

日本全国 IT 導入事例分析可視化システム

1. 背景

ニュース等のメディアを通して IT 業界の状況についてよく耳にするが、実際にはそれはある一部のデータのある一面を見ただけであり、IT 業界がどのような世界になっているのか、その全貌を眺めて見た人はいないだろう。本プロジェクトは、IT 導入事例を対象にし、それを様々な角度から分析することにより、それを明らかにしようというのが狙いである。急速な発展を遂げてきた IT 業界の中で、各 IT 企業は、生き残りをかけ、今どの分野が注目され、次はどの分野に投資すればよいのか、また、どの企業がどの分野に強く、どの企業がライバルであり、また協力して行けばよいのか、絶えずその判断を差し迫られている。本ソフトウェアにより、企業の動向や技術の動向などを把握することができ、方向性を示す一つの判断材料として各企業に情報を提供することができると考えている。かつてホームページ制作代行だけでも高額なビジネスになっていたように、IT 業界は進歩が速く、システムの投資時機がコストに大きく左右する場合が多い。本ソフトウェアを使用することで、IT 投資を行なう企業は、適切な IT 投資の時期や内容を決定する際の判断材料として活用することができ、効果的な IT 投資による競争力の確保が可能になる。

2. 目的

「日本全国 IT 導入事例分析可視化システム」は、年間十数兆円を超える日本の IT 投資の実態を Web に掲載されている事例から分析することを目的としたものであり、例えば、年次別、企業規模別、業種別、地域別など様々な角度から IT 投資がどのように行われているか、また推移しているかを可視化エンジン GAVE により、2 次元上に可視化することができる IT 導入事例を対象とした分析・可視化システムである。

3. 開発の内容

本プロジェクトの重要なポイントは、各 IT 企業の Web ページに掲載されている IT 導入事例の情報を収集するという点と、それを分析し、可視化する技術の2点である。多くの企業の Web ページには、IT 導入事例に関する情報が掲載されているが、それぞれの Web ページのフォーマットが異なる。そこで、Web ページの HTML フォーマットを解析し、それぞれのサイトに対応した情報収集ロボットのプログラムソースを自動生成するためのソフトウェアを開発した。そして、それらの情報収集ロボットにより IT 導入事例の情報を収集した。また、それらのデータをインタラクティブに分析することができる可視化エンジン GAVE(Galaxy Analysis and Visualization Engine)を開発した。

3.1 IT 導入事例文書解析ソフトウェアの開発

IT 導入事例は IT 企業の Web ページに掲載されており、その IT 導入事例文書から特定の情報(URL、導入元企業名、導入先企業名、導入日[年月日]、概要[300 文字以内]、キーワード、業種名、地域名、企業規模[従業員数])を抽出する「IT 導入事例情報収集ロボット」を作成する必要がある。しかし、各企業の Web ページは様々な HTML フォーマットになっており、各サイトに対応するロボットを作成しなければならないが、その数が多いため、作成するのにかなりの時間を要する。そこで、その手間を軽減するために図 1 のような「IT 導入事例文書解析ソフトウェア」を開発した。本ソフトウェアは、Web ブラウザと結果を出力するフィールドおよび検索フィールドで構成されており、下記のような HTML タグを解析できる機能を有している。

1. HTML タグおよび属性に該当するテキストの検索機能
2. テキストに該当するタグ構造の検索機能
3. タグ構造に該当するテキストの検索機能
4. 指定した正規表現に該当するテキストの検索機能
5. ロボットプログラムソース作成機能
6. キーワード抽出機能



図 1: IT 導入事例文書解析ソフトウェア

3.2 IT 導入事例文書共通収集フィルタの開発

3.1 の IT 導入事例文書解析ソフトウェアを使用すれば、各 IT 企業サイトに対応する収集ロボットを作成できるが、さらにロボット作成の手間を軽減するために、ロボットに組み込む共通的なフィルタを開発した。3.1 では、IT 導入事例文書(リンク先)から情報を抽出しているが、共通フィルタではその文書がリンクされている IT 導入事例リンク集の Web ページを中心に抽出する。たいていの事例リンク集はリンク元とリンク先で構成されており、リン

ク元には必ず URL および導入先企業名が存在する。また、概要、導入日もある場合があり、ない場合はリンク先から抽出する。

3.3 IT 導入事例データベース

検索エンジンなどを使用し、手作業で IT 企業サイトの導入事例リンク集を探し、それらの URL を 3.1、3.2 の収集ロボットに入力し、IT 導入事例の情報 (URL、導入先、導入日、概要、キーワード、業種名、規模、地域名) を収集した。2005 年 10 月現在の収集先企業は 37 社、事例の総数は 2790 件であり、そのうち、マイクロソフト・NEC・日本 IBM・富士通・日本 HP・日本ユニシスといった大きい企業のサイトでは毎月更新が行われており、それに合わせて毎月定期的に事例を収集している。現在のところ毎月 20 件ほど増えている。IT 導入事例データベースに関して調査した結果、本データベースが最大であることを確認している。

3.4 可視化エンジン GAVE の開発

本システムの可視化エンジンには、GAVE (Galaxy Analysis and Visualization Engine) を使用している。GAVE では、点の一つ一つ (ノードと呼ぶ) が IT 導入事例を示し、似通った事例は近くに表示されるようになっている。また、GAVE は、ただ表示するだけのエンジンではなく、ユーザのインタラクションが大きな意味を持つ。ユーザは、検索を行い、特定のノードをカラーリングし、様々な軸で操作することによって、リアルタイムにノードの動きを把握することができる。可視化される図は、ユーザの操作により様々に変化し、ユーザの好きな切り口で分析を行なうことができる。

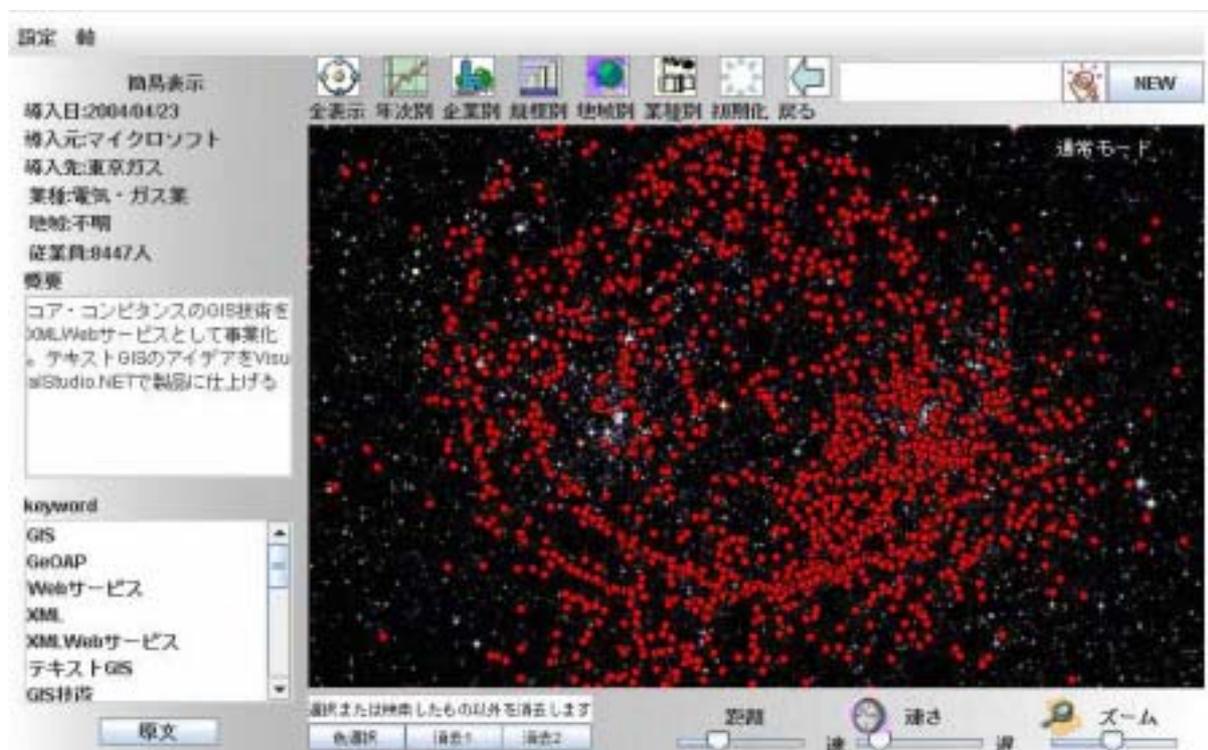


図 2: GAVE

4. 従来の技術(または機能)との相違

類似研究としては2004年9月1日にNRIサーバーパテント株式会社が「大量にある特許の特徴を瞬時につかめるグラフ機能」(<http://www.nri.co.jp/news/2004/040901.html>)を開発した。このシステムは、特許文献群を意味から分析し、その傾向を視覚的にグラフ表示するものである。これまでこのようなシステムは全く見受けられなかったが、近年、このようなデータを扱う可視化技術が注目されてきていることが伺える。我々のシステムでも、特許文献を扱うことが可能であるが、IT導入事例を対象にすることで違うアプローチができると考えている。また、このシステムは単に検索結果を2次元マップに表示するにとどまっており、まだまだ工夫の余地があると考えている。我々のシステムでは、単に可視化を行なうだけではなく、可視化を行なった後のユーザインタラクションに大きな意味があると考えている。様々な軸により、リアルタイムに様々な図に変化させ、その変化から多次的に情報を読み取る点が、類似研究との違いであり、優れている点であると考えている。

5. 期待される効果

本ソフトウェアを使用することにより、下記のことを知ることができる。

- ・競合事例の発見
 - ・IT業界の流行の発見
 - ・技術の動向を見る
 - ・他社の動向を見る
- など

よって、IT投資を行なう企業は、適切なIT投資の時期や内容を決定する際の判断材料として活用できると考えている。そして、IT業界の一つの指標になればよいと考えている。

6. 普及(または活用)の見通し

現在、下記の参考URLにてIT導入事例データベースおよび分析・可視化ソフトウェアを無料で公開している。そして、これからIT投資を行なう企業の方々に実際に使用してもらい、その要望に合わせてさらに必要な情報を追加していく予定である。

また、今後、対象とするデータをIT導入事例だけではなく、学会投稿論文、転職・就職事例、特許文献など様々なものに応用し、可視化エンジンを独立させ、様々なデータに適用できるようにする予定である。

IT業界の比較をするために、日本だけではなく、アメリカなど世界の事例を収集することも考えている。

7. 開発者名(所属)

- * 西村 啓, 池長 慶彦(早稲田大学大学院 理工学研究科 村岡研究室)
- * 早稲田情報技術研究所

(参考)

<http://jirei.muraoka.info.waseda.ac.jp/>