liding	IDS	Shidu(Jiagu)	note	
354	文書	集筭言	15	32	登	日入豆		
355	文書	集筭言	15	33	登	日藏土		
356	文書	集筭言	15	34	而			
357	文書	集筭言	15	35	無			
358	文書	集筭言	15	36	古		故	

## 4. Method

- In this search function, each character on the excavated materials has multilayered interpretation data.
- Suppose that there is a Liding sentence on an excavated material as “...又隈迺...”, and there are various Shidu interpretations about “隈” and “迺” by three researchers(A, B, and C). Here, our system finds this sentence using any search query as below.

- (1): To input “又隈”
- (2): To input “威苗”
- (3): To input “又魏”
- (4): To input “有\*逆”

	Liding	Shidu A	Shidu B	Shidu C
...	...	...	...	...
①	又	有	有	有
②	隈	魏	隈	威
③	迺	逆	苗	朝
...	...	...	...	...



Excavated material [1]

[1] Qinghua daxue cang zhanguo zhujian, “Bangjiachuwei” 清華大學藏戰國竹簡 邦家處位 slip number 10