

平成 29 年（2017 年）6 月 8 日

スケッチや画像から古典籍画像を検索できる最新 AI システムを開発

国立情報学研究所と国文学研究資料館による共同研究

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所（NII、所長：喜連川 優、東京都千代田区）コンテンツ科学研究系教授、佐藤 真一と同研究系特任研究員、松井 勇佑は、国文学研究資料館（国文研）^{(*)1} の山本 和明 特任教授と松田 訓典 特任助教との共同研究で、深層学習などの最新の人工知能（AI）手法により手描きのスケッチや画像から類似の古典籍画像を検索できるシステムを開発しました。本システムは 6 月 9 日（金）、10 日（土）に開催する「国立情報学研究所 オープンハウス 2017」でデモンストレーションを行います。

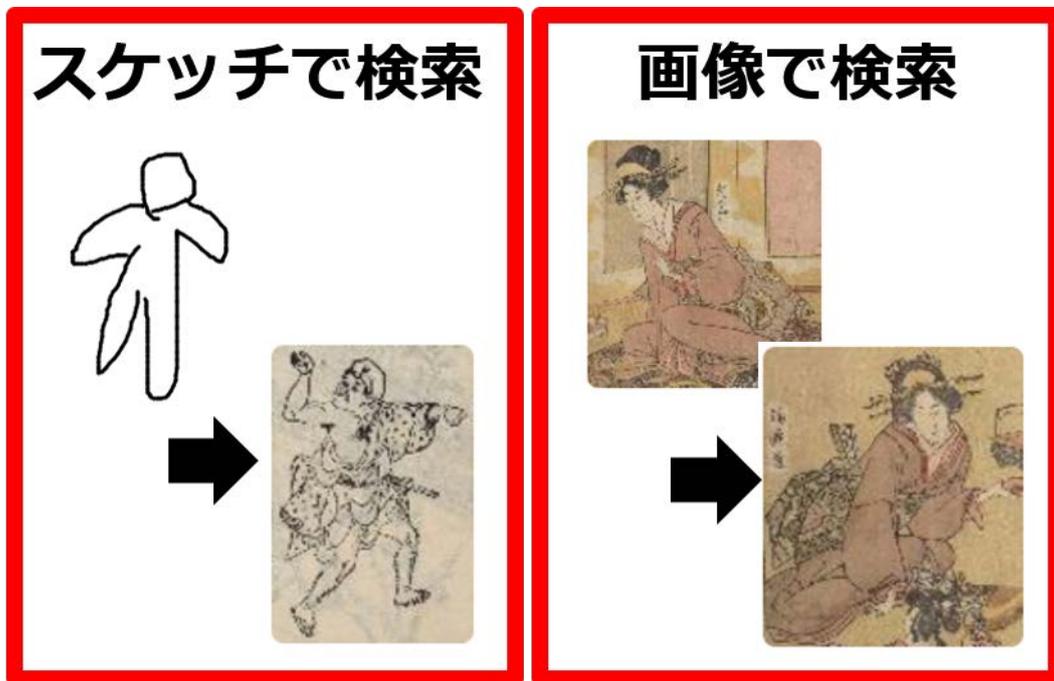


図 1 古典籍画像データに対する画像検索機能

本システムは、国文研との共同研究「画像検索のための構造化問い合わせ言語による歴史的典籍画像検索システム」の研究成果です。本共同研究は、国文研が NII との協働で公開を進めている歴史的典籍の画像データなどのオープンデータセット「日本古典籍データセット」^{(*)2} に対して、従来は不可能だった「画像内容を考慮した問い合わせ」ができるような画像検索システムの構築を目的としています。こうした画像検索を可能にすることで、古典籍を人々にとってより身近なものとし、さらには、古典籍に秘められた新知識の発見を支援することも目指しています。

今回開発した検索システムは深層学習および近似最近傍探索技術^(*)3)によって構成され、web ブラウザから実行できるウェブアプリとして提供します。

検索窓に手描きでスケッチをすると、似ている形状を含んだ古典籍画像をデータベースから検索します。検索にかかる時間は1ミリ秒以下で、画像中の検索対象領域が計百万まで増えても1ミリ秒程度で検索が可能です。検索結果の画像を検索窓にドラッグ&ドロップすると、その画像をもとに再検索が可能です。現在は「絵本和歌浦」「絵本時世粧」「絵本姫小松」「絵本玉かつら」「十二類絵巻」「絵本徒然草」の古典籍6冊のデータベースに含まれる計178枚の画像(1309領域^(*)4))を対象に検索を行うことができます。今後は、現在開発中の領域指定検索機能も組み込む予定です。

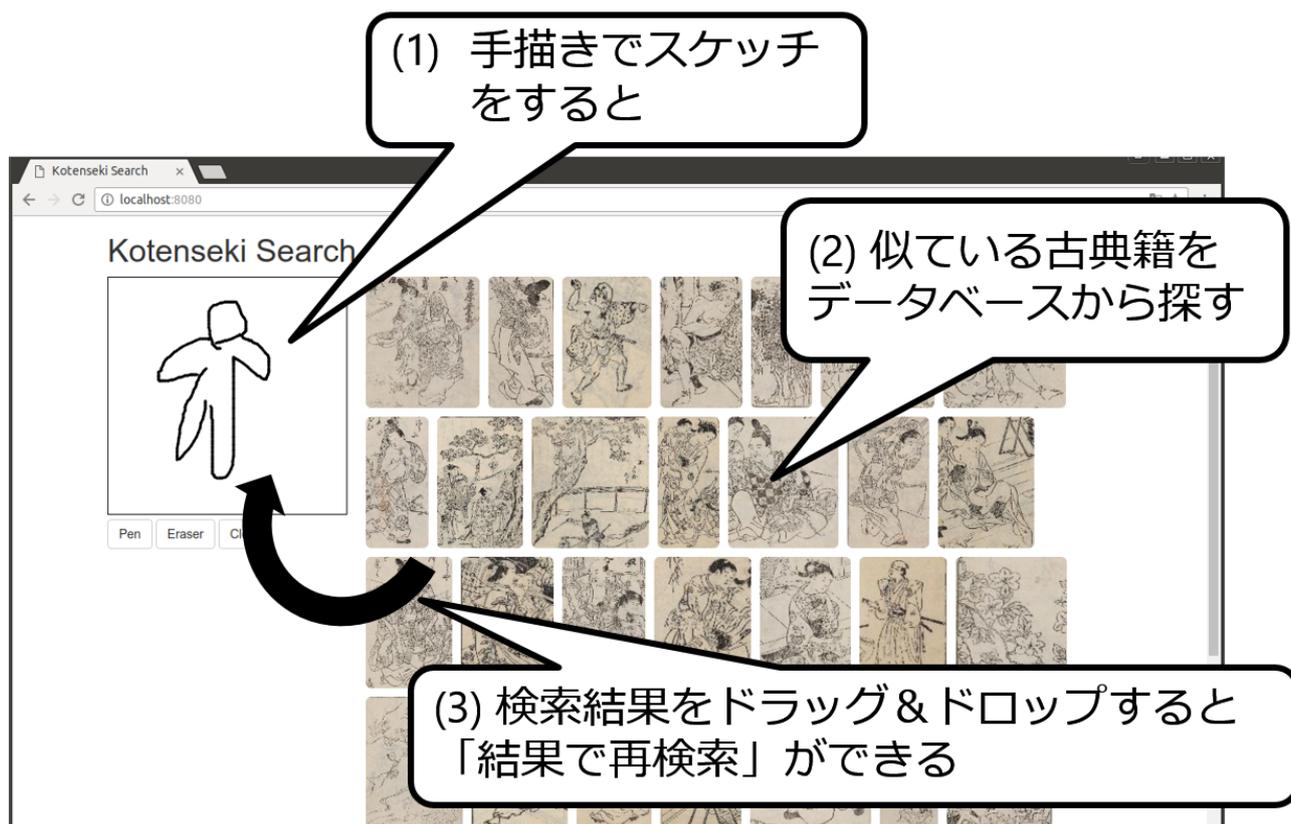


図 1 画像検索のイメージ

手描きのスケッチによる検索の場合、スケッチから即座に深層特徴量(画像情報を表現する高次元のベクトル)を抽出し、リアルタイムに検索を実行・更新することができます。

画像による検索では、検索窓の画像中の人物が着ている衣類の色に従って検索したり、人物の顔や画像全体の色彩が似ている画像を探したり、検索窓の画像中の川に反応して同様に川が描かれている画像を選んだりすることができます(別紙の検索結果例①~③参照)。

今後は、こうした検索を統合した枠組みとなる、画像検索の処理を構造的・統一的に記述する問い合わせ言語を開発・実装します。「平安時代の絵巻物の中から『光源氏を後ろから見つめる葵の上』の構図を持つ絵を探す」「天保時代の人々の生活を描いた画集の中から『男女が共に稲作に従事している姿』を探す」といった「画像内容を考慮した問い合わせ」による検索を可能にするほか、画像検索における統

一的な問い合わせ形式を提案することで、将来的には国内外の美術館に収蔵された古典籍画像の横断検索の実現を目指します。

画像検索のための問い合わせ言語

```
SELECT obj
FROM JeffPhotos
WHERE IsIn(obj.label, "dog")
```



検索された領域

「ジェフの写真フォルダーから
犬が写っている画像領域を検索する」

図 3 問い合わせの例

本共同研究は NII と国文研による所属機構や研究分野の枠を超えた共同研究で、情報・システム研究機構の未来投資型プロジェクトとして行われています。国文研は現在、文部科学省の大規模学術フロンティア促進事業「日本語の歴史的典籍の国際共同研究構築計画」^(*5) を推進しており、同計画において「日本語の歴史的典籍」約 30 万点を画像データ化し、既存の書誌情報データベースと統合させた「新日本古典籍総合データベース」の構築を研究基盤整備として行っています。

以上

〈メディアの皆様からのお問い合わせ先〉

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所
総務部企画課 広報チーム

直通：03-4212-2164 E-mail：media@nii.ac.jp

(*1) 国文学研究資料館： 大学共同利用機関法人 人間文化研究機構を構成する研究機関の一つ。ロバート・キャンベル館長、東京都立川市。国内各地の日本文学とその関連資料を大規模に集積し、日本文学をはじめとする様々な分野の研究者の利用に供するとともに、それらに基づく先進的な共同研究を推進する日本文学の基盤的な総合研究機関。<http://www.nijl.ac.jp/>

(*2) 「日本古典籍データセット」： 平成 28 年（2016 年）11 月 10 日付ニュースリリース「『日本古典籍データセット』公開で国文研と協働／国内研究機関のオープンデータの取り組みを支援・推進」(http://www.nii.ac.jp/userimg/press_20161110.pdf) 参照。

(*3) 近似最近傍探索： 最近傍探索は、検索問い合わせが与えられた時にもっとも近いデータ（最近傍解）をデータの集合から探し出す問題。文字認識や顔認識などに用いられる。近似最近傍探索では、厳密に最近傍を探索するのではなく、一定の誤差を許容することで計算量を減らしている。

(*4) 1309 領域： 現在は手動で抽出。今後は自動領域抽出を行う予定。

(*5) 「日本語の歴史的典籍の国際共同研究構築計画」： 国文研が中心となり、国内外の大学等と連携して[日本語の歴史的典籍](http://www.nijl.ac.jp/pages/cijproject/)に関する国際共同研究ネットワークを構築することが目的。<http://www.nijl.ac.jp/pages/cijproject/>

《別紙》

検索結果例①

Kotenseki Search



Pen Eraser Clear



黒い服を
正しく検索

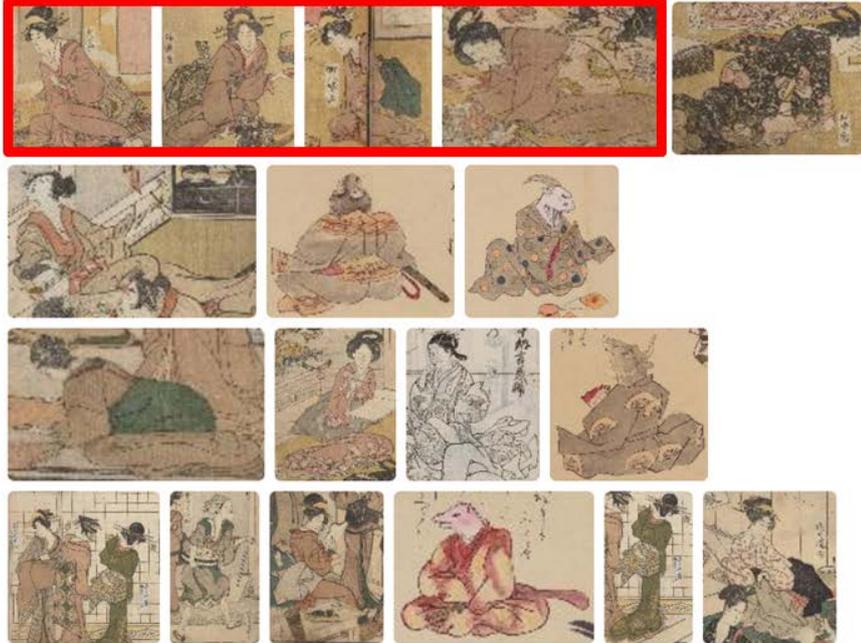
検索結果例②

似た色彩・人物

Kotenseki Search



Pen Eraser Clear



検索結果例③

Kotenseki Search



Pen Eraser Clear



「川」に反応