中国出土資料テキストデータにおける隷定・釈読データ横断検索システムの実装

片倉 峻平（東京大学大学院 人文社会系研究科）

概要：近年新たに発見されている中国出土資料は古代中国研究において大きな価値を持つ．とりわけそこに記される文章の解読は急務であり，研究に資するためにデジタルデータの整理が求められている．本報告では出土資料文章の解読情報をデータとして構築した上で，求めたい解読情報をそこから不足無く検索する方法を提案する．将来的には，報告者が構築中の出土資料デジタルアーカイブにこの検索システムを実装した上で公開することを計画している．
キーワード：検索システム，データベース，デジタルアーカイブ，中国出土資料，漢字

Implementation of a Search System across "Liding" and "Shidu" Interpretations of Text Data on Chinese Excavated Materials

Shumpei Katakura (Graduate School of Humanities and Sociology, The University of Tokyo)

Abstract: Chinese-excavated materials, which have been rediscovered these days, possess a great value for studies related to ancient China. Primarily, we should interpret the sentences on them and organize digital data about them for the sake of our studies. This study proposes a technique for creating data of interpretations concerning sentences on excavated materials and to search enough information from this data. In future, we plan to incorporate these data and search functions into our digital archive on development and release them to the public.
Keywords: Search function, Database, Digital archive, Chinese excavated material, Hanzi

１．はじめに

本報告は，中国出土資料に現れる古文字で記された文章において，各文字に解読データを重層的に付与した上で横断的に検索出来るシステムを実践するものである．現在報告者が構築中の出土資料デジタルアーカイブ[1]に本システムを搭載することで，諸説存在する古文字の解読情報を網羅的に検索することが可能となり，出土資料研究の進展を促すことを期待する．報告者は既に本システムの検索方法の概要について提案を行っており[2]，本報告はその内容も含めた上で更に検討を深めたものである．

報告では構築したシステムの一部を参加者に配布し，実際に操作を体験してもらうことを想定している．なおシステム自体は本稿執筆時点では目下構築中であり，報告時に配布する際には以下の内容を予定しているという前提で論述を進める．

２．中国出土資料

２．１ 中国出土資料の概要

1950 年代以降，中国大陸では戦国時代（前 5～前 3 世紀）の出土資料が発見され続けている[3]．「包山楚簡」「郭店楚簡」「清華大学蔵戦国竹簡（清華簡）」などに代表されるこれら同時代資料は古代中国研究全体にわたり非常に重要視されている．資料の書写材料はそのほとんどが竹簡

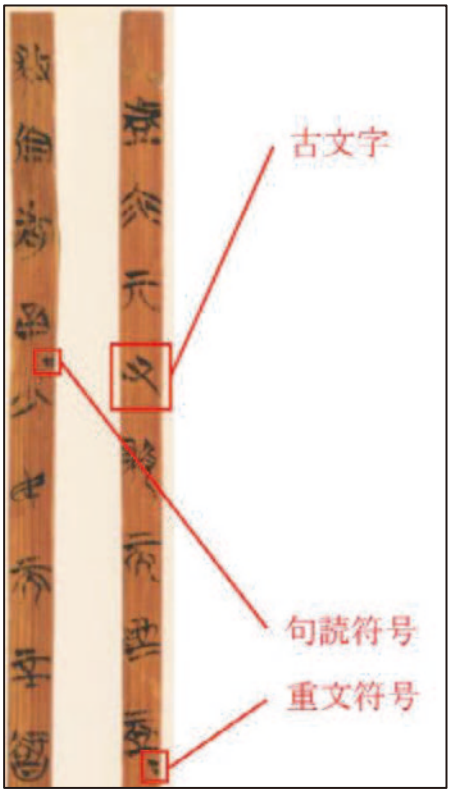

図 1 出土資料の図版（図版は清華簡「邦家處位」10 号簡[4]より）

（書写材料として用いられた竹の札）であり，その表面に文字や符号が墨で記されている．そこに現れる古文字には，現在の我々の用いている漢字とは用法が大きく異なるものも散見され，解読は容易ではない．そのため現在でも議論が紛糾している古文字は多く存在している．こうした難解な古文字の解読を進めるためには，類似した文字や文脈が他の資料でどのように現れ，各研究者からどのように解読されてきたのかという情報を大きく参照する必要が生じる．一方でこうした解読情報を網羅的にまとめているデータベースやウェブサイトなどは整っておらず，コンピュータを利用した円滑な情報参照環境は準備されていない．本システムの構築理由も，この状況を打開したいがためである．

出土資料は中国大陸や台湾などの所蔵機関が保存しており，出版物によりその図版を確認することが出来るものの，デジタル画像などの情報は一部資料（居延漢簡[5]など）を除いて未だオープン化されていない．そのためこれらオープン化されていない資料をウェブ上で容易に閲覧することは現時点では不可能となってしまっている．

２．２ 資料の文章構造

図 1 に出土資料竹簡の図版の一部を示す．基本的にはこのように，文字ごとにある程度の間隔を空けて縦一行で古文字が記されることが多い．稀に一つの竹簡に複数行が記されることもある．また句読符号や重文符号（合文符号とも呼ばれる）など文字以外の情報も併記される．句読符号は章句の区切りを示すものであり，点や鉤のような形で示されることが多い．図 1 では点で示されている．重文符号はその文字が複数回読まれる場合（重文）や，複数の文字が一つの文字に合成されるなどした結果一つの文字を異なる複数の文字として読む場合（合文）を示すものであり，下駄記号「=」に似た形で示されることが多い．重文符号の付いた文字が重文か合文かは文脈に応じて判断される．こうした符号情報は文字情報と並んで非常に重要であるため，データとして記述しておく必要がある．なお文字学上の定義とは別に，本報告では議論の便宜のために「同じ文字として複数回読まれる古文字」を「重文」と呼び，「異なる文字として複数回読まれる古文字」を「合文」と呼ぶこととする．

２．３ 古文字の解読

出土資料上の古文字を解読する際に用いられるのが，「隷定」「釈読」と呼ばれる情報である．「隷定」は古文字の構成要素を現在の我々が用いる漢字の字形（楷書体）に改めたもので，つまり形に関する解読情報である．「釈読」とはその古文字が現在我々の用いるどの漢字の意味に該当するかを判断したもので，つまり意味に関する解読情報である．図 2 にこれらの関係を示した．隷

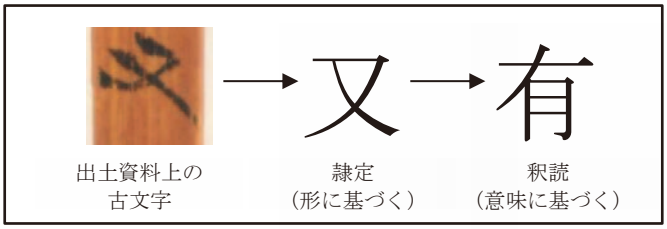

図 2　「隷定」と「釈読」の関係　（図版は清華簡「邦家處位」十号簡[4]より）

定は客観的な字形に基づく解読であるため研究者間の見解は一致しやすいが，釈読は意味に基づく解読であるため研究者それぞれの文章の読み方によって意見が大きく割れる場合も少なくない．隷定の見解が割れる場合は，筆法の複雑性や曖昧性からそれぞれの構成要素をどの形として捉えるかという見解の相違や，もしくはどこまで古文字の形を楷書体に反映させるかといった方針の相違が，その主な理由となる．

この解読情報は，資料の文字情報をテキスト情報化する際に中心となるものである．

３．文字情報のデータ記述

３．１ データ記述の基本ポリシー

本報告で提案するシステムに搭載してあるデータは，検索用途のために報告者が独自に作り上げたものである．現時点でのプロトタイプとしてのデータは単純な表に基づいており，古文字一つを単位としてレコードを構成し，その古文字に付随する様々な情報を各フィールドに付与している．フィールドは主に「メタデータ」と「解読情報」に分けられ，システム構築の上でより肝要となるのは後者である．以下では，この文字情報にまつわるデータのことを「古文字データ」と呼ぶ．

また，本システムでは，出来る限り検索漏れを無くすというポリシーを前提としている．検索漏れと検索ノイズはトレードオフの関係にあるが，検索ノイズが多くなろうとも，不足のない情報をユーザに提供することを旨とした設計となっており，古文字データ構築もこの方針に則る．

３．２ メタデータ

各レコードそれぞれに，登場する資料・篇・出現順・符号の有無などのメタデータを記述している．符号の有無は，その符号が置かれた直前の文字のレコードに示している．これは，重文符号がその直前の文字に対して機能していることを参考として，その方針を他の符号にも適用させたものである．

３．３ 解読情報

各レコードには「隷定」「釈読」の解読情報をそれぞれ記述する．なお，先に述べた通り釈読は研究者によって意見が割れることがしばしばであり，そのため各レコードには複数の研究者による各釈読案を併記しておくことが望ましい．

３．３．１ 隷定データ

隷定データは「隷定（単字）」と「隷定（IDS）」という二つのフィールドに記述を行う．隷定は古文字の字形およびその構成を楷書体に改めたものであるため，図 2 の「又」のように馴染みのある単字として記載出来る場合もあれば，図 3 のように馴染みのない複雑な形となる場合も多い．そのため隷定した文字の単字テキストデータが存在せずにテキストデータとして入力が出来ない場合も散見される．こうした場合は IDS（Ideographic Description Sequence）[7]を用いることで，複雑な隷定でもテキストデータとして検索可能とするようにしている．IDS というのは，漢字の配置構造と構成要素を並べることで一つの漢字を表現出来るような文字列のことであり，例えば「校」字は，IDS で「⿰木交」と表現することが出来る．図 3 の文字の場合は，そのレコードの「隷定（IDS）」フィールドに「⿱⿰采羊攵」という IDS でのデータを記述する．この時「隷定（単字）」フィールドは単字テキストデータが存在しないため空である．

隷定時の単字のテキストデータが存在する場合でも，構成要素での検索を可能とするため IDS データを併記することがある．例えば「忻」と隷定される文字は単字のテキストデータが存在するが，「心」「斤」という構成要素を持つ字でもあり，それぞれの構成要素での検索を可能とするため「隷定（IDS）」フィールドには「⿰心斤」と記述する．「又」のようにこれ以上の分解が難しく IDS 記述が困難な文字の場合は，「隷定（IDS）」フィールドは空となる．

先に述べたが，基本的に隷定は見解が一致するデータであるため，後述する釈読データのように（３．３．２），各隷定案ごとにフィールドを分けることは行わない．ただし例外的に以下に挙げるような状況が存在するため，それぞれの方針を示しておく．

隷定が不可能な古文字の場合は，隷定フィールドは単字・IDS ともに空欄とする．例えば図 4 の古文字は楷書体に該当する要素が存在せず隷定が不可能となっている．一方でこれが「冥」字に関わることは多く認められ[8]，すなわちまず「冥」に釈読出来ることは判明しており，そこからさらに発展的に様々な釈読が行われることが多い文字である．そのため釈読データを記述することで古文字情報を検索対象とすることは可能となっている．

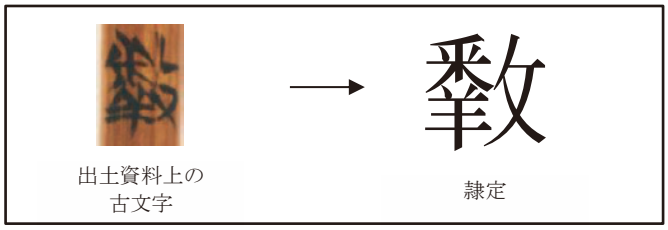

図 3　複雑な隷定の例（図版は清華簡「邦家處位」二号簡[4]より）

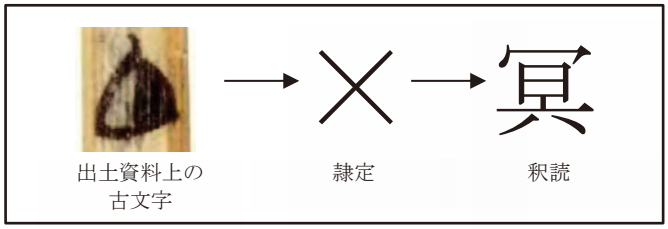

図 4　隷定の存在しない文字の解読　（図版は上海博物館蔵戦国楚竹書「容成氏」三十七号簡[6]より）

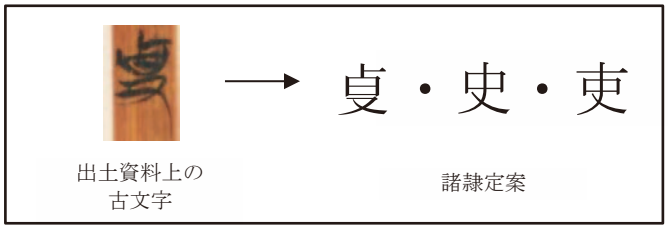

図 5　隷定の割れる古文字の例（図版は清華簡「邦家處位」一号簡[4]より）

隷定の解釈が複数存在する場合は，同一隷定フィールド内に複数の解釈を併記する．例えば図 5 に示す古文字の隷定は，上部・中部・下部の個々の構成要素をもっとも単純に組み合わせた「叟」のほか，全体の構成を考慮した上で単字に反映させて「史」や「吏」と記述される場合もある．この場合，この古文字のレコードでは，「隷定（単字）」に「史吏」，「隷定（IDS）」に「⿱卜日十」とそれぞれ入力しておく．（「叟」は単字のテキストデータを持たないため「⿱卜日十」と

表 1 各隷定データの記述例

	隷定（単字）	隷定（IDS）	釈読 α	備考
…	…	…	…	…
①	又		有	
②		⿱⿰采羊攵	弊	
③	忻	⿰心斤	忻	
④			冥	
⑤	史吏	⿱卜日十	使	資料 α では「史」と隷定され，資料 β では「吏」と隷定され…
…	…	…	…	…

IDS を用いる．）このような記述は古文字データの煩雑さを強くしてしまうが，今回の古文字データは本システムの検索用途に作成したもので検索漏れを防ぐことを第一としているため，この方針で隷定字の検索漏れに対応する．隷定の解釈が複数存在した場合のそれぞれの解釈の典拠は，同一レコード内の「備考」フィールドに記す．このフィールドは検索対象ではなく，検索結果を示す際に併せて提示することでユーザに典拠の情報を与えるものである．

表 1 に，上述した各例における古文字データの一部記述例をそれぞれ示した．①は隷定データが IDS に分解出来ない例，②は隷定データの単字のテキストデータが存在しない例，③は単字のテキストデータが存在し且つ IDS にも分解出来る例，④は隷定が不可能で釈読のみ存在する例，⑤は隷定の解釈が複数存在する例である．

３．３．２ 釈読データ

釈読データは，典拠となる資料ごとにカラムを設け，それぞれを個別に記述する．つまり典拠となる資料が増えれば増えるほど，釈読データとしてのカラムが増加することになる．

釈読情報として記述するデータは，結果としてどのような文字に同定されたのか，という最終情報のみとする．釈読に至る過程では，「古文字がどのような文脈で登場するのか」「どのような音韻情報を持ちうるのか」など様々な検討が行われ，これら情報も重要なものであるためデータとして登録することが出来れば有意義に働き得る．しかしこうした情報の記述量が増えれば増えるほど，それに伴いデータの構築難易度も高くなってしまう．そのため今回利用するデータにおいては，釈読の最終情報に加えその典拠（書籍・論文・BBS など）を特定出来る情報を併記するに留め，それぞれの解読過程の情報はユーザが必要に応じて参照出来るような環境を整えておく．釈読資料の典拠情報は，古文字データ内ではなくシステムの別の箇所でまとめて示すこととする．

釈読では，古文字の多くは伝世文献に既出の文字に解釈される．そのため釈読データは単字のテキストデータが既に存在している場合がほとんどで，隷定データのように IDS を用いて記述することは少ない．こうした前提から，釈読データのフィールドに記述を行う場合は，基本は一文字のみが登録されることとなる．

釈読データのフィールドが空欄の場合も多く存在する．これは，そのカラムが基づく典拠資料が部分的な情報しか提示しておらず，該当レコードでの釈読情報が不明な場合である．

古文字の隷定と釈読のデータが一致する場合は，釈読データには隷定データと同じ記述を行う．

３．４ 重文符号情報

ある古文字に重文符号が付される場合，その文字は重文もしくは合文と認められる．重文の場合は「同一文字（あるいは熟語）が複数回読まれる」，合文の場合は「一つの文字が異なる複数の文字として読まれる」という状況であり，それぞれに対応したデータ記述を行う必要がある．

本システムでは古文字一つを単位としてレコードを構成しているため，重文・合文であってもレコードは一つしか持たせない．

３．４．１ 重文の場合

例えば「AB₌C」（「＝」は重文符号）という文字列において「ABBC」と読まれる場合が重文である．このとき B の持つ文字情報は「B」だけに過ぎず，これを「BB」と複数読ませているのは重文符号の機能に依るものである．従ってこの場合，重文符号の存在を記述した上で，B のレコード内の隷定・釈読の各解読データフィールドには「B」一文字の情報のみ記述する．表 2 にはこの場合の記述例を示した．A'及び A"は，A に隷定される文字についての典拠資料 α，β の釈読をそれぞれ示している．（B と C においても同様．）

また「AB₌C₌D」という文字列において共に重文符号の付いた「BC」が熟語として重複し「ABCBCD」と読まれる場合もある．これも同様に B および C のレコード内の各解読データフィールドには「B」および「C」一文字ずつの情報しか記述しないこととする（表 3）．

表 2 重文を含む文字列「AB₌C」の記述例

	隷定（単字）	釈読 α	釈読 β	重文符号
…	…	…	…	
①	A	A'	A"	
②	B	B'	B"	○
③	C	C'	C"	
…	…	…	…	

表 3 重文を含む文字列「AB₌C₌D」の記述例

	隷定（単字）	釈読 α	釈読 β	重文符号
…	…	…	…	
①	A	A'	A"	
②	B	B'	B"	○
③	C	C'	C"	○
④	D	D'	D"	
…	…	…	…	

３．４．２ 合文の場合

例えば「AB₌C」という文字列において「B」が「XY」という字形を持ち「X」「Y」の合文と見なされて「AXYC」と読まれる場合がこれである．この時 B はレコードを一つしか持たない一方で「X」「Y」という複数の文字情報を有するため，重文符号の存在を記述した上で，B の各釈読フィールドには「XY」と二文字の記述を行う（表 4）．

合文は三文字の情報を含む場合もあり，その際は釈読フィールドに三文字の記述がなされる．

表 4 合文を含む文字列「AB₌C」の記述例

	隷定（単字）	隷定（IDS）	釈読 α	重文符号
…	…		…	
①	A		A'	
②		⿲XY	XY	○
③	C		C'	
…	…		…	

4．本報告で用いる古文字データ

本報告では，包山楚簡という資料を元に報告者が作成した古文字データを用いる．その総レコード数は資料の総文字数と同じ 13,000 弱である．この古文字データには，『包山楚簡解詁』[9]や『楚地出土戰國簡册[十四種]』[10]などを参照した解読情報を記述している．

5．横断検索システム

5．1 検索方法の概要

本システムは隷定及び各釈読データの横断検索を可能とする．例えば，隷定字が「…又隈迖…」と並ぶ出土資料文章があり，このうち「隈」字と「迖」字の釈読が典拠資料 α，β，γ により割れていると仮定する．（この文字列は説明の便宜のため作成したもので，実際の出土資料に登場するものではない．）この時の各文字のレコード例を表 5 に示す．

表 5 「…又隈迖…」の隷定・釈読データ

	隷定	釈読 α	釈読 β	釈読 γ
…	…	…	…	…
①	又	有	有	有
②	隈	魏	愄	威
③	迖	逆	苗	朝
…	…	…	…	…

この事例において，本システムでは以下のいずれの検索クエリでもこの文章の一部を見つけ出すことを可能とする（表 6）．

(1)：「又隈」で検索…①②の隷定を抽出．

(2)：「威苗」で検索…②の釈読 γ と，③の釈読 β を抽出．

(3)：「又魏」で検索…①の隷定と，②の釈読 α を抽出．

(4)：「有*逆」で検索…①の釈読 α，β，γ と，③の釈読 α を抽出．「*」は 1 文字以上の任意の文字列を示すワイルドカードとして機能している．

表 6 「…又隈迖…」データの横断検索イメージ

	隷定	釈読 α	釈読 β	釈読 γ
…	…	…	…	…
①	又	有	有	有
②	隈	魏	愄	威
③	迖	逆	苗	朝
…	…	…	…	…

(1)のような隷定だけの検索，**(2)**のような釈読だけの検索，**(3)**のような隷定と釈読をまたいだ検索，**(4)**のような文字をまたいだ検索をそれぞれ可能とする．**(1)**・**(2)**・**(3)**はそれぞれ熟語検索に有効であり，**(4)**は呼応・共起する単語を抽出する検索に有効となる．釈読データを多く登録すればするほど，ユーザにはその古文字に対する情報を不足無く提示することが出来る．

5．2 機能の概要

先の「又隈迖」が「X 楚簡という資料の中の Y 篇 2 号簡 3〜5 文字目に出てくる文字」である場合，**(3)**「又魏」での検索において表 7 のような検索結果画面を出力する．

この検索システムの持つ機能例は以下の通りである．

【機能 1】元データ全体が確認出来る．

検索結果画面では，ヒットしたフィールドのレコード全体をまず表示して，その中で検索対象となったものの背景色を変える．

こうすることでマッチした部分がレコードのどこに位置づけられているのかを直感的に理解出来るようにする．解読データのみならず出現資料情報や重文符号の有無などのメタデータも同時に確認出来る．

表 7 「又魏」検索結果画面の出力イメージ

資料情報（メタデータ）	隷定（単字）	隷定（IDS）	釈読 α	釈読 β	釈読 γ	重文符号	備考	画像情報
X 楚簡 Y 篇 2 号簡 3文字目	又		有	有	有			画像（又）
X 楚簡 Y 篇 2 号簡 4文字目	隈	⿰畏邑	魏	愄	威			画像（隈）

【機能 2】検索対象の関連データを参照出来る．
検索対象となった古文字のレコードだけでなく，その古文字に関連するデータを簡単に参照出来るようにする．
例えば篇全体の文章を知りたい場合は「X 楚簡 Y 篇」をクリックすればX 楚簡 Y 篇の全レコードを表示出来るようにしており，また前後の文脈のみを知りたい場合は「2 号簡」をクリックすればX 楚簡 Y 篇の 2 号簡に絞ったレコード群が表示される．隷定「隈」をクリックすると，「隈」と隷定・釈読されるデータが資料横断的に羅列される．
画像情報をクリックすれば，文字単位の切り抜き画像だけでなく簡全体の画像に飛べるようになっており，全体像を確認出来る．

【機能 3】検索時にオプションを選択出来る．
古文字データを検索する際には，出力形式を指定したり検索対象を絞り込んだりするようなオプションの選択を可能とする．
例えば，チェックボックスの ON / OFF などで画像情報の表示の有無を検索時に指定出来るような環境を整えておく．これは，画像を出すと動作が重くなったり画面を逼迫してしまったりするので，ユーザそれぞれの目的に応じて選択出来るようにするためである．
ほか，「隷定のみを検索対象とする」「釈読のみを検索対象とする」など解読データの種類による絞り込みを選択する機能も設ける．また，IDS フィールドを検索対象にするかどうかもオプション化することで，単字検索と偏旁検索の選択も可能とする．

５．３ 検索アルゴリズム

基本的な検索アルゴリズムとしては，まずクエリとして入力された文字列の最初の文字について，各レコードの隷定・釈読データに対して検索を行う．該当するデータを格納するレコードが存在した場合，クエリの次の文字列について，直後のレコードの隷定・釈読データに対して検索を行う．これを繰り返し，全ての文字列が連続したレコードに存在した場合，そのレコード群を検索結果として提示する．
表 6**(3)**「又魏」を例にこれを説明すると，まずはクエリの最初の文字「又」について検索を行い，レコード①の「隷定」フィールドにこれが存在することを確認する．この場合，クエリの次の文字「魏」が，レコード①直後のレコード②に存在するかどうかを検索し，これが「釈読 α」フィールドに存在することを確認する．クエリの文字列はこれで終了するため，検索結果としては表 7 のように連続したレコード①②を示すこととなる．
以上が基本的なアルゴリズムであるが，以下では例外的なものをいくつか示す．

５．３．１ 重文符号を持つレコードの検索

重文・合文においても不足なくデータを抽出するために，「重文符号」のフィールドに記述がある場合は，そのレコードでは特別な判断を行う．
ヒットしたデータが存在するレコードにおいて重文符号のフィールドに記述がある場合，まずそのレコードの各釈読フィールドに記述されている文字数が一文字であるかどうかを調べる．その結果，一文字である場合はそのレコードに該当する古文字は重文，複数文字が記述されている場合は合文であると判断する．これは，重文の場合は釈読フィールドには基本的に一文字のみが記される（３．４．１）という前提に依る．

５．３．１．１ 重文の検索

重文の場合は，その重文が単字で完結するものなのか（３．４．１ 「AB=C」→「ABBC」の例），それとも熟語に跨るものなのか（３．４．１ 「AB=C=D」→「ABCBCD」の例）を判断する必要がある．そのため，ヒットしたフィールドが存在するレコードの前後のレコードにおいても重文符号のフィールドに記述があるかを確認し，前後どちらにも記述がない場合は単字の重文，前後いずれかに記述がある場合は熟語の重文と判断する．
単字の重文の場合は，クエリの次の文字について直後のレコードに存在するか確認するという通常の手順に加え，再度同一レコードに存在するか確認するという手順も重ねて行う．表 2 を例に説明すると，「B○」（○は任意の一文字）という検索クエリの場合，一文字目の「B」がレコード②に存在し且つ②が重文符号に記述のあるレコードであるため，クエリの次の文字列「○」の検索対象範囲はレコード②と③となる．このアルゴリズムにより，「ABBC」と読まれる表 2 の部分において「BB」「BC」のどちらでも検索を可能とする．
熟語の重文の場合は，まず前後何行にわたって重文符号フィールドに記述のあるレコードが存在しているのかを確認し，クエリ最初の文字列がヒットしたレコードが連続する重文符号レコードの何番目なのかを判断する．その結果，連続する重文符号レコードの最終番目以外であれば，クエリの次の文字について直後のレコードに存在するか確認するという通常の手順を行う．重文符号に記述のある連続するレコード群の最終行であれば，クエリの次の文字について直後のレコードに存在するか確認するという通常の手順に加え，連続する重文符号レコードの開始行に存在するか確認するという手順も重ねて行う．表 3 を例に説明すると，「C○」という検索クエリの場合，

一文字目の「C」がレコード③に存在し且つ③が重文符号に記述のある連続するレコード群（②③）の最終行であるため，クエリの次の文字列「○」の検索対象範囲はレコード④と②となる．このアルゴリズムにより，「ABCBCD」と読まれる表 3 の部分において「CB」「CD」のどちらでも検索を可能とする．

５．３．１．２ 合文の検索

合文の場合は，クエリの次の文字について直後のレコードに存在するか確認するという通常の手順に加え，再度同一レコードに存在するか確認するという手順も重ねて行う．これは単字の重文と同様のアルゴリズムである．このアルゴリズムにより，「AXYB」と読まれる表 4 の部分において「XY」「YB」のどちらでも検索を可能とする．（なおこの場合，検索クエリ「XB」もヒットするため，検索ノイズとなってしまう．）

５．３．２ IDS 内の検索

ヒットしたデータが「隸定（IDS）」フィールドに存在した場合，クエリの次の文字について直後のレコードに存在するか確認するという通常の手順に加え，再度同一レコードに存在するか確認するという手順も重ねて行う．これは単字の重文と同様のアルゴリズムである．表 1 を連続したレコード群として説明すると，「又釆攴」というクエリでは文字列順に①②②のレコードに存在が確認出来るため，レコード①②を検索結果として提示する．同様に，「又釆忻」では①②③のレコードに，「羊心斤」では②③③のレコードに，「冥日十」では④⑤⑤のレコードにそれぞれ存在が確認出来るため，該当したレコードを検索結果として提示する．

６．デモシステムの公開方法

本報告では，研究利用目的としてデータを配布した上でスタンドアロンでも動作する公開方法を摂る．多くのデジタルアーカイブのような，ウェブ上にサーバを置きユーザが接続して利用するというやり方は採らない．これは，ネットワークセキュリティが未だ充分に担保出来ていないということに加え，図版データがオープン化されていないためにウェブ上での公開を行うと権利問題が発生する可能性があるからである．いずれこれら課題（特に権利問題）が解決された折には，ウェブ上でデジタルアーカイブとして公開を行い本システムの操作を可能とすることも想定している．

７．あとがき

本検索システムは古文字の解読情報を不足無く拾い上げることを可能としており，本システムを搭載したデジタルアーカイブが構築することで今後の出土資料研究の能率を大きく向上させることを期待している．また古文字データの作成も今後継続して行う必要があり，構築中の出土資料デジタルアーカイブには本システムと併せて搭載することで効果的に運用出来るよう考えている．

課題もいくつか残っている．例えば，異体字や隸定部品の抽象度に関わる問題である．表 1 の②の「攵」字などは「攴」と表されることもあるため，どちらの入力においてもヒットさせるような環境を用意しなければ検索漏れが生じてしまう．また「辶（辵）」は古文字では「彳」と「止」との組み合わせで表現されるため，「彳」+「止」=「辶（辵）」となるような検索環境も整える必要がある．こうした課題を解決するためには，鈴木(2018)[11]の行うように，従来存在している異体字テーブルに加えて古文字検索に対応した独自の異体字テーブルを搭載させる必要がある．

検索の元となる古文字データも，現在用いているような単純な表は，長期的な運用や拡張を見据えると最適なものだとは言えないだろう．本報告で示した方針に基づいてより優れたデータベースを追求する余地が存在する．

謝辞

本研究は科研費 21J12110 の助成を受けたものです．また大西克也先生（東京大学大学院）には主に出土資料・文字学分野に関して，大向一輝先生（東京大学大学院）・永崎研宣先生（人文情報学研究所）には主に情報処理分野に関して，大きくご助言を頂きました．深く御礼申し上げます．

参考文献

[1] 片倉峻平. 新規「新出土資料デジタルアーカイブ」の課題と提案. 日本漢字学会報 3. 日本漢字学会, 2021, pp.93-111.
[2] Shumpei Katakura. An Attempt at Creating Integrated Retrieval for Chinese Excavated Materials: An Implementation of a Search Function across Interpretations of Ancient Characters. JADH2021, Tokyo, 2021-09-08, https://www.hi.u-tokyo.ac.jp/JADH/2021/programme.html, (参照 2021-10-24).
[3] 中国出土資料学会編. 地下からの贈り物:新出土資料が語るいにしえの中国. 東方書店, 2014.
[4] 馬承源. 戰國楚竹書 : 上海博物館藏 2. 上海古籍出版社, 2002.
[5] “史語所藏居延漢簡資料庫”. https://wcd-ihp.ascdc.sinica.edu.tw/woodslip/, (参照 2021-10-24).
[6] 李学勤, 清華大学出土文献研究与保護中心. 清華大学藏戦国竹簡 8. 中西書局, 2018.
[7] “The Unicode Standard Version 6.0 – Core Specification”. http://www.unicode.org/versions/Unicode6.0.0/ch12.pdf, (参照 2021-10-24).
[8] 周波. 說上博簡《容成氏》的“冥”及其相關諸字. 復旦大學出土文獻與古文字研究中心, 2020-06-23, http://www.gwz.fudan.edu.cn/Web/Show/4588, (参照 2021-10-24).
[9] 劉信芳. 包山楚簡解詁. 藝文印書館, 2003.

[10] 陳偉等. 楚地出土戰國簡册[十四種]. 経済科学出版社, 2009.
[11] 鈴木慎吾. 『切韻』諸本テキスト一覧システムの構築について. じんもんこん 2018 論文集. 2018, pp.117-122.