

# 効果的な学生の就職支援：就職ナビサイトにおける企業情報横断検索システムの開発に関する研究

研究期間 平成 29 年度

研究代表者名 C.ソムチャイ

共同研究者名 岡村 拓弥

## 1. 研究背景

企業研究とは、自分の就職したい企業について調べる行為全般を指す言葉である。企業研究を行う利点としては様々な理由が挙げられる。まず、目指している職業には、ほとんどの場合複数の企業が存在しているため、それぞれの企業について、同業他社との違いを比較することで、その企業の強み・弱み・特徴などが見えてくる。就職後の待遇が良い企業を見つけることはもちろん、その企業について詳しく知っていれば、就職試験についても有利に動くことができる。

企業研究の手段としては、いくつかの方法がある。

- 大学の求人
- 合同企業説明会
- 就活ナビサイト

が代表的であると言えるだろう。

まず、大学に届いた求人情報の利点は、学科との合致性にあるだろう。自分の学んでいる学問に関係する系統の求人が多いため、大学で学んだことを活かせる仕事が多く見つかる傾向にある。欠点としては、利点と対照的に、学校の分野と関係のない求人はあまりなく、多様性に欠ける傾向にある。

次に、合同企業説明会の利点は、直接企業の社員と会話ができる点だろう。資料にはない様々な情報が得られるため、情報の「深さ」については特に優れている。しかしその分、実際に訪れることができる企業数は限定的であり、また、開催されている場所によっては交通費や時間の負担が大きくなる傾向にある。

最後に、ナビサイトを利用した企業研究についてである。他の 2 つに対して、閲覧できる企業数については大きなアドバンテージがある。多種多様な業種・職種を網羅し、勤務地も全国カバーしている。同時に起こる問題として、ヒット数が多いことが挙げられる。極端な条件で絞らない限り、数十件から百件以上の検索結果が得られることが多く、そのすべてをチェックしていくには大変な労力を要する。

このように、それぞれの手段に利点、欠点が存在しているが、この中でも特にナビサイトは、情報をカバーできる範囲の広さから、多くの学生が利用している。しかし、上記の通り、ナビサイト上での企業検索はヒット数が多いため、ユーザーに要求される時間的な負担は大きい。また、情報を閲覧するためにはその都度ページを開かなければならないため、複数の企業の情報を比較しようとするとその労力はさらに増してしま

う。

そこで、本研究では、データの自動収集と一覧表示、データソートなどの機能を持ったシステムを開発し、ユーザーの支援を目的とした研究をおこなった。

## 2. 問題定義と研究目的

### 2.1. 問題定義

ナビサイトでの企業研究において、問題となっている点を整理する。

#### 問題 1：すべての企業情報をチェックするのは難しい

サイトによって登録している企業数は異なるものの、すべてのサイトが 10000 件を超えている。そのため、条件をある程度絞って検索しても多くの企業がヒットしてしまい、すべてをチェックするには膨大な時間がかかってしまう。

#### 問題 2：企業情報の比較が難しい

企業検索を行う際、ナビサイトでは、検索の条件を入力した後、会社名の一覧が表示され、そのリンクを選択することで、選択した企業の情報が別タブ・ウィンドウで表示されるという流れになっている。このサイト構成は、選択した企業の情報が詳しくみられるという点では優れているが、多数の企業を比較しなければならない場合では、作業効率を低下させてしまう。

#### 問題 3：複数のナビサイトをチェックしなければならない

企業がナビサイトに登録するためには、登録料がかかるため、1つの企業はどこかひとつのサイトに登録するというケースが多い。そのため、ひとつのサイトだけでも多くの企業情報を閲覧しなければならないのに、さらに複数のナビサイトをチェックしなければならないという状況になっている。このため、一定の件数で情報収集をやめてしまう学生もおり、企業との出会いの場が失われている場合もある。

### 2.2. 研究目的

本研究では、上記の問題を解決するために、企業研究の効率化を行うシステムの開発を目的とする。

#### 提案 1：データのソート機能

すべての企業情報をチェックできないのは、企業の登録件数の多さが要因の一つであるが、直接の原因は、「検索後に」表示される件数が多いからである。その解決策の 1 つとして、データのソート機能が挙げられるのではないかと考えた。例えば、勤務地などの項目であれば、検索時に条件を入力することで件数を絞ることができるが、給与などの「数値」で表現される項目については、検索の段階では絞り込むことができず、問題 2 で取り上げたように、1 件ずつ開いてチェックする必要がある。

そのため、データソートが行えるようになれば、一定の数値に満たない企業をチェック

せずに済むため、通常の検索で絞り込んだ件数を、さらに少なくすることができる。

#### **提案 2：データの一覧表示**

問題 2 で取り上げた通り、多数の企業情報を比較する際、検索した企業の一覧から、1 件ずつ表示させなければならないため、非常に時間がかかってしまう。

そこで、企業情報を、一つの表に一覧で表示にすることができれば、余計な操作を省き、同時に、視覚的に比較しやすくなるのではないかと考えた。

#### **提案 3：データの統合**

複数のナビサイトを閲覧することは、多くの学生にとって、時間の消費量やモチベーションの面から、大きな負担となっている。そこで、複数のナビサイト上の企業情報を一元的に統合し、フォーマットを統一することで、問題を緩和できると考えられる。

#### **提案 4：データの自動収集**

提案 1～3 の手法を、企業研究の効率化のために提案したが、企業情報をユーザーが直接入力して行ってしまうと、その意味をほとんど成さなくなるため、企業データはシステムによって自動収集するべきである。

### **3. 本研究で検討すべき課題と対策**

「2.問題定義と研究目的」で取り上げた問題の、技術的な課題についてまとめる。

#### **3.1. 課題 1：収集対象の企業の選択**

ナビサイト上には、数万件もの企業データが存在するが、システムを利用するたびに全ての情報を収集すると、時間も容量も莫大なものになってしまう。そこで、次のような対策を提案する。

##### **対策：ブラウザ GUI による検索**

「Excel VBA と Internet Explorer の連携」の機能の中に、「Excel のフォーム上にブラウザを表示する」というものがある。この機能と、ナビサイトへのリンクのボタンを一体化した、「ブラウザ GUI」を開発し、ユーザーが必要な企業の情報を、実際にナビサイト上で検索してもらい、その検索結果に残った企業を対象にデータ収集することで、必要最低限な件数の情報を集めることが可能となる。



図 1：ブラウザ GUI

### 3.2. 課題 2：複数ウェブページにまたがる企業情報の抽出のルール化

「HTML ソースから見た、ナビサイトの構造」で、a タグによって各ページが繋がっており、情報ページに企業情報があることが分かった。本項では、その企業情報をどのようにして取り出すかについて述べる。まずデータ取得の最終目標である採用情報があるページにたどり着くために、リストページに表示されている、企業のトップページへの URL(以下企業 URL と表記)だけを取り出す必要がある。

#### 対策 A：企業 URL の特定

まず、企業 URL は a タグで記述されているという点に着目する。また、企業 URL 以外にも、企業の特集ページや、別サイトへのリンクなど、様々な a タグが存在するという特徴がある。つまり、リストページのすべての a タグから、企業 URL 以外の a タグ(以下ノイズと表記)をはじけば、企業 URL だけを取得できるということになる。説明の便宜上、工程をステップ①とステップ②に分けて説明する

## ステップ①

まず、おおまかにノイズをはじくために URL での選別を行う

たとえばリクナビの場合、情報ページの URL は

「<https://job.rikunabi.com/2017/company/top>」で始まるため、href に記述されている URL が一致しない a タグを除外する。

The screenshot shows a job listing for '株式会社 TSP' (TSP Co., Ltd.) with the following details:

- 業種: ソフトウェア
- 情報処理/通信/インターネット関連/コンピュータ・通信機器・OA機器
- 本社: 東京都
- 業務内容: 目録ソフトウェア開発] ◆モバイル・ウェアラブルの開発◆カーエレクトロニクスの開発◆ターミナル・POSレジの開発◆鉄道エレクトロニクスの開発◆メディカルエレクトロニクスの開発◆その他 RFID...

Callouts in the image highlight the following href values:

- href: <https://job.rikunabi.com/2017/company/top/r723700038/>
- href: <https://job.rikunabi.com/2017/company/top/r723700038/>
- href: <https://job.rikunabi.com/2017/mypage/reflection/regist/>
- href: <https://job.rikunabi.com/2017/company/seminars/r406620028/>

図 2: リストページの a タグの href

## ステップ②

さらに細かいノイズをはじくため、ステップ①を通過した a タグをさらに innertext の中身で判別する。リクナビの場合、①を通過できるのは、「企業名と企業 URL が入った a タグ」と、「企業へのアクセスを促すための a タグ」の 2 種類である。（本文編集時点でサイトの構成が変わり、「企業へのアクセスを促すための a タグ」が無くなったため、画像は疑似的に再現している。また、抽出処理は同じ原理で動いている。）このままでは同じ URL を 2 つ取得してしまうため、「企業へのアクセスを促すための a タグ」を除外する。この a タグは innertext が「詳細ページを見る」で統一されているのが特徴である。つまり、a タグの innertext が「詳細ページを見る」である a タグを除外すれば、①と併用することで、企業 URL のみを抽出することが可能である。



図 3: ステップ①を通過した a タグの innertext

ステップ①、②で取得した a タグの href と innertext を Excel シートに格納することで、それぞれ「企業 URL」「企業名」を保存することができる。また、例ではリクナビを取り上げたが、ノイズ除去に利用する URL や innertext の条件が異なるだけで、他のサイトでも同じ理論を利用している。

### 対策 B : 採用情報の特定

対策 B に関しては、特定までの段階をステップ①と②、③に分けて記述する。

#### ステップ①

A-1 で、情報ページの URL の取得に成功した。この URL を利用して採用情報を取得するまず、最初に気をつけなければならないのは、A で取得した URL を、そのまま利用するだけでは、採用情報が得られないことだ。なぜなら、取得した URL は、厳密には企業情報が載っているページで、採用情報が載っているページではないからである。



図 4 : 2 種類の URL

しかし、採用情報の URL には、取得 URL と共通する部分がある。それは、各企業を判別するための「番号」の部分である(以下企業番号と表記)。これは、研究対象のナビサイトすべてに共通する仕様である



図 5 : 企業番号の位置

そこで共通点に目をつけ、取得 URL から採用情報を生成する手法を確立した。まず、A で取得した URL を準備し、その中から、URL を構成する成分のうち、企業番号だけを取り出す。



図 6 : 企業番号の抽出

そして、あらかじめ Excel シート内に変換後の URL の原型を記述しておき、「企業番号」を埋め込むことで、採用情報が載っているページの URL を生成する。



図 7 : 番号の埋め込み

## ステップ②

生成された URL に対して、VBA を利用して自動でアクセスする。そして、アクセスしたページの中にある、採用情報を特定する手法を説明する。

まず、採用情報は文字列であり、table タグを利用して書かれている。また、記述のルールとして、以下のように並んでいる。



図 8: table タグの模式図

つまり、「採用人数」などの、表の見出しを table タグ内で文字列検索し、その次の文字列を取得すれば、採用情報を得られるということになる。

あらかじめ Excel シート内に見出しとなる文字列（入手したい項目）を格納しておき、それを参照して検索を行えば、必要な採用情報を取り出せるようになる。

The image shows two parts: an Excel spreadsheet and a job posting page. The Excel spreadsheet has columns for 'リクナビ' and 'マイナビ' with sub-columns for '変数名', 'プロシージャ', '役割', and '備考'. It lists various job details with IDs like Jm(0) through Jm(28). A red circle highlights the cell containing '勤務地' in the '備考' column of row Jm(24). A red arrow points from this cell to a job posting page. On the job posting page, under the '募集概要' section, '勤務地' is circled in red, and '東京、大阪' is circled in green. A green arrow points from the job posting page back to the Excel sheet, indicating a reference link.

図 9：シートからの参照

そして、図の 9. で示したように、要素は連番で構成されているため、文字列検索でヒットした要素の次の要素を取得すれば、その内容が取得できる。

HTML 文では次のようになる。

The image shows a screenshot of a job posting page on the left and its corresponding HTML code on the right. The screenshot includes a company logo, a description of the company, and a job listing section. The job listing section has a red circle around the word '勤務地' (Work Location) and a green arrow pointing to the text '東京、大阪' (Tokyo, Osaka). The HTML code on the right shows the structure of the page, with red circles around the '勤務地' and '東京、大阪' elements, and a green arrow pointing from the code to the corresponding text in the screenshot.

HTML Code:

```

<h3 class="company-title04">募集概要</h3>
<table class="company-data">
  <tbody>
    <tr>...</tr>
    <tr>
      <th class="company-data-th">
        <span id="03000005050457"></span>
        勤務地
      </th>
      <td class="company-data-td">東京、大阪</td>
    </tr>
    <tr>...</tr>
    <tr>...</tr>
  </tbody>
</table>
<h3 class="company-title04">給与・福利厚生(待遇)</h3>
<table class="company-data">...</table>
<h3 class="company-title04">参考データ</h3>
<table class="company-data">...</table>

```

図 10 : Web 上の位置と、HTML 文上の位置

### ステップ③

給与などの、採用条件に関する情報はステップ②で収集したが、会社の設立年や資本金等の情報は、企業のトップページに記載されているため、もう一度企業のトップページにアクセスし、ステップ②と同様の手段で収集を行う。

### 3.3. 課題 3 : それぞれのナビサイトの企業情報の統合

取得したデータは、リクナビ、マイナビ、キャリアスそれぞれのシートに保存されるが、このままでは、順次切り替えながら閲覧しなければならず、手間がかかる。

**対策 : 一つのシートに統合する。**

一度別のシートに保存することで、どのサイトからいくつのデータが取れたかの確認や、それぞれのサイトの仕様が変わった時の対処が簡単になる。

シートの名称はユーザーの任意に書き込みが可能で、一度ユーザーに名前を決めてもらうことで印象に残りやすく、後でチェックする際にも分かりやすくなる。

図 11：シート名記入フォーム

ただし、Excel の仕様上、同じシート名は使用できないので、重複を防ぐため、ユーザーの記入した名前の後ろに、現在時刻をシステムが自動で入力する。

例えば、「システムエンジニア」とシート名を入力し、「12 時 34 分 56 秒」に作成した場合「システムエンジニア-123456」というシート名で保存される。

また、統合されたシートの作成時にユーザーがメモを残せるようにできており

図 12：検索条件のメモ

「長崎の企業」「土日休みの企業」のように、検索したときの条件を簡単にメモできる。これにより、同じ業種で、細かい条件を変えて検索したときでも、後からわかりやすくなる。

Excel は、長い文章を表示させようとする、セルのサイズが大きくなり、全体として大変見づらいものになってしまう。そのため、セルのサイズはできるだけ小さくし、セルをダブルクリックすることで、セルの中身をメッセージボックスで表示できるようにした。

企業名	URL	採用人数	選考例	勤務地	勤務時間	採用ステップ	初任給	諸手当	昇給	賞与		
株式会社新日配薬品	https://pb	~10名	締切 随時	下関、山口、萩、岩	9:00~18:00	リー	230000円	要確認	年1回(5月)	年2回(7月)		
総合メディカル株式会社	https://pb	21~25名	—	職:本社および全	職:総合サポー	ントリー、会社説明会	230000円	営業、地域、家	年1回(4月)	年2回(7月)		
トランコム株式会社	https://pb	46~50名	締切 随時	関東・中部・関西及	9:00~18:00	要確認	224000円	現場手当、地域	要確認	年2回(7月)		
イー・キャランティ株式会社	https://pb	21~25名	締切 随時	名古屋、札幌	9:00~18:00	要確認	220000円	通勤交通費、年	1回	年2回		
(株)京橋エイジェンシー	https://pb	4~5名	書類選考および面	東京本社	9:30 ~ 18:00	書類選考および面接	220000円	職能手当、家族	年1回(4月)	年2回(夏)		
株式会社ヒューマンテクノシステムホールディングス	https://pb	16~20名	締切 未定	福岡県、東京都、愛	9:00~18:00	書類選考	218000円	通勤交通費支	年1回(4月)	年2回(6月)		
(株)ネットコムBB	https://pb	若干名	専門試験、筆記	東京営業所	09:00~18:00	専門試験、筆記	212000円	職能手当、役職	年1回(4月)	年12月)		
株式会社DNPアイディシステム	https://pb	2017年4月	若	東京本社(新宿)	9:00~18:00	休	211000円	社会保険(厚	昇給:年1回	社会保険		
九電ビジネスソリューションズ株式会社	https://pb	11~15名	締切 4月上旬	本社(福岡市)、北	(福岡市)、博田	(ア	206800円	時間外手当、請	要確認	年2回		
株式会社十八銀行	https://pb	未	—	—	—	—	205000円	通勤手当、居住	年1回	年2回		
株式会社南都銀行	https://pb	未	—	—	—	—	205000円	家族手当、ほ	年1回	年2回		
(株)三広	https://pb	未	—	—	—	—	205000円	交通費全額支	年1回(1月)	年2回(7月)		
IMD Alliance株式会社	https://pb	未	—	—	—	—	200000円	通勤手当あり	年1回	要確認		
宣研ロジエ(株)	https://pb	未	—	—	—	—	200000円	通勤手当、家族	年1回(4月)	月、12月)		
(株)全進堂	https://pb	未	—	—	—	—	200000円	職場体験	年1回(9月)	年2回(6月)		
(株)イーストウッド	https://pb	未	—	—	—	—	200000円	家族手当、1	回(4月)	年2回(6月)		
あおいたインフォメーションハウス(株)	https://pb	未	—	—	—	—	198600円	面談、筆記	198600円	通住手当、家族	年1回(4月)	年2回(7月)
川村産業株式会社	https://pb	未	—	—	—	—	192000円	家族手当	年1回(4月)	年2回(7月)		
(株)M&S UP	https://pb	未	—	—	—	—	190000円	筆記試験	190000円	役職手当、扶	養年1回(6月)	年2回(7月)
(株)アートプロセス	https://pb	未	—	—	—	—	183000円	面接	183000円	交通費など	年1回	年2回
(株)佐賀広告センター(佐賀新聞社グループ)	https://pb	未	—	—	—	—	180000円	筆記(通	180000円	営業外勤手当	年1回(4月)	年2回(7月)
有限会社 頼朝殿	https://pb	~5名	締切 随時	849-1615 佐賀県	17:00(休種1h)	要確認	165000円	宅利用の場合	年1回	年2回		
企業件数		22	作成日時	2016/12/14 11:37	詳細メモ	50未満で九州						

図 13 : ダブルクリックでの表示

#### 4. まとめと今後の研究課題

##### まとめ

本研究の目的は、就活ナビサイトを利用した企業研究における①登録情報数の多さ②データ比較の不便さ③複数のナビサイトをチェックすることのわずらわしさの3つの問題を解決するために、企業情報の自動収集・統合システムの研究開発を行った。その解決策として、①の問題は、データのソートによる選別や、選択した企業だけをまとめる機能などで対策を行い、通常の検索では絞り切れない項目に対しても高い効果を発揮することができた。②においては、一覧表示することで比較しやすくなり、ページをその都度開く必要もなくなったため、効率的な比較を可能にした。③の課題に関しては、データの自動収集と統合を行う機能によって、一つのワークシートで複数のサイトの情報を閲覧できるようになった。

本システムは、階層化されたウェブページ上のデータを自動収集し、シートに格納するという特徴を持つため、検索条件となる文字列などを変更するだけで多様なウェブサイトに対応が可能となっている。そのため、今後様々な研究が見込めるだろう。

##### 今後の展望

動作条件を変更することで様々なページに対応可能であるため、いくつかの研究が見込ま

れる。例えば、本研究の収集機能はそのまま生かし、研究対象だけを変更することで、「Amazon」「楽天」などのショッピングサイトの商品情報収集・統合や比較などは、サイトの構成も似通っているため、手軽に挑戦できるだろう。

「自動収集機能」に着目するならば、より多様なウェブサイトに対応するため、検索条件・抽出条件の体系化を図ることで、コンピュータに詳しくない人でも直感的に情報収集ができるプログラムが開発できるかもしれない。その場合は、HTML の階層の深さの設定や、取り出すタグの種類などをワークシート上で定義できるようにすることが、課題となるだろう。

「書式の統一機能」においては、すでに条件をある程度体系化しているため、より使いやすい形にまとめることができれば、データベースなどの、項目が入り組んだデータ形式に対して、「文字列操作」を一気に行えるプログラムの開発ができると考えられる。

## 5. 参考文献・URL

[1]近田 伸矢, 植木 悠二, 上田 寛(2013)

『Excel VBA で IE を思いのままに操作できるプログラミング術 Excel 2013/2010/2007/2003 対応』,インプレス.

[2]「VBA の IE 制御入門」

<<http://www.vba-ie.net/>>, (参照 2016-10-1) .

[3]「GoGo エクセルマクロをはじめよう！」

<<http://www.start-macro.com/55/>>, (参照 2016-10-1) .

[4]「by shigemk2」

<<http://www.shigemk2.com/>>, (参照 2016-10-1) .

[5]「HTML クイックリファレンス」

<<http://www.htmq.com/>>, (参照 2016-10-1) .

[6]「リクナビ 2017」

<<https://job.rikunabi.com/2017/>>, (参照 2016-10-1) .

[7]「マイナビ 2017」

<<http://job.mynavi.jp/2017/>>, (参照 2016-10-1) .

[8]「キャリアタス 2017」

<<https://job.career-tasu.jp/2017/top/>>, (参照 2016-10-1) .

[9]「【2017 卒向け】就活サイトをまとめてみました。」キャリアアカデミー池袋校

<<http://www.c-academy.co.jp/advice/254/>>,(参照 2016-10-1)