

令和 4 年 6 月 16 日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K12711

研究課題名(和文)大規模オントロジー間不整合の検出分析手法及び比較並列表示による可視化手法の開発

研究課題名(英文)Development of a parallel display method to visualise inconsistencies between large scale ontologies

研究代表者

伊藤 希 (Ytow, Nozomi)

筑波大学・生命環境系・講師

研究者番号：90251016

交付決定額(研究期間全体)：(直接経費) 1,500,000円

研究成果の概要(和文)：インターネット上に散在する様々な知識を集合知として活用する技術の一つに「オントロジー」がある。情報学でいう「オントロジー」とは、ある分野での用語をその用語があらわす概念に基づいて体系的に整理したものである。概念やその体系は、用語の利用者により異なり、また時代に応じ変化してゆくものであり、この結果として生ずる複数の「オントロジー」を比較する手法が集合知の活用には必要である。本研究では「オントロジー」間の比較検討を目的とした視覚化技術として、個々の概念は似通っているが体系が大きく異なる場合を扱う手法を開発した。

研究成果の学術的意義や社会的意義

生物多様性が急速に失われつつあると言われる一方で、生物多様性の評価に必要な情報は長らく専門家だけのものであった。地球規模生物多様性情報機構(GBIF)による生物分類情報の集積と公開により生物多様性データへのアクセスは一般に開放されたが、その集積期間はリンネ以降260年に及び、市民が利用するには整合性のとれた要約を提示する必要がある。本研究で開発された手法はそうした提示にを支援し、ひいては市民レベルでの生物多様性評価とその保護活動へと道を開くものである。

研究成果の概要(英文)：Ontology in sense of informatics is one of the techniques gathering knowledge scattered across the Internet into collective knowledge. It is a systematic organisation of terms based on the concepts they represent, which may differ by either user or organiser of terms or development of concepts which may result in different organisations. A visualisation method was developed to compare multiple ontologies, especially composed of terms of similar concepts organised into different hierarchies.

研究分野：生物多様性情報学

キーワード：生物多様性情報 オントロジー 可視化 不整合検出 並列表示

## 1. 研究開始当初の背景

本研究の核心は、「あいまいさを許容した不完全な概念を論理的に扱うにはどの様に表現すればよいか、その結果を視覚化する本文にはどうするのが良いのか」という「問い」にある。概念体系の具体例として生物分類体系を扱う。リンネを祖とする現在の生物命名法に基づいた生物分類体系は、260年以上に渡り数多の分類学者によって分散並列的に構築されてきた200万を超える概念を擁するオントロジーである。生物学上の全ての情報はその対象とする生物の指定に、この学名オントロジーに依存している。生物分類留意事項体系や、その基礎となる生物標本についての情報はこれまで博物館等に保管されそれへのアクセスも限られたものであったが、近年地球規模生物多様性情報機構(GBIF)に代表される様々な活動によりその電子化と公開が精力的に進められており、最近になって公開されたWeb APIによりGBIFが蓄積している生物標本データや分類体系データに容易にアクセスできるようになった。これにより、情報学的な意味での生物分類群オントロジー構築が現実的なものとなりつつある。

生物標本データを扱う上で分類群に関する情報は基本的かつ必須なものであるが、標本の蓄積が長きに及び一方で分類群や分類体系はそれらの標本を用いた研究により徐々に移り変わっており、分類群の指示範囲に一貫性は保証されていない。一方、標本データの利用という観点からすれば、その際どの分類群を用いるのが望ましいかは利用者の様態に依存する。分類学者や生態学者と政策決定者や一般市民では異なり得るし、地域的にも異なり得る。これらの需要を満たす上で必要となるのは、一つの「正しい」分類体系を用いて標本情報を整理することではなく、複数の分類体系間でのマッピングであり、その前段階としての複数分類体系の整合性に配慮した比較提示である。分類体系の比較は、分類学的再検討の一環として、分類学者により比較的小さな分類群について行われてきた。その際にデータベースを積極的に用いた研究は皆無ではないが、そこでは形質状態等を含む詳細な情報が用いられるとともに、従来、分類体系の比較は、分類学的再検討の一環として分類学者により手作業的に小規模な分類群について行われてきた。こうした比較をGBIFが提供する様な大規模データについて従来同様手作業で行うのは非現実的であり、機械的な処理による、完全ではなくともある程度実用的な比較提示を可能とする必要がある。そうした比較にあたっては、比較対象となる分類体系が「不完全」であることにも配慮する必要がある。ここで言う「不完全」とは、たとえば、時間的にかなり間を置いて発表された二つの分類体系を考えた場合、早い方の分類体系の発表後に発見された新種は後から発表された体系にしか含まれておらず、遅い方の分類体系を基準として考えると早い方の体系は「不完全」である様に見える、という意味である。しかし、このことは、新旧の体系の観点の違い、粒度の違いを排除しない。先に発表された分類体系では、後の「新種」を独立した種として認めない立場である可能性もある。その場合、先に発表された体系は不完全なのではなく、後に発表された体系とは異なる粒度を有しているということになる。こうした場合、分類学者の手作業的な比較では、どちらかの立場か、あるいはその比較を行なう者の立場の、いずれかの観点到コミットした比較が行なわれるが、機械的比較でいずれかの観点を選ぶのは現実的ではない。「不完全」さを許容しつつ、あるがままの比較を行ない得る手法が必要である。

## 2. 研究の目的

分類体系の比較結果の提示方法はいくつか試みられているが、分類学者への詳細な情報提供を意図したもの、あるいはデータビジュアライゼーションに関する研究を主眼としたものであり、操作の習熟に必要な期間を考えると、前述の様な多様な利用者が容易に使えるものではない。

本研究は、こうした、分類体系情報の需要と供給にみられるギャップを埋める手法についての検証を通じ、大規模「オントロジー」にも適用可能な、操作者に高度な習得を要求しない、直感的な体系間比較を可能とする手法を提供することを目的としたものである。特に、体系間に共通する類似の概念に着目した際に、その上位の構造が大きく異なる場合の体系並列表示に必要な手法開発を目標とした。

## 3. 研究の方法

生物多様性データに基づく体系の比較表示は、コンピュータスクリーンを用いて行われてきた。直感的な理解・把握の方法としては、こうした視覚的なものの他に、操作を通じた理解がある。たとえば、仮想空間中のオブジェクトに対する操作を通じた理解がこれに該当する。こうした方式を用いるものとして仮想現実空間を利用した体系構造のツリー表示による比較が考えられるが、これには仮想空間として没入環境を提供する施設が必要とされ、広く利用できる状況にはなかった。本研究開始年度に付加設備を全く必要としない、オールインワンタイプの仮想現実ゴーグルが一般に提供されるようになったのを受け、当該ゴーグルを用いた体系比較の可能性を検討した。具体的には使用感ならびにプログラミング環境、アプリケーション開発に必要な要素

などを評価した。これとあわせ、平面ディスプレイによる体系並列表示方法の開発もプログラミング言語 Java により行なった。

#### 4. 研究成果

オールインワンタイプの仮想現実ゴーグルの設定は容易であり、一般的な利用そのものには十分耐えるものであった。価格的にも博物館や研究機関で利用するには十分現実的なものである。表示される仮想現実空間オブジェクトの操作感は、ユーザが対象オブジェクトを軽いものと認識する場合には違和感がなかった。たとえば、仮想現実空間内で「紙飛行機を飛ばす」操作は違和感なく実行可能であった。一方、重量があると期待されるオブジェクトの操作には対象の「軽さ」による違和感があった。これは、フォースセンスのないコンシューマ用デバイスの限界であり、価格と機能のトレードオフを考えると止むを得ない部分である。しかし、ツリー構造の比較を行なう場合に、複数の概念体系を仮想空間のツリー構造として重ねて表示させたうえで、体系間の不整合性をツリー構造の絡まり具合として認識させるといった利用を考えると、抵抗感に相当するフォースフィードバックは必須であり、そうした利用には機能的に不十分であった。表示装置の分解能は、スペック的には現在普通に用いられているパーソナルコンピュータと同程度であるが、文字表示が多い場合には読み取りにくい場合が懸念された。一方、プログラミング環境としては、ネイティブ開発も不可能ではないのものの、ゲームプログラミング用システムの利用が推奨されている。ゲームプログラミングの学習曲線は急峻であり、「複数概念体系の表示とその操作的比較」というタスクをゲームプログラミングの文脈に落とし込む必要がある。総じて、表示装置として実用的ではあるが、概念体系比較というタスクそのものに用いるにはメリットとコストのバランスが悪いと評価された。生物多様性情報関連で仮想現実ゴーグルを利用するならば、一部で行なわれている博物館標本の仮想化といった動きと連動させた用途を検討の方が良いであろう。そうした用途での体系の比較表示には、仮想空間に直接表示させたツリー構造を利用するよりも、ポップアップさせた仮想平面を用いる方が、標本表示との混乱を避けられ望ましい。

ツリー構造を用いた概念体系の並列表示の際に、相互に対応する概念を表現するツリー構造中のノード（概念ノード）を整列させるために利用されている「整列ツリー」を拡張し、上位構造が整合しない体系中の複数概念を整列させるためのツリー構造を実装した。ツリー構造の上位下位に対応する上下と表示画面上の上下を区別するため、以下ではツリー構造の上位下位については高さ深さ、表示画面については画面上を北とする東西南北の方位で表現する。体系比較の際に着目する概念ノードを基準に、各比較対象体系から着目概念ノードの高さ方向のノード列を抽出して集約し、着目概念ノードを起点とした、元のツリー構造とは高さ深さが逆転したツリー構造を得た。この反転ツリーを、着目概念ノードから深くなる方向に対応する整列ツリーの北側に隣接させて南北逆転させた形で表示することで、ノードの南北方向の並びと元の比較対象ツリーでの南北方向のノード並びが一致した拡張整列ツリーを得た。左から右へ書く言語の場合、ディレクトリツリー構造のアナロジーを利用した画面表示では、ツリーの深さを西から東へのインデント量によって表現しており、同じ深さであれば東西の位置も同じになる様になっている。ノード間の親子関係は、親ノードの南側に接する、西側に一段シフトした子ノードの列として表現される。反転ツリーではこの南北が逆転するため、単にノードを表示しただけでは構造がわかりにくくなる。そのため、反転ツリーを通常のツリーに隣接させて表示する場合は、ノード間を繋ぐ線分などによって親子関係を明示する必要がある。反転ツリーでは西側にシフトすると深さが減る方向であり、その点も混乱の原因となり得る。反転ツリーをさらに東西方向にも反転させると深さ方向と東西方向が統一的になり理解しやすくなるが、親子関係を表現する接続線を反転ツリーで交差させずに表示させるためには、反転ツリーでは接続線をノードの東側に表示する必要がある。このため、通常ツリーでは同じ深さのノードは西端が揃うのに対し、東西南北二重反転ツリーではノードの東側が深さ方向に対応することになる。南北反転ツリーを用いた場合と東西南北二重反転ツリーを用いた場合を比べると、二重反転ツリーの方が違和感は少なく、南北反転ツリーを用いた表示は異質な印象を与えた。着目ノードの上位体系の表現としては、上位概念の名称によるもの（ヒトに対する脊椎動物）とその属性によるもの（「背骨がある」）がある。前者には二重反転ツリーが適している一方、後者に対しては南北反転ツリーの「異質さ」を積極的に利用することが可能である。反転ツリーと二重反転ツリーを東西に並列表示することで、概念の対象と属性を操作整合的に表示することも考えられる。

複数概念体系の並列表示手法として整列ツリーに反転ツリーを用いる方法を検討していたが、単なる整列ツリーにとどまらず、概念の属性表示にも利用可能であることが判明した。形式概念などの概念構造を表示する様々な場面への応用が期待される。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<https://github.com/nomenclator/taxonaut> 開発したプログラムのソースコード

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------