

# Google API を用いたグラフ構造の生成

## Generation of Graph Structures using Google API

村田剛志<sup>\*1 \*2</sup>  
Tsuyoshi Murata

<sup>\*1</sup> 国立情報学研究所  
National Institute of Informatics (NII)

<sup>\*2</sup> 科学技術振興機構  
Japan Science and Technology Agency (JST)

Visualization often plays an important role for human reasoning; suitable information visualization assists understanding of relations among objects. In the domain of the Web, several attempts have been made for visualizing pages in order to find related Web pages. Graph structure is often employed as a way for visualizing Web pages. This paper introduces previous approaches of Web visualization, followed by our own attempt based on Google API. Google API is an interface that allows user programs to access resources of Google search engine. Our visualization system utilizes Google as the data source for visualization, and the system has abilities of generating graphs of Web pages. The inputs to our visualization system are some target Web pages, and its outputs are graphs of Web pages related to the input pages with labels, which ease interpretations of relations among pages.

### 1. はじめに

発見や推論, 問題解決などの知的活動を行なうにあたり, 対象を視覚化することの重要性がしばしば指摘されている。視覚化は, データや処理結果を表示するインターフェースなど, 知的処理における補助的な役割としてみなされがちであるが, 対象における要素間の関係や構造など, 陽に示されていない情報を見出す上で重要である。一般に知的活動において適切な表現を用いることは, 対象からの情報獲得を促進するだけでなく, 推論や問題解決の方向性を制御し, 得られた結果を既存の知識と結びつけることを容易にする。人が問題解決においてしばしば図を用いるのもそのためである。

視覚化の手法として, 本稿においては対象における個々の要素を頂点, 要素間の関係を辺で表すグラフについて検討する。一口にグラフといっても, 重要な頂点の表現や, 頂点間の関係の表現など, 視覚化には多様な目的が考えられる。本稿では, Web ページのグラフ構造を例に, 既存の視覚化システムについて概観するとともに, 著者が構築した視覚化システムについても述べる。このシステムにおいては, Google API を用いてデータ獲得をすることによって, 対象である Web ページに関する情報やページ間の関連性を明確にしたグラフ構造を生成している。

### 2. Web における視覚化システム

#### 2.1 知的活動における視覚化

一般に視覚化を行なうことは, 単に見やすい表現を得るためだけのものではなく, 知的活動において必要な情報を抽出したり推論を制御したりする上で重要な役割を果たしている。そのような例として, 関連のあるデータは図形の中でグループ化されて比較的近くに存在するという性質 (locality [Larkin 87]) や, 与えられた状況についての図形を描く過程で, 大きさや空間的位置に関する情報などが自然と明らかになるという性質 (emergent property [Koedinger 92]) が挙げられる。また, 図形を用いた推論の過程で次の推論の対象を選ぶ方法として, 図形上で近くに

あるものに注目するというヒューリスティック[Narayanan 94]がある。Larkin は, 情報の探索, 認識, 推論の観点から sentential representation と diagrammatic representation とを比較し, 主に情報の探索において大きな効率の差があることを指摘している。視覚化を行なうことによって, 知的活動において有用なこれらの性質を利用することができる。

情報視覚化の研究[Stasko 98][Spence 01]は非常に数多くあるが, 物理的な実体のあるものを見やすく表示するコンピュータグラフィックスとは異なり, 抽象的な概念や関係を明確にするための視覚化においては, グラフ等の幾何学的な表現が用いられることが多い。実体を持たない対象における抽象的な関係を見出すために, キーワードや人名等を 2 次元平面上に配置する視覚化システムの研究が数多くなされている。そのような例として, KJ 法[川喜田 67]を始めとする発想支援や, SOM [Kohonen 95], 論文の共著関係などから人脈ネットワークを明らかにする Referral Web[Kauts 97], 文中の単語の共起関係から重要語のネットワークを生成する KeyGraph [大澤 99]などがある。これらの視覚化においては, 平面上の配置や, 要素間の距離や, 要素を結ぶ辺の太さによって対象間の関連性を表現している。

#### 2.2 Web ページの視覚化

Web ページを対象とした視覚化システムは, 1) 同一サイト内のページを視覚化するものと 2) 複数サイトのページを視覚化するものに大別できる。前者としては, Hyperbolic Tree によるディレクトリ構造の視覚化[Inxight 99]や, Mapuccino[Maarek 97]などがある。後者としては, vivisimo.com[Vivisimo] や Kartoo.com[Kartoo], 著者の視覚化システム[Murata 03]などがある。また, サイバー空間の視覚化を扱ったものとして[Dodge 01]がある。

#### 2.3 Google API によるデータ獲得

Web 情報は膨大であり, Web マイニング実験に必要なデータを獲得し, 維持することは容易ではない。サーチエンジンは膨大な Web ページを収集しており, そのデータを検索のためだけでなく知的システム実現のために活用する方法として Google API [Google 02]がある。これは, 米 Google 社が 2002 年に公開した SOAP および WSDL に基づいたインターフェースで,

連絡先: 村田剛志, 国立情報学研究所 知能システム研究系,  
〒101-8430 東京都千代田区一ツ橋 2-1-2,  
tmurata@nii.ac.jp

Google が収集した Web ページのデータを自分の好きなようにプログラミングして利用することができるものであり、多くの利用者がその活用を試みている[Calishain 03]. Google API を用いた視覚化システムとしては TouchGraph [TouchGraph] がある。また、キーワード連想を行なって視覚化する環境ソフトとしては Memorium [渡邊 02] がある。

### 3. Google API を用いた視覚化システム

本節では著者が構築した、Google API を用いてデータ獲得を行なう視覚化システムについて述べる。Google API は上述の通り、プログラムを用いて Google のリソースにアクセスすることを(1日1000回という制限つきで)許しているインターフェースである。Google は2004年4月現在42億ページ以上を収録する世界最大の検索エンジンである。これを情報源として有効に利用することは、クローラ等による Web データの収集よりも容易であるとともに、実際に知的システムを構築する上で重要である。

著者が構築した視覚化システムにおいては、Web ページと、ページを説明する情報(ページ情報)を頂点、ページとページ情報との関係を辺とするグラフを表示する。Google API を用いた検索の際に、以下のようなオプションを用いることで、入力した Web ページに関する情報を獲得している。

1. related: オプションによる関連ページの検索
2. info: オプションによるページ情報の検索

前者は、選択した Web ページの関連ページを検索し、新たな頂点としてグラフ上に表示させるものである。また後者は、選択した Web ページが Google に登録してある場合に、そのページに関する情報を新たな頂点としてグラフ上に表示させるものである。視覚化の実行例を以下に示す。



図1 関連ページの表示

図1は、www.maruzen.co.jp(中央付近)の頂点を入力した後に、その関連ページを表示させた結果を示している。これらの関連ページは Google 上で「related:www.maruzen.co.jp」を検索した場合の出力結果と同じであり、www.kinokuniya.co.jp や www.yaesu-book.co.jp, webcat.nii.ac.jp などの関連ページが頂点として現れている。

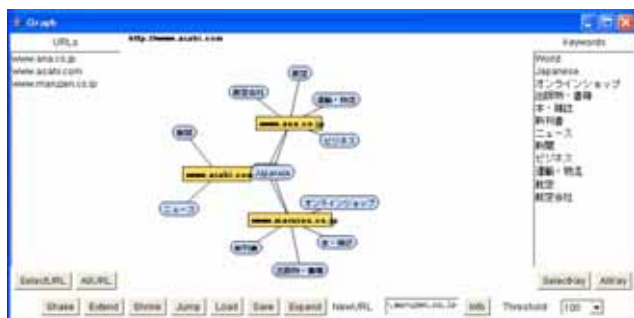


図2 Web ページとページ情報のグラフ表示

図2は、www.ana.co.jp, www.asahi.com, www.maruzen.co.jp の3つを入力した後に、それぞれのページ情報を表示させた結果を示している。www.ana.co.jp の周囲には、「航空会社」「運輸・物流」などのページ情報を表す頂点が現れている。これらのページ情報は Google 上で「info:www.ana.co.jp」を検索した場合に得られる情報を利用している。同様に、www.asahi.com のページ情報としては「ニュース」「新聞」などが、また www.maruzen.co.jp のページ情報としては「出版物・書籍」「本・雑誌」などが表示され、3つのページが「Japanese」「World」のページ情報の頂点でつながっている。

Web ページをグラフとして視覚化し、ページ間の関連性を見出すという点では、上述のシステムは TouchGraph などと同じである。しかし、ページ情報を頂点としてグラフ上に陽に表示し、その頂点を介してページ間の関連性を示している点が異なっている。筆者が過去に構築した Web コミュニティの視覚化システム[Murata 03]においては、ページ間にどのような関連性があるかについての解釈はユーザに委ねられていた。ページ情報を用いることはページ間の関連性の内容を明確にするだけでなく、全く異質なページ集合間の意外なつながりを見出す上で有効であると考えられる。

### 4. おわりに

知的活動における視覚化は、対象における構造や関係を見出す上で重要である。本稿では、グラフによる視覚化の試みとして、Google API を用いてデータ獲得を行なう視覚化システムについて述べた。

今後の方向性としては、上述のような異なる Web ページ集合間の関係の発見があげられる。またそれとは別の方向性として、対象における構造を考慮した視覚化手法の検討が考えられる。例えば、Web ページを頂点、ハイパーリンクを辺とするグラフ構造においては、スケールフリーなどの性質があることが知られている[Barabasi 02]。このような性質を反映させるような視覚化を実現することは、対象に内在する構造を把握して知的活動に役立てる上で重要である。

### 謝辞

本稿の視覚化システムを Java で作成する上で、京都大学の角康之先生のプロプログラムを参考にさせていただきました。ここに感謝の意を表したいと思います。

### 参考文献

- [Barabasi 02] A.-L. Barabasi: “Linked – The New Science of Networks”, Perseus Publishing, 2002.
- [Calishain 03] T. Calishain, R. Dornfest 著, 山名, 田中訳: “Google Hacks”, オライリー・ジャパン, 2003.
- [Dodge 01] M. Dodge, R. Kitchin: Mapping Cyberspace, Routledge, 2001.
- [Google 02] Google: “Develop Your Own Applications Using Google”, <http://www.google.com/apis/index.html>, 2002.
- [Inxight 99] Inxight: “Hyperbolic Tree”, <http://www.inxight.com>, 1999.
- [Kartoo] Kartoo: <http://www.kartoo.com/>.
- [Kauts 97] H. Kauts, B. Selman, and Mehul Shah: “The Hidden Web”, AI Magazine, Vol.18, No.2, pp.27-36, 1997.
- [Koedinger 92] K. R. Koedinger: “Emergent Properties and Structural Constraints: Advantages of Diagrammatic Representations for Reasoning and Learning”, Technical

- Report of AAAI Spring Symposium, Reasoning with Diagrammatic Representations, SS-92-02, pp.151-156, 1992.
- [Kohonen 95] T. Kohonen: "Self-Organizing Maps", Springer, 1995.
- [Larkin 87] J. H. Larkin and H. A. Simon: "Why a Diagram is (Sometimes) Worth Ten Thousand Words", Cognitive Science, Vol.11, No.1, pp.65-99, 1987.
- [Maarek 97] Y. Maarek, M. Jacovi, M. Shtalhim, S. Ur, D. Zernik, I. Z. Ben Shaul: "WebCutter: A system for dynamic and tailorable site mapping", Proc. of the 6th WWW Conference, pp.713-722, 1997.
- [Murata 03] T. Murata: "Visualizing the Structure of Web Communities Based on Data Acquired from a Search Engine", IEEE Transactions on Industrial Electronics, Vol. 50, No. 5, pp.860-866, 2003.
- [Narayanan 94] N. H. Narayanan, M. Suwa, and H. Motoda: "How Things Appear to Work: Predicting Behaviors from Device Diagrams", Proceedings of Twelfth National Conference on Artificial Intelligence, pp.1161-1167, 1994.
- [Spence 01] R. Spence: "Information Visualization", Addison-Wesley, 2001.
- [Stasko 98] J. Stasko, J. Domingue, M. Brown, B. Price: "Software Visualization", The MIT Press, 1998.
- [TouchGraph] TouchGraph: <http://www.touchgraph.com/TGGoogleBrowser.html>.
- [Vivisimo] Vivisimo: <http://vivisimo.com/>.
- [大澤 99] 大澤幸生, N. Benson, 谷内田正彦: "KeyGraph: Automatic Indexing by Segmenting and Unifying Co-occurrence Graphs", 電子情報通信学会論文誌 D-I, Vol.J82-D-I, No.2, pp.391-400, 1999.
- [川喜田 67] 川喜田二郎: "発想法", 中公新書, 1967.
- [渡邊 02] 渡邊恵太: "Memorium: 眺めるインタフェースの提案とその試作", <http://www.persistent.org/memorium.html>, 2002.