

ヲコト点データベースと検索システムの試作

堤 智昭, 土井 裕絵 (東京電機大学), 田島 孝治 (岐阜工業高等専門学校),
高田 智和 (国立国語研究所), 小助川 貞次 (富山大学)

本稿では, ヲコト点図のデータベース化とその検索システムについて報告する. データベース化したヲコト点図は, これまでに我々が提案してきた構造化記述手法にもとづいて電子化された主要ヲコト点 26 種である. 検索システムは Web アプリケーションとして実装し, インターネットを通じて試験公開を行った. さらに大学の訓点資料を扱う授業で本システムを利用し, アンケートによる評価を行った. その結果, 検索条件の入力方式改善の必要があるものの, 訓点学習に利用可能であるという評価が得られた.

Prototyping of Wokototen Database and Search System

Tomoaki TSUTSUMI / Hiroe DOI (Tokyo Denki University)

Koji TAJIMA (National Institute of Technology, Gifu College)

Tomokazu TAKADA (National Institute for Japanese Language and Linguistics)

Teiji KOSUKEGAWA (University of Toyama)

In this paper, we talk about prototype of database and search system of Wokototenzu. This database includes the 26 kinds of main Wokototen that computerized based on a proposed structuralization description method. This search system is web application. We published it experimentally through the internet. Then we used this search system at the university and estimated by a questionnaire. As a result, it was necessary to be Japanese word input method improvement of search condition, but we got a result that it's useful when learning Wokototen..

1. はじめに

日本を始めとする東アジア諸国には, 漢文で書かれた歴史的な資料が数多く存在している. かつての漢文は東アジアの共通書記言語であり, 社会・文化・政治など様々な事柄に関する記述が行われてきた. 漢文史料を研究することは, 東アジア諸国の歴史を紐解くために必要不可欠である.

漢文は, 中国語の文法に基づいた文語体の文章である. 中国語を母国語としない東アジア諸国では, 漢文を読者の母国語の文章として理解できるように訓読が行われてきた. この訓読を行うための補助的な役割を持つ符号を, 訓点と呼ぶ. 訓点には, 仮名や諸符号(句読点, 返り点, ヲコト点)など様々な種類が存在する.

訓点を書き加えられた漢文史料は, 東アジア諸国に数多く存在し, 漢文訓点資料と呼ばれ歴史・文学に関する研究対象となっている. 本研究は, 漢文訓点資料を対象とし, その中でもヲコト点と呼ばれる訓点が記された資料を対象とする.

これまでに我々は, ヲコト点を含む漢文訓点資料を電子化するための構造化記述手法を提案し, ヲコト点図データの電子化を行ってきた[1][2][3]. 築島裕[4]に記載された主要ヲコト点 26 種の電子化を試み, 提案する 7×7 マスのグリッド座標内にヲコト点図を配置可能なことを明らかにしてきた.

本稿では, これまでに電子化した主要ヲコト点 26 種のデータを活用し, ヲコト点の情報をデータベース化する試みについて報告する. 試作するデータベースは, ヲコト点の構成要素を用いて検索可能な, Web アプリケーションとして開発する. 本データベースは, 研究利用のみでなく漢文訓読資料の学習支援にも活用することも想定している.

2. ヲコト点の概要

2.1 訓点資料とヲコト点

ヲコト点は, 平安時代から鎌倉時代に広く用いられた訓点の一種である. ヲコト点は漢字の四隅を中心に加点され, 漢文を訓読する際, 助詞や助動詞の代用として用いられる. ヲコト点の読みは, 使用者の学派や時代によって異なり, 東大寺点や喜多院点など名前がつけられている. 近い学派同士の間点だとしても, その読みは若干の相違を持つことが多く, ヲコト点の種類は 100 種類以上存在すると言われている (築島裕[4]). 種類ごとに決められたヲコト点の読みは, どのような「形状」の点が, 漢字のどの「位置」に加点されるかによって決定する. そこで本稿では, ヲコト点の構成要素を「種類」「位置」「形状」「読み」の 4 つと定義する.

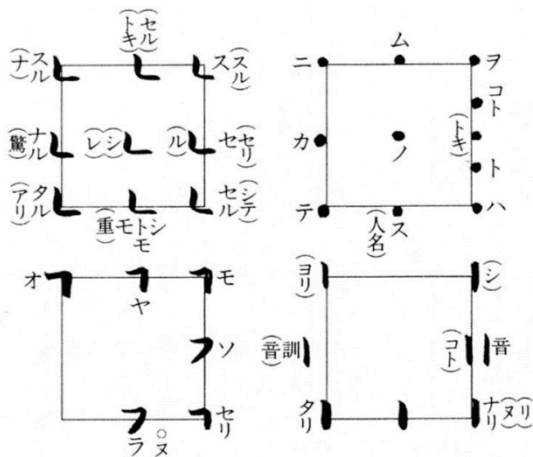


図 1 ヲコト点図の一部 ([4]より引用)

2.2 ヲコト点図

構成要素の違いによって様々な種類が存在するヲコト点には、ヲコト点の「種類」ごとに「位置」, 「形状」, 「読み」をまとめたヲコト点図と呼ばれる資料が存在する。ヲコト点図には、図 1 に示すようにヲコト点の形状と読み、付与された位置の組み合わせから、そのヲコト点がどのような読みを持つかが記載されている。ヲコト点図に書かれている黒い四角の枠は、任意の漢字をイメージしている。一つの枠には、比較的似通った「形状」のヲコト点が複数個まとめて図示される。このまとまりを「壺」と呼ぶ。ヲコト点図は複数の「壺」から構成される。

3 ヲコト点図のデータベース化

3.1 データベースの構造

本稿では、築島裕[4]に記載された主要 26 種のヲコト点図をデータベース化する。本データベースは、図 1 に示したようなヲコト点図データをテキスト形式として保存しつつ、位置や形状といったヲコト点を構成する情報を損なうことなく電子化する。そのために、図 2 に示すように、ヲコト点データベースはヲコト点テーブルとヲコト点の種類テーブルの二つを持つこととした。ヲコト点テーブルには、ヲコト点図に記載されたヲコト点一つ一つを 1 単位として保存する。ヲコト点の種類テーブルには、ヲコト点の種類(東大寺点, 喜多院点など)を 1 単位として保存する。

ヲコト点テーブルが持つ項目は、表 1 に示す 8 項目とした。ヲコト点の構成要素である「種類」「位置」「形状」「読み」については、それぞれ点の種類 ID, X 軸の位置・Y 軸の位置, 形状, 読みの項目に保存する。さらに、ヲコト点図の何壺目に記されていた点であるかを壺数の項目に保存し、ヲコト点一つ一つ固有の ID を点の ID の項目に保存した。点の ID は数字の連番とし、1 から順番に割り振った。

ヲコト点の種類テーブルが持つ項目は、表 2 に示す 3 項目とした。点の種類 ID の項目には「喜多院点」や「西墓点」といった点の種類名を保存する。群番号は、築島裕[4]に記載された各ヲコト点の群分類にもとづいて、1 から 8 までの数字を保存する。点の種類 ID は各ヲコト点の種類一つ一つに、1 から 26 までの値を重複しないように割り振り保存する。この値は、ヲコト点テーブルが持つ点の種類 ID と対応しており、図 3 に示すようにヲコト点テーブルからヲコト点の種類テーブルのデータを参照可能である。

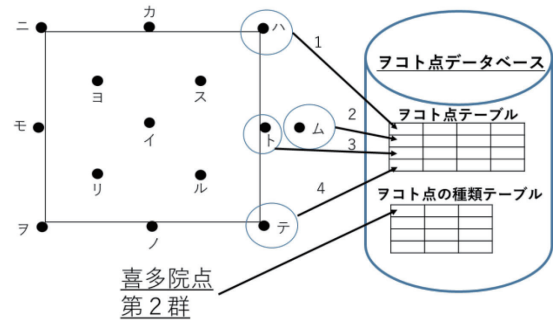


図 2 ヲコト点のデータベース化

表 1 ヲコト点テーブルの項目

項目名	説明
点の ID	ヲコト点一つ一つに振った固有の ID を保存する。
形状	ヲコト点を持つ形状 (「・」「 」) を保存する。
読み	ヲコト点を持つ読み (「ヲ」「コト」など) を保存する。
点の種類 ID	ヲコト点の種類データベースと連携するためのデータ (1 から 26 までの番号) を保存する。
X 軸の位置	-3 から 3 まで、X 軸方向のどの位置に点が存在したかを保存する。
Y 軸の位置	-3 から 3 まで、Y 軸方向のどの位置に点が存在したかを保存する。
壺数	何壺目に記載されていた点であるかの情報を保存する。

表 2 ヲコト点の種類テーブルの項目

項目名	説明
点の種類 ID	点の種類一つ一つに振った固有の ID を保存する。ヲコト点テーブルから点の種類テーブルを参照する時に使用する。
点の種類名	点の種類名 (「喜多院点」「西墓点」など) を保存する。
群番号	所属する群の番号 (1 から 8 群) を保存する。

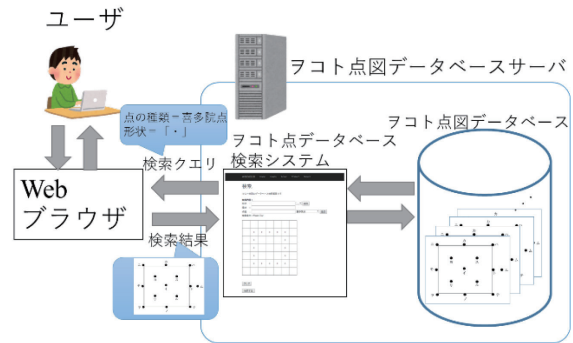
ヲコト点テーブル例

点のID	形状	読み	点の種類ID	X	Y	壺数
1	・	ニ	1	-2	-2	1

ヲコト点の種類テーブル例

点の種類ID	点の種類名	群番号
1	喜多院点	2

図 3 テーブルの項目と関係



3.2 入力データの正規化

ヲコト点の形状は星点（・）や線点（|）など多岐にわたる。これらは、文字ではなく記号であるが本データベースでは一般的な計算機が表現できる文字系の中で似通った形の文字で代用することとした。具体的には、Unicode の文字集合内の文字から似通った形を持つ文字を用いて代用した。Unicode の中では似通った形状が存在しない記号も使われている。例えば𐄂のような形状である。このような表現不可能な「形状」については、「|ト（よこ）」のように、二文字以上の組み合わせと、組み合わせる方向（縦、横）をカッコ書きで表す方式で表現することとした。

また、ヲコト点の種類名は同一のヲコト点図であっても、旧字体、新字体の二種類の表記方法が用いられる。例えば、築島裕[4]では「順曉和尚点」や「寶幢院点」のように旧字体で表記されている。本データベースでは、これらは全て新字体に統一し、保存することとした。

4. ヲコト点図データベース検索システム

4.1 システム構成

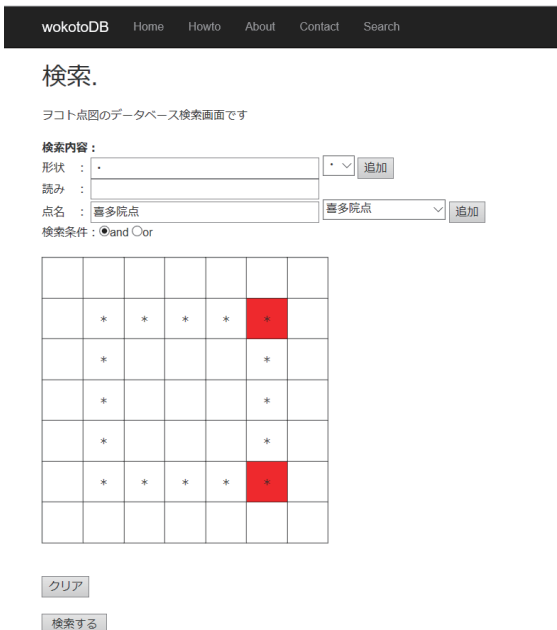
本ヲコト点データベース検索システムは、Web アプリケーションとして実装した。サーバの OS にはLinux ディストリビューションの一つであるCentOS7 を使用した。また、Web サーバにはApache2、及びKestrel を用い、アプリケーションの開発には、NETCore を用いた。本システムはhttp://cid.ninjal.ac.jp/wokototendb で試験公開を行っている。

本システムの概要を図 4 に示す。本システムでは、ユーザが Web ブラウザを用いてヲコト点図データベースサーバの検索ページにアクセスし、ヲコト点図の検索を行う。アクセスする Web ブラウザには、特別なソフトウェアやプラグインを追加する必要はなく、Google Chrome や Microsoft Edge など一般的な Web ブラウザを利用できる。

4.2 検索システムの設計

4.2.1 検索項目

本システムでは、検索に指定できる項目を「点の種類」「読み」「形状」、及び「位置」の 4 つとした。これら 4 項目は、ヲコト点図を調べると



きには形状と位置からどのような読みを持つかを調べる人が多いという、訓点研究者の意見をもとに決定した。

4.2.2 レイアウト

本システムの検索画面例を図 5 に示す。検索画面は、一番上部に黒色背景で示された検索システム内のメニューバーと、白色背景で示されたページの内容を表示する部分から構成される。メニューバーの一番右にある Search をクリックすると図 5 に示した検索画面を開くことができる。また、それ以外の WokotoDB、Home の文字をクリックすると本システムのトップページを、Howto をクリックすると本システムの使用方法を説明するページを、About をクリックすると、本システムについて説明するページを、Contact をクリックすると問い合わせ先に関するページをそれぞれ開くことができる。

(1) 検索画面レイアウト

入力フォームでは、テキストボックスとグリッドを用いて検索するヲコト点の条件を入力する。ヲコト点の形状、読み、点名はテキストボックス

に文字として入力する。ヲコト点の位置は7×7マスのグリッドから検索する位置をクリックして選択する。

それぞれのテキストボックスにはキーボードから検索条件を入力することができる。何もテキストボックスに入力されていない場合は、検索条件の指定が無いとみなす。また、読み、形状、点名に入力した検索条件全てを満たすものを検索する場合と、検索条件どれか一つでも満たすものを検索する場合の切り替えは、検索条件の **and**, **or** を選択することで切り替え可能とした。図5の例では、検索条件は、全てを満たす **and** 条件であり、形状は「・」、点名は「喜多院点」を指定し、読みの条件は指定されていない。そのため、「喜多院点」の中で「・」の形状を持つヲコト点、読みの内容にかかわらず全て検索される。

3.2 節でも述べたように本データベースでは一部のデータを正規化しているため、入力内容によってはユーザが意図する検索結果が得られない可能性が考えられる。例えば、点名について旧字体を用いて検索した場合は検索結果が得られない。そこで、形状と点名は、データベース内に保存されているデータの候補リストから選択できる方式とした。形状、点名のテキストボックスの右には、プルダウン式の入力候補一覧リストを設け、その中から検索したい候補を選ぶことができる。選んだ候補は、追加ボタンを押すとそれぞれのテキストボックスに追加される。複数の候補を追加することも可能である。その場合、テキストボックスには半角カンマ区切りで候補が入力され、それぞれの候補に該当するヲコト点全てが検索される。

点の位置を選択する7×7マスのグリッドは、選択されるとマスの色が赤色に変化する。選択されたマスは、もう一度クリックすると選択解除ができ、マスの色が白色に変化する。グリッドの下にあるクリアボタンを押すと、全ての選択が解除される。また、グリッドの中心5マスを囲っている「*」マークは、ヲコト点図の四角形(漢字の存在する範囲)を示す。図5の例では、グリッドの文字が存在している範囲の右上(2,-2)と右下(2,2)に存在するヲコト点のみが検索される。

(2) 検索結果画面レイアウト

本システムの検索結果画面例を図6に示す。検索結果画面は、最初に結果一覧を表示し、その下に結果を壺ごとに描画したものを表示する方式とした。

結果一覧の表は、点名、意味、形状、X、Y、壺数の項目を表示する。壺数を示す数字には、ページ内リンクを設置しており、クリックすると該当の壺の表示へ移動する。例えば、点名項目が喜多院点であり、壺数が1のリンクをクリックすると、喜多院点の1壺目の検索結果が描画された画面へ移動する。

壺ごとの表示は、図6の例のように、検索に該

図6 検索結果表示画面

当するヲコト点が存在した座標のセルに形状と読みが表示される。グリッドの中心5マスは薄紅色で表示し、ヲコト点図の四角形を視覚的に表現した。図6の例は、図5で示した検索条件(形状は「・」、点名は「喜多院点」、読みは任意、位置は(2,-2)か(2,2)のいずれか、を全て満たす)に合致するヲコト点を検索した結果である。この場合、喜多院点の1壺目(2,-2)の位置に存在する、形状「・」、読み「ハ」のヲコト点と、喜多院点の1壺目(2,2)の位置に存在する、形状「・」、読み「テ」のヲコト点が条件に合致したため表示された。

5. 評価実験

5.1 評価方法

本稿では、評価実験として訓点分野の研究者、及び漢文訓読の学習を行っているユーザに利用してもらい、使いやすいシステムとなっているかを確認した。実験は、大学の学部生を対象とし、訓点資料の読解を行う授業で実施した。授業の詳細は表3のとおりである。

アンケートの項目は本システムに関する内容と、アンケート対象者に関する内容について一度にまとめて Google フォームを用いて調査を行った。本システムに関する内容については以下のとおりとした。

- ・質問1: 本システムの使いやすさについて、5段階で評価してください

- ・質問2：本システムの見やすさについて、5段階で評価してください
- ・質問3：授業に有益だと思うか、5段階で評価してください
- ・質問4：本システムについて、当てはまる不満点（問題点）があれば選択してください
- ・質問5：その他を選択した方は具体的に不満点を教えてください

質問1から3は5段階の均等目盛り選択する方式とした。アンケート画面では、1から5までの数字とどれを選ぶかのラジオボタンが表示され、1に近づくほどマイナス評価、5に近づくほどプラス評価となる。質問4では、より具体的に問題点を洗い出すため、以下に示す4つの項目について問題となるかどうかをチェックボックス式で選択する方式とした。

- ①紙媒体のヲコト点図より扱いづらい
- ②紙媒体のヲコト点図より調べる時間がかかった
- ③検索の方法が分かりづらい
- ④検索結果が分かりづらい

また、それ以外にも気づいた問題点を自由に記述できるようにするため、⑤その他の項目を用意した。⑤その他を選択した場合、質問5でその具体的な内容を記述する方式とした。また、質問の最後に「感想、ご意見等あればお聞かせください。」という項目を用意し任意に回答を求めた。

アンケート対象者に関する内容については、以下のとおりとした。

- ・質問6：授業担当の先生の授業を何回受けていますか？
- ・質問7：本ページをどのような端末を使って利用しましたか
- ・質問8：その他を選んだ方は具体的に書いてください
- ・質問9：コンピュータ(スマートフォンを除く)の扱いについて、当てはまるもの全てにチェックをしてください

質問6は回答者の学習度合いを把握するために行った。入力は、授業を受けている回数を選択する方式とした。選択肢は以下5つである。

- ①初めて
- ②2回目
- ③3回目
- ④4回目
- ⑤5回以上

質問7は、回答者の本システムを利用する環境を把握するために行った。入力は、利用した環境をラジオボタンにより選択する方式とした。選択肢は以下の5つである。⑤その他を選んだ回答者は質問8で具体的に記述する方式とした。

- ①デスクトップPC、ノートPC（タッチパネルではない）
- ②ノートPC(タッチパネル式)
- ③タブレットPC

表3 授業の詳細

項目	説明
授業名	日本語学講読
対象学年	学部2,3,4年生
主な対象学部	人文学部, 経済学部, 人間発達科学部
授業資料	漢書楊雄伝天曆2年点(10世紀の漢籍訓点資料)

- ④スマートフォン
- ⑤その他

質問9は、回答者のコンピュータの扱いに関する技能を確認するために行った。回答は、4つの項目それぞれに該当するかどうかをチェックボックス式で入力方式とした。項目は以下のとおりである。

- ①Webサイトの閲覧ができる
- ②SNS (FacebookやTwitter等)の利用ができる
- ③Officeソフト (Word, Excel, PowerPoint等)を使ってレポートや発表資料が作れる
- ④簡単なプログラミングができる

5.2 アンケート結果

対象とした授業の時間中に、本システムを利用してもらいアンケートを行ったところ、授業履修者25名中、14名の学生から回答を得た。質問1, 2, 3, 4, 7, 9の回答結果をそれぞれ図7, 8, 9, 10, 11, 12に示す。質問6は全ての回答者が5回以上と回答した。質問5には、「追加ボタンを押さないと入力したことにならないのは分かりづらい。2つ以上選択するときのみ追加ボタンを押すという形にしてほしい。すべての位置という選択肢がないため全箇所を検索する方法がわかりづらい。検索結果のサムネイルが出力できるとわかりやすい。」という回答を得た。

質問1, 2の結果では、回答が2から5の値に分布していることから、本システムは普段紙媒体を使って訓点を学習している人でも極端に使いづらいことはないと考えられる。また、質問3の結果が4, 5の値に分布していることから学習者にとって有益なシステムであるという評価が得られた。一方、質問4の結果よりヲコト点の検索の仕方や結果が、紙媒体のヲコト点図を用いるよりもわかりづらい傾向があると考えられる。特に、質問5の回答より、検索条件の入力方式の改善が必要であると言える。

また本システムは、質問7の回答結果よりスマートフォンを用いた利用が非常に多いことがわかった。質問9より、回答者の大半が一般的なコンピュータリテラシーを備えていることがわかる。しかし、本システムのようにインターネット接続を用いて調べ学習を行う場合、コンピュータよりもどこでも容易にインターネット接続が可

能なスマートフォンが多く利用されていると考えられる。

5 まとめ

本稿では、ヲコト点図のデータベース化とその検索システムについて報告した。データベース化したヲコト点図は、これまでに我々が提案してきた構造化記述手法にもとづいて電子化された主要ヲコト点 26 種である。本システムを、ヲコト点の学習者に利用してもらい、使い勝手を中心とした評価項目についてアンケートを実施した。その結果、本システムの有用性に高い評価が得られ、訓点学習にも利用可能であると判断できる評価が得られた。一方、検索条件の入力がわかりづらく、方式の改善が必要であるという評価も得たため、今後改善を行う。また、本システムは、スマートフォンから多く利用されていることがわかった。そのため、今後はスマートフォン向けにシステムの GUI の改良を行う。また、スマートフォン向けアプリケーションとして実装することも今後検討していく予定である。

謝辞

本研究は JSPS 科研費 17K1850606 の助成を受けたものである。また、人間文化研究機構広領域連携基幹研究プロジェクト「異分野融合による総合書物学」の国語研ユニット「表記情報と書誌形態情報を加えた日本語歴史コーパスの精緻化」による成果の一部である。

参考文献

- 1) 田島孝治, 堤智昭, 高田智和: ヲコト点電子化のためのデータ構造と入力支援システムの試作, 人文科学とコンピュータシンポジウム「じんもんこん 2012」じんもんこん 2012 論文集, Vol.2012, pp211-216(2012).
- 2) 堤智昭, 田島孝治, 高田智和: 点図情報入力支援ツールによるヲコト点図の電子化, 人文科学とコンピュータシンポジウム「じんもんこん 2015」, じんもんこん 2015 論文集, Vol.2015, pp.185-190(2015).
- 3) 堤智昭, 田島孝治, 高田智和, 小助川貞次: コンピュータを用いた主要ヲコト点の関係性の解析, 人文科学とコンピュータシンポジウム「じんもんこん 2016」, じんもんこん 2016 論文集, 2016, 139-146 (2016-12-02).
- 4) 築島裕: 訓点語彙集成<第1巻>, ヲコト点概要, 汲古書院(2007) .

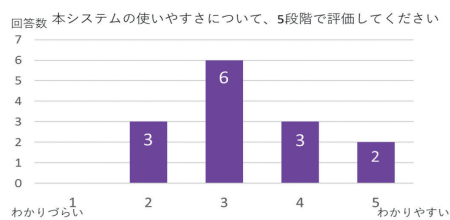


図 7 質問 1 の回答結果

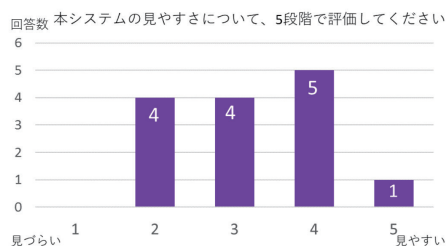


図 8 質問 2 の回答結果

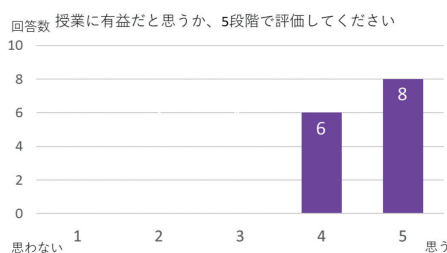


図 9 質問 3 の回答結果

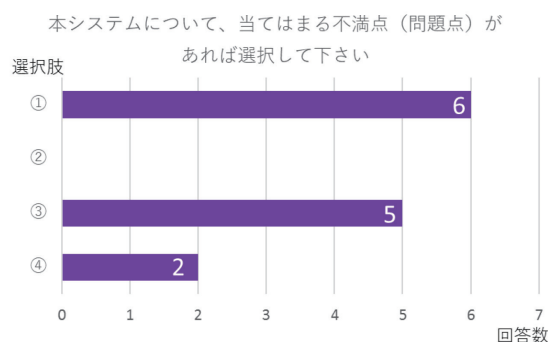


図 10 質問 4 の回答結果

本ページをどのような端末を使って利用しましたか

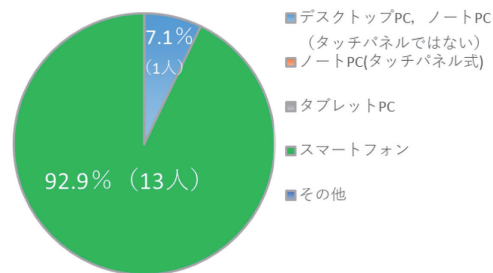


図 11 質問 7 の回答結果

コンピュータ(スマートフォンを除く)の扱いについて、当てはまるもの全てにチェックをしてください

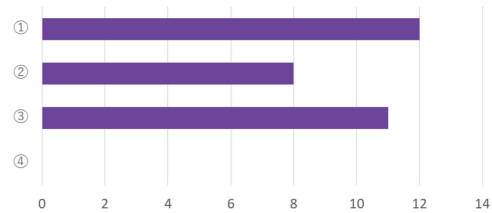


図 12 質問 9 の回答結果