

企業情報検索支援 Web アプリケーションの開発

Development of Companies Information Search Support Web Application

五十嵐 雄太*1 新美 礼彦*1 小西 修*1
Yuta Igarashi Ayahiko Niimi Osamu Konishi

*1 公立はこだて未来大学 システム情報科学部
Future University-Hakodate, School of Systems Information Science

We consider “Industry Map” is very effective searching method for people. Because it is not searching of one concrete object like keyword searching but viewing at the whole image searching. However, it has problem that it is difficult to know that each industry contains companies, because “Industry Map” does not contain typical famous companies listed on industry list. In this paper, we propose new “Industry Map” which combined the number of the search results of the company name by using Google as the number of search results. The proposed system estimates a topicality of companies and famous companies by the number of search results. And the proposed system displays both companies which included in each industry and the industries list. We purpose that it is easy to search company’s information for job hunting. We verify practical effectiveness of searching by the system that uses the number of search results.

1. 背景と目的

就職活動を行う際、情報収集源として書籍、インターネット、TV、新聞などがあり、業界・企業に対して多くの情報を取得することが可能である。企業・業界について研究する際、こういった業界があり、こういった企業があるのかを知りたい場合がある。その時の情報収集の手段として「業界地図」[三原 07]が挙げられる。これはキーワード検索のように一つの具体的な対象を検索するのではなく、検索結果の一覧を視覚化し全体の概観を眺めながら企業検索できるというのが特徴である。この検索手法は業界・企業に関して知識がないユーザにとって非常に効果的な手法であると考えられる。しかし、「業界地図」の場合、業種一覧から業種を選択しそれに含まれる企業の全体像を提示するという手順なので、この業種にはどんな企業が含まれているということがユーザは想像しづらい。そのため業種について知識のない者にとって、それを掴むために多くの業種を閲覧するという手間がかかってしまう。

そこで本研究は、「業界地図」で企業を評価する指標として用いられる収益性、規模などの数値データ、企業詳細情報を Web 上から取得し DB に蓄える。それらのデータを視覚化することにより新しい業界地図の作成を行う。また google での企業検索結果数（以下より検索数データと略す）を組み合わせた表示を行う。これを用いる理由としては 3.4 の「google の検索結果数の指標について」という項目で述べる。このデータを用い、業種一覧と同時にその業種に含まれる代表的な企業を同時に表示することにより、ユーザの企業情報検索支援を行えることを目的とする。

2. 関連研究

視覚化を用いたツールおよび検索支援ツールについて以下に述べる。

2.1 視覚化を用いたアプリについて

(1) 日経マップ 日経平均 225 銘柄の業種ごと、企業ごとに株価の上げ下げを可視化したサイトである。[日経マップ] 株価の上げ下げを色で、企業の規模を面積で表現している。株価情報の更新は 5 分おきに約 20 分遅れの情報を公開している。

(2) grokker2 Web 検索結果のクラスタリングおよび検索結果の視覚化を行っている Web アプリケーションである。[grokker2] ユーザのあいまいなキーワード検索に対して、そのキーワードに関連する語を表示している。さらにこれらの語をクラスタリングのラベルとして用い、グループ化している。これは従来の Web 検索結果のように文字列のリストを表示するのではなく、クラスタリングによって分けたグループを可視化することにより検索支援を行っている。

2.2 関連研究と本研究の特徴

日経マップでは企業の規模を方形の面積で表し、株価の上下を色で表現している。また全体の概観を感覚的に把握できる表示を行っている。そのためどの産業の銘柄に上げ下げが多いのかや時価総額の大小を一目で感覚的に知ることができる。しかし、親ノード（業種）も小ノード（企業）も方形で表現しているため、階層があることを把握しづらいという欠点がある。grokker2 のアプリについて、階層を持つノードはすべて円で、これ以上階層を持たないノードについては方形で表現しており、ユーザにとって階層の有無を把握しやすいという利点がある。また円ノードに含まれる代表的な小ノードが表示されているため、どのノードにどんなものが含まれているかの把握が行える。

本アプリケーションは日経マップの利点である、色、大きさによって企業・業界を視覚化することにより感覚的に企業を把握することができることと、grokker2 の利点である階層構造の有無を把握する表示方法および親ノードが持っている代表的な小ノードの表示を行うことにより企業情報検索支援を行う。

3. 提案するアプリケーション

提案するアプリケーションの構成について以下に述べる。

連絡先: 〒 041-8655 北海道函館市亀田中野町 116-2

公立はこだて未来大学 システム情報科学部

新美 礼彦

TEL:0138-34-6222 FAX:0138-34-6301

E-mail:niimi@fun.ac.jp

3.1 アプリケーション構成

本システムは視覚化を行うために必要となる情報を Web 上より情報抽出を行い、DB に蓄える部分と、DB に蓄えたデータを視覚的に表現する部分の 2 つから構成されている。ユーザは Web ブラウザから本システムにアクセスし、企業探索を行うことができる。(図 1 参照)

本アプリケーションは Java で実装されており、Web ブラウザから利用可能である。システム内にサーバと企業情報を蓄えた DB を置く。また googleAPI を用い、google での検索結果数を DB に蓄える。

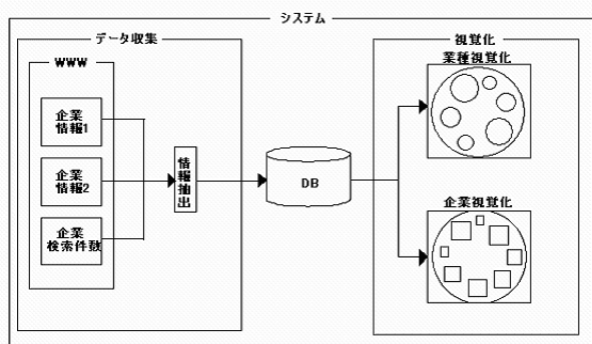


図 1: システム構成図

3.2 扱うデータについて

本アプリケーションで扱うデータの種類は、大きく 2 つに分けられる。一つは企業を視覚化するために用いるデータで、一つは企業の詳細情報である。前者の方のデータは、NIKKEINET [NIKKEINET] というサイトで企業を「総合得点」、「規模」、「収益性」、「安定性」、「成長性」の五つの指標を基にした企業評価を行っている。それぞれ 1000、100、100、100、100 点満点で、各企業ごとに点数がつけられている。後者は東洋経済企業データバンク [東洋経済] から企業の業種、事業内容といった企業の詳細データが掲載されているサイトである。

本システムの情報元としては、先の節で挙げた 3 つの Web サイトから情報抽出を行った。NIKKEINET では、企業名と、証券コード、4 つの指標値と総合得点、東洋経済 Web からは業種、事業内容といった企業概要、Google からは企業名の検索件数の抽出を行った。企業検索件数を含めると抽出する項目は 15 個となる。大業種については日本経済産業 [日本標準産業分類] を参考にし、中業種から自動的に推測を行った。情報抽出は、システム側で指定した情報のみを抽出し、データベースへのデータ挿入を行った。

実際の動作としては、まずシステム側で指定した URL (NIKKEINET など) にアクセスし、そのページ自体を html ファイルとして保存する。再びデータ構築を行う際までのバックアップ用ファイルとして用いるためである。

指定した情報名の抽出には Java 言語のパターンマッチを用いて、html ファイル内のタグに着目し、必要な情報の抽出を行い、データベースへのデータ挿入を行う。企業検索件数については GoogleWebAPI を用い、取得を行っている。取得した検索件数は各業種ごとに csv ファイルとして保存している。検索ワードとしては NIKKEINET で抽出した企業名に (株) や株式会社などを除いたワードで検索を行っている。

3.3 表示方法

先に挙げた企業評価の点数を用いて、業界・企業を視覚的に表現する。視覚化の手法として関連研究で記載した日経マップと grokker2 の 2 つの視覚化手法を組み合わせる。提案する表示方法について以下に図を載せる。図 2 がユーザがこのサーバにアクセスすると最初に表示される画面のモデル図で、図 3 が特定の業種を選択した場合に表示される画面のモデル図である。

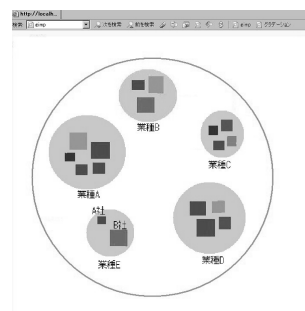


図 2: 業種一覧表示

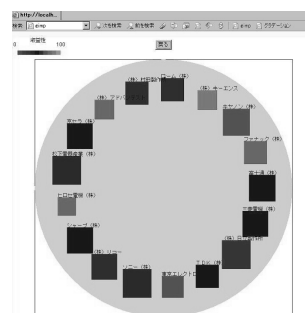


図 3: 企業一覧表示

図 2 では大まかな業種の一覧を表示する。円ノードは業種を表している。円ノードに含まれている方形のノードは代表的な企業を表示する。代表的な企業の選定に、検索数データを用いる。そこから階層を辿り、最終的な階層の表示が図 3 である。大きな円が業種を示しており、方形が企業を示している。

本システムでは企業に直接アクセスするのではなく、「大業種」「中業種」「企業」といったように、段階ごとのアクセスを行い、企業探索を行う。このようにカテゴリーを設け階層構造で表現することにより、莫大なデータの整理を行い、必要な情報にアクセスするためにフィルタをかけ、効率的な検索を行うことができる。

業種分類について、中分類は東洋経済 Web を参考にし、大業種については日本標準産業分類を参考にしている。本システムでは中分類については、29 種、大業種については 5 種類を扱っている。

3.4 google の検索結果数の指標について

ユーザが最終的に辿りつきたい情報は企業情報である。そのため業種一覧を閲覧しても、どの企業がどの業種に属しているのかが分からず、一つ一つの業種を見て、企業を探すという手間がかかってしまう。そこで業種と同時にそれに属するユーザに知られている可能性の高い代表的な企業を表示することにより、その手間を省くことができると推測される。

代表的な企業の選定として検索数データを用いる。検索結果数が多ければ、認知度の高い企業であると推測できる。また、企業に関して大きなニュースがあった場合、Web上の検索結果数は増加する。その増加数が多ければ話題性のある企業だと推測し、ユーザにとって知られている企業だと推測される。この情報を考慮した表示を行うことによりユーザの検索支援を行えると考えられる。

検索に用いる企業名は、表記を変えて数パターン試し、最も検索結果の多かった表記を採用した。

4. 実装

企業評価において総合得点、規模、収益性3つの指標を用いた業界地図の作成を行った。結果の表示は以下に示す図である。図4が業種で大分類で分類した画面で、図5が業種で中分類で分類した画面、図6が業種で分類した画面ある。また、図3が特定の業種を選択した時の画面である。図7が企業を選択した時に表示される企業詳細情報画面である。

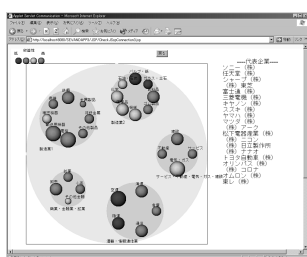


図 4: 大業種一覧の画面

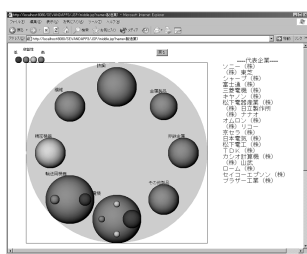


図 5: 中業種一覧表示

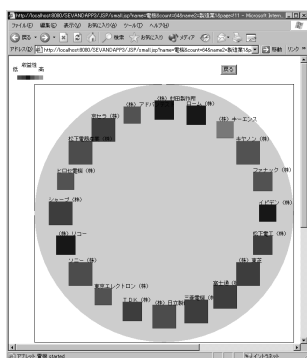


図 6: 現システムの業種一覧表示

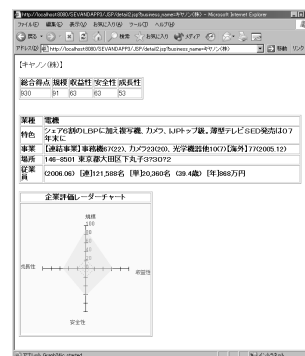


図 7: 企業詳細情報表示

本システムでは、Webブラウザからのアクセスする。ユーザのキーワード入力ではなく、マウスアクションによってブラウジングを行うシステムである。本システムでは「大業種一覧」「中業種一覧」「企業一覧」といったようにステップを踏み、企業を検索する。(図8参照)本システムを起動した初期画面では図4のような表示となる。5つの円が大業種を表しており、その円に含まれている球が中業種を表している。中業種の大きさによって規模を表現し、球の色によって収益性の高低を表している。収益性の高低の判断としては、同図の左上にある4つの球によって判断することができる。

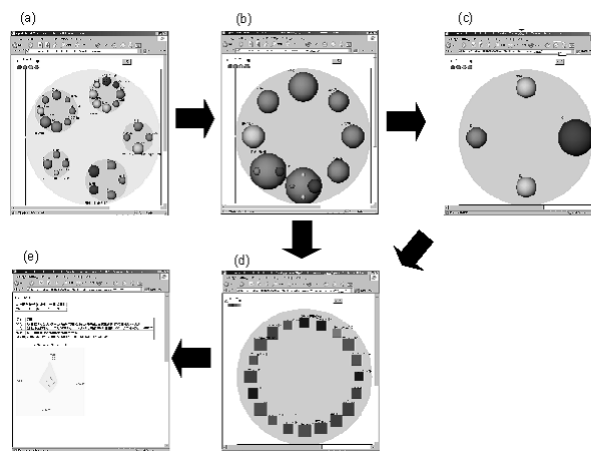


図 8: 画面遷移図

5. 実験・評価

提案した行表示法の有効性を検証するために、まず、データ収集に関する部分、企業の検索件数を取得する部分、表示を行う部分の3つの動作確認を行った。

この実験では、指定したURLから情報抽出に関する動作確認を行った。情報を抽出するサイトとしてはNIKKEINET、東洋経済Webからの情報抽出を行った。NIKKEINETのサイトには1ページにつきおよそ50社分の企業情報が掲載されている。そのページから1社ずつの会社情報を抽出するため、東洋経済Webサイトにアクセスし情報抽出を試み、実際に500社分のデータが収集できることを確認した。実験では、一度に必要な全データを取得するため、データ取得時に相手

サーバに負荷をかけてしまう事があった。

企業の検索件数を取得する部分に関して、検索に用いる企業名の表記を変えて数パターン試し、最も検索結果の多かった表記を採用した。実験では、googleAPIを用いたため、サーバの状態によっては、googleの連続クエリ使用の制限に引っかかってしまい、取得プログラムが中断することがあった。

表示に関して、画面遷移が正常に行われているかの検証を行った。検証結果として多くの企業を含む業種については中業種一覧を細分化して表示という手順を踏み、企業へアクセスすることを確認した。また企業数が少ない場合については、中業種一覧の後、企業にアクセスすることを確認した。本実験では29業種/100社のアクセスを行い、すべて正常な画面遷移を行うことができた。

作成したシステムを実際に大学4年生および大学の就職担当職員に使用してもらい、使用感を尋ねた。表示方法については、おおむね好感を得られたが、「既存の就職支援サイトと連携できないのか」、「分類が利用者の直感とあわないのでは」とのコメントがあった。

6. 考察

本システムの特徴として以下のものが上げられる。

1. 企業・業界に対して全体を視覚的に見渡せる表示を行っているため、企業同士の比較を行いやすい。
2. 企業の詳細情報を掲載し、経営状態を指標で表すことで、経営知識がない人でも企業の経営状態について簡単に知ることができる。
3. 参照しているWebサイトが優良企業ランキングで掲載されている企業を対象として行っているため、それ以外の企業に対しては検索を行うことができない。
4. 企業・業種の視覚的な表現により、容易に規模、収益性の2点についての経営状態をつかむことができる。また、業種内での主流企業の把握も行いやすい。
5. 代表企業の表示により、業種に対してのイメージがつかみやすい。
6. データを更新する際に、特定サイトに負荷をかけてしまう。
7. システムにインタラクティブ性がないため、すぐ飽きてしまう。

企業一覧を面積の大きさや色で表示することにより企業比較を行いやすく、ユーザにとって知られている可能性の高い企業以外にも注目を引くようになっているので検索の支援をするにあたり効果的であると考えられる。さらに企業の詳細情報を掲載することにより、規模、収益性以外の情報を収集できる。

また、代表的な企業の表示については、ユーザが業種と企業名どちらのほうが知っているかと比較した場合、業種よりも企業名のほうが知っている可能性が高い。そのため業種一覧と同時にそれに含まれる企業名を表示することによる検索支援は有用性が高いと考えられる。

対象企業や業種分類に関しては、登録データを増やしたり、ユーザになじみのある分類を取り入れるなどの改良を行う必要があると思われる。

データ更新時の相手サイトへの負荷に関しては、現在取得している就職関係のデータはそれほど頻りに更新されるデータではないため、データ取得に複数回に分けるなど、ある程度の時間をかけて行うことで回避できると考えられる。

7. 結言

本研究では、検索数データを用いたWeb上での業界地図の提案を行った。本研究の目的は、企業のデータの可視化表現と代表的な企業の表示により検索支援を行えることであった。提案した業界地図を実際に構築し、その評価を行った。

就職活動を始めたばかりの学生にとって、視覚的に企業・業種を表示することは、今まで知らなかった企業の発見や業種内での主流企業を判断することができるため、視覚的表現は企業探索において有効な手段であると考えられる。一方、特定のサイトに負荷をかけてしまったりインタラクティブ性がないといったGUIに関する問題点が挙げられる。負荷をかけないために時間制限を設けてのアクセスや、リアルタイムの情報の付加、システムに動的な要素を取り入れることにより、インタラクティブ性を向上させ、より有用性の高いシステムが構築できると考えられる。

参考文献

- [三原 07] 三原 淳男, 業界地図がまるごとわかる本 2007 年度版, 高橋書店 (2007).
- [日経マップ] 日経マップ, <http://n225.jp/>
- [grokker2] grokker2, <http://www.grokker.com/>
- [NIKKEINET] NIKKEINET, <http://www.nikkei.co.jp/>
- [東洋経済] 東洋経済企業データバンク, <http://www.toyokeizai.net/data/>
- [日本標準産業分類] 日本標準産業分類, <http://www.stat.go.jp/index/seido/sangyo/index.htm>