

知識創出を目指した集合知再構成手法の提案

川戸 祐介[†], 松村 敦[†], 宇陀 則彦[†]

[†]筑波大学大学院 図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2

E-mail: [†]{kawato, matsumur, uda}@slis.tsukuba.ac.jp

本研究は複数の関連する集合知の情報を再構成することで、単体の集合知の利用ではできなかった新しい知識創出を目指した。まず、ソーシャルブックマーク、百科事典作成、Q&A サイトの各集合知を利用する場合の問題点を指摘した。次に、複数の集合知を組み合わせ、これらの問題を解決可能な、回答の補完、百科事典の項目に対する意見付与、意味的な類似キーワードの付与の3例を示した。これらを実現するための、集合知の再構成手法を検討し、プロトタイプシステムの実装を行った。

Proposal of a Reconstruction Method of Collective Intelligence for Knowledge Creation

Yusuke Kawato[†], Atsushi Matsumura[†] and Norihiko Uda[†]

[†]Graduate School of Library, Information and Media Studies, University of Tsukuba

In this research we propose a new reconstruction method that intends to create a combination of knowledge from collective intelligence. At first, we examined some problems on use of single collective intelligence. To solve these problems, we put multiple collective intelligence together and reconstruct. We report this method and the prototype system.

1. はじめに

近年、インターネット回線の普及やパソコン、携帯電話を初めとしたインターネットへの接続機器の普及などにより、個人個人が自由に Web 上のコンテンツを入手したり、公開したりすることが活発に行われている。こうした動きの中で、集合知と呼ばれる新しい形成法によって構築されたコンテンツが注目を浴びている。集合知の特徴の一つとして、誰もが自由に知識を形成できるとい

う点がある。不特定多数の人がコンテンツを構成しているため、多種多様な個別の知識が蓄積され、その結果、個人で持つには、到底不可能な量になるとともに、それぞれが得意分野の知識を持ち寄ることで、質的にも高度な知識になる。

そこで、このような集合知の特徴をいかした、サービスが増えてきた。例えば、(1) 特定のウェブページに対して、ソーシャルブックマークサイトにおけるブックマーク数をまとめて表示すると

いったサービス[1], (2) ソーシャルブックマークを用いてサーチエンジンの検索精度を向上させる研究[2], (3) 1次資料とウェブ資源を結びつけるために Wikipedia を用いる研究[3]といったものがある。(1), (2) では, 集合知が持つ情報の一部を利用して既存のサービスの情報量, 質の向上を行っている。(3) では, 既存の知識と集合知の知識の融合を目指している。

しかし, これらの研究やサービスでは, 集合知に含まれるごく一部の情報を用いただけであり, 集合知が持つ多種多様な知識を十分に活用しているものとはいえない。そこで本研究では, 各集合知の性質を踏まえ, 複数の集合知を融合することで, 知識創出を目指した。

2. 集合知

2.1 定義

集合知とは, 優れた個人の意見よりも, 集団の意見を集約したものが正しい場合があるという, 群衆の知恵のことである。集合知としての群衆の知恵が形成されるための条件として以下の4点を挙げられる[4]。

- ・ 個が互いに独立していること
- ・ 個が分散していること
- ・ 個が多様であること
- ・ 個を集約する優れた仕組みがあること

インターネットにおける個とは, それぞれコンテンツを製作するユーザである。インターネットにおいて集合知が注目された理由には, コミュニケーションコストが極めて低く, 上記の条件を矛盾なく成立させ, 多数の参加者を集めるようなシステム設計が可能であるということがある[5]。本研究においても, これら4点を矛盾なく成立させて集合知を利用することを目指す。

2.2 集合知のタイプ

現在, 集合知と呼ばれるサービスは多く行われている。本研究では, Q&A サイト, 百科事典作成, ソーシャルブックマークという利用目的の異なる3つの集合知を再構成の対象とした。

・ Q&A サイト

Q&A サイトとは, 不特定多数の人に対して, 質問と回答を提供するサイトのことである。質問文に対して, 他のユーザは自由に回答し, これらの質問文や回答文が蓄積され, 公開されていくことで同様の疑問を持つ人にとって有用な情報とな

る。このようなシステムの代表例として, Yahoo! 知恵袋[6]やOKwave[7]などがある。Q&A サイトの特徴として, 各質問に対して, 人が直接答えることから, 質問の意図に沿った回答が得られる可能性が高いという点があげられる。しかし, 逆に答える人が決まっていないために, 必ず回答がつく保証はないのが問題点である。

・ 百科事典作成

百科事典型の集合知に Wikipedia[8]がある。Wikipedia では, 各ユーザが自由に項目や説明文, 関連項目などの入力, 修正を行うことができる。不特定多数によって構築されているため, 特に統制されることなく, 話題になった項目が自由に書き込まれる。その結果, 英語版では約 203 万語, 日本語版では約 42 万語という多くの見出し語が存在する(2007年10月現在)。しかし, その一方で, 不特定多数が自由に記入することが可能であるため, 専門家のみによって構築された従来の百科事典と比べると信頼性にかけてという問題が存在する。

・ ソーシャルブックマーク

ソーシャルブックマークは, インターネット上でブックマークを共有する仕組みである。ソーシャルブックマークを提供している Web サイトとして, はてなブックマーク[9]や, Buzzurl[10]などがある。これらのソーシャルブックマークではタグと呼ばれるキーワードを自由に付与することで組織化を行う。これによって, 共通のタグを介して関連する項目を調べることが可能である。反面, ユーザはタグを自由に付与することができるため, 語の統制がとれず, 同じ意味であってもタグが違えば, 関連した情報を辿ることができないという問題点がある。また, 共通のタグを所有する情報を調べることはできるが, 付与したタグの意味に沿った関連情報を調べることはできないといった問題点もある。

3. 集合知再構成手法

3.1 提案手法

以上の検討から, 本研究では, 各集合知の問題点を踏まえた再構成手法を提案する。各集合知における問題点を表1に示す。これらの問題点に対し, 本研究では, 他の集合知の情報を再構成し提示することで解決しようと考えた。

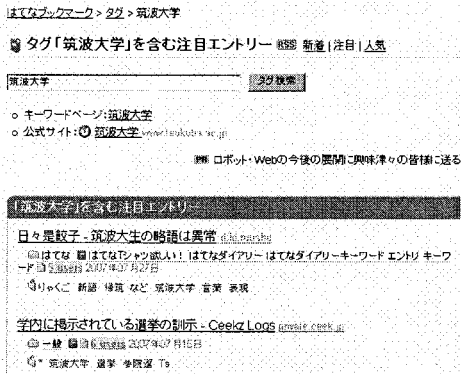


図5 「筑波大学」のソーシャルブックマークでの検索結果

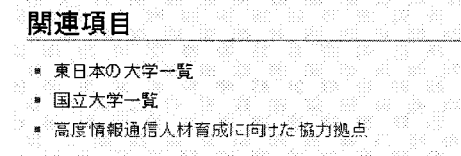


図6 百科事典の「筑波大学」の中の関連項目

以上3例について、再構成に利用した集合知をまとめたものを表2に示す。

4. 集合知再構成システム

4.1 システム概要

以上述べた再構成の具体例を実現するシステムについて述べる。本研究では、複数の集合知を用いて再構成を行う。このとき、具体例で示した3例では各集合知に含まれる情報全てを用いたわけではなく、一部の情報を用いて再構成を行う。

表2 具体例と再構成に用いる集合知

	ソーシャルブックマーク	百科事典作成	Q&Aサイト
回答の補完		○	○
百科事典に対する意見付与	○	○	
意味的な類似キーワードの付与	○	○	

これより、本システムでは、各集合知の構造を踏まえた情報取得を行い、必要な情報のみを取り出す。そして、これらを取り出した情報を再構成し提示するシステムの実現を目指す。図7にシステムの概要を示す。

本システムの構成要素は、データ収集部、データ再構成部、データ表示部の3つである。各集合知の構造を踏まえた情報取得を行い、必要な情報のみの取り出しをデータ収集部にて行う。これら取り出した情報の再構成をデータ再構成部にて行う。最後にユーザへの提示をデータ表示部が行う。

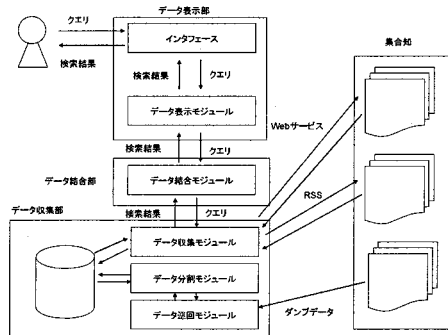


図7 システムの概要

4.2 再構成の仕組み

4.2.1 回答の補完

まず、回答のない質問をYahoo!知恵袋からRSSを取得し、パースすることで質問文のみを検出する。次に、検出した質問文の中から、主要なキーワードを形態素解析によって特定し、そのキーワードの説明をデータベースから探す。このデータベースには、あらかじめWikipediaからダンプデータを取得し、構造をパースして、項目名と項目の内容を格納しておく。もし、キーワードに対応する項目があれば、項目の内容を付与し、なければ関連度の高い順に情報を提示する。ここでの関連度は、Wikipediaの提示する関連度を用いる。回答の補完の流れを図8に示す。

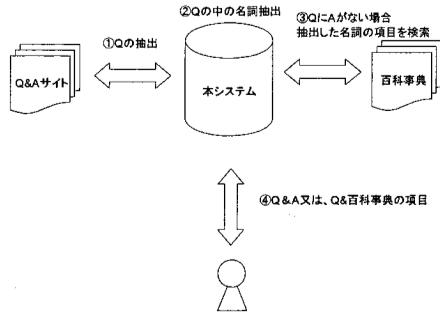


図8 質問への一時回答の流れ

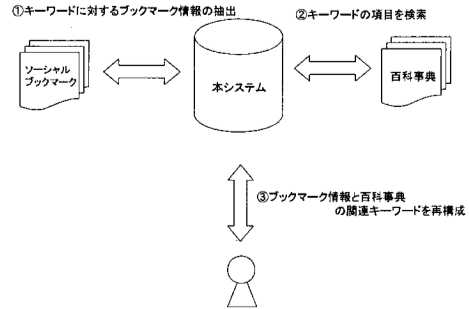


図10 意味的な類似キーワードの提示の流れ

4.2.2 百科事典に対する意見付与

ユーザが利用している項目から Wikipedia の URL を特定し、その URL に対してのブックマーク数をはてなブックマークから Web サービスを用いて取得する。ブックマークのタグについては、RSS をパースし、取得する。百科事典に対する意見付与の流れを図9に示す。

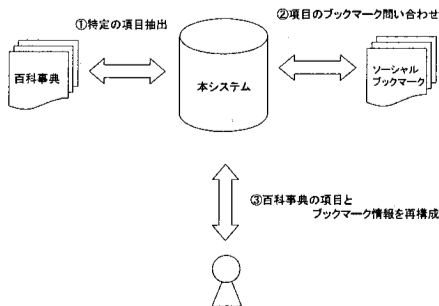


図9 百科事典の項目に対する関連意見の流れ

4.2.3 意味的な類似キーワードの付与

RSS を用いて特定のタグが付与されているブックマーク先、ブックマークタグをはてなブックマークから取得する。また、そこで取得に用いたキーワードの類似キーワードをデータベースから取得する。データベースには、あらかじめ Wikipedia からダンプデータを取得し、構造をパースして、項目ごとにリンク先、被リンク先を集計し、格納しておく。回答の補完の流れを図10に示す。

4.3 実行例

図11はシステムにアクセスしたときのインタフェースである。この時に、ユーザはキーワードを入力する。次に図12のように入力したキーワードについて、各集合知における情報を表示する。また、各集合知に含まれる要素を選択するコントロールパネルを右下に表示する。ユーザは図12の右下のコントロールパネルから再構成する集合知の情報をチェックボックスで選択し、集合知統合利用ボタンを押す。するとシステムは、ユーザの選択に基づいて、表示部にデータを再構成する。再構成されたデータを受け取り、図13のようにインタフェースに表示する。

5. おわりに

本研究では、単体の集合知における問題点について検討を行い、それを解決するために、複数の集合知を組み合わせた再構成手法を提案した。また、システムのプロトタイプを実装し、集合知の再構成の具体例について検討を行った。本システムによって、各集合知が持つ情報を、部分的に再構成して表示することが可能となり、問題解決のみならず、ユーザが自由に新しい知識を創出するためのプラットフォームを実現できるようになるのではないかと考える。今後は、本システムを用いた評価実験を行い、ユーザが実際にどのような新しい知識を創出するのか、またそこから創出される知識モデルがどのようなものとなるのかについて検討を行っていききたい。

参考文献

- [1] <http://www.addclips.org/>
- [2] 山家雄介, 中村聡史, Adam Jatowt, 田中克己: Web 検索のランキング精度向上のためのソーシャルブックマークの利用. DBSJ Letters, Vol. 6, No. 1, pp. 177-180, (2007).
- [3] 田村悟之, 清田陽司, 増田英孝, 中川裕志 図書館における自動レファレンスサービスシステムの実現に向けて～Web 上の二次情報と図書館の一次情報の統合～. 情報処理学会情報学基礎研究会 2007-FI-086, pp. 1-8, (2007).
- [4] スロウィツキー, ジェームズ. 「みんなの意見」は案外正しい (小高尚子訳). 角川書店, 東京, 2006.
- [5] 大向一輝: Web2.0 と集合知, 情報処理, Vol. 47, No. 11, pp. 1214-1221, (2006).
- [6] <http://chiebukuro.yahoo.co.jp/>
- [7] <http://okwave.jp/>
- [8] <http://ja.wikipedia.org/>
- [9] <http://b.hatena.ne.jp/>
- [10] <http://buzzurl.jp/>

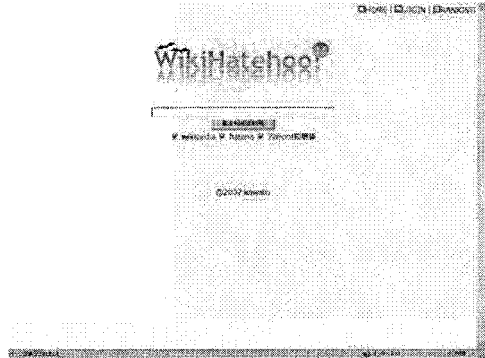


図 11 システムトップインタフェース



図 12 集合知選択インタフェース

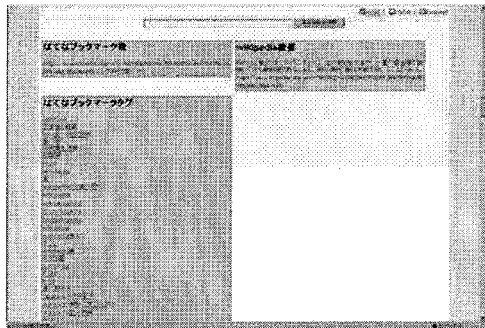


図 13 集合知統合インタフェース