データサイエンス時代の歴史情報基盤の構築

佐野智也，増田知子
Email：tomoya@law.nagoya-u.ac.jp
デジタルアーカイブ学会第3回研究大会
(2019年3月16日)

CC BY

「日本研究のための歴史情報」プロジェクト

法学、政治学、経済学で扱う
歴史資料（明治～占領期）のDB化と研究利活用

- 『人事興信録』データベース
- 「SCAPIN」データベース
- 『法律新聞』データベース
- 裁判例データベース(明治・大正編)

http://jahis.law.nagoya-u.ac.jp/

プロジェクト課題・メンバー

News

2018.12.14
法情報処理研究開発プロジェクト（リーダー・外山勝彦、名古屋大学情報基盤センター教授）において、次の論文が The Best Paper Award, The 31th International Conference on Legal Knowledge and Information Systems (JURIX 2018)を受賞しました。
Takahiro Yamakoshi, Takahiro Komamizu, Yasuhiro Ogawa, Katsuhiko Toyama: "Japanese Legal Term Correction using Random Forests"

2018.12.8
法制史学会中部部会第85回例会（於 名古屋大学CALEフォーラム）において、増田知子教授が大久保泰甫（名古屋大学名誉教授）著『ボワソナードと国際法：台湾出兵事件の透視図』（岩波書店、2016）の書評報告を行いました。

2018.9.28
歴史情報プロジェクトと多言語法令情報基盤構築プロジェクト（代表：松浦好治・名古屋大学特任教授）の合同研究会において、許 順福氏（華南師範大学法学院講師）が報告を行いました。

2018.8.3
『人事興信録』（第四版）データベースの公開にあたり、名古屋大学からプレスリリースを行いました。

過去の一覧

データベース

『人事興信録』データベース

『人事興信録』（人事興信所）のうち、第四版を全文検索できます。

⇒ 使い方はこちら

SCAPIN-DB

占領期（1945～1952年）において連合国最高司令官（SCAP）から日本政府に対して出された指令を全文検索できます。

『法律新聞』データベース

『法律新聞』1～92号（明治33年9月24日～35年6月23日）を全文検索できます。

開発中のデータベース

目録・画像情報

プロジェクトでは、以下の資料の目録および画像を作成しています。

- 入江俊郎関係文書・佐藤達夫関係文書
- ノッター・ファイル
- 木内信胤関係文書
- 「人事興信録購讀者芳名表」

研究業績一覧

[種別] タイトル	著者名/発表者名	発行所/発表雑誌等/発表学会等
[論文] 近代日本の『人事興信録』（人事興信所）の研究（4）	増田知子, 佐野智也	名古屋大学法政論集280号(2018.12)203-259頁
[学会発表] 条文の変遷を追跡できる情報基盤の構築－民法（債権関係）改正への対応に向けて－	佐野智也	情報ネットワーク法学会第18回研究大会 立正大学(2018.12.9)
[論文] 近代日本の『人事興信録』（人事興信所）の研究（3）	増田知子, 佐野智也	名古屋大学法政論集278号(2018.6)181-237頁

テキストデータの重視

デジタルデータ

- 画像データ
- テキストデータ

テキストデータの利用価値は高い

- 全文検索
- テキストマイニング

→　書籍や画像とは異なるレベルの研究手法

明治～占領期の資料のテキストデータ

ほとんど作られていない！！

- 当時の印刷技術
- 資料の経年変化
- 旧字体

↓

- OCRの精度が低い
- 校正に知識が必要

↓

高コスト化

テキストデータ作成の問題点

『人事興信録』第四版：

本文1992頁、1頁あたり約2300文字

- 大規模なテキスト化を受注可能な信頼できる業者の選定
- コストの抑制

「SCAPIN」

原本の印字が不鮮明なもの、文書全体がかすれてほとんど読めないものなどを含む2635件の文書

- 量よりも原資料の質が問題
  （OCRができないと、コストの増大につながる）

テキストデータ作成の方法

『人事興信録』
株式会社凸版印刷「高精度全文テキスト化サービス」を利用（作業現場も見学の上で決定）

「SCAPIN」
OCRによる元データの精度によって方法を選択
- 活字化した資料があるもの → 外注
- OCRの結果が良いもの → 外注
- OCRの結果が悪いもの → 専門の研究員による手入力

校正の底本は国会図書館所蔵版に統一

テキストデータと画像データの位置づけ

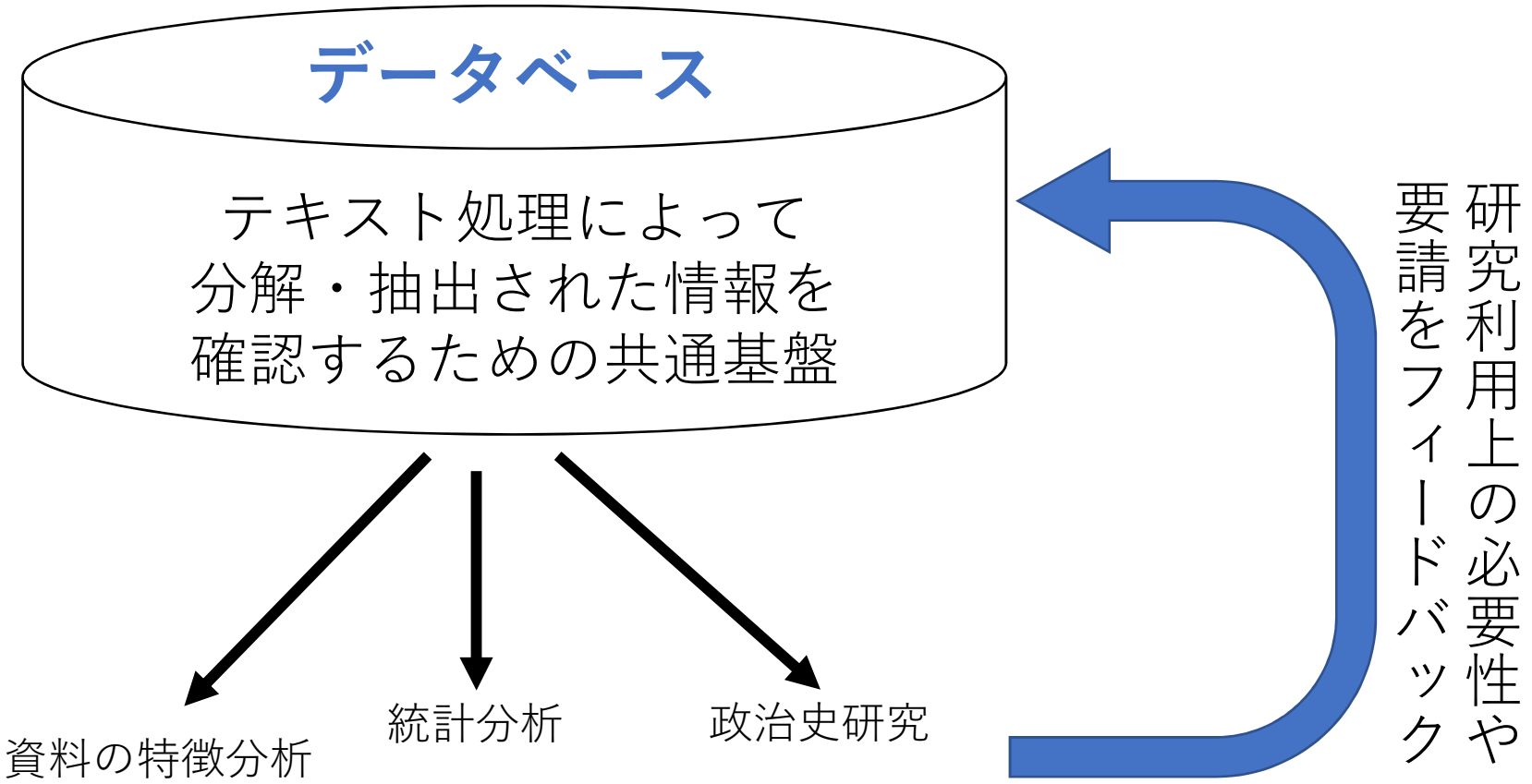
テキストデータとしての活用を重視

- テキストデータの精度は、99.98%以上
- 字体や体裁の再現にはこだわらない
  - 漢字の字体については、JIS第四水準の範囲内
  - 作成コストやテキスト処理の負担を軽減

画像データの役割

- 体裁などの原本の確認
  - テキストと原本を照合できるよう、データベース上では、ページ単位あるいは文書単位で、原資料の画像を簡単に確認できるようにしている

人事興信録研究におけるDB



人事興信録DBの特徴・機能

- 細かい検索項目
- 新字体・旧字体の差異があってもマッチング可能
- 氏名に特有な差異（「渡邉」と「渡邊」など）があってもマッチング可能
- 『人事興信録』に採録されている親や兄弟を表示

日本研究のための歴史情報　『人事興信録』データベース　第4版 [大正4(1915)年1月]

全項目一括検索　and

採録者名　and　性別　選択なし

生年月日　日付指定　範囲指定　西暦　年　月　日　年齢　年齢指定　範囲指定　西暦　1915　年で　～　歳

位階　選択なし　勲等　選択なし

功級　選択なし

爵位　選択なし　その他・資格　選択なし

身分　選択なし　在籍地　選択なし

旧身分・家柄　and

職業　and

親　and　親との続柄　選択なし

家族　and

記述部分　and

出身校　and　政党　選択なし　or

所得税　～　営業税　～

住所　and　電話番号　and

参照人物　and　参照次数　～　全て

項目間の検索　and

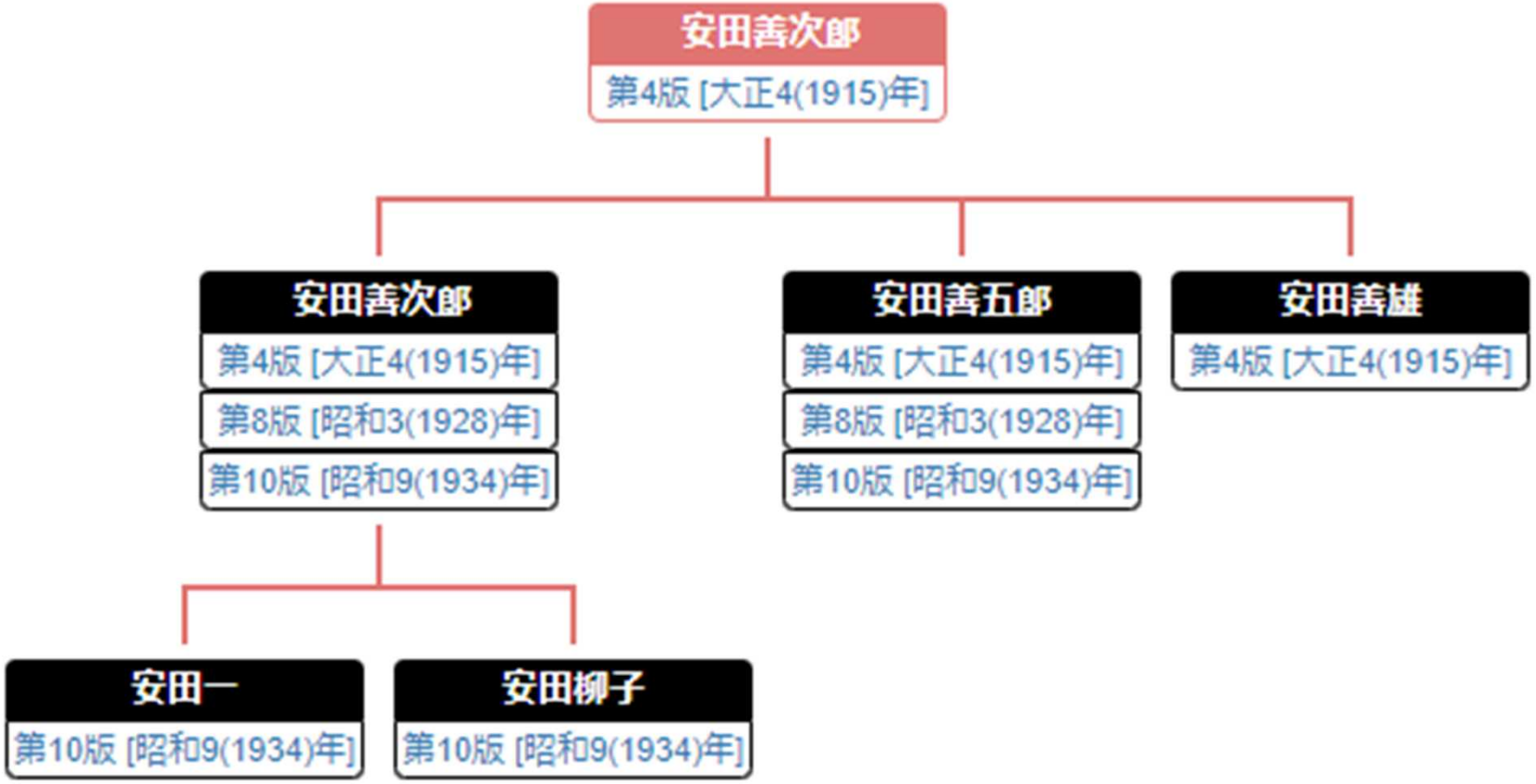
検索　クリア

安田善次郎 (男性)

人事興信録

実親子関係がある採録者

位階・勲等・功級	従六位、勲二等
爵位・身分・家柄	東京府平民



画像を表示

SCAPIN研究におけるDB

SCAPINは、内容自体があまり知られていない
↓
データベースの目的：検索して内容を読むこと
　様々な角度から内容を読解できるように資料を関連付けしている
- 資料で明示されている関連SCAPINへのアクセス
- 日本語訳資料
- 判明した具体的な立法との関係
- 『日本管理法令研究』に示されている、管理法令日誌の同日の日付の出来事を表示

SCAPIN-26: PROTECTION OF ALLIED AND AXIS PROPERTY

参考原資料

NDLデジタルコレクション
『SCAPINS』

Amended by

SCAPIN-270
SCAPIN-1845

See also

SCAPIN-1670
SCAPIN-1697

関連法令

昭和20年 大蔵省令 第80号　*1

*1 佐藤達夫関係文書 [1436]

『日本管理法令研究』

【日本語訳】 12巻（11a）
「管理日誌」 1945年9月13日

OFFICE OF THE SUPREME COMMANDER
FOR THE ALLIED POWERS

AG 091.112 (13 Sep 45) MG
(SCAPIN-26) 13 September 1945.

MEMORANDUM FOR IMPERIAL JAPANESE GOVERNMENT

THROUGH Central Liaison Office, Yokohama

SUBJECT Protection of Allied and Axis Property

1. The Imperial Japanese Government will impound immediately and report within fifteen days all property and other assets, books of account and other records owned or controlled, in whole or in part, directly or indirectly, by the Governments or any nationals, within or without Japan, of Germany, Italy, Bulgaria, Finland, Thailand, Rumania and Hungary.
2. The Imperial Japanese Government will cause to be preserved in good order and condition all property and other assets, books of account and other records, owned or controlled in whole or in part by the Governments or any national of (any or) the United Nations on December 7, 1941 and make a complete report to the Supreme Commander within one week.
3. Acknowledgement of this directive is requested.

FOR THE SUPREME COMMANDER:

/s/ Harold Fair.
/t/ HAROLD FAIR,
Lt. Colonel, A.G.D.,
Asst. Adjutant General.

おわりに

課題

明治～占領期については、多くの資料がテキストデータ化されていない

- 法律の基本資料である官報や帝国議会議事録すら、テキストデータにはなっていない

付記

トライアル版として『人事興信録』第八版を加えたデータベースを公開しています（4月正式リリース予定）

http://jahis.law.nagoya-u.ac.jp/who/search/

謝辞

本研究は、JSPS科研費基盤A（課題番号16H01998）の助成を受けています