

特許情報検索と解析の将来展望

—マクロ分析とセミミクロ解析とコア解析の連携分析(Cytoscape 活用)—

Future perspective for Patent Information Retrieval and Analysis

-Combination analysis among macro-analysis, semi-micro-analysis and core-analysis by use of the landscape tool such as "Cytoscape"-

一般社団法人情報科学技術協会 パテントドキュメンテーション部会主査 **桐山 勉**

PROFILE: 2004年からINFOSTA-SIG-パテントドキュメンテーション部会のコアパースン。2006年2月に帝人知的財産センターを定年退職し、2006年4月からJapioの特許情報研究所の客員研究員を務める。2011年よりIPI-Award Selection Board(国際特許情報賞選考委員)も務める。特許情報解析の専門家でもある。2013年4月より、はやぶさ国際特許事務所の顧問も務める。

1 はじめに

2013年6月17日の報道ステーションにてiPS細胞研究でノーベル賞受賞の山中伸弥教授が米国サンフランシスコの研究所からlive中継放送に出演された。アナウンサーの古館氏が「なぜ頻りに米国に出張されるのですか」と質問すると、「出張を止めると、最新の重要な情報が入ってこなくなり、自分の新しいアイデアも生まれなくなるから」という返事をされた。この言葉に強いインパクトと刺激を受けた。日本の特許情報検索担当者(所謂、サーチャー)も同じように世界の同業の専門家とface to faceで意見交換をすることが必要ではないだろうか。

日本の特許マップ解析の第一人者だけでなく世界の特許情報解析の第一人者とも笑顔で面談し、意見を交換することが必要ではないだろうか。

2 世界を知り、世界と共に生きる

世界に一歩足を踏み出すと、日本国内のシンポジウムと特許情報フェアとは全く違った雰囲気と顔ぶれがある。日本ではあまり聞かれない専門用語が頻りに目と耳に入ってくる。

その一つがHype Cycle by Gartner¹⁾である。PIAC2011 in CHINAにて教わった。もう一つの事例がNetwork Patent Analysis(NPA分析)である。GH(GRIFFITH HACK)特許事務所のMike Lloyd氏

の米国のPIUG²⁾2011およびPIUG2012の発表³⁾にて教わった。人間の記憶力をはるかに超えた領域でのBig Dataの分析である。世界には、凄い特許情報の専門家がいるものだ。

海外の特許マップの論文・発表プレゼンの中で筆者が特に注目しているのはLucy Akers氏と製薬会社Bristol-Myers Squibbの特許解析チームによる発表記事⁴⁾である。理由は、米国と欧州の全体を見回して広く特許マップToolを研究しているからである。そして、彼らは数年前から医療分野にて遺伝子DNA解析に利用されている可視化ToolのCytoscapeの存在を知り研究しているからである。

さらに導かれるが如く遭遇したのが、Cytoscapeをうまく活用して起業化しているLinguamatics社と、その販売責任者のDr. Susan M LeBeauだった。1年前にはLinguamatics社のI2E特許検索システム⁵⁾に筆者は全く気付かなかった。今年のPIUG2013においてBristol Myers Squibb社の特許解析の発表⁶⁾を聴いてその価値が漸く理解できた。

筆者をNPA分析に駆り立てた五つ目の理由は、帝人(株)を定年退職してから7年も経過して、大企業の特許情報グループのように高い検索システムを活用できるお金がないことである。そのために、無料のNPA分析Toolをどうしても使いこなす必要があり、Cytoscapeに真剣に取り組まざるを得なくなった。

ここに到達するまでには、国内の安藤俊幸氏の論文⁷⁾と特許情報のテキストマイニングという著書⁸⁾およびそれらの引用文献から基礎知識を学んだ。

3 無料の Cytoscape を活用する

幸運なことに、「Cytoscape による特許情報のネットワーク解析とビジュアル化」という論文⁹⁾を発表した今津均氏が国内にいた。実際に、Cytoscape のオリジナル URL¹⁰⁾ から無料の（最新版は Ver3.00 であるが）使い易い Ver2.82 をダウンロードした。

また、Ver2.82 に沿った詳細な日本語マニュアルが、「バイオインフォマティクス砂漠」¹¹⁾ というサイトに公開されている。日本人であり米国西海岸に在住の小野圭一郎氏がこの Cytoscape の応用開発に深くかかっている。

筆者が Cytoscape に注目している理由が他にもある。人間は多次元の図を実は頭の中で描いているが二次元の図にすると相関関係が対等で描けられない。だから、関係するものを距離でなく線で結びつけるのが良いと思うようになったからである。例えば、4つの頂点の関係が対等である正三角錐の頂点は平面図にすると等距離には表示できない。多面的な相関関係を平面図の距離だけで表しては誤解を招きやすいと考える。

Cytoscape の良い点は、全体の可視化だけではなく、その結果のデジタル出力にある。可視化だけなら、「R」と「KHcoder」などが既に知られているが、図として出力するだけである。Cytoscape は図でなくメタデータとして出力するのである。

今津氏の論文を追試して特開平 04-164895 の関係の 12 件の特許の NPA 図を描くところまでは、何とか追試ができた。Cytoscape のサイトと Linguamatics 社のサイトからサンプルファイル (cys 形式) を入手した。取りあえず「先ず慣れること」から Cytoscape に入門した。何も理論的なことは判らないので、人の 3~5 倍の時間が掛った。

少しだけ判ったことは、Cytoscape は相関関係を可視化する Tool にすぎないため、Landscape 図の上で描かれる点（ノード）と点（ノード）とその相関関係の 3 つが必須要素であることが判った。この 3 点のデータがないと何も描けないことが判明した。Cytoscape で扱えるデータ・ファイルを作成するまでが難しく、入門者にとっては敷居がかなり高いことを知った。Cytoscape に取込むまでのデータの流れとその後の流れを図 1 に示す。

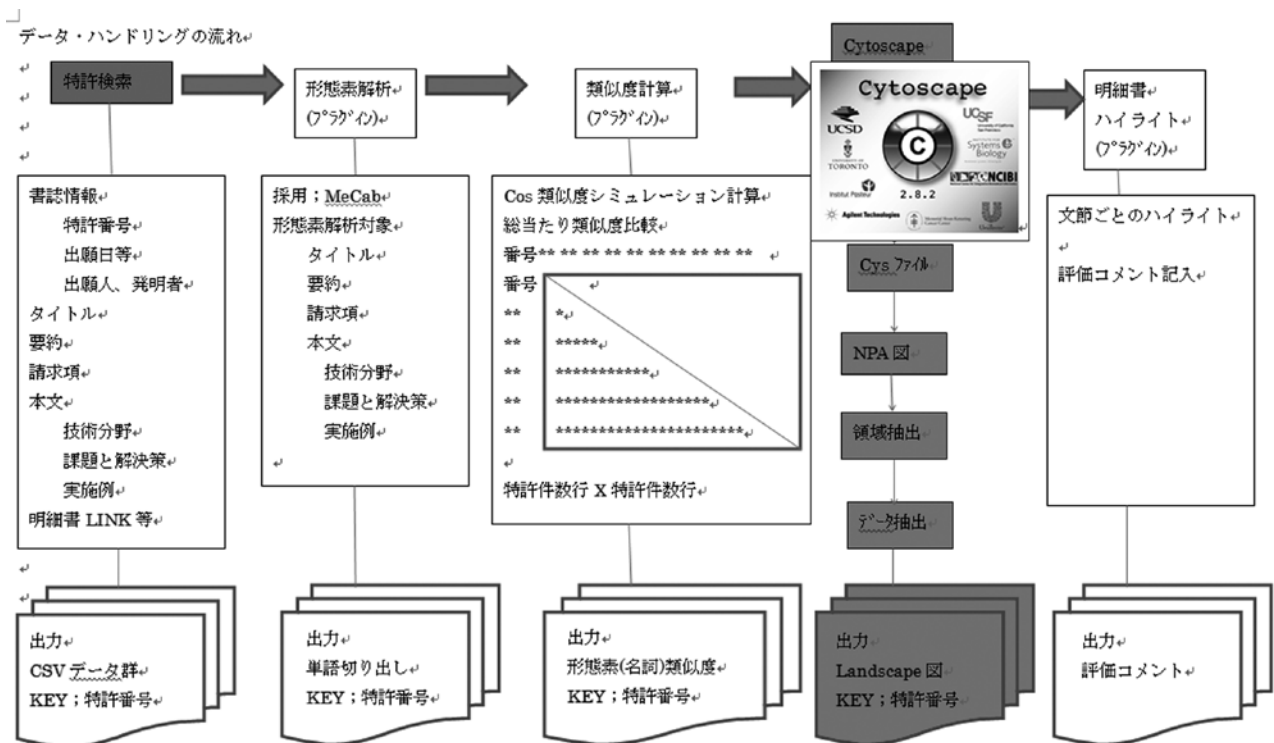


図 1 Cytoscape に持っていくまでのデータの流れと、Cytoscape の機能

さらに判ったことは、この Cytoscape は拡張性があるために、プラグイン・ソフトを作りさえすれば、色々と組み込むことができる。その特長に注目して、特許情報検索システムを組み込んだのが Linguamatics 社であることが、筆者にも漸く見えてきた。

筆者の PC には知人に特別に作って頂いたプラグイン・ソフトが組み込まれており、市販のものではない。市販の特許検索システムから書誌情報とタイトルと要約とクレーム等を CSV 形式でダウンロードしている。そして、タイトル + 要約、またはタイトル + 要約 + 請求項のどちらかの組合せを選んで、形態素解析を行ってから個々の特許情報の類似度を予め計算している。この類似度計算プラグイン・ソフトは特別に自前仕様で作られたものである。

個々の特許情報の間の類似度比較は cos 類似度というもので計算している。そして、個々の特許情報のノードとノード、および、その相関関係を予め計算している。それを Cytoscape のインプット・データとして取り込んでいる。その結果を Landscape 図として

Cytoscape で描かせているだけである。

以上のデータの流れから、Cytoscape によって取り込むまでに、ノードとノード、およびそれらの相関関係の中間ファイルを準備作成するまでがプロセス上のキーポイントである。

Cytoscape の特長は、Landscape 図の上でマウスにより四角の枠で領域を囲むことによりデータを抽出する。その抽出されたデータを出力させる機能を標準で持っている。また、オプションとしてデータ内の検索機能、所謂、Enhanced search のプラグインが用意されている。全体像の中の部分を Cluster として部分分離して表示する AllegroMCODE というプラグインも用意されている。

筆者の PC には、抽出された出力の個々の特許情報の明細書部分をデジタルで表示させ、自分が注目している用語または記述部分を一括ハイライト表示させる機能をプラグイン・ソフトとしてさらに追加している。図 2 に、参考事例として筆者の PC 画面にて「電動車椅子およびロボット歩行機」に関する Landscape 図とそこから一

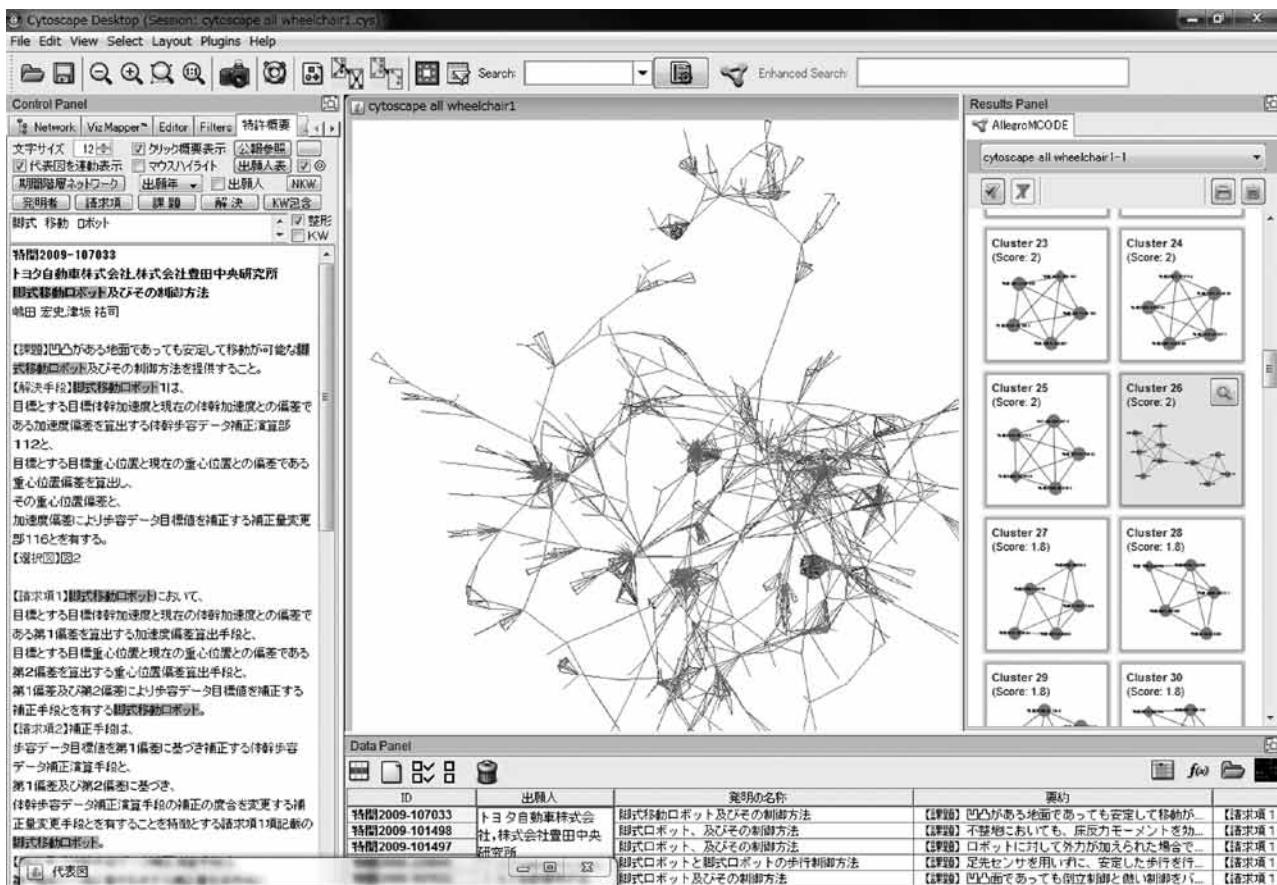


図 2 無料の Cytoscape と追加のプラグインにて稼働する筆者の PC 画面 (参考例)

部だけ抽出してリスト表示させたものを示す。さらに、そのうちの1件の特許の明細書の本文を色々な文脈で分離させた文章群において自分の求める専門用語または文章表現にてハイライト表示をさせたものを示す。

筆者のPCでは、(Cytoscape という Tool は拡張性が高く、色々なプラグイン・ソフトを組み込むこと) 図2にて示唆している。日本の優秀なプロバイダー大手が、Linguamatrixs 社のように取り組めば、日本特許情報を Cytoscape で可視化できることを示唆している。

また、アジア特許情報研究会の山村健一氏 (TTDC) が中国人発明者の解析を INFOPRO2013 にて発表すると聞いている。すなわち、Cytoscape は言語として英語だけでなく、日本語でも中国語 (簡体字) でも扱えることを示唆している。

4 3つの視カレベルでの連携活用必要性

先達の知恵として、自然界には3つの視力が存在することが判っている。一つは、鳥の目で上空から鳥瞰図のように眺望し、かつ、上空から地上の小さな虫も見逃さない目である。二つ目は、魚の目であり、海の潮流を正確に捉え、かつ、餌を見つける目である。三つ目は、一つの花の雌蕊と雄蕊の周りから甘い蜜を吸い取る目である。図3で示したごとく、鳥の目は技術の全体像である Landscape 図を観るためのマクロ分析の目、魚の目は技術の潮流と傾向を把握するセミマイクロ解析の目、そして、虫の目は、個々の明細書を読み込むコア解析の目と筆者らは呼んでいる。

現在、日本の特許情報は全文がデジタル形式で、毎週提供されている。一方、個人が自由に使える PC も極めて高性能になっている。そのために、3~6万件の特許

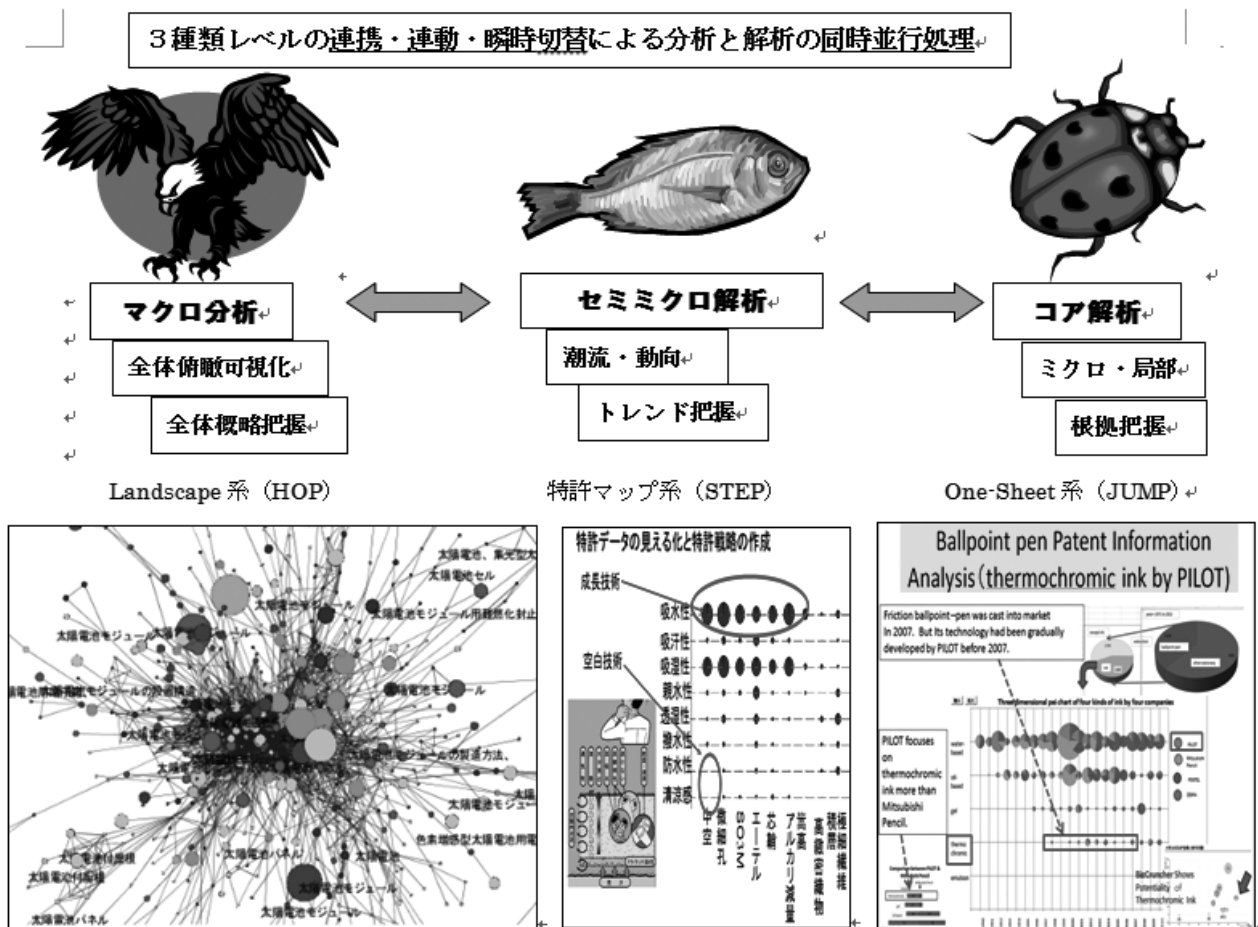


図3 3種類レベルでの連携活用の必要性

情報を個人が自分の PC でハンドリングすることができる時代である。ここで述べる Cytoscape も、1 万件以上の日本特許情報をデジタルハンドリングすることが容易である。

筆者は今までに全体俯瞰可視化をするために創知社の X LUS と、富士通社の ATMS/Analyzer を触った経験がある。今では、定年退職をしたので無料の Cytoscape を活用せざるを得ない。読者が自分の分析テーマに適したシステムを活用されることを薦めたい。コンピュータを使った類似度のマクロ分析は、全ての研究テーマに対してオールマイティーに分析できるものではないが、研究テーマによってはかなり有効な場合もある。R&D の時間効率化の観点から、有効なテーマに対しては PC により特許情報シミュレーションの利用を推奨する。まさに Hype Cycle 曲線の極大値の後の谷からどの程度まで浮かび上がるかは本人の努力次第である。

筆者は、セミマイクロ解析の段階では市販特許マップ・ソフトを頻繁に活用している。筆者は、2006 年以降から現在でもレイテック社主催の PAT-LIST 研究会を支援しているために、日常的に PAT-LIST-JPS を日本特許の解析・可視化に活用している。中国語の特許情報解析には PAT-LIST-GLS という特許マップ・ソフトを使い、三カ国語（中国語 - 英語 - 日本語 + 代表図）による解析を行っている。

コア解析としては、1 件 1 件の明細書を読み、自社分類（または自分のカテゴリー分類と技術分類と機能分類）を付与してから、これらの自社分類と自分のコメント評価などを組み合わせて、特許マップ図を作成している。

これらの3種類のレベルの連携活用は極めて重要である。筆者は、特に、「ユビキタス御前デモ」（筆者の新造語）と称して、いつでもどこでもクライアント依頼を受けて、クライアントの PC にてパワーポイントで説明しながら特許マップ・ソフト PAT-LIST を同時に動かしている。そして、特許マップ図の上から具体的なディスカッションの個々の明細書を 30 秒から 3 分でリンク表示をさせ、問題の個所をハイライトで表示させている。特許マップを社内に普及させるためには、この「ユビキタス御前デモ」を実践することが必要となる。

筆者は、特許情報解析のために二画面ディスプレイを

使用することを推奨している。拙宅でも 19 インチの二画面ディスプレイ環境で仕事をしている。左画面には Cytoscape が稼働し、右画面には PAT-LIST が稼働している。さらに、特許検索システムもパワーポイントも Windows で同時並行で稼働している。筆者の業務から二画面 PC 環境を取り上げたら単なる素人老人の仕事になり下がるが、皆様のご支援とそれら Tool のお蔭で IP-Information Scientist の端くれを演じることができている。

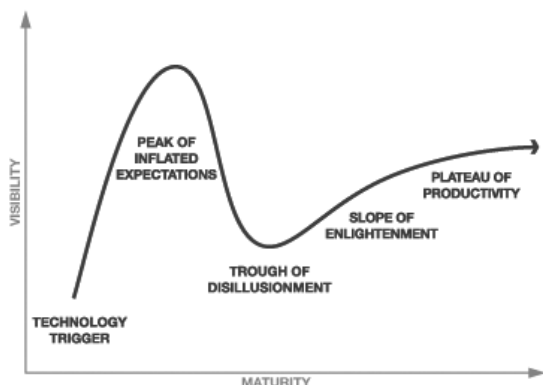
5 さいごに

冒頭の山中伸弥教授の頻繁な米国出張の話ではないが、私たち特許情報専門家のサーチャーも世界に出て、世界を知り世界と共に生きることが必要と考える。日本にも米国と欧州のように All JAPAN PIUG ができないだろうか。中国には PIUG チャプター組織¹²⁾があり、2013 年 8 月 27 日時点で登録者が 35,193 名を超えている。しかるに、日本人の米国 PIUG 登録者は約 20 名である。筆者は、国内の色々な特許情報専門家が透明な雨傘のような緩やかな連携を構築できるように熱望している。毎年、米国 PIUG 年次大会と欧州の IPI-ConfEx と中国の PIAC に参加しながら、狭い日本国内にて特許情報専門家の分野別勉強会のバラバラな状態を懸念している。世界の良い点の取り込みこそが日本人サーチャーの進歩と発展の鍵と信じる。

参考文献

- 1) Hype Cycle by Gartner

<http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>



- 2) PIUG;Patent Information Users Group,
<http://www.piug.org/>
- 3) Let us guide you through the global IP landscape, Use of Network Patent Analysis(NPA) for advanced analysis of patent data; Mike Lloyd,
GRIFFITH HACK, PIUG2012年次大会, (Denver, USA) (30April 2012)
- 4) Enhancing patent landscape analysis with visualization output, Lucy Akers et al, World Patent Information, Vol.32 ,p203-220,(2010)
- 5) Linguamatics 社の I2E 検索システム;
<http://www.linguamatics.com/>
- 6) Leveraging Text Analytics Tool in Patent Analysis to Empower Business Decisions ; Yun Yun Yang, Bristol-Myers Squibb, PIUG2013年次大会, (Alexandria, USA) (30April 2013)
- 7) テキストマイニングと統計解析言語 R による特許情報の可視化; 安藤俊幸, 情報管理, Vol.52, No.1, p20-31 (2009)
- 8) 特許情報のテキストマイニング; 豊田裕貴、菰田文男編著、ミネルヴァ書房発行 (2011年3月)
- 9) Cytoscape による特許情報のネットワーク解析とビジュアル化; 今津均, 情報管理, 54(8), 463-475(2011)

<http://dxdoi.or/10.1241/johokanri.54.463>

- 10) Cytoscape 公式 URL;

<http://www.cytoscape.org/>

- 11) バイオインフォマティクス砂漠;

<http://wiki.livedoor.jp/bioinformatics/>

- 12) 中国 PIUG チャプター組織 (米国 PIUG の下部組織); <http://www.piug.org.cn/bbs/>