

法帖文字データベースの設計

山本 毅雄、朱 平*

図書館情報大学 図書館情報学部

法帖（書道の手本）中に含まれる漢字、およびこれに関連する種々の情報を格納するデータベースの論理設計をおこなった。東洋の伝統芸術である書道には、歴史の流れに伴って、書道文字と書作品に関する大量・複雑な情報が存在する。これらの情報の相互関係を明らかにし、拡張ER（実体・関連）モデルを利用して、正規化および一貫性制約に留意しつつ、法帖中の各文字レベルでのデータを持ち、これと作者・作品・時・刊本その他の情報を関連づけて記述しうるデータベース概念モデルを設計した。さらに、このうち文字データないし数値データで表現できる部分について、論理データベースモデル（関係データベースモデル）への変換をおこなった。

DESIGN OF A CALLIGRAPHIC CHARACTERS DATABASE

Takeo YAMAMOTO and Ping ZHU*

Department of Library and Information Science
University of Information Science

Tsukuba Academic City, Ibaraki 305, Japan.

A logical design of a database for Chinese characters in classical calligraphic texts and for related information is presented. Throughout the long history of Oriental calligraphy, a large and complex body of complicated information has been accumulated on calligraphic works and on characters contained thereof. Definitions were given for some of the relevant concepts. By using the Extended Entity-Relationship Model, a conceptual database model was designed which can hold data on each character in classical calligraphic texts (hojo in Japanese), and which describes relationship among the character and the calligrapher, the individual calligraphic work, time, publication and so on. The normalization and integrity constraints were discussed. Textual and numeric part of the data model was converted to a logical database model (relational model).

* 現所属：株式会社エルマ(ERMA Inc.)

1 はじめに

書道は、中国・朝鮮半島・日本の東アジア諸国において、古くは共通していたが、現在ではそれぞれに特色をもって発展している芸術である。芸術を美しく書くこと・これを見ることを楽しむ類似的芸術形態は、他にもイスラム諸国、インド、あるいは西欧(calligraphy)など、世界の多くの文化圏に存在している。

これらの文字芸術に関しては、歴史の流れに伴って、大量・複雑な情報が発生し、保存されてきている。

本研究では、たとえば博物館や図書館、出版社、個人蒐集家などが所蔵する法帖(書道の手本)のデータベース化に際しての、データベース構造の設計の問題をとりあげる。これは当然、画像を含むマルチメディアデータベースとなるが、本研究では画像データの問題については触れず、法帖と法帖文字に関する種々の情報を、各文字のレベルに至るまで微細に解析し、これを格納して利用可能とするデータベースの論理設計を行う。これらの情報の相互関係を明らかにし、拡張ER(実体・関連)モデル[1]、[2]、[3]を利用して、正規化および一貫性制約に留意しつつ、法帖中の各文字レベルでのデータを持ち、これと作者・作品・時・刊本・文字種その他の情報を関連づけて記述しうるデータベース概念モデルを設計する。さらに、このうち文字データないし数値データで表現できる部分について、論理データベースモデル(関係データベースモデル)への変換を行う。

2 拡張ERモデルによる概念設計

2.1 マクロERスキーマ

実世界に存在する法帖文字および書道に関するさまざまな情報と、それらの間の関連は、2.1.1~2.1.3節に示すような状況である。これを整理し、実体型と関連型の名称を与えて、図1のようなマクロERスキーマ[3]を設定した。ここで、長方形は実体集合をあらわす。菱形は関連集合をあらわし、その構成要素である実体集合と線で結ばれている。拡張ERモデルを利用して、作者と作品の直積関係からなる作品単位の実体集合や、文字と同種文字の直積関係からなる文字種の実体集合を表わしている。各実体集合の定義を2.1.4節に示す。

2.1.1 法帖とは

法帖とは、書の鑑賞あるいは学習のために、名家の書の真跡、あるいはその由緒ある複製等を集めて、何等かの技術で製本したものをいう。

毛筆で書いた文字を書といひ、その書き方を学び研究することを書道という。書道の歴史は、中国の殷王朝(紀元前約1500年)に始まるといわれている。

後漢時代から、名家の墨跡(書跡)を珍重し、保存して鑑賞・研習することが始まった。それらの媒体としては、古くは絹帛・紙・竹片などが用いられた。このうち、竹片に書かれたものを除いて、個々の書作品を古人は「帖」と呼んだ。これが、一人の作者の一つの書作品に対応する。ただし、作者は不明のものがあり、また正確には不明であるが、伝承によってある書家に属するとされているものも多い。

もとの墨跡(真跡)は散逸しがちであるので、これらを金属・石材・木板などに写し取ってして刻し、これを紙に刷り取った「拓本」が作られた。碑文、石刻、鐘銘文などのように、もともと岩石または金属上にあらわされた文字(金石文)も、多くは拓本として流通した。また、真跡の上に紙を重ねて輪郭を写し取り、後に内部を墨でぬり潰す、1枚の紙を剝がして2枚にする、など巧緻な方法によって複製したり、書家が真跡を手本にして

書く(臨書)ことが行われた。

これらの複製ないし模写は、もとより真跡よりは価値が下がるが、時代がたつて真跡が失われたり、得難くなると、複製や模写によってのみわずかに古人の作品の面影が伝えられる状況となり、それ自体が真跡に近い高い価値を有するようになった。王羲之の「蘭亭序」など主要な書作品には、後代の有名な書家による多くの臨書があり、これら後代の書家の作品として独立の価値を認められているものも稀でない。

こうして、これらの複製や模写が、さらに上記のような種々の方法でくりかえし複製され、模写される結果となった。これらの作業は、まがいものを作る目的で密かに行われた場合と、皇帝や有力政治家、富豪などの命によって公けに行われた場合があった。いずれにしても、時代が下つて原本が散逸するにつれ、その価値が増大するのが一般である。しかし時には、碑文の発掘や古文書の発見などにより、より原本に近い形のもの新たに発見されることもある。

法帖は、歴史上の時々、これらの真跡ないし複製を集めあるいは新規に作成して、編集・製本して作られた。ここでは、この作業をした人(依頼・命令・後援したものを含む)を、「出版者」とよぶことにする。これらは、古くは手作業で作られたため、部数も一部ないしきわめて少なかったが、最近では写真版など近代的複写・印刷技術を用いて大量複製される場合もあり、どちらも「法帖」と呼ばれている。ここでは、両者を特に区別しないて扱う。

2.1.2 法帖の内容

法帖には、一人の作者(書家)の、一つの作品を中心に編集されているものが多いが、それには限らない。

古い著名な法帖には、多くの場合、編者またはその後援者(出版者)、あるいはその後代々の所有者などによって、序や跋が付加されている。これには対象作品内容や作者そのもの、あるいは書風についての解説・批評、釈文、伝来の由来や編纂の経過など、さまざまな興味ある内容を含むことが多い。すなわちこれらは、個々の書に対応する情報というよりは、作者と作品の組合せに対応する情報である。

これらの序や跋自体が書作品として珍重される場合もある。例えば元代の書家・趙子昂は、蘭亭序を愛する余り、13種の跋を書いており、これを集めた法帖が作られている。

1種類の書作品について、種々の伝本を集め、これらの由来を述べ、その真贋や良否を論じた法帖もよくある。また、一つの法帖に含まれる文書は必ずしも1種類とは限らず、例えば「蘇東坡李白詩卷」のように、同一書家のいくつかの書作品を集めたものがある。複数の書家のさまざまな書作品を集めた法帖もある。

2.1.3 法帖に含まれる文字

ここでは、漢字の草書や、かなの連綿体のように文字の切れ目がはっきりしない場合を除き、篆書、隸書、楷書、行書などの書体で書かれた法帖について考える。法帖は手書きであるから、同一の文字が何回か現われても異なった形をしているのが常である。法帖に含まれる情報が十分に利用しうるためには、これら文字のレベルでの情報が得られなくてはならない。

法帖の各文字が、(ある)辞書のどの文字(文字種)に当るかを定めるのは、困難なことである。古い法帖では、虫食い・破損・褪色などによって難読の字がある。また、度重なる複製や、後代の注解や書入れなど、人為的な要因によって読めなくなる場合もある。作者の墨跡が完全に残されていれば大丈夫というわけではない。書

き誤りもあるし、また、環境に応じて字形を整えあるいは変化させるため、筆画を通常の字形と変えることはしばしば行われる。さらにこれらの文字が後代の手本となり、学習され実用によって、文字の形を時代によって変化させる一因となる。

このように、文字の形（および意味）は時代によって変化しており、きわめて多様性に富む。したがって少しサンプルを大きくとれば、現代の有限な文字セット、例えば JIS 漢字コードなどに含まれない文字が出現するのはほぼ確実である。そればかりか、説文解字・康熙字典などの古い辞典、あるいはこれらを総合した現代の大きな辞典にも含まれない文字も存在しうる。実際には、意味や形の類似をもとに、法帖中の大部分の「文字」データの各々に対して、これがある文字セットの中の「文字種」と、多対 1 の対応をつけているといえる。

各文字の形は、もちろん基本的な情報である。データベース化には、画像入力機器によってデータを取得し、2 値化・雑音消去・輪郭抽出などの処理を加えて記録する。2 値化の際のしきい値の決定、雑音とデータの判定、他の字の字画や紙面のよごれ・しみ、所蔵印や書込みなどをいかに区別するかが問題となる。

2. 1. 4 実体集合の定義

時：書の作者の生没年、あるいは書作品およびその各種表現形態の製作年

作者：書作品の直接の筆者である書家、あるいは書跡のもとと書作品の筆者、ないし伝承によりこれらの筆者と信じられている人。たとえば、「唐人模写蘭亭序」や「宋定武拓蘭亭序」では、王羲之が作者となる。「趙子昂臨蘭亭序」では、趙子昂を作者とする。

作品：一人の作者（書家）が何かの目的のため、一回に書いた（揮毫した）墨筆。上の例では、「蘭亭序」が作品である。

作品単位：ある作者が、ある作品を書いたこと。

書風：ある作品単位における作者の文字の書きぶりを、書の研究者ないし専門家が評論したもの。評論は経験的・心理的なものであるから、人によって違う見方もありうる。また、一人の作者でも、作品によって異なる書風になることもありうる。

釈文：ある書作品について、その各文字ごとに可能な限り現在使用されている文字を対応させて解釈したもの。

書跡：書家の墨筆。あるいはそれを伝える何かの形態で、1 件しか得られないもの、たとえば石碑、拓本、古法帖など。

刊本：手本や鑑賞用に、書跡を集めて出版したもの。
個々の書：書跡あるいは刊本から取った、一つの作品と対応する書の、コンピュータ処理可能な表現形態。

文字：個々の書から分離した書道（手書き）文字。

文字種：ある既存の文字セットに従った場合の、個々の書内に存在している文字の種類。

2. 2 グローバル ER スキーマ

上で得られた実体集合と関連集合をくわしく分析し、その属性を定義した。また、属性間の依存性をしらべ、正規化を行った。その結果として、新たな実体集合と関連集合が定義された。また、実体型の基本キーを定め、関連型の写像の型（M : N など）を求めた。

この結果得られたグローバル ER スキーマ（情報構造レベルの拡張 ER 図）を、図 2 に示す。一般には、属性名を実体集合あるいは関連集合と値集合（楕円）を結び直線上に書くが、ここでは簡略形を用いて、属性名と値

集合名が同じ場合これを楕円内に示すだけとし、属性名と値集合名が同じでない時だけ属性名を直線上に別記する。基本キーとして選択した属性には下線をつける。

2. 3 意味的リレーション

上で得られたグローバル ER スキーマに従って、これらの実体集合や関連集合が属性によって値集合に写像される様子を、実現値によって具体的に示すため、意味的リレーションの表を作った。表 1 に実体リレーション、表 2 に関連リレーションの、それぞれ一例を示す。

2. 4 存在制約およびキー制約

例えば実体集合「個々の書」の要素 s が存在するためには、その属性集合「書の番号」「書の名」「書の画像」の s に対する値がすべて存在する必要がある。このような存在制約を、各実体集合および関連集合に対して求めた。

また、「文字」という実体集合は、弱実体集合である。これは、実体集合「個々の書」、及び関連集合「個々の書・文字」に依存する。

実体集合「作品単位」には「作者」と「作品」の基本キー制約、「文字種」には「文字・同種文字」で結合される「親文字」と「同種小文字」の基本キー制約を課す必要がある。

3 関係データベースの論理設計

上で求められた拡張 ER モデルを、関係モデルに変換した。データ型が数値ないしテキストデータである場合は、簡単に関係表に変換される。そして関数従属性から、基本キーが計算できる。表 3、表 4 に、それぞれ表 1、表 2 から変換された関係表を示す。

4 おわりに

コンピュータディスプレイおよびコンピュータ印刷技術の発達によって、文字の形の美が大きな問題となり、関連の諸分野は新たな重要性を帯びてきた。3500 年を上回る書道の伝統は、今後見直され、新たな形で発展して行くことと思われるが、そのためにも、これまで伝えられた情報を整理し、これを十分に活用することが重要である。

本研究では、これらの情報の相互関係を明らかにし、拡張 ER（実体・関連）モデルを利用して、正規化および一貫性制約に留意しつつ、法帖中の各文字レベルでのデータを持ち、これと作者・作品・時・刊本その他の情報を関連づけて記述しうるデータベース概念モデルを設計した。さらに、このうち文字データないし数値データで表現できる部分について、関係データベースモデル（論理データベースモデル）への変換を行った。

将来、このモデルにもとづく法帖文字のデータベース実現によって、書に関する研究への利用や、書作品のコンピュータグラフィックスへの利用が容易になることが期待される。

文献

[1] Peter Pin-Shan Chen, "The Entity-Relationship Model Towards a Unified View of Data," ACM Trans. Database Systems, Vol. 1, No. 1, p.481-487 (1976).
 [2] C. S. dos Santos, E. J. Neuhold and A. L. Purtado, "A Data Type Approach to the Entity-Relationship Model," in P. P. Chen(ed.), Entity-Relationship Approach to systems Analysis and Design, p.103-119, North-Holland(1980).

[3] 酒井博敬、「情報資源管理の技法－ERモデルによるデータベース設計」、オーム社(1987)。

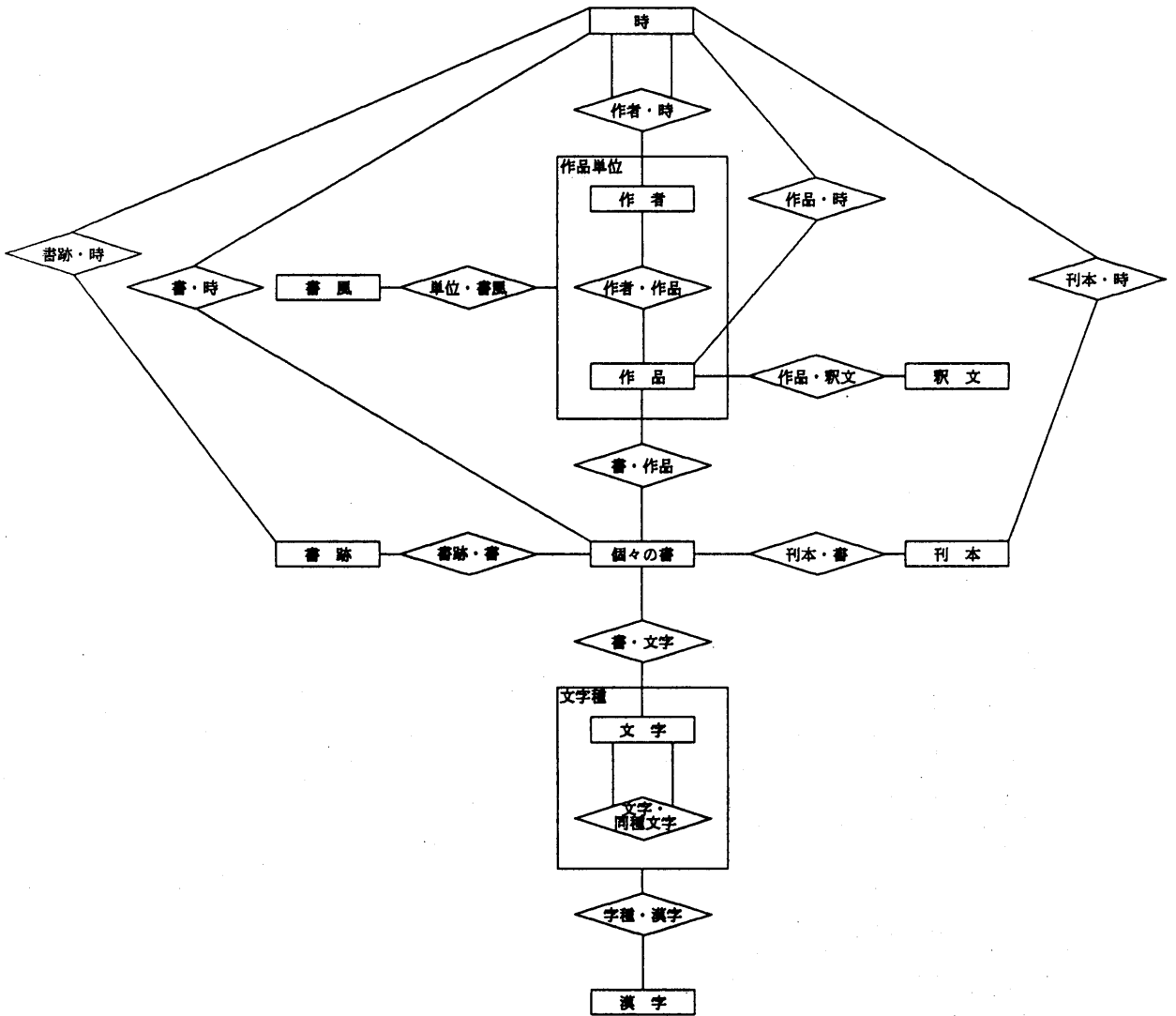


図1 法帖文字データベースのマクロERスキーマ(拡張ERモデル)

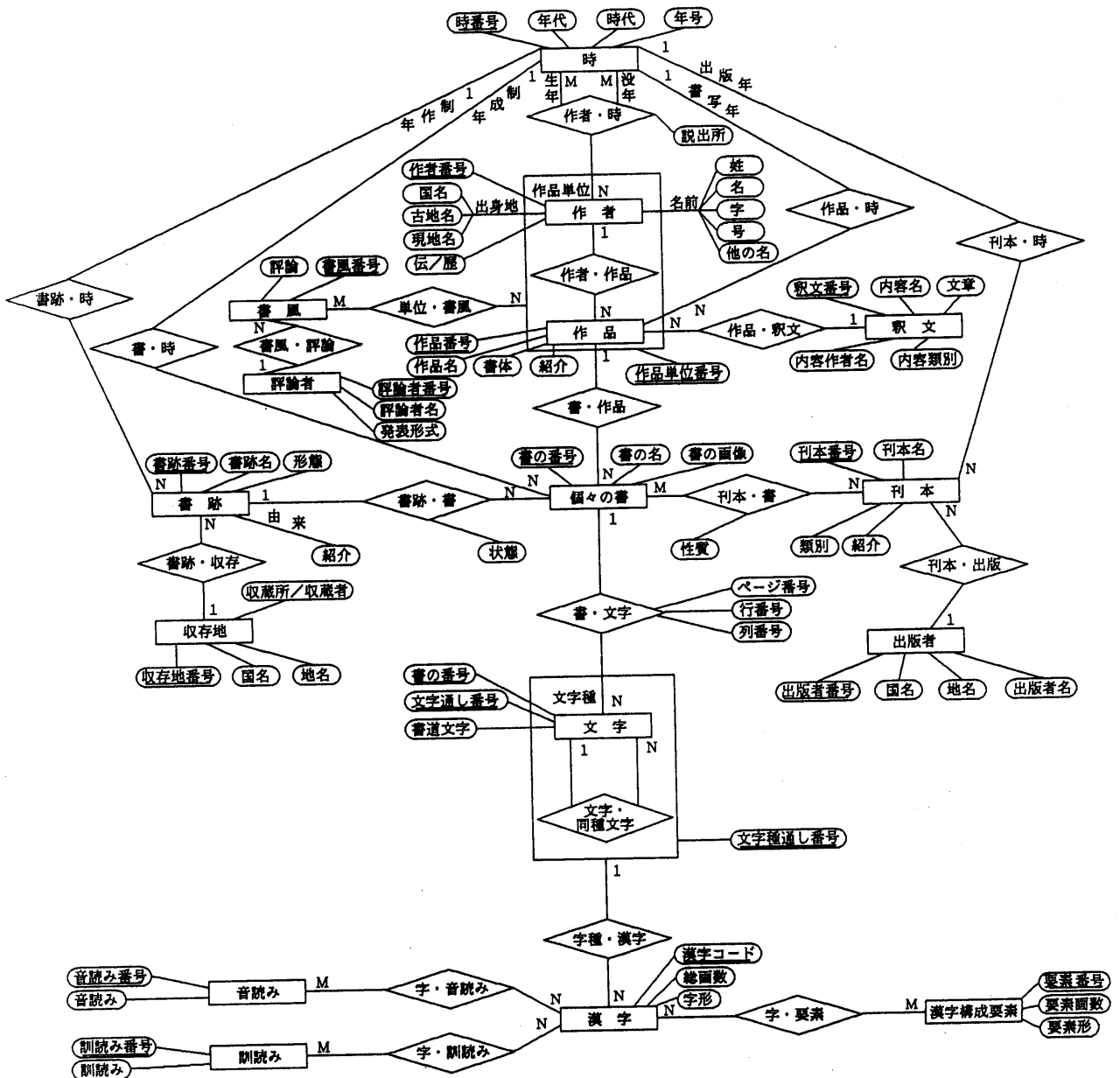


図2 法帖文字データベースのグローバルERスキーマ (拡張ERモデル)

表1 実体リレーションの例（意味的リレーション「作者」）

属性 値集合 実体	基本キー	名前				出身地			伝/歴
	姓	名	字	号	他の名	国名	古地名	現地名	小伝/略歴
E1	王	羲之	逸少		右軍	中国	琅邪	山東臨沂	秘書郎から出発して会稽王友、臨川太守となり、征西將軍として武昌に鎮していた……そして護軍將軍に進んだ。永和七年(351)、右軍將軍、会稽内史となり、郡治山陰県へ赴任し、在任四年で辞任した。その後も山陰県に住み、名士とともに山水の間に遊び、服食養生の道を楽しんだ。
E2	趙	孟頫	子昂	松雪	文敏	中国	呉興	浙江湖州	(略)
E3	蘇	軾	子瞻	東坡		中国	眉州眉山	四川省	(略)
E4	三津首	広野		伝教大師	最澄	日本	近江滋賀	比叡山	(略)
E5	佐伯氏	真魚		弘法大師	空海	日本	讃岐多度		(略)

表2 関連リレーションの例（意味的リレーション「刊本・個々の書」）

		← 関連の基本キー →			
実体集合	役割	刊本	個々の書		
実体属性	値集合	刊本番号	書の番号	性質	関連属性
関連		H1	S5	墨跡	
		H1	S6	拓本	
		H1	S7	拓本	
		H2	S1	模本	
		H2	S2	拓本	
		H2	S3	臨書墨跡	
		H3	S10	古筆	

表3 関係の例（関係「作者」）

階級	姓	名	字	号	他の名	国名	古地名	現地名	略 歴
E 1	王	羲之	逸少		右軍	中国	琅 邪	山東臨沂	秘書郎から出発して会稽王友、臨川太守となり、征西將軍として武昌に鎮していた……そして護軍將軍に進んだ。永和七年(351)、右軍將軍、会稽内史となり、郡治山陰県へ赴任し、在任四年で辞任した。その後も山陰県に住み、名士とともに山水の間に遊び、服食養生の道を楽しんだ。
E 2	趙	孟頫	子昂	松雪	文敏	中国	呉興	浙江湖州	(略)
E 3	蘇	軾	子瞻	東坡		中国	眉州眉山	四川省	(略)
E 4	三津首	広野		伝教大師	最澄	日本	近江滋賀	比叡山	(略)
E 5	佐伯氏	真魚		弘法大師	空海	日本	讃岐多度		(略)

表4 関係の例（関係「刊本-個々の書」）

刊本番号	書の番号	書の性質
H 1	S 5	墨 跡
H 1	S 6	拓 本
H 1	S 7	拓 本
H 2	S 1	模 本
H 2	S 2	拓 本
H 2	S 3	臨書墨跡
H 3	S 1 0	古 筆