

動的な視覚化手法を用いた特許情報検索システム

淡谷 浩平 奈須 庄健 井前 吾郎 重野 寛 松下 温
慶應義塾大学理工学部情報工学科

特許調査では検索漏れに厳しいことや2次情報(特許権者、特許分類等)が重要視されるという要求条件がある。本論文ではこれらの要求条件を満たすため、磁石が砂鉄を引き寄せるメタファを検索過程に導入した視覚化モデルである「磁石モデル」を提案する。本モデルでは検索キーワードを磁石、個々の特許文書を砂鉄とみなし、同一平面上にくっつかの二次情報を配置する。そして特許文書が各2次情報から出現し、検索キーワードとの関連度に応じてそのキーワードの近くまで引き寄せられるアニメーションを表示する。このアニメーションにより、キーワードと特許文書の関連度だけでなく、同時に各文書の2次情報という確かな情報を提供することで、キーワードとの関連度が低く算出された文書にも注目の機会を与える。

Retrieval System for Patent Information using Dynamic Visualization

Kohei Awaya Shoken Nasu Goro Inomae Hiroshi Shigeno Yutaka Matsushita
Department of Information and Computer Science, Faculty of Science and Technology, Keio University

In patent investigation, it is severe to the omission in retrieval and the second information, such as Inventors, Assignee and Current U.S Class is treated importantly. In this paper, we define a concept of "Retrieval Awareness" as awareness of the hidden information while retrieving and propose "Magnet Model" which introduce the relation between a magnet and iron sand in order to these requirement. This model provides "Retrieval Awareness", such as the relationship between the keyword and the patent document, each document's second information and the relationship between the keyword and second information by using dynamic visualization.

1. はじめに

近年、先進諸国を始めとして多くの国家がプロパテント政策を推進してきており、各企業においても知的所有権の権利取得や保護を強化する動きが高まっている。また、特に昨今ビジネスモデル特許や遺伝子配列特許等の新しいタイプの特許の出現もあり、特許出願量に加え内容の多様性も増してきている。このため研究者や技術者をはじめこれまでよりも多くの人々が特許情報を把握する必要性が増している。

特許は権利関係に大きな影響を及ぼすことが多いため調査においては検索漏れに厳しいという要求条件がある。また特許文書は全文テキストである1次情報に加え、特許権者、発明者、出願日、公開日、国際特許分類、米国特許分類など、他の種類の文書よりも多くの2次情報を含んでおり、これらが調査において重要視される。そのため、Web文書や技術文献等を対象にした文書検索や文書調査のシステムでは対応しきれない部分がある。

そこで本研究ではこれらの要求条件を満たすため、文書と2次情報との関係を明確にし、効率的に調査を行うためのアプローチとして、磁石が砂鉄を引き寄せるメタファを検索過程に導入した視覚化モデルを提案し、そのプロトタイプを実装した。

2. 特許調査

現在最も一般的に用いられている特許検索ツールは米国特許庁サイト(United States Patent and

Trademark Office)の検索システム[1]やQPAT[2]であり、これらはキーワード検索を基本としている。仮にQPATで特許権者がKeio University、発明者がAwayaの特許を調査する場合以下のように入力する:

'Awaya' / inv and 'Keio University' / pan
すると検索結果がリスト表示で返され、そのリストにラベルがふられる。もし1度目の入力だった場合そのラベルはS1である。この検索結果に対してさらに検索式を定義する場合、例えば

("visualization") AND S1
と入力すればS1の検索結果の中でvisualizationを含む文書を検索することができる。

しかし、このようなシステムでは特許空間の把握が困難であり、また2次情報はあらかじめ設定することしかできない。さらに、検索式をうまく設定しないと結果が膨大になりユーザによる把握が困難になったり、絞りすぎにより検索漏れが発生してしまったりする。

自然言語処理などの技術が向上し、文書の関連度などはある程度高い精度で算出できるようになってきている。しかし、それでもやはりコンピュータによる処理だけで完全に文書の意味、内容を判断することは難しい。そんな中、情報視覚化の技術を用いて文書検索を支援するシステムが数多く研究されてきている。

3. 関連研究

2次元の視覚化手法を用いた文書検索システムに DocSpace[3], VIVE[4]等がある。

DocSpaceは文献データベースのインタラクティブな視覚化であり、多数の文献とキーワードをその関連性に基づいて配置する視覚的分類技術に、ユーザの操作による視覚化結果の動的な更新を導入したシステムである。文書とキーワードの距離をばねモデルで表し、キーワードを動かすとそれに伴い文書も移動する。

VIVEでは、いくつかの検索キーワードを2次元空間上に配置するとそれらを頂点とした多角形が表示され、それらに囲まれた領域に検索された文書がノードとしてプロットされる。配置は自然言語処理などを用いて求められた文書とキーワードの関連度から決定され、キーワードを動かすとそれに伴って文書が移動することで、インタラクティブな検索が可能となる。

しかしこれらの手法には以下のような問題点があると我々は考える。

- 文書の配置が、複数あるキーワードの内のどのキーワードに依存しているのか、つまり文書がどのキーワードと関連しているのかが動かせなければわからない
- キーワードと文書の関連度に基づいて配置が決定されているため、他の情報（著者など）を配置することが難しい。特に特許文書については2次情報（特許権者、発明者、米国特許分類、国際特許分類等）が多く存在するためこの点は問題となる
- 文書の詳細情報（アブストラクトや要約）を表示できるスペースが少ない

また3次元の視覚化手法を用いた文書検索システムに、VR-VIVE[5], NIRVE[6]等がある。

VR-VIVEはVIVEを3次元のバーチャルリアリティ空間に拡張したシステムである。ユーザの指定した4つのキーワードに囲まれた四面体の領域内に情報がマッピングされる。ユーザは頂点を移動したり、3次元空間内で視点を変えたりしながらインタラクティブに検索を進めることができる。

NIRVEの3次元システムは検索の行程をあらわすものである。球状の物体の表面にいくつかの箱が設置され、一つの箱はいくつかのキーワードにより検索された結果の文書群が入っている。緯度はその検索キーワードの数を表し、緯度が高いほど多くのキーワードにより検索されていることになる。つまり、緯度の低い箱が仮に2つのキーワードで検索された箱だとすると、その上の箱は次にどんなキーワードを入力するとどのような結果を示すのかを表すことになる。

このような3次元空間を用いた視覚化手法には、

- 3次元空間での迷子問題

- 回転操作などを加えないと見えない部分が出てくる
- 表示が複雑過ぎて、特に情報分野に関する非熟練者にとって理解しにくいなどの問題点がある。

4. 提案

前節までの考察を元に我々は特許情報検索システムに必要な特徴として次のようなものを考える。

- 情報検索技術に対する非熟練者でも容易に理解できる表示
- 検索キーワードと文書がどの程度関連しているのかの明確な表示
- 2次情報の利用
- タイトル・要約などの文章の表示

そこで我々は、2次元空間において動的な視覚化手法を用いることで扱える情報を増加する手法と、仮想3次元空間を有効活用して文書を表示するスペースを確保する手法を提案する。

4.1 磁石メタファ

文書と検索キーワードの関連度を明確に示すため、特許文書と検索キーワードを、図1のように磁石と砂鉄のような関係として視覚化する。仮に“security”という検索キーワードを配置すると“security”と関係のあるノード（特許文書）が“security”に引き寄せられるアニメーションを表示する。この際、検索キーワードとの関連度が深いノードほど内側にまで引き寄せられる。このような表示法を用いることで文書と検索キーワードとの関連度を明確に示すことが可能となる。

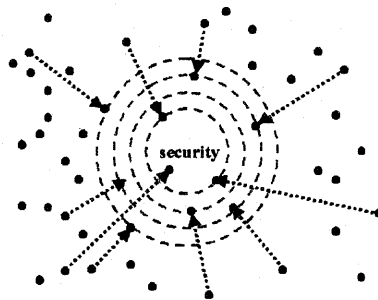


図1 磁石と砂鉄のような関係

4.2 Retrieval Awarenessの提供

特許文書の特徴として、2次情報が多く含まれていることを述べた。一般的な特許調査においては、ある注目している企業の保持している特許やある分類に属する特許を調べるために、始めに特許権者や米国特許分類等を指定して、その中でキーワード検索を行っている。

しかし、これではキーワードで検索された特許群にどのような企業のものがあるのか、またどの

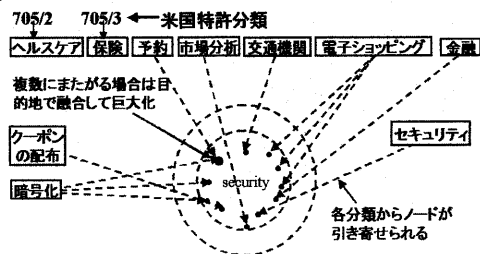
ような分類のものがあるのかという情報をユーザに伝えることができない。入力したキーワードに関連する特許が意外な企業の特許であったり、意外な分類に所属している可能性も十分に考えられるため、このような気づきの提供は有用である。

しかし、これまで出現元を表示してどのような分類からの出現が多いのか、またどんな分類からの出現があるのかという情報を付加的に提供するシステムはほとんどなかった。

そこで我々は、2次情報を平面上に配置し、入力されたキーワードに関連のある特許文書をその2次情報から出現させることで、2次情報を提供する手法を提案する。この検索過程において得られる情報を我々はRetrieval Awarenessと定義する。図2は2次情報として米国特許分類を配置した例であり、各特許ノードは各分類から出現し磁石に引き寄せられていく。米国特許分類や国際特許分類は一つの特許に複数の分類が割り当てられることが多くあるため、複数の分類にまたがるものはそれぞれの分類から出現し、目的地で融合・拡大する。また、移動時に気になった動きをした特許を移動後に確認するため、特許ノードをクリックすると出現元の分類に対して線がひかれ、分類をクリックするとその分類から出現した特許ノードに対し線がひかれる機能を備える。

なお、表示には2つのモードがあり、1つは分類を始めから配置し、配置された2次情報に含まれる特許ノードのみが出現する、つまり2次情報で絞込みを行うモード、もう1つは配置した検索キーワードと関連のある特許の出現元をすべて表示するモードを選択することができる。後者において、全ての2次情報を表示することはスペース的に難しいので、2次情報の表示を切り替える。つまり、仮に米国特許分類705/2に3つの特許が含まれる場合、それらを放出し終わったら705/2は縮小するか画面上から消えることで他の分類を表示するスペースを確保する。

図2では米国特許分類のみを例として表示したが、他に特許権者、発明者、国際特許分類を選択でき、それらを混在して配置する事も可能とする。



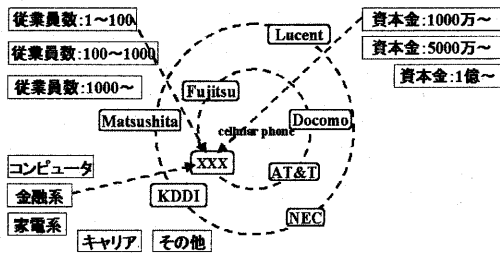


図4 企業情報の表示

企業情報をノードとした場合、出現元には企業の情報（従業員数、資本金、業種等）を配置する。図4は携帯電話：“cellular phone”を磁石として配置した例だが、これにより特許空間の中で各企業がどれだけ携帯電話に関わっているか、携帯電話に関連した企業にはどんなものがあり、またどの程度関連しているのかという情報を伝えることが可能である。

4.5 仮想3次元空間の利用

情報検索の分野に情報視覚化を導入する際に問題になることのひとつとして、一般的なキーワード検索のリスト表示のように、文書のアブストラクトや要約などを表示しにくいことがある。タイトルだけでなくアブストラクトなり要約などの文章を表示することはGoogleなどの検索エンジンで実証されているように非常に有用であり、求められる機能である。ノードをクリックするとその要約が表示されるシステムもあるが、いちいちそのノードをクリックすることがユーザにとって大きな負担となってしまう。

本視覚化のように概観での理解が容易に行えれば角度を変えても情報を認識することは可能であると我々は考える。図5の上の表示を下の表示に切り替えても必要な情報を得ることは可能である。

そこで、本システムでは2次元空間で操作するモードと、仮想3次元空間に視覚化の平面とリスト表示の平面を同時に表示するモードを実装した。仮想3次元空間においては視覚化の平面を図5のように傾けることでスペースを確保し、そのスペースにリスト表示の平面を配置する。この手法により、アブストラクトや要約などの詳細情報を効率的に表示することが可能となる。

なお、本研究ではリストに表示するのはタイトルとアブストラクトとし、要約の作成については対象外とする。

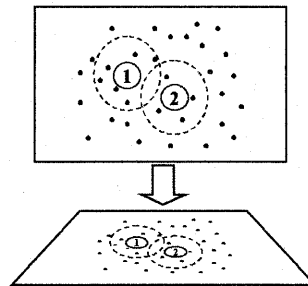


図5 平面に対する視点の移動

5. システムの実装

本節ではここまで説明してきた概念に基づいて実装したシステムについて述べる。特許文書は米国特許庁サイトからダウンロードしたものを用いた。文書はHTML形式となっており、本実装においては1996年1月1日から2001年11月までの米国特許分類705がふられた米国特許全4273件を扱った。

本研究は基本的に視覚化手法の提案であるため、各特許文書における出現単語のスコアを算出する手法には言及していないが、本実装においては基本としてTF法を用い、更に我々の研究であるReferencesを用いた特許文書の分析手法[8]を用いることでその精度を高めた。

5.1 視覚化の全景

図6が実装画面の例である。右下のテキストフィールドに検索キーワードを入力し、setボタンをクリックするとそのキーワードの磁石が出現する。次に左下にあるInventors, Assignee, CCL, IPCボタンをクリックすると、それに対応したリストが図6のように立ちあがる。リストから選択すると2次情報が画面上に表示されるので、マウスと用いてキーワードと2次情報の配置を適当に決定する。

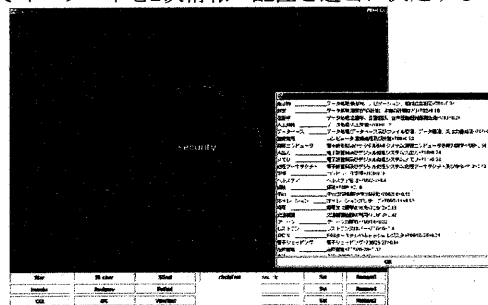


図6 システムの全景

5.2 ノードの移動

2次情報の配置が終了したら左下にあるstartボタンをクリックする。すると、図7のように各2次情

報から特許ノードが出現し、配置したキーワードに引き寄せられて行くアニメーションが表示される。グレースケールのためわかりにくい但实际上には2次情報にはそれぞれ色がつけられており、ノードは出現元と同じ色をしている。図の例でい

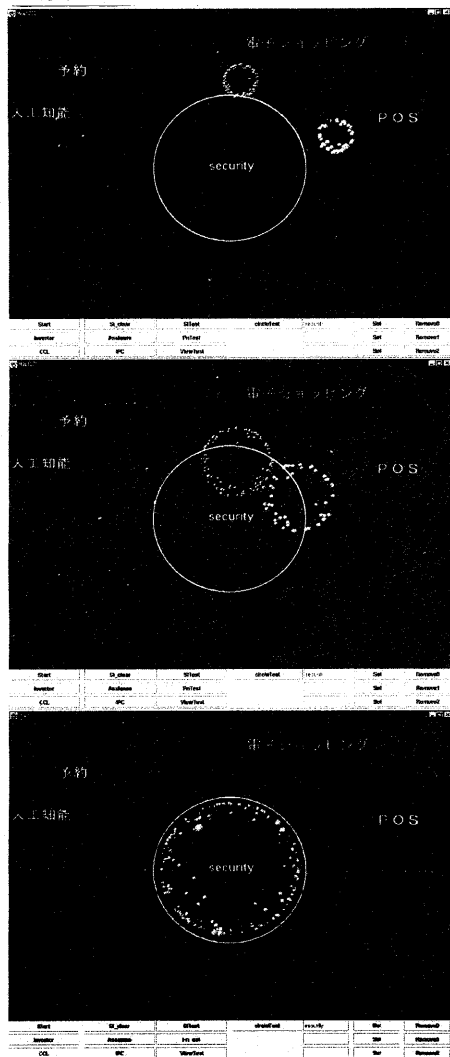


図7 ノードの移動

ば、この動きを見ているだけで

- 予約はsecurityと関係有りそうなのに特許が無い
- 人工知能は特許が少ないけどある
- レストランには重みが低く数も少ないが特許がある
- 各分類の量の把握

等の情報を得ることができる。このような情報が我々の定義した言葉でいうRetrieval Awareness (検索アウェアネス)である。

5.3 仮想3次元表示

図8は視覚化部分とリスト表示を仮想3次元空間で連携した時の例である。リストには特許番号、タイトル、Abstract 2行分が表示されており、特許7件分の情報が掲載されている。各情報は線で視覚化部分とつながれており、視覚化システムとの連携を保っている。表示されている7件の情報は現在画面に表示されている特許のスコアが上位7位のものであり、YahooやGoogleで次の検索結果を表示するのと同じように8位から14位までの表示が可能である。現段階の実装では49件まで表示できるよう設定してある。

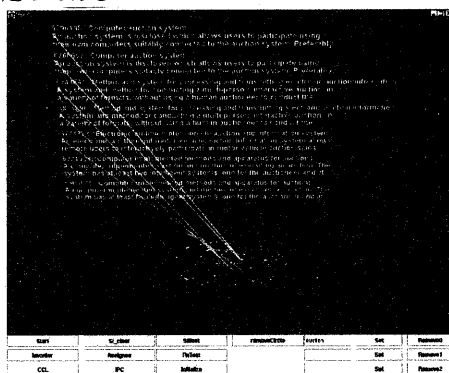


図8 視覚化とリスト表示の連携

6. システムの評価

前節まででRetrieval Awarenessの有用性を示してきた。しかし、どんなにノードの移動が有効な意味を持ち多くの情報をユーザに提示しても、それをユーザが認識できなければ何の意味も持たない。ユーザが情報を受け取れない可能性としては、そもそもノードの移動に情報をのせることに無理があること、ノードの移動に目や認識力がついていけないことが考えられる。

そこで、評価実験としてユーザが実際に情報を受け取れているのか、また2次情報の数とノードの移動速度の関係について被験者20人に対して実験を行った。実験は我々が準備した条件下(キーワードと2次情報を任意に決定し、2次情報数5,7,10,13個の場合について行う)で被験者にノードの移動を見てもらい、以下の質問に回答してもらいその正当率を評価の基準とした。

- ノードが出現しない2次情報
- 出現特許数が最も少ない2次情報
- 出現特許数が最も多い2次情報
- キーワードとの関連が深い2次情報

なお、本視覚化システムはJava3Dで実装しており、アニメーションは javax.media.j3d の PositionPathInterpolator を利用している。この PositionPathInterpolator ではアニメーションの設定を javax.media.j3d.Alpha という値に設定している。この Alpha のパラメータの中に increasingAlphaDuration (増加に要するミリ秒)があり、これはノードの出発点と目的地が決まっている状態で、その間の移動に何秒かけるかという値である。つまり値が大きいほどノードの移動速度は遅く、短いほど速いということになる。本実験ではノード移動時間が 2000, 4000, 7000, 10000, 13000 の場合において評価を行った。

図9,10は質問事項別に評価結果をまとめたグラフであり、スペースの関係で4つのうち2つのみ掲載する。どちらのグラフにおいてもノード移動時間が2000、つまり最も移動速度が速い場合は、特に2次情報数が13個の場合で正答者数はかなり少ない。しかしながら、速度を落としていけば正答者数が増えて行く様子がわかる。移動時間13000においてはほぼ全員が正答を示していることから、配置2次情報数に応じて移動速度を適切に設定すれば、ほぼ確実にユーザに情報を提供できることを確認できた。本実装ではノード移動時間はユーザが任意に設定できるが、本実験結果をデフォルト値として設定することで、システム初心者に対する違和感を減少できると考える。

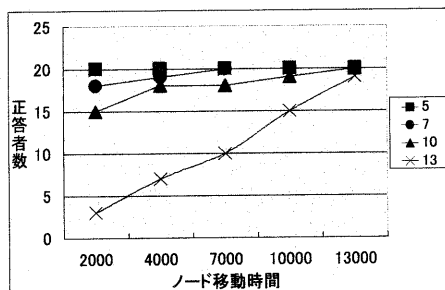


図9 ノードの出現に関する結果

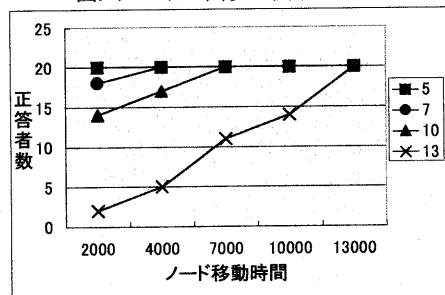


図10 最も少ない2次情報に関する結果

7. まとめ

本研究では、2次元空間において動的な視覚化手法を用いることで扱える情報を増加する手法と、3次元空間を有効活用して文書を表示するスペースを確保する手法により、2次情報を多く含む特許文書を効率的に調査できるシステムを提案し、プロトタイプを実装した。

更に評価実験を行うことでユーザがキーワードと特許文書の関連度のみならず、各文書の2次情報や、キーワードと2次情報の関連などの情報を実際に獲得できることを確認した。

今後、特許を専門的に扱っているようなより多くのユーザに本システムを試用していただくことでシステムを洗練させていき、また技術論文、Web文書、映画や音楽情報等への応用にも取り組んで行く予定である。

参考文献

- [1] United States Patent and Trademark Office's web page, <http://www.uspto.gov>
- [2] A patent search engine owned by Questel-Orbit, <http://www.qpat.com>
- [3] Tatemura, J.: Visualizing Document Space by Force-directed Dynamic Layout, Proc. 1997 IEEE Symposium on Visual Languages (VL'97), pp.119-120 (1997).
- [4] Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B. and Williams, J.G.: Visualization of a document collection: the VIVE system, Information Processing & Management, Vol.29, No.1, pp.69-82, 1993
- [5] Benford, S.D., Snowdon, D.N., Greenhalgh, C.M., Ingram, R.J., Knox, I. and Brown, C.C.: VR-VIVE: A virtual environment for co-operative information retrieval, Proc. Eurographics'95, pp.349-360, 1995
- [6] J. Cugini, S. Laskowski, M. Sebrechts.: Design of 3D Visualization of Search Results: Evolution and Evaluation, Proceedings of IST/SPIE's 12th Annual International Symposium: Electronic Imaging 2000: Visual Data Exploration and Analysis (SPIE 2000), San Jose, CA, 23-28 January 2000.
- [7] Shneiderman B.: Dynamic queries for visual information seeking, IEEE Software, Vol.11, No.6, pp.70-77, 1994
- [8] 奈須庄健, 淡谷浩平, 出口智恵, 重野寛, 松下温. ユーザの視点に基づくビジネスモデル特許の視覚化手法の提案と実装. 情報処理学会第62回全国大会特5, pp.111-112, March 2001