

(続紙 1)

京都大学	博士 (情報学)	氏名	周 慧珍 (Huichen Chou)
論文題目	Analyzing and Designing the Open Collaboration of Knowledge Content Creation (知識コンテンツ作成のためのオープンコラボレーションの分析と設計)		
(論文内容の要旨)			
<p>The goal of this thesis is to support the continuing development of knowledge contents created through open collaboration. This thesis considers different types of open collaboration involving different levels of variety in participants and different objectives of content output. When the collaboration involves a wide range of participants working together for a single knowledge content, the quality of the output often cannot be guaranteed. It is not fully understood how different teams yield knowledge content output of good quality. Therefore, this thesis presents methods to deepen understanding and extract collaboration patterns of quality knowledge content creation. For the collaboration where outputs are independent from each other but shared to allow others to use and to build on, copyright restrictions are a hindrance. To cope with this issue, this thesis presents a method to allow copyright-restricted knowledge contents to be shared and reused among participants of the open collaboration. This thesis consists of six chapters.</p> <p>Chapter 1 outlines the thesis including the objectives of this research, issues, and approaches.</p> <p>Chapter 2 discusses the existing problems in the different types of the open collaboration of knowledge content creation. We use Wikipedia as an example of calling people to work together on a single output, and teaching materials as an example of making use of existing knowledge contents to create independent outputs. This chapter first reports the previous research efforts of studying the quality issues in Wikipedia that focus on the factors including the number of editors and editing activities. Yet it is unclear how quality knowledge contents can be created from different teams of a variety of people. Therefore, it is difficult to extract collaboration patterns in quality knowledge content creation. To address the copyright issue in the open collaboration of teaching material creation, this chapter then explores several possible solutions including the use of blockchain technology. However, existing blockchain-based systems mainly focus on supporting the collaboration of creating a single knowledge content. The above limitations of existing work on creating knowledge contents via open collaboration motivate our research in this thesis.</p> <p>Chapter 3 explains the research of understanding how open collaboration that calls people online to work together can create quality knowledge contents. This research uses Wikipedia Good Articles (GA) as a case study and focuses on how different teams create quality knowledge contents by analyzing who contributes to the collaborative creation and how they are involved in the collaboration process. We first employ factor analysis to identify editing abilities and then use these editing ability scores to distinguish editors. Then, we generate a sequence of editors in the GA creation process to analyze the collaboration patterns. We report the analysis of three different categories of Wikipedia articles: “US state parks,” “children’s books,” and “chemical compounds and materials.” The result shows that there is a common collaboration pattern in most of the GA examples analyzed in this research. That is, prior to achieving GA nomination, an editor with strong content-shaping ability expands the article and continuously works on it. We also found that GA creation relies only slightly on the sheer number of editors as most editors contribute little to the editing activities. The findings deepen the understanding of how diversified teams create quality knowledge contents in the open</p>			

collaboration.

Chapter 4 proposes a method of extracting collaboration patterns of quality knowledge content creation and uses Wikipedia GA as a case study. The analysis in Chapter 3 revealed the existence of some common collaboration patterns in the Wikipedia GA creation. However, it is necessary to develop a general method to comprehensively discover various collaboration patterns to provide reference for creating more quality knowledge contents. In the proposed method, the dynamic time warping algorithm is first used to cluster articles based on the GA size evolution. Then, factor analysis is used to differentiate editors based on their editing activities. After that, the mean sequence of each cluster is calculated to identify the three phases that characterize the evolution of an article: growth, plateau, and decline. Finally, the composition of the active editors of each phase is incorporated to illustrate the overall collaboration patterns. The proposed method successfully identified the collaboration patterns in the three Wikipedia categories used in Chapter 3, which can be used as reference for creating more quality knowledge contents.

Chapter 5 describes a system design that supports the open collaboration of using existing knowledge contents to build into one's content with copyright sharing. The research considers teaching materials as an example of knowledge content sharing with copyright restriction and exploits the advantage of blockchain technology to support the collaboration. The proposed system bonds participants with a smart contract; it securely registers the records of the authorship of originally created teaching materials and the contribution distribution of collaboratively created teaching materials that partially reuses existing contents. Such records can be used as authorship evidences to claim economic benefits when a material is used. We implement our design on an Ethereum platform and use teaching material creation scenarios to demonstrate its effectiveness and practical potential in supporting the open collaboration that reuses copyright-restricted knowledge contents.

Chapter 6 concludes the thesis by summarizing the contributions to support the continuing development of knowledge contents created through open collaboration and suggesting possible future directions.

(続紙 2)

(論文審査の結果の要旨)

本論文は、持続的な知識コンテンツ作成支援において、不特定多数の作業によるコンテンツの作成と既存のコンテンツの再利用による新しいコンテンツの作成といった2種類のオープンコラボレーションを対象として、高品質なコンテンツを作成するためのコラボレーションパターンの分析と抽出、著作権のあるコンテンツを再利用するためのコラボレーションシステムの設計に関する手法を提案するものであり、得られた主要な成果は以下の通りである。

1. 高品質なコンテンツ作成のためのコラボレーションパターンの分析

不特定多数の作業による知識コンテンツ作成においては、コンテンツ品質の保証が重要な課題である。そこで、Wikipedia Good Articles (GA)を具体例として、不特定多数の作業が如何に高品質なコンテンツを作成しているか理解するための手法を提案した。具体的には、因子分析を用いて編集活動に基づき作業者を分類し、コンテンツ作成過程における各種の作業者の参加順序を生成し、コラボレーションパターンを分析した。また、3つのWikipedia GAカテゴリを用いた分析によって、多くのGAでは共通のコラボレーションパターンが存在することを示した。

2. 高品質なコンテンツ作成のためのコラボレーションパターンの抽出

知識コンテンツを持続的に作成するためには、高品質なコンテンツ作成に関するコラボレーションパターンを抽出し有効に活用することが重要である。そこで、Wikipedia GAを具体例として、コラボレーションパターンを網羅的に抽出する一般的な手法を提案した。まず、動的時間伸縮法を用いてコンテンツ作成過程のサイズ変化に基づきGAに対してクラスタリングを行い、Growth、Plateau、Declineといった3種類のフェーズの組み合わせにより各クラスタを表現した。次に、因子分析を用いて作業者を分類し、各GAクラスタのフェーズ毎の作業者構成状況に基づきコラボレーションパターンを抽出した。最後に、3つのWikipedia GAカテゴリに対して、提案手法により網羅的にコラボレーションパターンを抽出できたことを確認した。

3. 著作権のあるコンテンツの再利用のためのコラボレーションシステムの設計

既存の知識コンテンツの再利用による新しいコンテンツの作成においては、著作権のあるコンテンツを自由に再利用できないという問題が生じる場合がある。そこで、教材作成を具体例として、教材の著作権を保護するためのオープンコラボレーションシステムを設計し構築した。具体的には、著者の教材への貢献度を改ざんできないように、ブロックチェーン技術を用いて、教材作成と再利用に関するスマートコントラクトを設計しイーサリアムプラットフォーム上にシステムを構築した。また、構築したシステムの処理性能を検証し実用上問題ないことを確認した。

以上、本論文は、知識コンテンツ作成のためのオープンコラボレーションにおける、高品質なコンテンツ作成のためのコラボレーションパターンの分析と抽出、著作権のあるコンテンツの再利用のためのコラボレーションシステムの設計に関する手法を提案した結果をまとめたものであり、学術上・実応用上寄与するところが少なくない。よって、本論文は博士(情報学)の学位論文として価値あるものと認められる。また、令和4年5月19日、論文内容とそれに関連した事項について試問を行った結果、合格と認められた。また、本論文のインターネットでの全文公表についても支障がないことを確認した。

要旨公開可能日： 年 月 日以降