

演劇資料アーカイブに対する年代推論システム

吉岡 卓 森井マスマ 谷 聖一 紅野 謙介
日本大学文理学部 日本大学文理学部 日本大学文理学部 日本大学文理学部
戸田 誠之助
日本大学文理学部

舞台の痕跡が記された上演台本は、演劇研究において重要な資料である。しかし上演台本には書籍の奥付にあたるものがないために、資料として利用する際に必要な上演年月日や幕数などの記述を欠いている場合が多い。そこで本稿では、デジタル化された台本のアーカイブを対象とし、アーカイブ内の確定した年代情報をもとに、年代の未確定な箇所を推論するシステムを構築した。

Temporal reasoning system for the digital library of theater literature

Suguru Yoshioka Masumi Morii
College of Humanities and Sciences College of Humanities and Sciences
Nihon University Nihon University

Seiichi Tani Keisuke Kohno
College of Humanities and Sciences College of Humanities and Sciences
Nihon University Nihon University

Scinosuke Toda
College of Humanities and Sciences
Nihon University

Acting scripts are important data and widely studied in artificial intelligence and theater, because traces of theater are set down on scripts. But, many of the acting scripts lack the imprint. Although information of publication data is necessary, theater researcher can not utilize them from acting scripts such as playday, acting configuration and so on. So, we show our temporal reasoning system based on the digitalized acting script that infers a playday for acting scripts with no date.

1 はじめに

演劇研究の1つとして上演台本を扱う研究がある[2, 5]。舞台上の演技は上演される時々において消滅するが、台本にはその痕跡が残され、資料として継続していく。ただし紙媒体の台本を永続的に保存するには限界があり、デジタル化する事で保存および検索が容易となる[1]。しかしひとまとまりであるはずの台本が分散していたり、ある部分が欠落している場合、それぞれの年代を確定しそれが上演された当時の状態に戻すことは困難である。そこで我々は、デジタル化された台本のアーカイブを対象とし、アーカ

イブ内の確定された年代情報をもとにして、未確定な台本の上演年代を推論し補完するシステムを構築した。我々は特に新派劇の台本に対する推論システムを作成した。このシステムを用いることで、上演日時を推定し、散在していた台本をまとめる事が期待できる。明治の一時期に歌舞伎をしのぐ人気を博し、その後も多くの観客を魅了してきた新派劇であるが、現在その衰退は著しい。しかしそうした状況であるからこそ、新派劇を伝統芸能である他の演劇ジャンルと同様に、保存し研究していくことは必要である。そのために基礎資料の収集と整理は急

務であり、本稿では我々が提案するシステムがそうした状況に対して大きな有用性をもっている事を示す。

次章では、本研究で対象とする新派劇について述べる。3章では、我々が構築した推論システムについて示す。4章ではシステムによる推論結果の詳細な検証を示す。最後に、本研究のまとめと今後の課題について述べる。

2 演劇資料のデジタル化

近年、様々な研究施設において演劇資料のデジタル化が進められている。本学で現在進行中の「デジタルアーカイブ・インフラストラクチャーの構築と高度利用」¹においても、演劇分野のひとつである新派劇に関する資料の保存を試みている。中でも我々が注目しているのは、喜多村緑郎文庫の台本である。喜多村緑郎とは、明治から昭和にかけて現役の役者として活躍し、新派の芸の確立に貢献した有名な人物である[6]。収集された台本を計算機に保存しテキストベースのデジタルアーカイブとすることで、情報の検索、発信が容易になる。我々が作成した喜多村緑郎文庫に対するデジタルアーカイブシステムを図1に示す。

新派劇は明治20年代に興った演劇のジャンルであり、能楽・文楽・歌舞伎などと比較すれば、格段に新しい時代のものである。そしてその新しさと大衆性ゆえに、研究対象としてはほとんど顧みられることがなく、資料の収集および整理が立ち遅れているのが現状である。とりわけ新派劇の台本は、上演ごとに役者や劇場の規模、上演時間の長短に応じて書き換えられるため、上演された回数分だけ台本が存在することになる。しかし同一演目に対して複数存在する台本のうち、上演年代が特定できているものはごく少数に限られている。そこで研究者は対象台本の年代特定という作業を経てはじめて、それらを資料として活用することができるのだが、それにはかなりの手間が必要である。具体的にはまず既刊のいくつかの年表[3, 4]を使って、その演日の上演年月日をすべて把握する。その上で場割や配役を参考にして合致するものを見つけ出



図1: 喜多村緑郎文庫アーカイブ

す。しかし場割や配役といった情報はすべての年表に完備されているわけではなく、もしそれが欠けている場合は、さらに当時の演劇雑誌や新聞、あるいは劇場で販売された絵本筋書(上演プログラム)などを参照する作業が必要になってくる。

そこで我々は、演劇の台本に関するデジタルアーカイブを作成し、特に年代に着目することで、不明な年代を計算機を用いて推定する事を試みた。計算機を用いることで、大規模なデータを即座に処理することが可能となる。詳しくは次章にて示す。

3 計算機による上演年代の推論

情報科学の分野では情報の復元が広く研究されている。特に自然言語処理や人工知能の分野では、少ない情報から多くの情報を推論する研究が行われており[7]、このようなシステムは一般的に推論システムと呼ばれる。本稿ではテキストベースのアーカイブを対象とすることから、自然言語処理の技術を用いて年代推論システムを構築する。なお本稿では、特定の対象に応じて集められたデジタルデータをアーカイブ、そこに含まれる各要素をコンテンツと呼ぶこととする。

3.1 演劇資料アーカイブの特徴

本節では喜多村緑郎文庫アーカイブの特徴について示す。本学において構築しているアーカ

¹私立大学学術研究高度化推進事業 学術フロンティア推進事業 2003-2007 による私学助成を得て行われている。

図 3: 表 (CSV) 形式による喜多村緑郎文庫アーカイブ

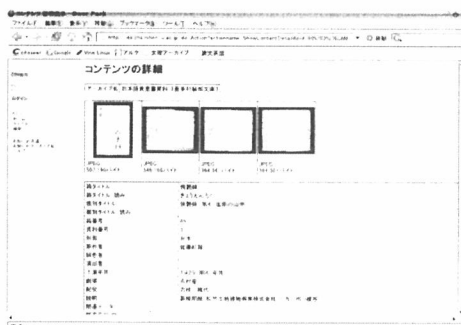


図 2: アーカイブ内のコンテンツの例

イブシステム²のコンテンツの例を図 2 に示す。喜多村緑郎文庫アーカイブにおいて、そこに含まれる各コンテンツには以下に示すようなインデックスが割り振られている。

箱タイトル, 箱タイトル(読み),
個別タイトル, 個別タイトル(読み),
箱番号, 資料番号, 形態, 原作者,
脚色者, 演出者, 上演年月, 劇場,
配役, 説明, 関連データ

ここで、箱タイトルとは上演された台本のタイトル、つまり演目と呼ばれるものである。また個別タイトルとは、それに上演時の幕数や場割の情報を加えて細分化したものである。箱番号と資料番号にはそれぞれ箱タイトルと個別タ

イトルを分類するために割り振られた数値が入力されており、必ずしも時系列順に並んでいるわけではない。形態とは演劇資料の分類が記述されており、具体的には台本、書拔、口上などに分かれる。演劇を研究する上で特に重要なのが台本と書拔である。台本とは作品を上演する際に用いる資料である。書拔とは役者によって違いはあるが、喜多村緑郎の場合、台本から自分の役柄の台詞のみを抜き出したものであり、資料の価値が非常に高い。説明には、資料自体の物質的な特徴が記述されている。主に紙の種類、記述に用いられた用具の種類、特筆すべき特徴が調査者の視点によって記述されている。関連データには他ページへのリンクをはる事が可能である。しかし、現在のところ本アーカイブでは用いられていない。よって本稿でも関連データについては扱わない事とする。

また本学のデジタルアーカイブシステムでは、図 1 に示したアーカイブの一覧を一括してダウンロードする事が可能である。ダウンロードしたアーカイブの一覧を図 3 に示す。図 3 に示したファイルは CSV 形式のファイルであり、我々はこのファイルを本研究の対象として扱う。

ここで、いくつかのインデックス間の関係を図 4 に示す。図 4 における個別タイトルが本学のアーカイブシステムにおけるコンテンツとなっている。図 4 において、例えば箱 1 の演目を行う場合、資料 1 の台本と資料 4 の台本を用いて上演が行われる事がある。また、同じ箱 1 の上演に対して、資料 2 の台本とその資料 2 に対する書拔(資料 3)を用いて行われる事もある。よっ

²<http://da.chs.nihon-u.ac.jp/da/>

喜多村文庫アーカイブ

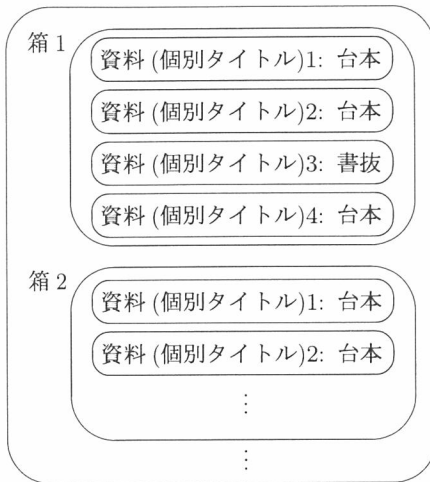


図4: 箱と資料(コンテンツ)の関係

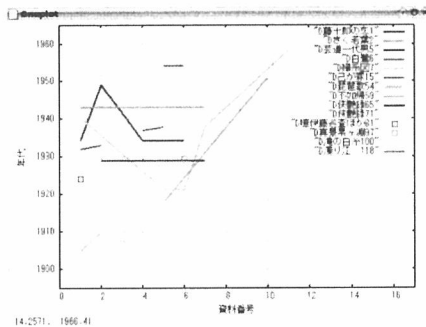


図5: アーカイブ内に含まれる台本

て個別タイトルを持つ情報、特に時間的な情報と個別タイトル間の関係は演劇研究において重要な項目であり、演劇研究者は書抜と台本を照らし合わせる事でより構造的な演劇の研究が可能となる。

そこで我々は、アーカイブに含まれている上演年代の確定した資料に対して、形態と資料番号に応じたグラフを出力した。その結果を図5と図6に示す。図5は台本に対する出力結果であり、図6は書抜に対する出力結果である。それぞれ横軸が資料番号、縦軸が時間軸となっている。例えば、図5から箱番号1「藤十郎の恋」の資料のうち形態が台本のものを抜き出すと以下に示す上演年代が得られる。

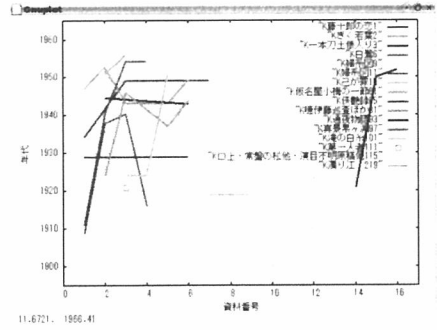


図6: アーカイブ内に含まれる書抜

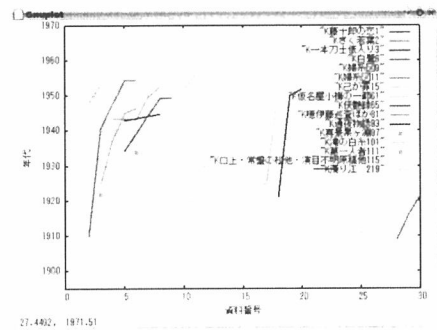


図7: 番号付け変え後の書抜

資料番号	上演年月
1	1934年3月
2	1949年3月
4	1934年3月
5	1934年3月
6	1934年3月

しかし前述のとおり番号の大小には意味がない。そのため図5と図6に示すような資料分類では、どの書抜がどの台本に対応するのか判断するのに大変労力がかかる。すなわち時系列に沿うように台本を分類する事が必要となる。そこで書抜に対して年代順に新しく資料番号を付け変えたものを図7に示す。図7における資料番号は時系列によって番号が付けられており、図6と違って資料番号が大きくなるごとに年代も上がっているのが分かる。台本に対しても同様の処理を行い、特に藤十郎の恋に対する台本と書抜を合わせた出力結果を図8に示す。

表 1: 演劇資料アーカイブからの抜粋

タイトル	上演年月	説明
婦系図 1 幕	?	罫線用紙 (硯成社) にカーボン複写
婦系図 6 幕 12 場	1940 年 5 月	罫線用紙 (硯成社) にカーボン複写

表 2: 上演年月に対する推論方法

タイトル	上演年月	説明
婦系図 1 幕	1940 年 5 月	罫線用紙 (硯成社) にカーボン複写
婦系図 6 幕 12 場	1940 年 5 月	罫線用紙 (硯成社) にカーボン複写

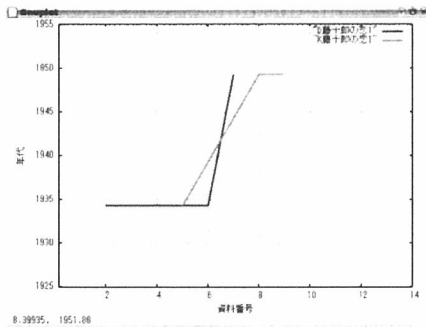


図 8: 番号付け変え後の台本と書抜

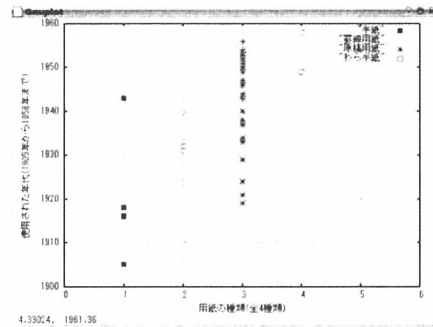


図 9: アーカイブ内に現れる用紙

このように、時系列に沿った資料番号を新しく付け変える事で書抜と台本の関係が鮮明となり、演劇研究に寄与する事が期待できる。またこの事は台本の資料的価値を比較する事にも有用となりえる。しかしそのためには台本の上演年月を数多く定める必要がある。そこで我々は不明な年代を推論するようなシステムの作成を試みた。

3.2 年代推論システム

年代推論システムの構築に対して、特にタイトル、上演年月、説明のインデックスに着目した。演劇資料アーカイブからそれらのインデックスを抜粋した例を表 1 に示す。表 1 において「婦系図 1 幕」がいつ上演されたかの情報がアーカイブ内において抜けている。しかし、既に年代が確定している「婦系図 6 幕 12 場」の説明文中の台本情報と「婦系図 1 幕」の台本情報は似通っている。そこで、表 2 の太字で示したような資料の情報に着目することで、不明確な上演

年月を推論することが期待できる。よって本研究では、説明文中における用紙の種類と印刷方法に着目して推論するシステム構築をめざす。

準備として、演劇資料アーカイブに含まれる紙の種類と印刷方法に対する抽出実験を行った。実験に用いたシステムは Perl ver.5.8.2 を用いて Vine linux 5.2 CR 上に実装した。我々は演劇資料として用いられる事の多い紙として、半紙、罫線用紙、原稿用紙、わら半紙に着目し、演劇資料アーカイブから抽出した。抽出結果を図 9 に示す。図 9 に示すように、半紙とわら半紙の使用年代は時間的に重複しない。すなわち、年代の不明な台本を推論する手がかりとして用いる事ができる。次に印刷方法として、筆書き、カーボン複写、万年筆書き、謄写版印刷に着目し同様の処理を行った。抽出結果を図 10 に示す。図 10 に示すように、筆書きと謄写版印刷に対しても時間的な重複がみられなかった。すなわち、筆書きで書かれた演劇資料と謄写版印刷による演劇資料では上演年代も離れていると考える事ができる。また表 1 に示すように、演劇資料として

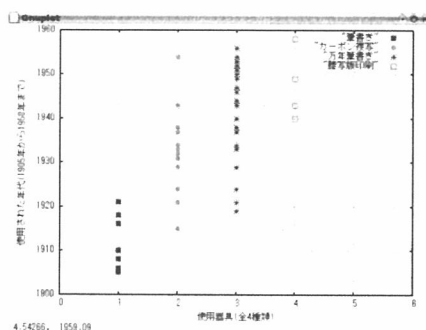


図 10: アーカイブ内に現れる印刷方法

原稿用紙と罫線用紙が用いられている場合、用紙の種類に加えて括弧付きで会社名や種類が記述されている。そこで括弧内をパターンマッチングで抜き出し、図 9 に示す分類をさらに細分化するための特徴抽出を試みた。まず原稿用紙の製作会社や種類に着目した結果、以下に示す紙の種類が得られた。

1	10-20	1	10-24
1	1020	3	10X20
1	10 20	3	20X10
1	20-10	1	A4
1	B4	5	LINEN PAPER
2	MARUZEN	7	NOBLE
2	OS	1	SUCCESS
1	TOKYO	2	YM
1	NOBLE	1	NOBLE 15X10
1	NOBLE.C4	1	OS
4	なし	1	クリーム
9	コクヨ	9	喜多村用箋
1	高知堂	1	新警察社
1	深川 松屋文具店	3	大阪 柳屋
1	日章	1	日章 NO41
1	白木屋製	6	鳩居堂
3	文房堂製	3	緑樹菌用紙乙号
11	緑樹用紙		
1	チューリップの絵入り		
1	小田原高等女学校作文用紙		

ここで左側の数値はアーカイブ内に含まれている数、右側が原稿用紙の種類を表す。演劇資料アーカイブに含まれる原稿用紙の種類は全 37 種類、原稿用紙が用いられたコンテンツの数は 94 個であった。同様に罫線用紙について調べた結果を以下に示す。

2	12	2	なし
3	協信社印行	1	銀座伊東屋と、山
22	硯成社	2	五虎印
1	光英社	2	甲
1	十キ	5	十二
1	十二行	1	小林製
10	松竹株式会社	4	松竹興業株式会社
20	松竹合名社	1	人形町 上田屋製
1	大阪戎橋せのや	2	木内文芸部
1	十二、牛込ウス井製		
1	松竹株式会社演劇企画部		
3	松竹土地建物興業株式会社		

すなわちアーカイブ内に含まれる罫線用紙の種類は 21 種類、全体で 86 箇所において用いられていた。

この結果、原稿用紙と罫線用紙において同じ名前のものは「なし」だけであり、会社名や種類が重なることは無かった。よって原稿用紙で種類の無い物を「原稿用紙種類なし」、罫線用紙で種類の無い物を「罫線用紙種類なし」という文字列に置き換えることで、会社名が分かれば一意的に用紙の種類が区別できる事が分かる。上記に示した詳細な用紙情報を用いる事で、不明な年代を推定する事が可能となる。詳しくは次節にて示す。

3.3 詳細な用紙情報を用いた推論

本研究で作成したシステムは、アーカイブ内の上演年月に日時が埋まっているとき、演劇資料として用いられた紙の情報と上演年月をルール化する。次に、上演年代の不確定な資料の紙情報に対してルールを適用し推論を行う。このような推論を可能とするために、準備として年代が確定しているコンテンツを抽出した。抽出された紙の種類を図 11 に示す。

図 11 に示すように、紙の種類として 30 種類、それらの使用年代として 68 種類の結果が得られた。紙の種類による年代的な広がりがないため、紙の種類を判別できれば資料の使用年代を容易に推測できる事が分かる。例えば、図 11 において「十キ」製の罫線用紙は 1910(明 43) 年に用いられているため、年代の不明な演目の台本が「十キ」製であれば同じ年代であると推論する。我々の推論システムは、年代が特定されていない個別タイトルの説明文に対して、既に得られた紙の種類と用具が同種類の様式であれ

図 12: 年代補完済みアーカイブ

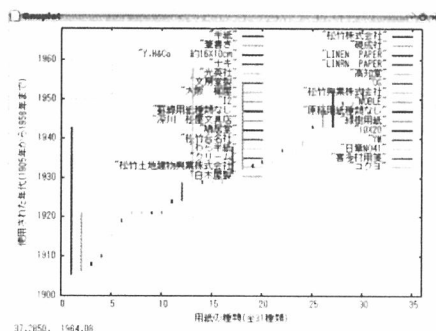


図 11: 詳細な用紙分類と年代分布

ば年代を埋める。このように埋められた年代を(候補1)とする。

また説明の欄に年代が載っているにも関わらず、記入漏れがあるものは(候補2)として同時に抜き出す。具体的には、説明文中に年代が書かれており上演年月のインデックスが空欄の場合にシステムは年代抽出を試みる。我々のアーカイブでは、説明文の中に年代が含まれる時は必ず年号も一緒にあらわれる。そこで年号と文字列「年」にかこまれた年数を最小マッチングして抜き出し、年号に応じて西暦に直すことで上演年月の欄に補完する。ただし上演年月に補完する際に「(候補2)」という文字列を併記する。このようにして得られた喜多村緑郎文庫アーカイブを図12に示す。

4 推論の結果と検証

本章では、我々の推論システムを用いた結果と検証を示す。システムによる推論結果は以下の通りである。

台本総数	年代記入済	候補1	候補2
247	82	103	5

すなわち、元の演劇資料アーカイブには33%しか年代が書かれていなかったが、我々のシステムによって77%の年代がアーカイブに含まれた。そこで、補完された年代の詳細な検証結果を以下に示す。

正解	区間内	判別不可能	不正解	その他
21	5	74	4	4

補完した年代の正否に関しては、本学の演劇資料アーカイブ以外の資料も参考に演劇研究者によって検証が行われた。紙の種類によっては複数の年代で用いられることがあり、そのような場合は複数の年代が候補としてアーカイブに含まれている。その時間区間の中に入っているものは区間内としている。例えば、松竹合名社の罫線用紙は1924(大13)年と1929(昭4)年において使用されていた。このとき、年代が不明で松竹合名社の罫線用紙を用いている資料には、上演年代として両方の年代が補完されている。そこで他の演劇資料によって、補完した台本の上演が1926年であると判断される場合は区間内とカウントした。つまり紙の使用年代を絞りこむ事で正解となりうるものが含まれる。また、正解

だと思われるが資料不足で断定はできない場合を判別不可能とした。多くの台本はここに分類されるが、本学以外の図書館や関連機関が所蔵する資料との相関によって今後判別できる可能性がある。その他に区分されたものとして、手紙、メモ、大道具帳がある。これらに使用された紙は喜多村緑郎文庫の「上演台本」とは用途が異なるため、同種の紙であっても年代が異なる可能性が高い。よってその他に含まれる紙のみのルール化が別に必要であると考えられる。さらに上演台本に2種類の紙が含まれているため、正誤の判断がつかなかった物もその他に分類した。不正解であった資料は、図11に示す紙の使用年代の区間外となる場合である。例えば「徳富蘆花/不如帰6幕」は硯成社の罫線用紙とカーボン複写によって作成されている。図11によれば、硯成社は昭和7年から昭和29年の間に用いられた事が分かる。しかし「不如帰6幕」の脚色者「柳川春葉」が携わった上演年代は、他の年表資料を参照してみると、大正2年、大正9年、大正10年の3つしかない。すなわち本システムによって補完された年代よりも古い時代であり不正解となる。本研究では紙の種類に応じて年代を推論したが、脚色者や演出者に対してもルール化を行い、ルール間のバイアスを調節することで正解データを推論する必要がある。

さらに、台本によっては旧暦で年代が書かれているものもある。例えば「徳富蘆花/不如帰」の台本は「申の卯月」に書かれたとされている。申年といっても、1908年、1920年、1932年など一般にはそれが何年であるか特定するのは困難である。しかし「徳富蘆花/不如帰」の台本は筆書きで書かれており、図11における筆書きの範囲によって1908年の申年である事が自動的に推論された。また、この結果は他の演劇資料を用いて正しい事が検証された。よって、我々の推論システムは不明な年代を補完するだけでなく、旧暦で書かれた曖昧な時間表現を特定するのに応用することも可能である。

5 おわりに

我々は演劇資料アーカイブに対して、演劇の台本自体の特徴に着目し、そこから年代を特定するような推論システムを作成した。また旧暦

に対する年代の推論も可能とした。このシステムを用いることで、散在している台本を構造的に収集し比較する事が可能となり、演劇研究に役立てる事が期待できる。また我々の推論システムは喜多村文庫以外の年代測定にも役立てる事ができる。

今後は候補としてあげた年代をさらに絞りこむため、説明文以外の情報にも着目する必要がある。新派劇の台本は、松竹大谷図書館や国立劇場伝統芸能情報館、および早稲田演劇博物館などに所蔵されている。台本は上演に携わる役者全員に配布されるため、重複して存在する一方で、部分的な脱落がみられる場合もある。現在それぞれの機関で所蔵している台本の重複状況については把握されていないが、このシステムを利用して上演年代を特定することができれば、資料の欠如を相互の機関で補い合うことが容易となり、調査、研究における利便性は飛躍的に上がることになる。またGUIによるユーザフレンドリなインターフェース開発を行う事で、人間とコンピュータとのインタラクション向上も目的とする。

参考文献

- [1] 赤間亮：文系利用者側の視点で構築するデジタルアーカイブ—画像資料とデータベースの有機的・可変的リンカー、人文科学とコンピュータシンポジウム論文集, pp. 32-39, 2000.
- [2] 『霜夜鐘十字辻笠』上演用台本、歌舞伎学会 現行レパートリーを考える会、歌舞伎 研究と批評, Vol. 36, pp. 13-31, 2006.
- [3] 松竹百年史 演劇資料、松竹株式会社, 1996
- [4] 新派 年表、劇団新派編、大手町出版, 1978
- [5] 阪急池田文庫所蔵 歌舞伎台本日録、阪急学園池田文庫, 1970.
- [6] 柳永二郎：新派の60年、河出書房, 1948.
- [7] S. Yoshioka, K. Kaneiwa and S. Tojo : Occurrence Logic with Temporal Heredity, *Proceeding of the IICAI-03*, pp.1296-1309, 2003.