

上智大学「日本の人名データベース」の構築について Constructing the Sophia University “Japanese Biographical Database”

ボルン 礼於

Leo Born

ハイデルベルグ大学 計算言語学部, Im Neuenheimer Feld 325, ハイデルベルグ, ドイツ
Heidelberg University, Department of Computational Linguistics, Im Neuenheimer Feld 325,
Heidelberg, Germany

あらまし: 上智大学の「日本の人名データベース」は、日本の歴代人物の伝記と、その個人を超える人間関係のネットワーク情報の提供を目的として作成された。本データベース構築は、頼春水(1746~1816)の研究から始まり、現在は人物約2,200人、人物間のイベント約7,000件を年月日・身分・出典などで検索し閲覧することができる。ハーバード大学のChina Biographical Databaseの構築をモデルにして、本データベースの改善の上、データベースをアクセスするウェブアプリケーションの開発をすすめている。将来的に人物のデータを蓄積するとともに、各人物のもつ姻戚・交友関係を詳細にモデル化することで、通史的検討に止まらない社会変遷の模様を追っていきたい。本報告では、本データベース構築の概要と、様々なオープンソースツールを用いたウェブアプリケーションの作成及びネットワーク可視化方法、今後の課題について述べる。

Summary: The Sophia University “Japanese Biographical Database” (JBDB) was constructed with the purpose of providing bibliographical information on Japanese historical figures and their personal, social, and political networks. Starting with research on Rai Shunsui (1746-1816), the database currently encompasses entries on ca. 2,200 individuals and ca. 7,000 events pertaining to these individuals and their interactions, allowing to search and visualize these entries by date, social status, sources, and other filters. Building upon the architecture of the Harvard University *China Biographical Database*, we are currently refining our own database, while also developing a new web application to access the database. With the addition of further biographical data, we aim to study patterns of societal changes that emerge out of familial and social relations among people. This paper focuses on the database design, the development of the web application with particular emphasis on the visualization component, and future directions for the project.

キーワード: データベース, ネットワーク研究, 集団履歴調査, 可視化

Keywords: database, network studies, prosopography, visualization

1. はじめに

現在、本研究グループは近世日本の人物についての伝記や、関係を取り扱うデータベースを構築し、あわせてデータベース上で動作される情報検索と可視化用のウェブアプリケーションを作成中である。本研究グループでは社会の構成要素として「集団」に着目し、その実態をとらえるため、集団履歴調査(プロソポグラフィ)やネットワーク分析(Social Network Analysis)の手法を検討している。歴史学における先行研究では、主に欧州史[1]や中国史[2]に焦点が置かれているが、本デー

タベースは、これまで集団履歴調査を含む分析の手法があまり適用されていない近世日本という研究対象にしている。これにより、近世日本の研究にネットワーク分析が導入される契機になるのではないかと考えている。

本プロジェクトの第一歩は、儒学者頼春水(広島藩儒)を取り巻くネットワークに着目することである。具体的には、ネットワーク分析や集団履歴調査の手法を用いた分析を行うため、頼春水の日記や書簡など、史料上で確認できる人名のデータベースを作成する。本プロジェクトは、個人の思想

トップ階層	テーブル数	件数
人物	5	2,209
文通	5	1,918
集会	4	4,982
アクティビティ	3	6,736
場所	8	5,699
文献	1	27

表 1. データ移動以前のデータ数

的評伝の形成のみならず、近世の人物の交流の様態をネットワーク分析の手法を用いて解明することで、伝記的研究からだけでは見えてこない同時代の学者交流および知の体系の蓄積像を明らかにしようとするものである。

そのため、過去 5 年にわたって人物数約 2,200 人、人物間のイベント数約 7,000 件がデータベースに入力された。このデータベースをもとに、ある人物が行動する時代的・場所的背景、またその人物が社会的にはたす役割を分析することが可能である。

本報告では、上智大学の「日本の人名データベース」の構築やウェブアプリケーション、およびネットワーク可視化ツールの概要について述べてみたい。

2. データベース構築

本プロジェクトは日本の歴代人物の伝記と、個人を超える人間関係のネットワークに関する情報を提供することを目的としているため、データの管理システムとして PostgreSQL[3]の関係データベースを用いた。データ入力は 27 件の参考文献に基づき手動で行っているが、デジタル化された文献を取得できる場合も考慮し、自動的な情報抽出方法も検討中である。

始めは、6 つのデータ内容で分類されたテーブル階層を中心に、合計 26 データベーステーブルを作成し、5 年間の期間でそこにデータが入力されてきた。その 6 つのトップ階層の統計を表 1 で示す。

しかし、データ記入の作業によって、特に件数が多いアクティビティという階層に問題があることが明らかになった。アクティビティというのは個人の家族、仕事、あるいは趣味に関わる情報をまとめて集めたテーブルである。そのためデータ属性や必然的なキーが大幅に異なっている。この問題を解消するために、データベース構築を改善し、データ移動を重点化することが決定された。

東亜史学圏でパイオニア的な役割を持つ、ハーバード大学の China Biographical Database (CBDB、



図 1. トップページ

「中国伝記データベース」)[4]をモデルにして、本データベース改善をすすめている。2016 年後期にハーバード大学から CBDB のデータベーステンプレートを頂き、それをもとに今までのデータを新しい構築にマッピングし、データ移動を行なっている。具体的には、データ移動以前の 26 テーブルのデータを、CBDB のテンプレートに基づいた 90 テーブルに移す作業である。

今まで表現の面で限界があった点も細かく区別をできるようになり、特に家族関係と仕事に関するテーブルが最も詳細である。その他にも、今までアクティビティで保存された住所変更や、身分変更も専用テーブルに記入できるようになる。今までモデル化されていなかった物事の購入・交換・獲得、又は執筆作品についての情報も、新データベース構築で適切にモデル化されている。

更に、「イベント」というものも、個人のイベントと数人が関係するイベントに適当にモデル化されている。それによって、各データカテゴリーに応じた属性を定義することができ、クエリと可視化に関する表現力が拡張される。

データ移動と同時期に、データベースにアクセスするウェブアプリケーションをモダン化する必要性に応じて、新ウェブアプリケーション開発を開始し、2017 年秋に「日本の人名データベース」(JBDB)として公開した[5]。以下でそのウェブアプリケーションについて述べてみたい。

3. ウェブアプリケーション

データベースはインターネット上で公開するものであるため、Node.js で動作する JavaScript のウェブアプリケーションを実装している。オープンソースツールを利用し、バックエンドは Express.js に基づく LoopBack[6] というフレームワーク、フロントエンドは AngularJS 1.X[7] で実行されている。



図 2. 人物のリスト



図 3. 人物の詳細ページ

LoopBackはオブジェクト関係マッピング (ORM) の導入と、REST APIの作成を容易にするため、データベース用とサーバー用のクエリデザインを統一させている。AngularJSではプラグインを取り込み、様々な機能を提供することができる。UI開発に主として利用しているプラグインはAngularJS制御用BootstrapのUI Bootstrap[8]である。多様な端末に対応させるため、レスポンシブデザインやWebフォントを可能にするフレームワークを導入することが重要である。

ウェブアプリケーションはトップページ (図1) にアクセスするとデータベースUIへのリンクと、スクロールダウンすると、研究グループの簡単な紹介が表示される。アプリケーションUIは2ヶ国語 (日本語と英語) であるため、スイッチ (トップページでは右上に配置) をクリックすることでUI言語を変更することができる。既定言語は日本語であるが、変更の場合、他の設定と一緒に言語の設定はクッキーに保存される。

データベースUIは既定では人物リストを表示し (図2⁽¹⁾)、メニューバーでは人物、イベント (現在、文通と集会を含む)、出典のリストを選択できる。イベントと出典も、以下で述べる人物の表示と同様、基本情報が記載されているリスト表示と、詳細ページに分かれている。

人物リストで示されている情報は漢字とローマ字の書き方で名前、出身地 (藩・国・郡・村)、生まれの身分 (武士と非士の身分を合わせて18の選択肢)、生没年、データの更新日時、そしてログインユーザーのみが選択できるアクションである。生没年はUI言語の設定により、和暦又は西暦での表示である。生没年をマウスでホバーすると相互の紀年法が表示される。

データの閲覧にはログインすることは不要だが、データ編集にはユーザーアカウントが必要で

ある。現在、プロジェクトメンバーのみにアカウントがあり、そのメンバーには特別なアドミン権限が与えられている。なお、大学の授業でも活用していくために (5「おわりに」参照)、学生用の制限付きのアカウントも準備中である。

画面のヘッダーにはウェブアプリケーションのロゴと題名以外に、アプリケーション全画面に共通のナビゲーションメニュー (右側) を配し、ログインユーザー専用のメニュー (左側) をクリックでドロワーとして提供できるようにしている。

各データリストについては属性に対する簡易検索機能が用意され、リスト表示の対象を絞り込むことができる。例えば人物データは人名、または生まれの身分で検索することが可能である。検索結果を保存する簡易機能もあるが、現在.txtフォーマットに限られている。⁽²⁾

表示されたリストの項目をクリックすると閲覧ページに飛ぶ形態となっている (図3)。人物の場合は、表示される情報は三つの画面に分けてあり、既定では基本情報が表示される。データリストに記載されている情報以外に、人物の画像、他名のリスト (通称や諱)、データ入力に参考された文献のリスト、そしてコメントボックスが表示されている。基本情報の画面から関連するイベント (その人物が関わっているイベントのリスト)、そして履歴と家族関係に関わる情報が記載されている画面に飛ぶ仕組みであるが、後者のデータはまだ移動中であるため、非公開である。編集作業も図4に似た画面で行われ、多くの属性 (身分や出身地) はドロップダウンリストを利用し、入力する。

UIの目標はスムーズなUXを提供することであるため、多数のサードパーティライブラリーを加えたコードデザインを用意した。以上述べた通り、中心はBootstrapだが、それ以外に導入されているプラグインは表2で示す。

プラグイン名	機能
bootstrap-drawer [9]	メニューオーバーレイ用
angular-translate [10]	多言語 UI 用
angular-table [11]	ソート機能付きテーブル用
AngularJS Dropdown Multiselect [12]	ドロップダウン用
ng-file-upload [13]	ファイルアップロード用
Lightbox2 [14]	画像表示用
Angular File Saver [15]	検索結果保存用
Angular-xeditable [16]	インライン編集用
Animate.css [17]	一般 UI アニメーション用
AngularJS Slider [18]	スライダーUI 用

表 2. 利用中のサードパーティープラグイン

4. ネットワーク可視化

データベース構築の目的は、個人を超える情報の上で、集団ネットワークを明らかにすることである。そのため、様々なネットワーク的アプローチの調査手法を研究ツールとして提供し、インタラクティブなネットワークを可視化させる必要がある。

本ウェブアプリケーションのコードデザインの特徴は、基盤となるデータベースレイヤー上で行われる、モジュール化された多様コンポーネントの総合化にある。ネットワーク可視化は、モジュール化されたコンポーネントの一例として挙げられるが、地理情報システム (GIS) 等もそのようなコンポーネントとして考えられる。

本ウェブアプリケーションにおける、ネットワーク可視化ツールのコンポーネントは、オープンソースである vis.js [19] という可視化ライブラリーを用いて実装している。D3.js [20] 等の他の可視化プラグインよりも、vis.js は機能が少ないように見えるが、ネットワークとタイムラインの可視化を中心としているため、本プロジェクトには最適である。

また、ネットワーク可視化において、vis.js は様々なカスタマイズのオプションを提供しており、多様な物理的シミュレーションアルゴリズムを導入しており、特にダイナミックなネットワーク可視化用に便利である。データベース内のデータ量は今後拡大されるため、全ネットワーク共通のノード数とエッジ数は確定数ではない。



図 4. ネットワーク可視化ツールのフロントページ

4.1. 可視化ツールの使い方

ネットワーク可視化ツールのネットワーク画面の説明に具体例として、頼春水の息子である頼山陽 (1780~1832) のネットワーク可視化を取り上げてみたい (図 5)。

可視化ツールはフロントページ (図 4) にアクセスすると可視化設定画面が示される。その設定画面において、人物選択の手法とネットワークの表示に関する設定が行われる。

人物選択には現在以下の三つの手法がある。選択された手法によって、ネットワークは異なる人物に基づいて可視化される。

一) 「全員」を選択すると、データベース内の全人物が考慮される。⁽³⁾ この設定ではデータベースの完全ネットワークが作成される。

二) 「名前で選択」を選択すると、人物検索窓と人物リストが示される。そのままリストから人物を選ぶこと、又は検索にて人物リストを絞り込むことで、考慮される人物が絞り込まれる。この設定では、選択された人物から始まり、その人物との関係性を持つ他の人物がデータベースからクエリされる。ネットワーク可視化はクエリ結果と元々選択された人物のセットの和集合に基づく。図 5 はこの設定で「頼山陽」を検索して選んだことに基づき作成された。

三) 「出典で選択」を選択すると、出典のドロップダウンが表示される。既定設定としては、選択された出典に伝記情報が書かれてある人物から始まり、その人物との関係性を有する他の人物がデータベースからクエリされる。ネットワーク可視化はクエリ結果と選択された出典に現れる人物のセットの和集合に基づく。

現在、クエリで人物の間関係性を定めるには、イベント (文通と集会) のみが考慮される。しかし、今後はネットワークを家族関係や仕事関係の情報にも基づくようにする予定である。

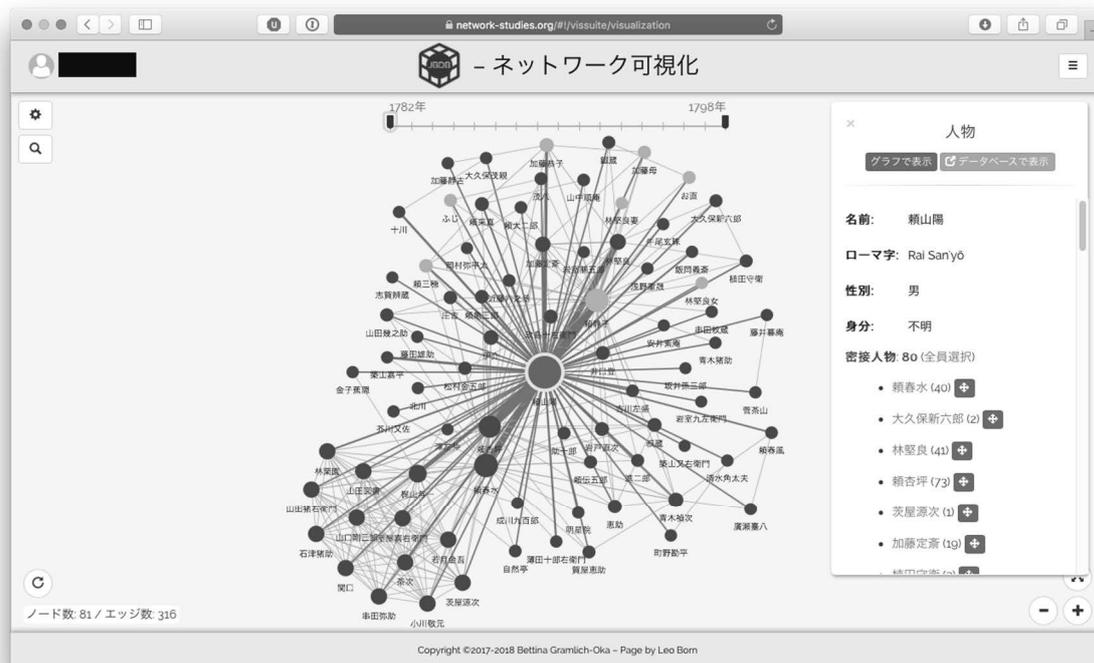


図 5. 頼山陽のネットワーク可視化

人物の選択手法が決定された後、ネットワークでノードの附属グループをどの属性に基づいて強調表示するか選択できる。附属グループを示す属性は現在、性別と生まれの身分のみだが、データベース内のテーブル属性に任意に基づくため、他の属性も今後追加することができる。図 5 は性別のグループ附属にて作成された。(4)

可視化のレンダリングが終了した状態は、図 5 で示している。画面の左下にはノード数とエッジ数を示すオーバーレイがある。ネットワーク自体はインタラクティブで、自由にノードを動かす、ズームレベルや表示フレームも自由に変更することが可能である。

ひとつのノードを選択すると、繋がっているノードとそのエッジが強調される上に、選択されたノードの詳細情報が別のオーバーレイで表示される。オーバーレイは人物の基本情報と、どの他の人物と繋がっているかをリストで示し、その関係人物ををクリックで全員選択して、グループとして動かせる。さらに、密接人物のリストからは、他の人物との共通イベントを Bootstrap モーダルで表示し (図 6)、画面フレームを他の人物の方に動かすことができる。人物オーバーレイやモーダルリストから新しいタブで、直接データベースでのエントリーに飛ぶ形態となっている。ネットワークを変更状態から初期状態に戻すには、画面の左下

のリセットボタンをクリックする。最後に、画面の左上の検索メニューによって、ネットワークで表示されている人物を検索することができる。

以上で述べた基本設定以外の複数の追加設定 (図 7) について、次で説明していきたい。

4.2. 共通追加設定

この共通追加設定は、人物の選択手法に関わらず選択できるものである。図 5 は追加設定の「接合強度を可視化」、「人物の重要性を可視化」、「タイム・スライダーを追加」で作成された。

最初の設定、「接合強度を可視化」では、ネットワークで二人の人物の間の接合強度が可視化される。接合強度とは、ある二人の人物の間の接合数 (現在、共通のイベント数) にて定められる。接合数が多い人物ペアの間のエッジは接合数が少ない人物ペアの間のエッジより太くレンダリングされる。

次の設定、「人物の重要性を可視化」では、ノードの大きさが人物の重要性 (接合度) によって可視化される。重要性とは、ある人物のネットワークでの次数 (エッジ数) で定められる。接合数が多い人物は接合数が少ない人物より大きく示される。この設定が選択されない場合でも、可視化画面の設定メニュー (歯車シンボルをクリックしてアクセス) でスライダーを使い、ネットワークでの人物



図 6. ネットワークで選択した人物の詳細モーダル

を最低字数で強調表示することができる。この設定を選択せずに、次のタイム・スライダーの設定を選択すると、次数スライダーとタイム・スライダーの使用は選言的となる。

第3の共通追加設定、「タイム・スライダーを追加」では、ネットワークでのエッジ（現在、イベント）を年でフィルターできるスライダーが可視化画面に追加される。このタイム・スライダーを使うことで、ネットワークの構造を年ごとに分析することができる（図8）。

最後の共通追加設定、「グラフで未接合のノードを表示」では、エッジのないノードも表示される。この設定により、データベース内で他の人物と関係ない人物も考慮させることができる。人物全員とこの設定を選択すると、データベース内の全人物がネットワークでレンダリングされる。

共通設定とは言え、追加設定の使用やコンビネーションには制限がある。選択手法ごとに述べてみたい。「全員」の選択手法では、タイム・スライダーが低速であるため、利用不可である。ノード数、又データベース内のイベント年の範囲において、計算が複雑になるためである。

「名前で選択」の手法では、未接合のノードを表示することが不可能である。この設定のクエリ構築では、選択された人物と必ず関係を持つ他の人物のみが考慮されるため、未接合のノードは発生しない。

最後に、「出典で選択」の手法では、タイム・スライダーと未接合ノードの追加設定は選言的である。タイム・スライダーを表示することで、ネットワークをイベントの日時で閲覧するため、イベントなしの未接合ノードは無用となる。また、未接合ノードを示す設定では、日時でフィルターすることが困難である。

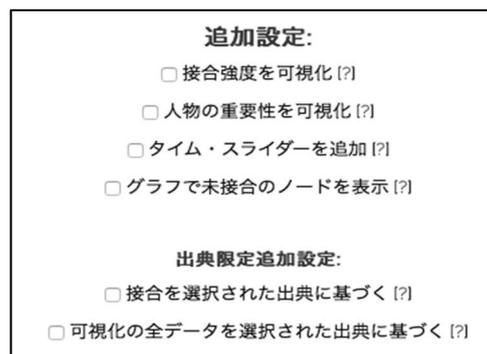


図 7. ネットワーク可視化ツールの追加設定

4.3. 出典限定追加設定

以上で述べた共通の追加設定以外に、「出典で選択」の手法に限られる、二つの追加設定がある。この設定では、出典の範囲を確定することによって、ネットワークを段階で出典中心的に作成することになるので、この追加設定は選言的である。

一段目は、出典限定追加設定が選択されていない、デフォルトの状態である。この段階では、これまで述べた通り、選択された出典に伝記情報が記される人物から始まり、その人物との関係性を有する人物がクエリされ、クエリ結果と選択された出典に現れる人物のセットの和集合に基づき、ネットワークがレンダリングされる。

次の段階は、「接合を選択された出典に基づく」の第1の選択肢である。この設定では、ネットワークのエッジ（接合）に基づくデータが選択された出典に限られる。選択された出典に基づくイベントに参加する人物のセットと、既定設定と同様に、選択された出典に伝記情報が載る人物のセットの和集合に基づき、ネットワークが表示される。このような仕組みでは、伝記情報が記されてある人物だけではなく、イベントのコンテキストのみで選択された出典で現れる人物も考慮される。

最後の段階は、「可視化の全データを選択された出典に基づく」の第2の選択肢で、出典において一番狭い範囲でネットワークを作成する設定である。この設定では、ネットワーク全体に基づくデータが、選択された出典に限られる。選択された出典に伝記情報が記されてある人物のみが考慮され、更にある二人の人物の間の接合も選択された出典に限られて定められる。この設定では、選択された出典のみのデータがネットワークのインプットとなるため、出典自体がもたらすネットワークが可視化されることになる。

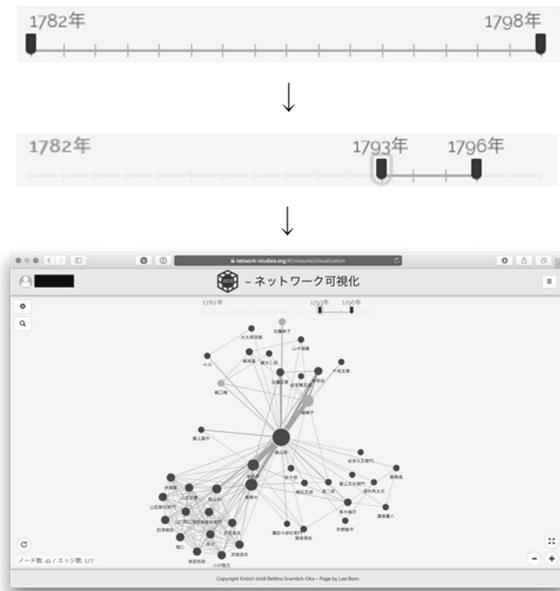


図 8. 図 5 でタイム・スライダーを使用した結果

5. おわりに

本プロジェクトの一部をなす「近世日本の知的ネットワークの研究」では、頼春水のグループを含め、様々な集団に焦点を当てる。そして、ネットワークにかかわった人々の書きものを幅広く取り上げ、総合的な分析を加えていく。人物が藩政や国政にかかわっている場合、各藩内また藩の枠を超えて広がる思想的ネットワークを研究することは、近世日本で政治改革がどのように論じられ、実施されたかを理解する重要な手掛かりとなるだろう。ネットワーク的アプローチは、近世日本の政治が幕府、諸藩、またそれ以下のレベルで、具体的にどう動いていたのかを理解する上でも有用な見方を提供してくれる。

データベースの構築の上、ウェブアプリケーション及びネットワーク可視化ツールの運用によって、この目標に近付くことができる。ウェブアプリケーションのデザインの利点としてモジュール化されたコンポーネントの導入が挙げられる。今後、ネットワーク可視化用のコンポーネント以外にも、空間的なデータの分析に用いられる地理情報システム (GIS) をコンポーネントとして取り入れる見込みである。

さらに、本研究プロジェクトでは、長期的には同様の研究手法を用いて、近世日本以外にも異なる時代や地域を対象とした研究を進めていく予定である。このような時空間情報及び文化横断的アプローチに基づくツールの提供は研究だけでなく、教育にも役立つだろう。

注釈

- (1) 図表ではログインされた状態が表示されるため、画像は編集されている。
- (2) CSV と XML フォーマットでも保存できることを予定している。
- (3) この点は可視化の追加設定にて制限される場合もある (項 4.2 参照)。計算の複雑性とパフォーマンスのため、既定では未接合のノード (データベース内で他の人物と関係ない人物) が表示されないが、追加設定にて未接合のノードも考慮させることができる。後者の方を選択する場合のみにデータベース内の全人物がネットワークで考慮される。
- (4) グループ数は属性にて異なり、確定ではないため、強調用の色の選択はランダム化されている。

参考文献

- [1] Padgett, J. F. and Ansell, C. K.. *Robust Action and the Rise of the Medici, 1400-1434*. The American Journal of Sociology, 1993, vol. 98, no. 6, p. 1259-1319.
- [2] Berman, M. L.. *Boundaries or Networks in Historical GIS: Concepts of Measuring Space and Administrative Geography in Chinese History*. Historical Geography, 2005, vol. 33, p. 118-133.
- [3] PostgreSQL. <https://www.postgresql.org/>, (参照 2018-01-30).
- [4] Harvard University, Academia Sinica, and Peking University. *China Biographical Database*. <https://projects.iq.harvard.edu/cbdb/>, (参照 2018-01-30).
- [5] Sophia University. *Japanese Biographical Database*. <https://network-studies.org/>, (参照 2018-01-30).
- [6] LoopBack. <https://loopback.io/>, (参照 2018-01-30).
- [7] AngularJS. <https://angularjs.org/>, (参照 2018-01-30).
- [8] UI Bootstrap. <https://angular-ui.github.io/bootstrap/>, (参照 2018-01-30).
- [9] bootstrap-drawer. <https://github.com/clineamb/bootstrap-drawer/>, (参照 2018-01-30).
- [10] angular-translate. <https://angular-translate.github.io/>, (参照 2018-01-30).
- [11] angular-table. <https://github.com/samu/angular-table/>, (参照 2018-01-30).

- [12] AngularJS Dropdown Multiselect. <http://dotansimha.github.io/angularjs-dropdown-multiselect/docs/#/main>, (参照 2018-01-30).
- [13] ng-file-upload. <https://github.com/danialfarid/ng-file-upload>, (参照 2018-01-30).
- [14] Lightbox2. <http://lokeshdhakar.com/projects/lightbox2>, (参照 2018-01-30).
- [15] Angular File Saver. <http://alferov.github.io/angular-file-saver/>, (参照 2018-01-30).
- [16] Angular-xeditable. <https://vitalets.github.io/angular-xeditable/#>, (参照 2018-01-30).
- [17] Animate.css. <https://daneden.github.io/animate.css/>, (参照 2018-01-30).
- [18] AngularJS Slider. <http://angular-slider.github.io/angularjs-slider/>, (参照 2018-01-30).
- [19] Almende B.V. *vis.js*. <http://visjs.org>, (参照 2018-01-30).
- [20] D3.js. <https://d3js.org>, (参照 2018-01-30).