

科学知識の獲得支援のための“map of scientific knowledge”を目指して
- 視覚化手法のレビューと評価、開発に向けた取り組み -

東京大学大学院 総合文化研究科 広域科学専攻 博士課程2年
科学技術インタープリター養成プログラム 2期生

堀部直人

指導教員: 岡本拓司 准教授

2010.02.28

科学知識の獲得支援のための“map of scientific knowledge”を目指して —視覚化手法のレビューと評価、開発に向けた取り組み—

科学技術インタープリター養成プログラム二期

堀部直人 nhoribee@08.alumni.u-tokyo.ac.jp

指導教員：岡本拓司准教授

要旨

科学コミュニケーションと言いつても、その活動内容は様々である。私は、これら活動が3つに分類可能だと考える。(i)科学者から社会へのコミュニケーション、(ii)非科学者が民主的な意志決定の主体として科学では解決できない問題群と関わっていくためのコミュニケーション、(iii)非科学者が、科学の発達した社会を生きていくための科学リテラシーを身につけるためのコミュニケーションである。科学コミュニケーションが求められる背景を考えると、(ii)や(iii)のコミュニケーションに非科学者が積極的に参加しなければならないという義務はない。非科学者は科学コミュニケーションに価値があると思えるときに参加すればよいのである。したがって、参加を促すためには科学コミュニケーションのコストパフォーマンスを高めていく必要がある。コストパフォーマンスを高める方法として、「視覚化」により、科学知識獲得に伴う時間的・精神的コストを削減することを考えた。

情報整理のための視覚化としては、年表・系統樹・一覧表・路線図といった時空間軸を考慮したものや、一家に一枚シリーズ・大統計大マンダラ・生態学の系譜といった主観を含めてまとめていったものが存在する。科学知識を視覚化する一般的手法の開発を目指し、その足がかりとして、科学計量学に注目した。

科学計量学は、計量的な手法によって学術の動向を整理しようという学際的な領域である。例えば、Leydesdorffらによる“map of science”はジャーナル間の引用-被引用関係をもとに、172の学問領域(ISI subject categories)の繋がりを視覚化している。また、Visualization and Information Retrieval (VIR) Research Groupは“map of science”よりも解像度の高いマップをダイナミックに生成するソフトウェア(“PNASLink”)を開発している。

科学計量学による視覚化の目的は、研究領域の整理、研究コミュニティの整理、資金配分・評価であり、科学知識獲得に有効かどうかは自明ではない。そこで、アンケート評価を行った。これにより、科学計量学による視覚化は未知の学問領域を学ぶ手助けになりそうだという弱い傾向と、より有効な視覚化のためになにが足りないかを明らかにできた。

アンケート結果を基に、科学リテラシーの標準となることを目指したScience for All Americansに対して科学計量学の手法を適用することで、科学リテラシーの視覚化を試みた。残念なことに、おそらくは読み込んだ単語数の少なさが原因で、妥当性の低いリンクで術語が繋がる、少数の語同士だけでリンクしたクラスターが複数形成されるなど有用なマップは得られなかった。

有効なマップ作成のためには、(1) 読み込むデータ数を増やし、(2) 解像度を考慮した術語の選定を行うと同時に、(3) 多義語をうまく処理し、(4) リンク数の制限や、向き、太さを考慮し、(5) 使いやすいユーザーインターフェースを設計する必要がある。なによりも重要なのは、科学計量学をつかってどんな視覚化が可能かを考えることではなく、どのように視覚化すれば科学知識獲得が支援できるかをきちんとデザインし、そのデザインを実装するために科学計量学の手法を利用していくことである。

目次

1	序論	3
2	視覚化手法	5
2.1	明確な軸に沿った整理	5
2.1.1	地図・年表・系統樹・周期表	5
2.2	統計的・計量的な整理	5
2.2.1	自己組織化マップ	5
2.2.2	科学計量学	5
2.3	主観を交えた整理	6
2.3.1	学問の見取り図	6
2.3.2	一家に1枚シリーズ	6
2.4	考察	7
3	アンケート調査	9
3.1	アンケート結果	9
3.1.1	map of science	9
3.1.2	PNASLink	9
3.1.3	自由回答欄	9
3.2	アンケート結果のまとめと考察	10
4	開発に向けた取り組み	14
4.1	方法と結果	14
4.2	考察	14
5	プログラムを受講して	19
5.1	プログラムを通して影響され、変化したこと	19
5.2	今後の抱負	19
	参考文献	20
	Appendix	20
	解析用プログラム	21
	アンケート	23

1 序論

科学コミュニケーション元年ともよばれる 2005 年、東京大学、早稲田大学、北海道大学の 3 大学が、文部科学省科学技術振興調整費をうけて科学技術コミュニケーション人材の養成プログラムをスタートさせた。科学技術コミュニケーション人材の育成という根本理念を共有しながらも、各プログラムは異なった特徴を有する。これは、科学コミュニケーションが広大な裾野をもつ領域であるためだ。広大な裾野のどこを、どのように扱うかを考えるため、まずは科学コミュニケーション活動について整理するところからはじめたい。

「科学コミュニケーション」のもっとも包括的な定義としては『科学と科学技術と技術に関するコミュニケーション一般』[7]がある。実際、サイエンスカフェ、サイエンスショー、出張授業、コンセンサス会議、テクノロジーアセスメント、サイエンスショップなど目的や対象の異なる活動のすべてが、科学コミュニケーション活動であるとされる。

筆者は、これら活動が以下に挙げる 3 つの要素の組み合わせとして分類・整理可能であると考ええる。それは、(i) 科学者から社会へのコミュニケーション、(ii) 非科学者が民主的な意志決定の主体として科学では解決できない問題群と関わっていくためのコミュニケーション、(iii) 非科学者が科学の発達した社会を生きていくための科学リテラシーを獲得していくためのコミュニケーション、である。例えばサイエンスカフェであれば“(i)+(ii)”あるいは“(i)+(iii)”、出張授業であれば“(iii)”あるいは“(i)+(iii)”、コンセンサス会議であれば“(ii)”あるいは“(i)+(ii)”といった具合になる。

近年科学コミュニケーションの必要性が叫ばれているが、その必要性は自明ではない。(i)については科学者の知的生産物に対する責任、あるいは研究予算を受けたことに対する責任からその必要性、あるいは義務が生じると考えられる。一方、(ii)や(iii)については、自らの活動を支持してもらいたいという科学者の要望、あるいは民主的な意志決定に参加して欲しいという為政者の要望からその必要性が主張されているようであるが、非科学者の側に参加の義務はない。民主的な意志決定に参加することや科学リテラシーの獲得に価値があると感じたときに科学コミュニケーション活動へと参加すればよいのである。

では、(ii)や(iii)の科学コミュニケーションを促進するにはどうしたらよいだろうか。功利的な考え方ではあるが、非科学者が科学コミュニケーション活動に参加することのコストパフォーマンスを高めることが必要だろう。つまり、参加により得られた科学知識が生活に役立つといった利便性、科学知識を獲得することで得られる喜び、科学コミュニケーション活動に参加することの楽しさといった利益と、知識獲得に伴う精神的な苦勞、時間、金銭などのコストとのバランスの改善である。近年盛り上がりを見せている科学と芸術の融合といった催しは利益を高める戦略と考えることができる。一方コストに関してはあまり関心が払われていない。そこで本研究では、コストを削減するという方向から科学コミュニケーション活動のコストパフォーマンスを高めることを考える。

科学コミュニケーション活動においてコストと捉えうるものとして、情報の取捨選択にかかるコスト(例えば、なにについての話を聞きに行くか、どこに聞きに行くかなど)、参加に伴う時間的・金銭的成本、知識獲得に要する時間的・精神的コストが考えられる。筆者は、情報の視覚化によって情報の取捨選択にかかるコストや知識獲得にかかるコストを削減することができる考え、注目した(例えば図 1,2)。

本研究の目的は、「知識獲得の支援に有効な視覚化手法の開発」である。

次節でまず、情報整理のための視覚化手法の具体例を紹介する。ここでの目的は、科学知識獲得を支援するような情報整理の方法を考察することである。第 3 節では、科学計量的手法による情報整理に知識獲得を支援する効果があるかどうかを調査するアンケートの結果を紹介する。続く第 4 節では、科学リテラシーの獲得支援を目的とした視覚化を実際に行った結果を示し、よりよい視覚化手法についての考察を行う。

砂漠で桜が咲けば食糧危機も怖くない!?

今年の桜が咲く少し前のことです。「自分で勝手に動き出す不思議な油」を研究しているデンマークの研究者がやってきました。僕が「ハエの動き」を研究していることがきっかけとなり、僕と先生と彼とで、一緒に新しい実験を始めることになりました。

桜が満開になる頃には実験も軌道に乗り、毎日のように議論し、ご飯を一緒に食べ、桜を眺め、と楽しく濃厚な時を送りました。そして桜が散る頃、彼はデンマークへと帰っていきました。E-mail やインターネット電話でいつでも連絡が取れるとはいえ、桜が散っていくのと相まって、とても名残惜しい別れでした。

桜は年度の変り目のほんの短い間に一斉に咲き、そして散っていくため、出会いや別れの象徴となっています。思い返してみると、僕と研究の出会いにも桜がかかわっていま

した。今回は僕が研究者になろうと考え始めた高校生の頃の話をしてみたいと思います。

今ニュースを読めると、地球温暖化だとか、環境問題だとかがよく取り上げられています。僕が中学生・高校生の頃は、地球規模での人口増加とそれに伴う食糧危機という問題がよく取り上げられていたように思います。

僕はその頃すでに、なぜかはよく分からないけれど理系少年でした。なので、政治や経済の面から食糧危機の解決を目指すという考えはほとんど浮かばず、品

種改良によってたくさんの実をつける植物をつくってみたいと考えていました。

この漠然とした考えが具体的な目標へと固まっていたのは高校2年生の時です。生物の授業で遺伝子の仕組みを習い、遺伝子を直接操作することでいままでの方法よりもはるかに効率よく品種改良ができるのではないかと考えたのです。

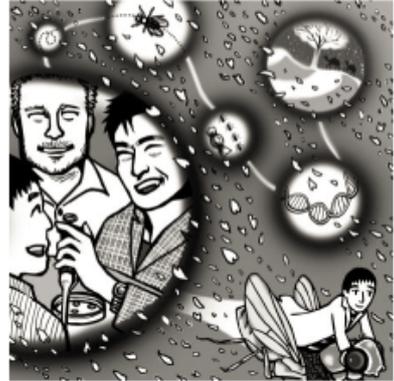
遺伝子組み換え植物の研究をしたい、これがたぶん僕が研究者になろうとした最初のきっかけです。そして高校生の僕は、卒業文集だかどこかでは忘れてしまいましたが、「砂漠で咲く桜をつくりたい」と将来の夢を書きました。砂漠で咲く桜をつくり出す技術があれば、食糧問題もきっと解決

できると考えたのです。

ところが前回紹介したように、僕はいま昆虫の行動を研究しています。そのいきつは次回以降、詳しく紹介していきます。砂漠に咲く桜をつくる研究とは今後も縁がなさそうですが、自分の研究やデンマークの彼との共同研究ではきれいな花を咲かせようと思います。



研究用のハエを飼育する恒温器の前で



イラスト：阿部真理子（東京大学大学院総合文化研究科博士課程）

東京大学の研究室の様子や科学の話題を紹介するコラムです。堀部さんのコラムは6月ごろまで不定期に連載する予定です
Science Potのウェブサイト www.sasaku.com/science_pot/science_pot.html

図 1: 朝日中学生ウイークリー紙上で連載企画。堀部が文章を担当し、科学技術インタープリター養成プログラム3期生の阿部真理子氏がイラストを担当。イラストが文章の内容を効果的に要約すると同時に親しみやすくもしている。朝日中学生ウイークリー2009年5月3日の記事より。

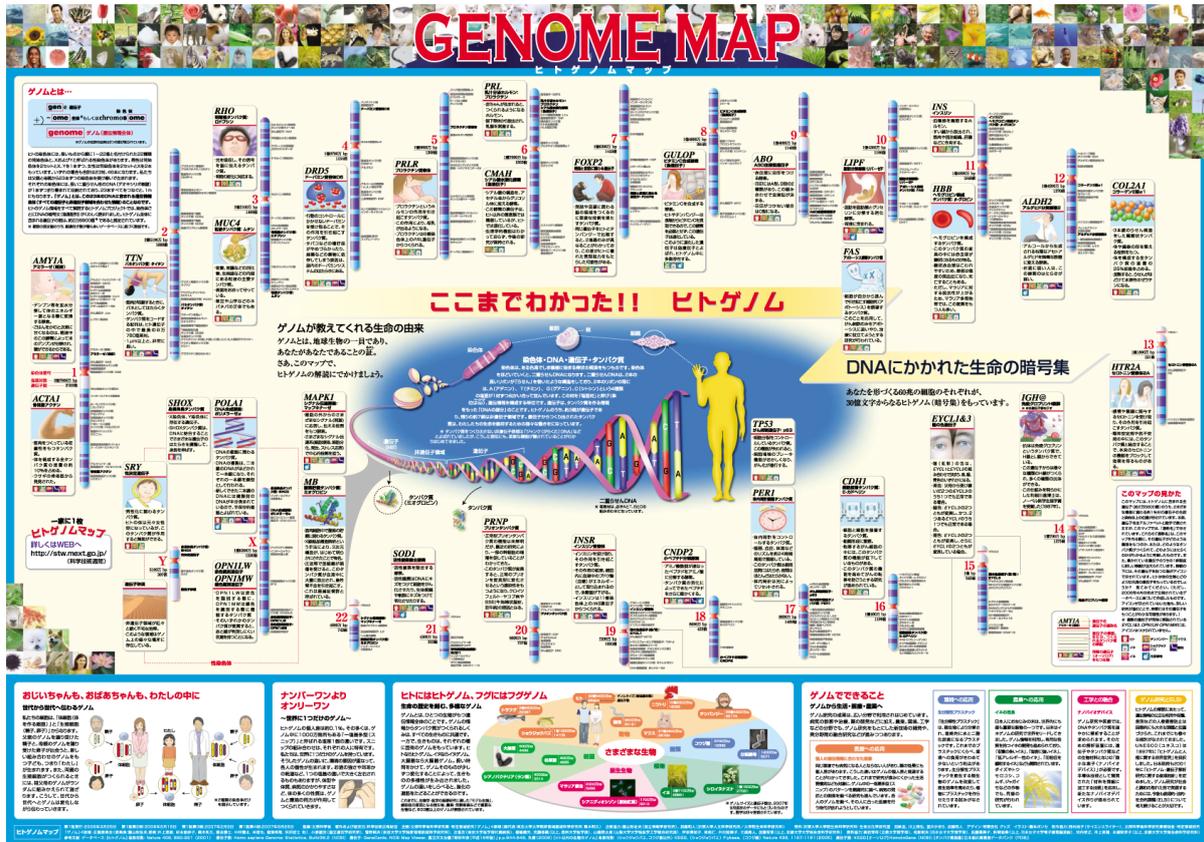


図 2: 科学と親しむ機会を多くの人に提供することを目的に作成された『一家に1枚シリーズ』のヒトゲノムマップ。ゲノムについての解説文章と、染色体・重要な遺伝子の一覧が掲載されている。

2 視覚化手法

本節では情報を整理し、視覚化して提示する方法についてレビューを行う。整理の方法として、(1) 明確な軸に沿った整理、(2) 軸が不明確であるが統計的・計量的手法を用いた整理、(3) 軸が不明確であり整理者の主観に依存するところの多い整理、の3パターンが(3)の整理方法からみいだされた。それぞれについて具体例とともに特徴を紹介する。

2.1 明確な軸に沿った整理

2.1.1 地図・年表・系統樹・周期表

ある地域を縮小し、記号や文字を利用して平面上に描いたものが地図である。地図には主として2次元の空間軸が存在する。ただし、路線図のように正確な位置関係よりもむしろ接続関係を重視して作成されるものもある。地表に存在するものをまんべんなく記した一般図と、特定の主題について専門的に表す主題図とに大別できる。

対象物を空間軸ではなく時間軸に沿って並べたものが年表や系統樹である。年表は時間軸に沿って直線的に並べるが、系統樹では進化の道筋に従って、枝分かれした樹木の形で並べ整理を行う。元素を原子番号の順番で並べる周期表のように時空間軸以外の軸を用いた整理も可能である。

2.2 統計的・計量的な整理

2.2.1 自己組織化マップ

整理した結果を2次元平面上に描き出すことを考えると、同時に利用できる軸の数はせいぜい2~3個までである。そのため、複数の軸が存在する場合には何らかの工夫を行ったうえで、2次元平面上に描き出す必要がある。

自己組織化マップ (Self-organization map) はそのような方法の一種である [1]。これは、正方形格子あるいは六角格子上にデータがもつ次元数と同じだけの次元数を持つベクトルを用意しランダムな初期値を与え、(i) 提示されたデータと最も似ているベクトルを選び出し、(ii) そのベクトルを提示されたデータに近づくよう値を修正するとともに、(iii) そのベクトルの近傍のベクトルも提示されたデータに近づくよう修正を行う。この(i)-(iii)までを繰り返すことで似たベクトルが近傍に存在する格子が完成する。この格子上にデータを投影させることで、多次元の情報を2次元平面に落とすことができる(図3参照)。ただし、多次元の情報をむりやり2次元に落とし込むため、データを提示する順番や初期値に依存して異なる結果が得られることも多い。

2.2.2 科学計量学

科学計量学とは『科学技術活動を定量的に扱おうとする研究一般』[6]である。研究者、学術雑誌、学問領域、術語などを論文の共著関係、単語の共出現、引用-被引用関係、ファンドの情報などを定量的に扱っていくことで、研究領域の整理、研究コミュニティの整理、資金配分や評価に応用する。学問分野や学術用語の視覚化も試みられている。筆者はこの手法に注目しており、次節ではアンケートによる評価を行っている。以下に具体例を詳しく紹介する。

Map of science 物理学や化学といった学問領域が科学の視覚化において最も解像度の低い階層であり、学問領域を科学計量学の手法を用いて視覚化したのが Leydesdorff[2] による “map of science”(図 4) である。(ここで解像度とは視覚化の際に扱われるデータの階層を意味する。例えば、「科学 ⊃ 物理学 ⊃ 古典力学 ⊃ 運動方程式」といった包含関係においては、より右側に位置する項目ほど解像度が高い。) 学問領域の細分化が進んでいるため、その正確な数の把握は困難であるが、ここでは ISI subject categories に登録されている 172 の学問領域を、ジャーナル間の引用-被引用関係を利用して視覚化している。172 の学問領域に絞ったとはいえ、まだ解像度が高いと感じた彼らは、172 の学問領域をさらに解像度の低い 14 の領域に束ねることでより視覚的に把握しやすい図を作成している¹。この図からは、化学 (Chemistry) が生物学 (Environmental sciences, Agriculture, Biomedical sciences) と物理学・工学 (Physics, Engineering, Materials sciences) とを結ぶハブになっていること、生物医学研究 (Biomedical sciences) が非常に盛んであることなどが読み取れる。このように、“map of science” は、低い解像度に抑えることで全体像を示す視覚化方法といえる。

PNASLink 専門書の索引語程度の解像度を高解像度としよう。この解像度での視覚化として、Visualization and Information Retrieval (VIR) Research Group の PNASLink[4] を挙げるができる。これは *Proceedings of the National Academy of Sciences of the United States of America* 誌に掲載された論文において共出現する術語のペアのうち、重要なペアを結ぶことにより視覚化を行うものである。PNASLink の最大の特徴は、ユーザーの要求に応じてインタラクティブに視覚化するソフトウェアとして実装している点にある。例えば、図 5 は筆者が “evolution” という語を入力して描かせた結果である。

2.3 主観を交えた整理

2.3.1 学問の見取り図

学問分野のような複雑なデータを客観的な方法で整理するのは難しく、主観に基づいた整理が行われることもある。私的な整理のおもしろいものとしては、三中信宏氏による「大統計大マンダラ」[5] や門司正三氏による「生態学の系譜」[8] などが挙げられる。いずれも個人の主観に強く依存しており、クセが強く、おもしろいと感じる人がいる反面、異論・反論を唱えたい人もいるだろう。

2.3.2 一家に 1 枚シリーズ

科学技術週間を多くの人に知ってもらうとともに科学に親しむ機会を提供するために、文部科学省が中心となって作成を行っているのが『一家に 1 枚シリーズ』である。いままでに周期表、ヒトゲノムマップ (図 2)、宇宙図、光マップ、天体望遠鏡 400 年、という名称のポスターが作成されている。基本コンセプトとして以下のことが掲げられている。

1. 大人から子供まで部分的にでも興味を持たせるもの
2. 見た目がきれいで、部屋に貼っておきたいもの
3. 基礎的・普遍的な科学知識を中心とするもの
4. 身近な物や事象との関連付けをして、親しみをもてるもの

¹ ウェブ (<http://users.fmg.uva.nl/lleydesdorff/map06/index.htm>) 上で公開されているマップでは、ノードをクリックすることで低解像度から中解像度の学問領域へと移動できる。これにより興味のある領域の詳細を把握することが可能となっている。

主観で描かれる部分が多いとはいえ、多くの研究者やデザイナーが関わっていることもあり、非常に親しみやすいものとなっている。欠点を挙げるとするならば、時間と費用がかかること、それゆえ様々な分野を網羅するのが難しいことなどであろう。しかし、理化学研究所の「ゆめみる脳科学地図」のような類似の企画が催されることもあり、各研究機関がアウトリーチ活動として自らの研究をデザイン性の高いポスターとして発信するようになるとこの点が解消されるかもしれない。

2.4 考察

科学知識の獲得支援に有効な整理手法はどのようなものだろうか。科学知識を空間軸に沿って配置するのは困難であるが、系統樹のようなものを考え、時間軸上で配置することは可能であろう。門司正三氏の「生態学の系譜」では厳密ではないが時間軸が設定されている。しかし、時間軸上に並べることで知識の獲得支援にはならないだろう。おそらく、一家に1枚シリーズのような構成が知識獲得支援には有効となるだろう。しかし、一家に1枚シリーズは多くの人の手と時間、金銭がかかるため、マイナーな学問領域の地図が作られることはあまり期待できない。したがって、なんらかのデータベースから自動で地図ができあがる手法を考える必要がある。そのようなものとして有効そうなのは、統計的・計量的な整理を行う科学計量学である。次節で、その可能性について調べるアンケート調査の結果を紹介する。

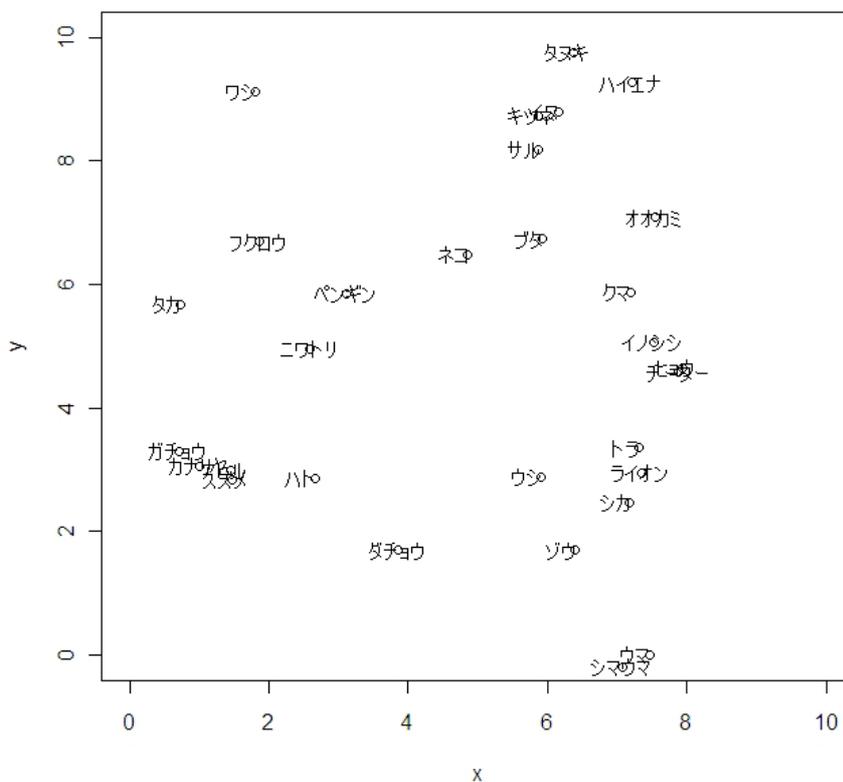


図 3: 30 種類の動物を 14 の特徴に基づき、自己組織化マップを用いて整理した。形態の似た動物が近くに表示されている。豊田秀樹 編著『データマイニング入門』(東京出版) のデータをもとに筆者が作成。

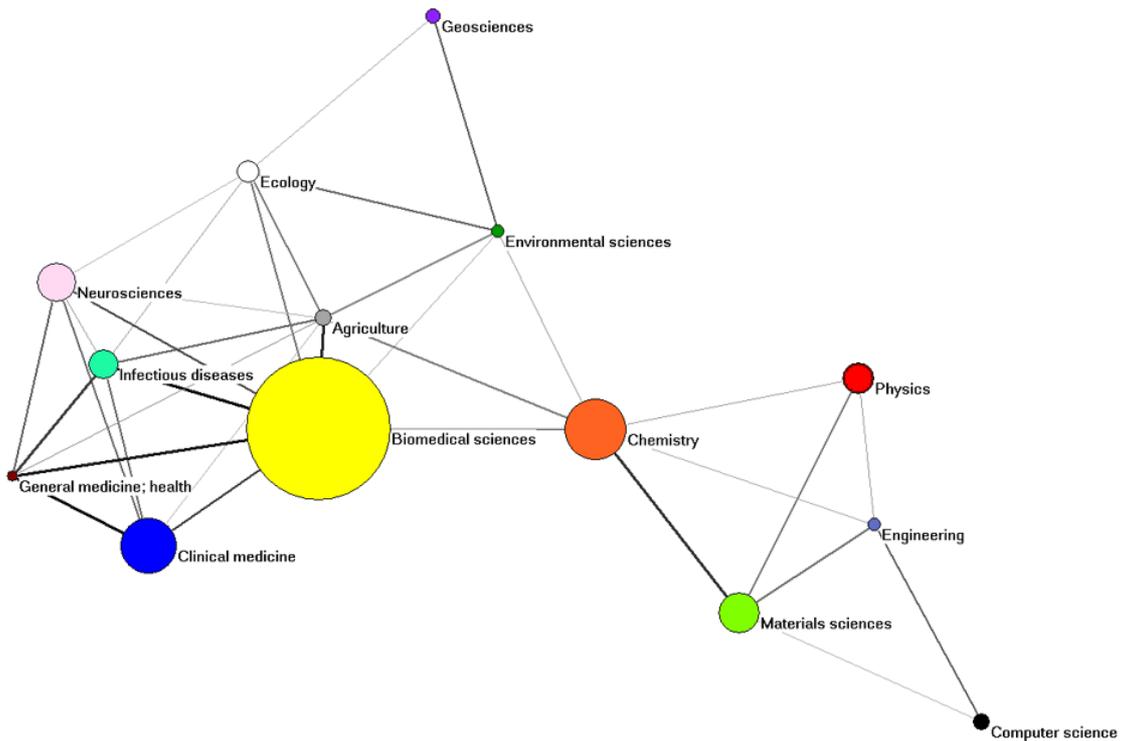


図 4: Leydesdorff らによる “map of science”。ここでは 172 の学問領域をより解像度の低い 14 の領域に整理したうえで図示している。(http://users.fmg.uva.nl/lleydesdorff/map06/index.htm より)

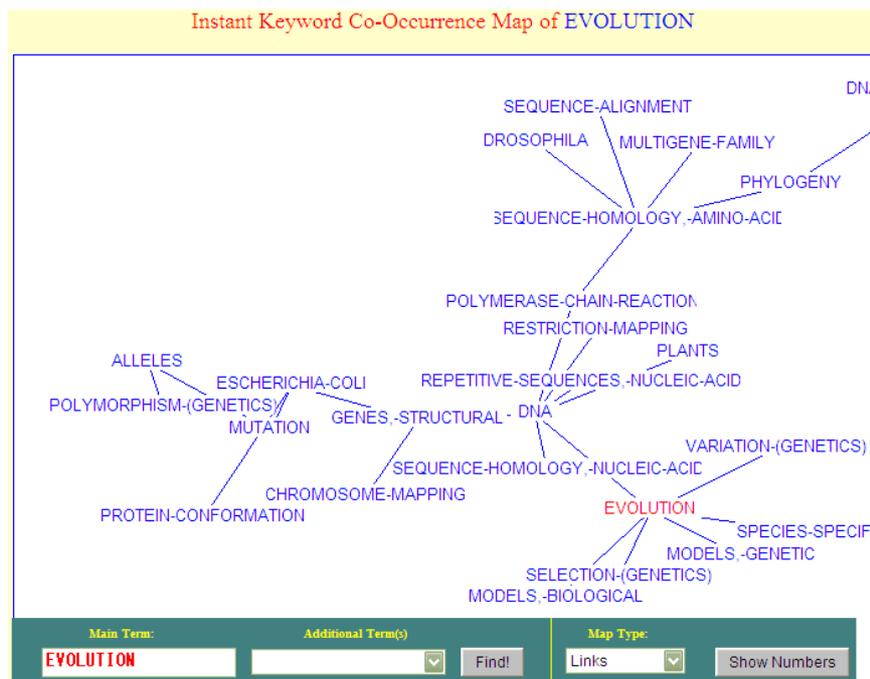


図 5: PNASLINK による “evolution” という術語を中心としたマッピング。ウェブ上に公開されているソフトウェアを使用して作成。(http://project.cis.drexel.edu/pnas/)より)

3 アンケート調査

様々な視覚化手法の中から、科学知識の獲得支援に役立つような手法として科学計量学に注目した。しかし、科学計量学にもとづく視覚化は、研究領域の整理、研究コミュニティの整理、資金配分・評価のために行われることが多く、科学知識の獲得支援に有効であるかどうかは自明ではない。そこで、アンケートによる評価を行った。評価の対象としたのは Leydesdorff らの “map of science” [2] と VIR Research Group による “PNASLink” [4] で、前者は学問領域を、後者は学術用語を視覚化している。評価は (i) 視覚化がおもしろいか、(ii) その視覚化が妥当であるか、(iii) 視覚化することで知識が整理されるか、(iv) 視覚化されたことにより新たな発見があったか、という4つの観点から行った。また、その学問領域、学術用語に対する習熟度も同時に尋ねており、習熟度との関連も評価した。(実際のアンケート項目については Appendix を参照。) 学部3年生から博士課程修了者まで、計37の回答を得た。

3.1 アンケート結果

3.1.1 map of science

おもしろさ、妥当性、知識が整理されるか、新しい発見があったかの集計結果を箱ひげ図で示した(図6,7)。全体像、評価者の専門領域、非専門領域いずれについてもおおむねおもしろく、妥当であるという評価が得られた。一方、知識が整理されるか、新しい発見があったかという点に関してはさほどでもないという回答であった。

専門領域と非専門領域についての回答を比較することで、習熟度と視覚化の評価の関連を調べたものが図8である。おもしろさと妥当性については習熟度によって評価が変わることは少ない。一方、知識が整理されるかと新しい発見があったかという点に関しては、なじみのない領域ほど評価が分かれる、すなわち、「そう思う」という回答が増えると同時に「全くそう思わない」という回答も増える傾向が見られた。

3.1.2 PNASLink

より解像度の高い視覚化となっている PNASLink については、表示結果はおおむねおもしろく、妥当であり、新しい発見がありそうという評価が得られた一方、知識の整理にはさほど役立たないという結果だった(図9)。これは、習熟度が高い語の視覚化を行った際にも1-2割という比較的高い割合で知らない術語が出現すること(図10)、リンクされているという情報だけではその意味の推論が難しいこと、術語の解像度がそろっていないことなどが原因として考えられる。

習熟度との関連では、習熟度が高いほどおもしろいと感じ、視覚化が妥当であると評価される傾向が見られ、妥当性との相関については統計的に有意だった(図11)。また、習熟度が低いほど知識が整理され、新しい発見があった、と評価される傾向が見られたが、統計的に有意ではなかった(図12)。

3.1.3 自由回答欄

自由回答欄に寄せられたコメントから代表的なものを抜粋(文体を一部改変)して紹介する。

- 興味を持たないであろうと思われる分野が目に入ってくることで興味を持った。
- 視覚化することで情報がとらえやすかった。
- 研究の広がりを感じた。
- ノードの意味や、ノード間をつなぐ意味まで知りたくなる。
- 半歩だけ他分野に飛び込む場合には大変有効だと感じた。

- リンクがありきたり。
- リンクが多すぎてどこが重要なのが分からなくなってしまう。
- ネットワークの図が視覚的に見にくい。
- 「概念」「手法」「材料」といったキーワードがごっちゃになっていて整理された気がしない。
- つながりがどれくらい太いのかが不明。
- 知っている分野については、特に目新しい印象はなかった。
- 特定の分野をテーマ別に地図にすれば頭の整理、知識の習得に役立つと思う。
- 写真やイラストを使って、術語に簡単なコメントをつけた方がよい。
- 方向性のあるリンクを使ってはどうか。
- 階層性が分かると整理しやすくなる。
- 3次元にしたり、自由にズームイン・アウトができるとよい。
- 日本語版で試してみたい。

3.2 アンケート結果のまとめと考察

科学計量学による視覚化そのものの評価はおおむね肯定的であり、妥当で親しみやすいものとなっていることがうかがえる。その一方、知識の整理と新たな発見についてはさほどの支持は得られなかった。視覚化対象の解像度がそろっていないこと、リンクに方向や意味が与えられていないこと、リンクの数が多すぎるなど視覚的に見やすすくないことなどが原因として考えられる。

とはいえ、習熟度の低い学問領域や術語に対しては視覚化の有効性が示唆されており、上記原因として考えられる事項を解決していくことでなじみの薄い学問領域や術語の学習を支援するための視覚化手法を開発できるだろう。次節で実際の取り組みを紹介する。

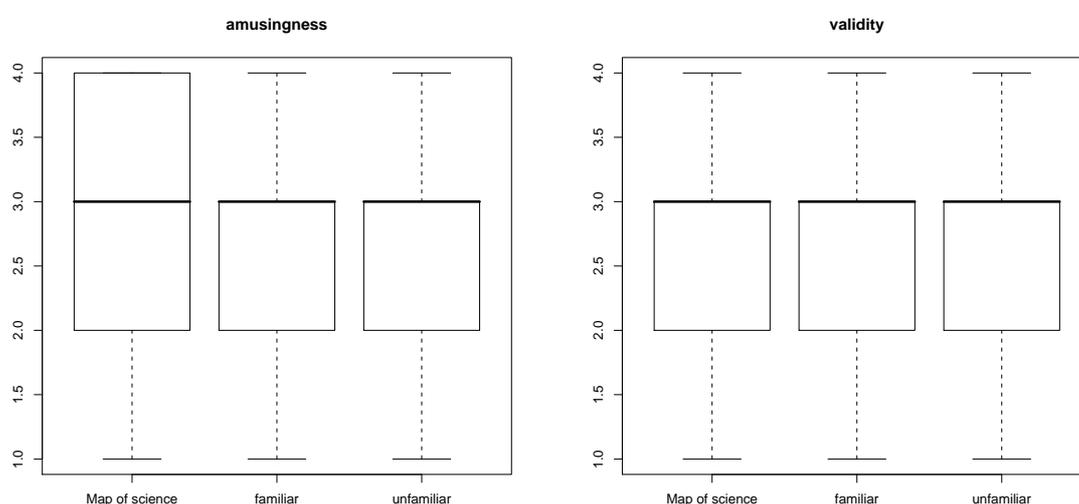


図 6: “map of science” のおもしろさ (左) と妥当性 (右)。全体像、評価者の専門領域、非専門領域の順。

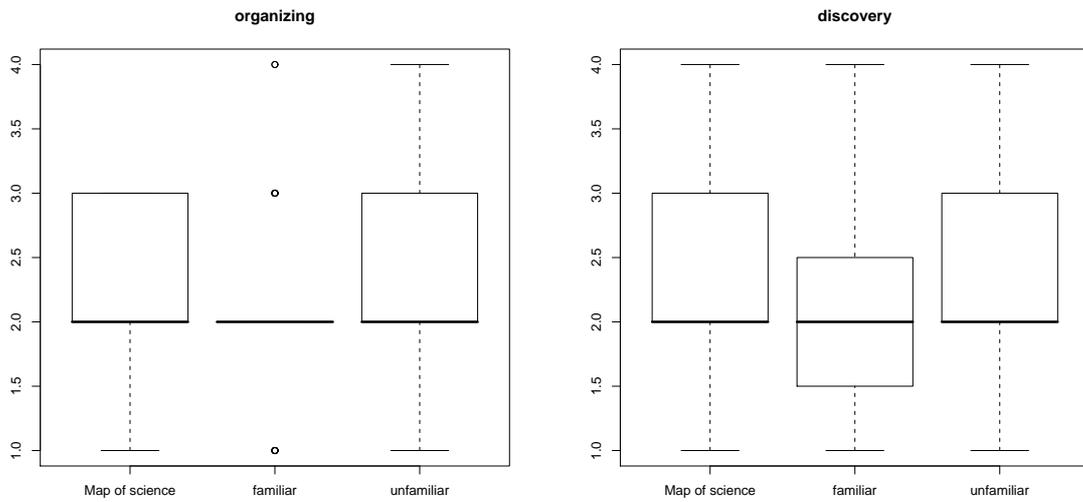


図 7: “map of science” で知識が整理されるか (左) と新しい発見があったか (右)。全体像、評価者の専門領域、非専門領域の順。

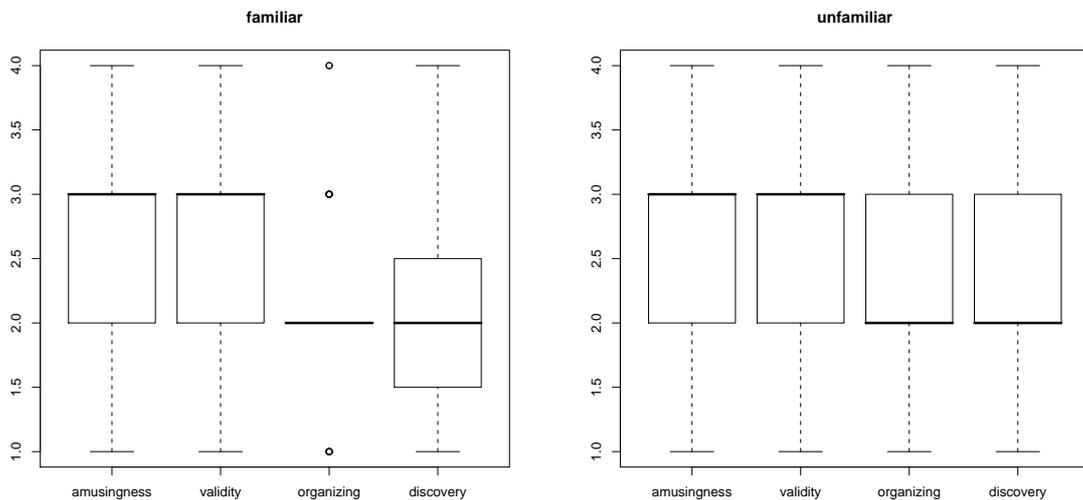


図 8: 専門領域と非専門領域について、4つの評価項目で比較したもの。評価者の専門領域 (左) と非専門領域 (右) となっており、それぞれの図について左からおもしろさ、妥当性、知識が整理されるか、新しい発見があったか。

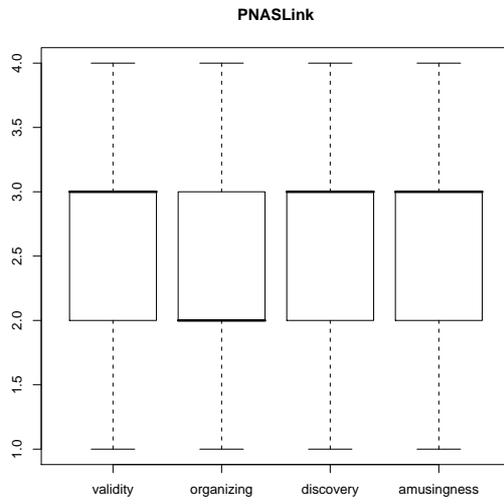


図 9: “PNASLink” の評価。左から順に妥当性、知識が整理されるか、新しい発見があるか、おもしろいか。

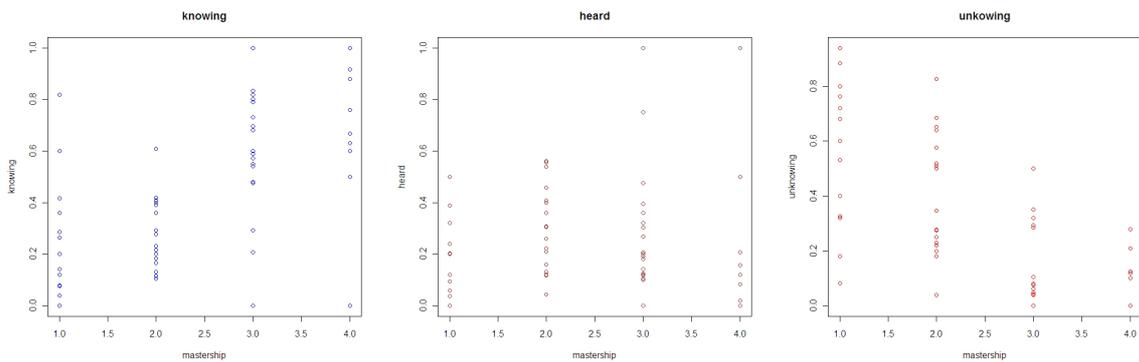


図 10: “PNASLink” で表示された術語のうち知っている語の割合 (左)、聞いたことのある語の割合 (中央)、知らない語の割合 (右)。

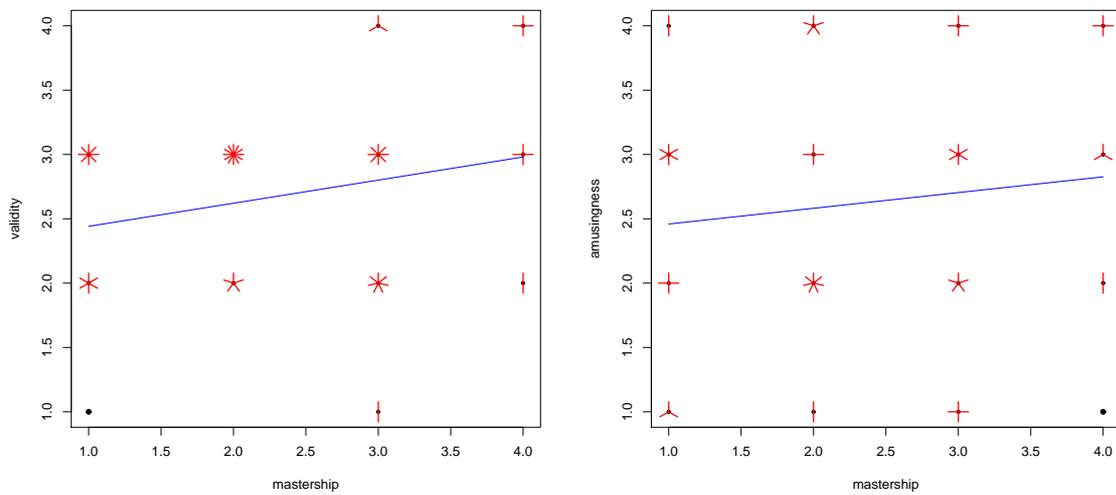


図 11: 術語に対する習熟度と妥当性 (左)、おもしろさ (右) の関係。両者ともに正の相関が見られるが、おもしろさについては統計的に有意ではない。

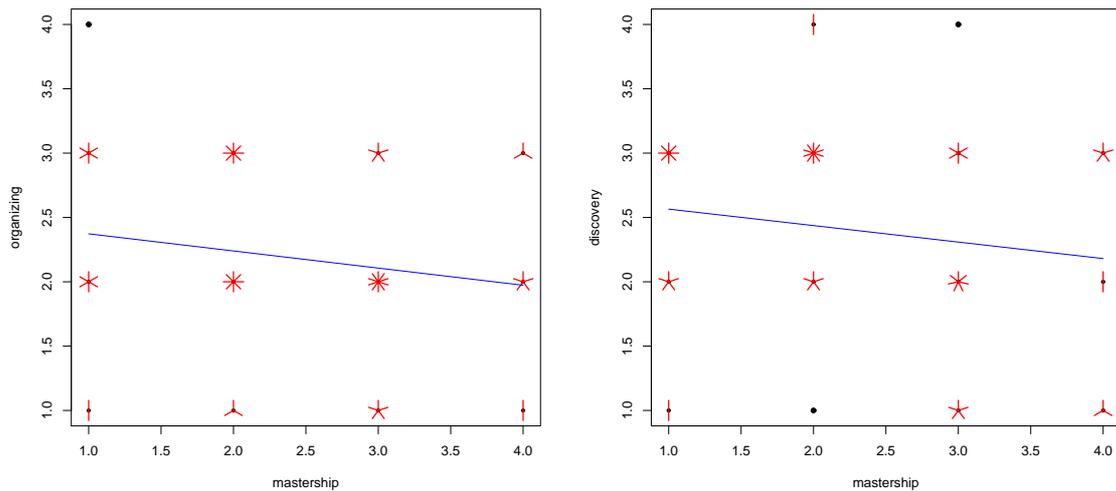


図 12: 術語に対する習熟度と知識が整理されるか (左)、新しい発見があるか (右) の関係。両者ともに負の相関が見られるが、どちらも統計的に有意ではない。

4 開発に向けた取り組み

4.1 方法と結果

科学知識獲得支援のための視覚化の一般的手法についての考察するため、科学リテラシーの標準となることを目指した Science for All Americans[3](以下 SFAA) を対象として選び、科学計量学の手法を適用することで科学リテラシーの視覚化を試みた。

まず、(i)SFAA をテキストデータとして保存し、(ii) 単語を切り出して単語ごとに頻度をカウントした。次に、(iii) 視覚化の際に表示させたい術語を選定していった。どの術語を選ぶかという点に明確な基準はなく、学問領域や術語を示す名詞のうちで重要そうなものを抜き出していった。最終的に 245 の術語が選ばれた (表 1)。そして、(iv) 選ばれた術語がどの章に出現したかを記録し、(v) 同じ章に出現したかどうかの情報から術語間の距離を表す行列を計算、(vi) 行列をもとにグラフを描く、という作業を行った。

図 13 は単純に共出現情報のみを利用した場合の距離行列である。全体的に色が薄く、数値の小ささがうかがえる。そのため、閾値を低く設定してグラフを描く必要があるが、その場合ほとんどの術語が他の術語と相互に結合してしまい有効な視覚化とはならない(図 16 参照。これは近接指標を用いた場合に得られた同様の問題を持つマップである。)。そこで、科学計量学で用いられる指標を利用して距離行列の補正を行った。代表的な補正方法は以下の通りであり、補正の結果を図 14,15 に示す(ただし、 C_i はある単語がいくつの場所(本研究では章の数)で登場したか。 N は登場しうる場所の数。)。これより上記問題を回避していると考えられる指標は包含指標であることが判る。包含指標を用いて補正した距離行列をもとに描いたのが図 17 である。

包含指標 $\left(I_{ij} = \frac{C_{ij}}{\min(C_i, C_j)} \right)$ 頻出語と稀な語との関係をみる。

近接指標 $\left(P_{ij} = \frac{C_{ij}}{C_i \times C_j} \times N \right)$ 稀な語同士の関係をみる。

Jaccard 指標 $\left(J_{ij} = \frac{C_{ij}}{C_i + C_j - C_{ij}} \right)$ 中程度に出現する語同士の関係をみる。

同等性指標 $\left(E_{ij} = \frac{C_{ij}^2}{C_i \times C_j} \right)$ 中程度に出現する語同士の関係をみる。

4.2 考察

多少なりとも有用な科学リテラシーマップが描かれることを期待したが、図 16,17 にあるとおりあまりうまくいっていない。これはおそらく、読み込んだデータの少なさによるものである。Map of science や PNASLink では数千から数万の論文データが利用されているが、本研究では SFAA の章を単位として 16 のデータしか用いていない。したがって、アンケート評価に用いた先行研究の 100 ~ 1000 分の一程度のデータ量ということになる。このため、Jaccard 指標や同等制指標のような中程度に出現する語同士の関係をみる際に有効な指標ではなく、包含指標や近接指標といった出現が稀な語をみる際に使用される指標を使わざるを得ず、多数のノイズがのった視覚化となった。

アンケート結果と作成作業から考えると、よりよい視覚化を行うためには、(1) 読み込むデータ数を増やし、(2) 階層性を考慮した術語の選定を行うと同時に、(3) 多義語をうまく処理し、(4) リンク数の制限や、向き、太さを考慮し、(5) 使いやすいユーザーインターフェースを設計する必要がある。

読み込むデータ数を増やすためにウェブ上に公開されているデータを寄せ集めたとしても集まるのは今回用いたデータ量のせいぜい 10 倍だろう。それに対し、ノイズの少ない視覚化のために

求められるのは100~1000倍のデータ量である。電子データとして利用可能な教科書や専門書のデータセットの存在が必要であり、プロジェクト・グーテンベルクやGoogleの書籍デジタル化プロジェクトなどの利用を考えたい。術語の選定は今回手動で行ったために基準にむらがあり、また作業量も多かった。PNASLinkのように、Medical Subject Headingsが提供するMesh termsを利用するなど外部の辞書を利用するのが現実的であろう。また、Mesh termsではXMLを用いた様々なタグが用意されており、これを用いて解像度を定義していくことができるかもしれない。ただし、術語の辞書ができたとしても、スペルだけから文章中の熟語の意味を確定することはできないこともある。例えば“solution”という語は、化学の分野では「溶液」を意味するが、数学では「解」であり意味が全く異なる。文脈に依存した術語の処理を実装する必要があるが、これはかなり難しいため専門化の力を借りるべきだろう。アンケートにより指摘されることが多かったのが「術語同士がただつながっているだけでは意味がつかみづらい」ということだった。複数の閾値を設けてリンクの太さを変更する、全体でのリンクの数を制限する、 C_i , C_j の大小を用いてリンクに向きを設けるなどの改善が可能だろう。また、視覚化すること自体はおもしろいのだが、実際の視覚化(特にPNASLink)はあまり見やすくはないという意見も多く寄せられた。術語の繋がりを単に表示するのではなく、その術語に対応する画像を配置する、解像度を自由に変更可能といったインターフェースの改善を行っていく必要がある。

そしておそらく、この適切なインターフェースがどのようなものを設計することが一番重要である。言い換えるならば、どのように視覚化ができるのかではなく、どのように視覚化すると知識獲得に役立つかをデザインしていくということである。手法に捕らわれてしまい、この当たり前のことが抜けていたように思う。科学計量学は適切にデザインされたインターフェースを実現するための一手段であることを忘れてはならない。

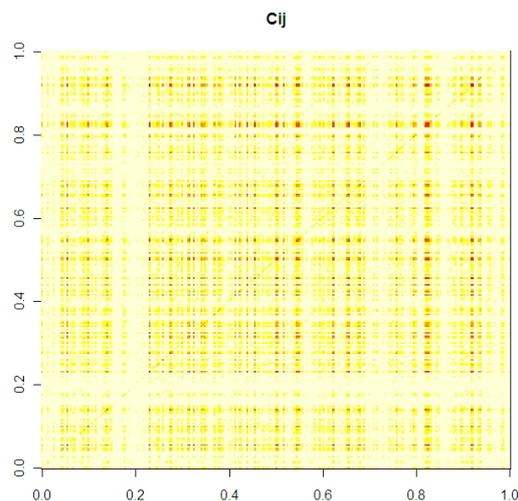


図 13: 単語間の距離行列を視覚化したもの。縦軸横軸には245の単語が並び、交差する点の色が単語間の距離を表す。赤だと近く、白だと遠い。

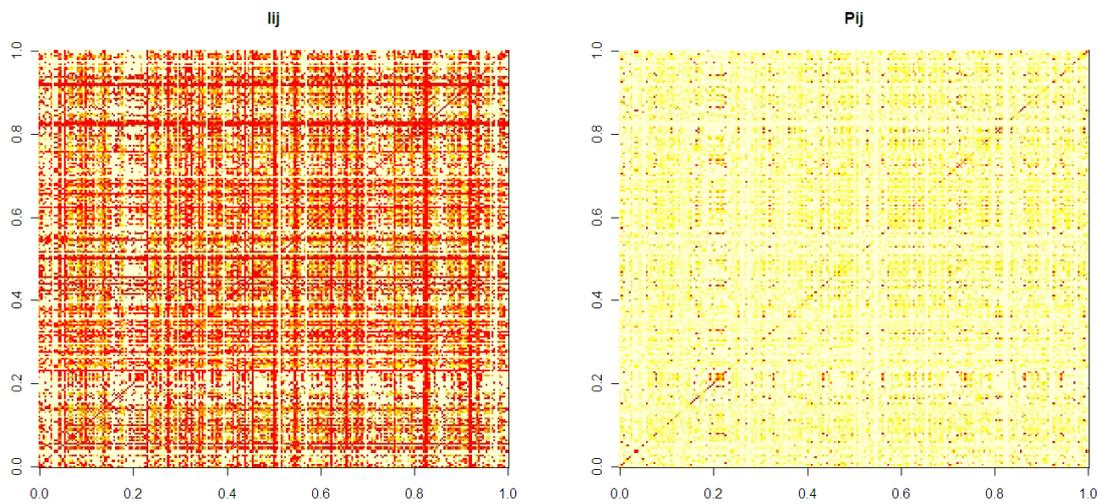


図 14: 単語間の距離行列を視覚化したもの。包含指標 (左) と近接指標 (右)。縦軸横軸には 245 の単語が並び、交差する点の色が単語間の距離を表す。赤だと近く、白だと遠い。

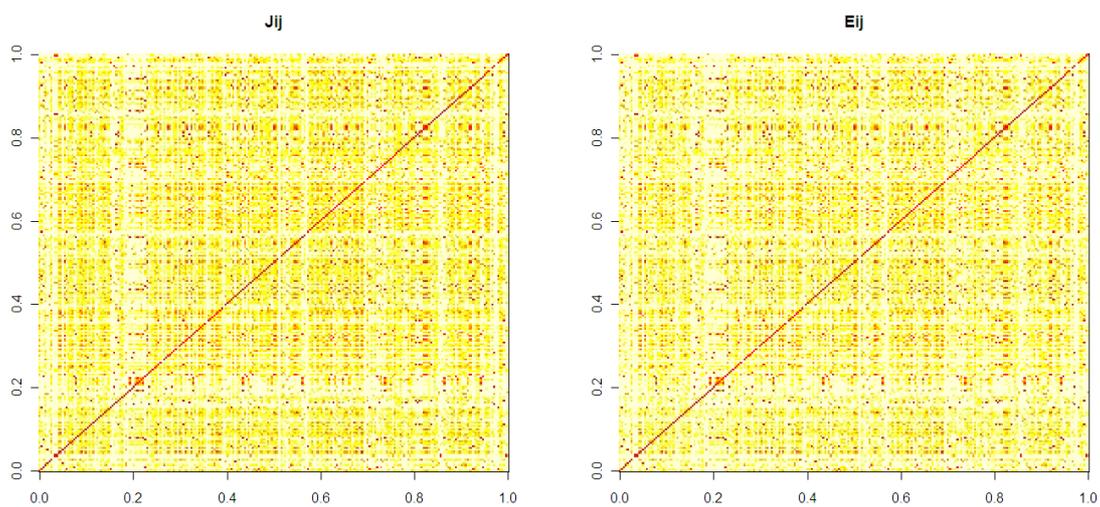


図 15: 単語間の距離行列を視覚化したもの。Jaccard 指標 (左) と同等性指標 (右)。縦軸横軸には 245 の単語が並び、交差する点の色が単語間の距離を表す。赤だと近く、白だと遠い。

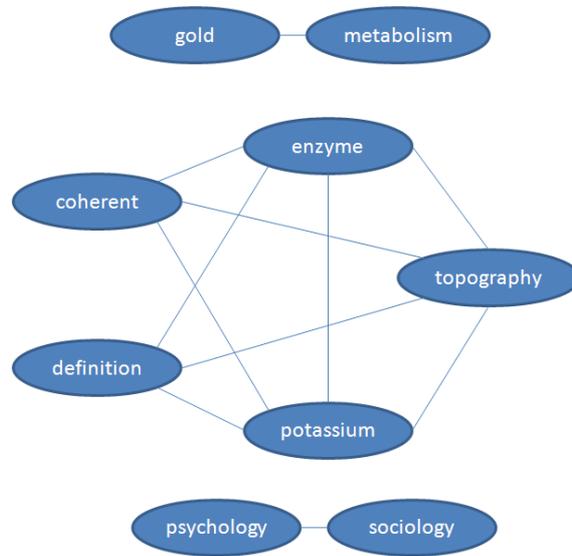


図 16: 近接指標を用いて描いた科学リテラシーマップ。少数の語が密に結合したクラスターが複数生じる。

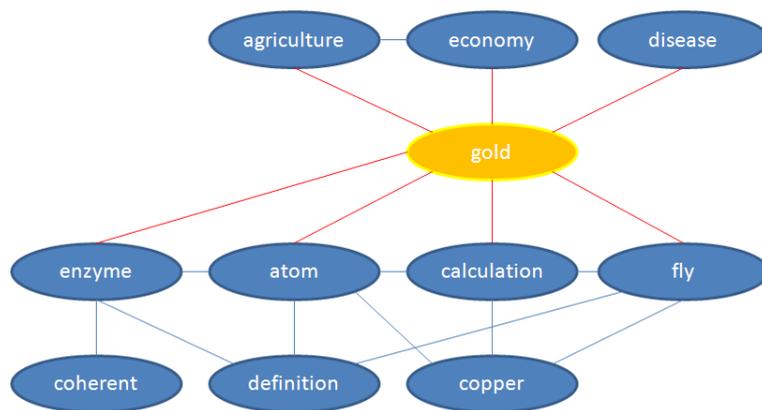


図 17: 包含指標を用いて描いた科学リテラシーマップ。あまり関連性が高いとは思われない語同士であってもリンクが存在する。その傾向は“gold”で特に顕著である。

表 1: 術語のリスト

agriculture	crossbreeding	ethics	iron	orbit	sensor
alchemy	crystal	evolution	isotope	organism	sex
alcohol	death	experiment	jupiter	origin	simulation
algebra	decent	eye	language	oscillator	skepticism
algorithms	decimal	feedback	linear	oxygen	sociobiology
antibiotics	decode	fermi	literacy	ozone	sociology
antibodies	decomposer	fiber	magnesium	pasteur	sodium
aristotle	decomposition	flux	mammals	pest	software
astrology	deduction	fly	manipulation	philosophers	spectrum
astrophysics	definition	formulas	map	philosophy	sperm
atom	deletions	fossil	material	physicians	spin
bacteria	descendants	fuel	mathematics	physics	statistic
battery	deert	fungi	measurement	physiology	stomach
behavior	desertification	galaxy	medicine	plutonimu	sulfur
biotechnology	destruction	galileo	membrane	pollen	sun
bird	development	gasoline	memory	pollution	symbol
blood	diameter	geology	mendel	potassium	symmetry
brain	dimensions	geometry	metabolism	predator	symptoms
caffeine	disaster	glod	metal	prediction	synthesis
calcium	disease	graph	microorganism	prey	system
calculation	disturbance	hazards	microscope	primate	technology
cancer	diversification	health	microwaves	principle	tectonics
carbon	diversity	heart	mineral	probability	telescope
catastrophe	dna	helium	momentum	probe	television
cell	doctor	heredity	moon	protein	temperature
chaos	dynamite	history	motor	protons	topography
chemistry	earth	homicides	muscle	psychology	transition
chromosomes	earthquake	hydrogen	mutation	recombinations	ultraviolet
climate	eclipses	hypothesis	nail	recycle	universe
coalition	economy	ice	nature	replication	uranium
coheret	ecosystem	immunity	neon	resource	uranus
compass	einstein	industry	nerve	risk	vegetation
complexity	elements	inertia	networks	robot	velocity
concepts	embryo	infection	neutrons	rockets	venus
computer	encoding	infrared	newton	rodents	watt
computation	energy	inheritance	nitrogen	salt	wave
conservation	engine	innovation	nuclear	science	weapons
convection	envelope	insect	number	scientist	zinc
vonvergence	environment	intelligence	nutrients	sea	zoology
convex	enzymes	invasion	offspring	selection	zoo
copper	equation	ions	optimum	semiconductors	

5 プログラムを受講して

5.1 プログラムを通して影響され、変化したこと

インタープリター養成プログラムでのこれこれという体験からこんなことを学び、その成果は今後の自分の研究にも活かされていくだろう、とでも答えられればよいのかもしれない。たしかに、少しはプレゼンや文章がうまくなったし、「伝えるためには伝える対象について十二分に知っておく必要がある」だとか「伝えるではなく伝わる」といったノウハウを身につけることができたが、それらは些細なことにすぎず、もっと大きく成長したという実感がある。しかしそれが具体的にになにか、と聞かれてもうまく言葉にできない。

5.2 今後の抱負

科学知識獲得を支援する視覚化手法の開発は今後も続けていきたい。科学コミュニケーション活動として大きなイベントを開催したり、今現在行っているよりももっと積極的に情報発信を行っていったりということはあまり考えていないが、今後の研究や研究に限らない場での「ものの見方」のすべてが、インタープリターとしての活動である。

参考文献

- [1] T. Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 2000.
- [2] L. Leydesdorff and I. Rafols. A global map of science based on the isi subject categories. *Journal of the American Society for Information Science and Technology*, 60(2):348–362, 2009.
- [3] F. J. Rutherford and A. Ahlgren. *Science for All Americans*. Oxford University Press, 1991.
- [4] H. D. White, X. Lin, J. W. Buzydlowski, and C. M. Chen. User-controlled mapping of significant literatures. *Proceedings of the National Academy of Sciences of the United States of America*, 101:5297–5302, 2004.
- [5] 三中信宏. 大統計大マンダラ, 1995.
- [6] 藤垣裕子, 平川秀幸, 富澤宏之, 調麻佐志, 林隆行, 牧野淳一郎. 研究評価・科学論のための科学計量学入門. 丸善株式会社, 2004.
- [7] 藤垣裕子, 廣野喜幸. 科学コミュニケーション論. 東京大学出版会, 2008.
- [8] 門司正三. 生態学総論. 生態学講座. 共立出版, 1976.

Appendix

以下の資料を附す。

- 解析用プログラム
- アンケート

解析用プログラム

```
#####  
#wordmetric.R  
#created 20100125 by Naoto HORIBE  
#####  
library(tm)  
  
dir()  
SFAA=Corpus(DirSource("./SFAA/"))  
meta(SFAA,tag="TOC",type="local")=c(  
"INTRODUCTION",  
"THE NATURE OF SCIENCE",  
"THE NATURE OF MATHEMATICS",  
"THE NATURE OF TECHNOLOGY",  
"THE PHYSICAL SETTING",  
"THE LIVING ENVIRONMENT",  
"THE HUMAN ORGANISM",  
"HUMAN SOCIETY",  
"THE DESIGNED WORLD",  
"THE MATHEMATICAL WORLD",  
"HISTORICAL PERSPECTIVES",  
"COMMON THEMES",  
"HABITS OF MIND",  
"EFFECTIVE LEARNING AND TEACHING",  
"REFORMING EDUCATION",  
"NEXT STEPS")  
  
match00=TermDocMatrix(SFAA[1])  
match01=TermDocMatrix(SFAA[2])  
match02=TermDocMatrix(SFAA[3])  
match03=TermDocMatrix(SFAA[4])  
match04=TermDocMatrix(SFAA[5])  
match05=TermDocMatrix(SFAA[6])  
  
match06=TermDocMatrix(SFAA[7])  
match07=TermDocMatrix(SFAA[8])  
match08=TermDocMatrix(SFAA[9])  
match09=TermDocMatrix(SFAA[10])  
match10=TermDocMatrix(SFAA[11])  
  
match11=TermDocMatrix(SFAA[12])  
match12=TermDocMatrix(SFAA[13])  
match13=TermDocMatrix(SFAA[14])  
match14=TermDocMatrix(SFAA[15])  
match15=TermDocMatrix(SFAA[16])  
  
wordlist00=colnames(match00)  
wordlist01=colnames(match01)  
wordlist02=colnames(match02)  
wordlist03=colnames(match03)  
wordlist04=colnames(match04)  
wordlist05=colnames(match05)  
  
wordlist06=colnames(match06)  
wordlist07=colnames(match07)  
wordlist08=colnames(match08)  
wordlist09=colnames(match09)  
wordlist10=colnames(match10)  
  
wordlist11=colnames(match11)  
wordlist12=colnames(match12)  
wordlist13=colnames(match13)  
wordlist14=colnames(match14)  
wordlist15=colnames(match15)  
  
termlist=read.table("termlist.txt",sep="\n",colClasses = "character")
```

```

mat_appearance=matrix(ncol=length(termlist[,1]),nrow=16)

for(i in 1:length(termlist[,1])){
  mat_appearance[ 1,i]=length(grep(termlist[i,],wordlist00))
  mat_appearance[ 2,i]=length(grep(termlist[i,],wordlist01))
  mat_appearance[ 3,i]=length(grep(termlist[i,],wordlist02))
  mat_appearance[ 4,i]=length(grep(termlist[i,],wordlist03))
  mat_appearance[ 5,i]=length(grep(termlist[i,],wordlist04))
  mat_appearance[ 6,i]=length(grep(termlist[i,],wordlist05))
  mat_appearance[ 7,i]=length(grep(termlist[i,],wordlist06))
  mat_appearance[ 8,i]=length(grep(termlist[i,],wordlist07))
  mat_appearance[ 9,i]=length(grep(termlist[i,],wordlist08))
  mat_appearance[10,i]=length(grep(termlist[i,],wordlist09))
  mat_appearance[11,i]=length(grep(termlist[i,],wordlist10))
  mat_appearance[12,i]=length(grep(termlist[i,],wordlist11))
  mat_appearance[13,i]=length(grep(termlist[i,],wordlist12))
  mat_appearance[14,i]=length(grep(termlist[i,],wordlist13))
  mat_appearance[15,i]=length(grep(termlist[i,],wordlist14))
  mat_appearance[16,i]=length(grep(termlist[i,],wordlist15))
}

mat_appearance[mat_appearance>0]=1
Ci=colSums(mat_appearance)
Cij=matrix(ncol=length(termlist[,1]),nrow=length(termlist[,1]))

for(i in 1:length(termlist[,1])){
  for(j in 1:length(termlist[,1])){
    Cij[i,j]=sum(mat_appearance[,i]*mat_appearance[,j])
  }
}

Iij=matrix(ncol=length(termlist[,1]),nrow=length(termlist[,1]))
Pij=matrix(ncol=length(termlist[,1]),nrow=length(termlist[,1]))
Jij=matrix(ncol=length(termlist[,1]),nrow=length(termlist[,1]))
Eij=matrix(ncol=length(termlist[,1]),nrow=length(termlist[,1]))

for(i in 1:length(termlist[,1])){
  for(j in 1:length(termlist[,1])){
    Iij[i,j]=Cij[i,j]/min(Ci[i],Ci[j])
    Pij[i,j]=Cij[i,j]/(Ci[i]*Ci[j])
    Jij[i,j]=Cij[i,j]/(Ci[i]+Ci[j]-Cij[i,j])
    Eij[i,j]=Cij[i,j]^2/(Ci[i]*Ci[j])
  }
}

Cij[is.nan(Cij)]=0
Iij[is.nan(Iij)]=0
Pij[is.nan(Pij)]=0
Jij[is.nan(Jij)]=0
Eij[is.nan(Eij)]=0
Cij[is.na(Cij)]=0
Iij[is.na(Iij)]=0
Pij[is.na(Pij)]=0
Jij[is.na(Jij)]=0
Eij[is.na(Eij)]=0

windows();image(1-Cij,main="Cij")
windows();image(1-Iij,main="Iij")
windows();image(1-Pij,main="Pij")
windows();image(1-Jij,main="Jij")
windows();image(1-Eij,main="Eij")

Iij[Iij[,]!=1]=0
image(1-Iij,main="thresholding(Iij)")
write.table(Pij,file="Iij.txt")

```

“科学の地図” の評価に関するアンケート

東京大学科学技術インタープリター養成プログラムの研究として、学問領域や専門用語同士の関連を視覚化することの効果について調べています。視覚化することが知識の整理や獲得の支援となるかどうかを検討するため、ぜひ意見をお聞かせください。アンケートは匿名式です。統計処理の結果を学術目的で公開させていただきますが、個別の回答を公表することはありません。

東京大学大学院総合文化研究科
科学技術インタープリター養成プログラム
堀部直人
nhoribee@08.alumni.u-tokyo.ac.jp

1. Map of Science

<http://users.fmg.uva.nl/lleydesdorff/map06/index.htm> にアクセスしてください。Leydesdorff による Map of Science が表示されるはずですが、この Map of Science は、14 の学問領域を関連の近さ、領域の活発さをグラフとして表示したものです。この図を見て以下の質問にお答えください。

1.1. この図を見て「おもしろい」と思いましたか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.2. この図にある学問領域の整理は、あなたの知識やイメージから判断して妥当だと思いますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.3. この図をみることで、知識が整理された/されると感じますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.4. この図をみることで、新しい発見があったと感じますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

次に、あなたの専門とする学問領域の一つを選びクリックしてください。より詳細な学問領域について地図が表示されます。先ほどと同様以下の質問にお答えください。

1.5. あなたが選んだ学問領域を教えてください _____

1.6. この図を見て「おもしろい」と思いましたか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.7. この図にある学問領域の整理は、あなたの知識やイメージから判断して妥当だと思いますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.8. この図をみることで、知識が整理された/されると感じますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.9. この図をみることで、新しい発見があったと感じますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

今度は、あなたとあまりなじみのない領域の一つを選びクリックしてください。再び以下の質問にお答えください。

1.10. あなたが選んだ学問領域を教えてください _____

1.11. この図を見て「おもしろい」と思いましたか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.12. この図にある学問領域の整理は、あなたの知識やイメージから判断して妥当だと思いますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.13. この図をみることで、知識が整理された/されると感じますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

1.14. この図をみることで、新しい発見があったと感じますか？

- そう思う まあそう思う あまりそう思わない 全くそう思わない

2. PNASLINK

<http://project.cis.drexel.edu/pnas/> にアクセスし、loginに“mapping”、passwordに“experiment”と入力してください。あなたの専門とする分野から学術用語を一つ選び入力し、search in: で MeSH term にチェックを入れ submit をクリックしてください。術語間の関連が表示されます。この結果について以下の質問にお答えください。（※入力する術語は Mesh term に登録されている語である必要があります。表示されない場合は他の術語で試してください。例えば Evolution, DNA, Brain などが有効です。左のカラムの Hot Topics も参照ください。）

2. 1. 入力した術語を教えてください。 _____

2. 2. この術語についてあなたは十分に習熟していると思いますか。

そう思う まあそう思う あまりそう思わない 全くそう思わない

2. 3. この図にある術語間の関連は、あなたの知識やイメージから判断して妥当だと思いますか？

そう思う まあそう思う あまりそう思わない 全くそう思わない

2. 4. この図をみることで、知識が整理された/されると感じますか？

そう思う まあそう思う あまりそう思わない 全くそう思わない

2. 5. 表示された術語のうち意味を含めて知っている、聞いたことはあるが詳しくは知らない、聞いたことがない術語の数をそれぞれ教えてください。

十分知っている _____個 聞いたことがある _____個 聞いたことがない _____個

2. 6. この図をみることで、新しい発見があったと感じますか？

そう思う まあそう思う あまりそう思わない 全くそう思わない

2. 7. 他の語でも試してみたいと思いますか？

そう思う まあそう思う あまりそう思わない 全くそう思わない

次に、あなたの専門としない分野の学術用語を選び、同様のことを行い、以下の質問にお答えください。

2. 8. 入力した術語を教えてください。 _____

2. 9. この術語についてあなたは十分に習熟していると思いますか。

そう思う まあそう思う あまりそう思わない 全くそう思わない

2. 10. この図にある術語間の関連は、あなたの知識やイメージから判断して妥当だと思いますか？

そう思う まあそう思う あまりそう思わない 全くそう思わない

2. 11. この図をみることで、知識が整理された/されると感じますか？

そう思う まあそう思う あまりそう思わない 全くそう思わない

2. 12. 表示された術語のうち意味を含めて知っている、聞いたことはあるが詳しくは知らない、聞いたことがない術語の数をそれぞれ教えてください。

十分知っている _____個 聞いたことがある _____個 聞いたことがない _____個

2. 13. この図をみることで、新しい発見があったと感じますか？

そう思う まあそう思う あまりそう思わない 全くそう思わない

2. 14. 他の語でも試してみたいと思いますか？

そう思う まあそう思う あまりそう思わない 全くそう思わない

3. 最後に、差し支えない範囲で以下の質問にお答えください。

3.1. あなたの専門分野を教えてください。 _____

3.2. あなたの学年 or 最終学歴を教えてください。 _____

3.3. “科学の地図”の可能性や図を見て感じたことなどがあれば自由にご記入ください。

ご協力どうもありがとうございました。