

氏 名 Rathachai Chawuthai

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 1887 号

学位授与の日付 平成28年9月28日

学位授与の要件 複合科学研究科 情報学専攻  
学位規則第6条第1項該当

学位論文題目 Linked Open Data-based Knowledge Management Process  
for Biodiversity

論文審査委員 主 査 教授 武田 英明  
教授 佐藤 健  
准教授 北本 朝展  
准教授 大向 一輝  
准教授 市瀬 龍太郎  
主任調査員 川村 隆浩 科学技術振興機構  
(電気通信大学 客員准教授)

論文内容の要旨  
Summary of thesis contents

Thanks to the advancement of Linked Open Data (LOD) technology, it enables data around the world to become explicitly and implicitly exchangeable through the Internet, so the motivation to empower the capability of Knowledge Management (KM) for global accessibility comes to be achievable. However, the shift is not just simply replacing a legacy storage with a graph database. There are challenging issues about how to deal with the change in knowledge, how to have human-/machine-readable knowledge graphs, how to realize that the structure of a knowledge graph is suitable for finding new knowledge, and how to have learners learn knowledge from a knowledge graph conveniently. Thus, this thesis takes this opportunity to study the role of LOD in any KM process through some key KM activities by using biodiversity as domain knowledge. For the Ph.D. course, knowledge capture, knowledge exchange, knowledge discovery, and knowledge presentation are studied; because they are common KM activities of and primarily found in any KM processes. To inform how importance the LOD in KM process is, three projects are introduced for the according activities.

Linked Taxonomic Knowledge (LTK) takes responsibility of capturing and exchanging of the knowledge by initiating an RDF data model to capture the change in biological taxonomy with appropriate context and publish the data with a simple and lightweight structure. This project introduces three types of taxonomic identifiers for light-weight representation. There are the nominal entity, the simple nominal entity, and the contextual nominal entity. The nominal entity includes taxon concepts and taxonomic names; the simple nominal entity is used for encapsulating a taxon concept and a scientific name within a single URI; and the contextual nominal entity represents a taxon concept, a scientific name, and an aspect of time within a single URI. This project also initiates three types of knowledge graphs that are the Event-Centric model, the Transition model, and the Snapshot model. The Event-Centric model is firstly created for capturing the change in taxonomy. It includes operations of change, relations between background knowledge of the changes, aspects of time, and references. Next, the Transition model is used to present the chronological change between taxa. Last, the Snapshot model presents temporal information of a taxon. The Event-Centric model is designed for presenting the change in taxonomy, so the model is complex by design but flexible for any other applications. Thus, the Transition model and the Snapshot model is designed to be light-weight representation for exchanging data with LOD cloud.

Link Prediction on Interspecies Interaction (LPII) is accounted for the knowledge discovery. By analyzing the knowledge graph of fungus-host interactions and then gives biologists a recommended list to discover more interspecies interactions. LPII uses three different knowledge graphs for making prediction, that are a bipartite graph, a projection network, and a taxonomy. This project introduces a hybrid

(別紙様式 2)  
(Separate Form 2)

recommender system including scoring functions based on Collaborative Filtering, Community Structure, and Biological Classification. These scoring functions work with the bipartite graph, the projection network, and taxonomy getting from LOD respectively. It has been found that the linear combination of three scoring functions is more accurate than other combinations, and some missing relation have been found from the new discovery of the National Museum of Nature and Science in Tokyo and some literatures from external publications.

Biodiversity Knowledge Graph Visualization (BViz) is in charge of the knowledge presentation by simplifying and rearranging a knowledge graph for delivering a human-friendly node-link diagram to users. BViz introduces three modules that are the Graph Simplification, the Triple Ranking, and the Property Selection. First, the Graph Simplification uses Semantic Web rules to merge some same-as nodes, removing inferred transitive links, and removing inferred type-hierarchy. Second, the Triple Ranking helps to rank all triples in the query graph from common information to topic-specific information. Last, the Property Selection allows users to display or hide some triples containing specific URIs or namespaces. A web application is implemented based on these methods in which all of them can be controlled by users with the interactive user interface. Thus learners can query a knowledge graph, simplify the graph, rearrange the graph from common information to topic-specific information, and select some parts of the graph based on user preference.

The results of these projects demonstrate that the combination of LOD and KM can address the issues and provides contributions to biodiversity domain by both publications and applications. Moreover, the knowledge and experience gained from this study are intended to be a guideline for creating LOD-based KM systems to any other domains.

本博士論文は、生物多様性に関わる研究におけるナレッジマネジメントのコンピュータによる支援に関する研究である。生物多様性に関わる研究者は観察等の研究活動を通じて知識を発展させる。コンピュータによる支援は、コンピュータによる知識の取り込み（知識獲得）、取り込んだ知識を多様に利用可能にする（知識交換）、この知識から新しい知識を発見する（知識発見）、コンピュータにある知識を人間に提示する（知識提示）といった機能が必要となる。本論文では、このナレッジマネジメントのプロセスを知識獲得、知識交換、知識発見、知識提示に分け、それぞれを **Linked Open Data (LOD)** に基づくシステムで支援することを提案している。博士論文は 8 章からなる。第 1 章と第 2 章では導入と関連研究について述べている。知識獲得、知識交換、知識発見、知識提示に関する研究が第 3 章から順に第 6 章である。第 7 章で議論を、第 8 章で結論を述べている。

第 3 章と第 4 章では、知識獲得と知識交換の支援についての研究である。LTK というシステムを提案し、タキソノミーの変化を表現するオントロジーとそれに基づくタクソンの LOD による表現方法を提示している。これによりタキソノミーの変遷を正確に記述できる。生物学におけるタクソンに関する知識は動的な地知識である。すなわち、新しい発見により変更される知識である。このような動的な知識をコンピュータで記述するために、**event-centric model** という知識の変遷を記述するモデルを提案している。さらに統合や分離といった変化を記述するためのオントロジーと、学名とシノニム等を一元的に扱うための **contextual nominal entity** という概念を導入している。さらに **transition model** と **snapshot model** という **event-centric model** から推論で導出する簡易モデルを用意することで、変遷のみの表現とある時点でのタキソノミーの表現が可能になっている。これらのモデルとオントロジーはすべて LOD の方法によって用意されている。このモデルを実際の特定分野のタキソノミーに適用することで有効性を示している。

第 5 章は知識発見についての研究であり、LPII というシステムにおいて生物間相互作用の発見をタキソノミーの利用を含む異なる 3 つのアルゴリズムの組み合わせを提案し、実データによる実験で効果があることを示している。生物間の相互作用は一般に関係は莫大な組み合わせが存在するが、実際に観察する機会は限られている。ここでは、既存の生物間相互作用のデータから機械学習の手法で未知の関係を予測するというを行っている。アルゴリズムとしては、協調学習、コミュニティ発見を利用した関係発見、タキソノミーを利用した関係発見の 3 つを用意した。菌類のデータに基づいて実験を行った。3 つのアルゴリズムを適切に組み合わせることが有効なことがわかった。とくに関係が密な部分では協調学習を、関係が疎な部分ではタキソノミーを利用した関係発見が有効なことがわかった。

第 6 章では知識提示についての研究であり、LOD のグラフをいかに提示するかを検討し、一般的な見方と専門的見方で異なる簡略化をする方法を提案し **BViZ** というシステムを実装している。LOD はグラフであるので、規模が大きくなると人間にとっては理解が難しくなる。そこでいかに提示する情報の量を減らすかということが課題となる。ここでは、**TFIDF** の考え方を援用し、データ全体に頻出する項目を中心に表示する一般的な見方と、注目している部分の中で他ではあまり見られない部分を提示する専門的見方の 2 つを考え、これによって提示するデータを変える手法を提案している。これによりシンプルなグラフ構造を提示することが可能になっている。

(別紙様式 3)

**(Separate Form 3)**

博士論文の内容は、1本の査読付きジャーナル論文、2本の査読付き国際会議論文、1本の査読付き国際会議ポスター・デモ論文などで発表されており、当該分野のコミュニティからも評価されている。以上より、本論文は博士論文として、十分な水準であると審査委員全員一致で認められた。