

# 共同編集型比較表システム Wikiparison

沼 晃 介・河 又 翔 平・久保田 夕 貴・  
 小 野 慎 也・紀 中 一 明・須 田 健太郎・  
 湯 山 諭・片 野 実・平 井 宏 祐・  
 世 戸 口 文 耀 (専修大学ネットワーク情報学部)

## Consumer Generated Comparison Table System “Wikiparison”

Kosuke NUMA, Shohei KAWAMATA, Yuki KUBOTA, Shinya ONO, Kazuaki KINAKA,  
 Kentaro SUDA, Satoshi YUYAMA, Makoto KATANO, Kosuke HIRAI, Takeaki SETOGUCHI  
 (School of Network and Information, Senshu University)

In this research, we propose a consumer generated comparison table system named “Wikiparison.” With the help of user’s collective intelligence, our system collects and restructures distributed Web information into a simple table format so that user can compare various data. Our Wikiparison system manages compared data in the form of a triple {Item, Key, Value}, which can be translated into RDF triple. This means that our system can share stored data as linked data. This paper describes the implementation and its result of our system.

キーワード：比較表，消費者生成メディア，ソーシャルウェブ，リンクトデータ，セマンティックウェブ

**Key words** : Comparison Table, Consumer Generated Media, Social Web, Linked Data, Semantic Web

## 1. はじめに

情報発信の簡易化に伴い、大量のデータがウェブ上に蓄積され続けている。これらの情報にアクセスすることも容易になっている反面、その中から個々の最新データを的確に抜き出し比較することは困難といえる。これは情報を掲載する行為が多様であることと、そうして公開される情報に多くのノイズが含まれていることが主な原因だと考えられる。そこで我々はあらゆる情報を不特定多数のユーザが同一フォーマットで編集、作成することで情報の構造化と比較を可能にするウェブサービス、共同編集型比較表サービス Wikiparison を提案、実装した (図 1)。

Wikiparison はあらゆる分野の比較表をユーザが簡単に作成ならびに共同編集可能な CGM 型ウェブサービスである。Wikipedia のようにユーザが共同編集を行うことにより、各個人が持つ情報を効率よく集約し、表形式での構造化を可能とする。本稿では情報を表形式にまとめ、構造化することで情報を効率的に収集し比較するために、共同編集型比較表を提案する。また、Wikiparison 内のデータの再利用を可能としたトランスクルード機能の活用とその効果、考察をし、特に近年のウェブ研究におけ



図1 Wikiparison のトップページ画面

る潮流である Linked Data との関連について検討する。

## 2. 共同編集型比較表

### 2.1 比較表形式の表現

複数の情報を見通しよく対比する目的で、比較表形式は有用な表現手段のひとつである。図表や図解の効果はライティングやプレゼンテーション手法の領域でも重要視されている通りであるが、特に表表現に関しては、項目間のデータやその特性の共通性や差異を比較したり、特性の存在や不在を可視化したりする利点が指摘されている (Editage [2012])。また教育心理学の領域においても、表を始めとする図表が理解を促進ことが報告されている (Hawk [1986])。オンラインにおける情報収集においても、表形式の表現は情報の理解を助けるものと考えられる。

しかし現在のウェブにおいて、自分が欲している情報がひとつに纏まっていること、ならびにそれらがユーザにとって見やすい形式で存在していることは少ない。ネット通販大手の Amazon においても、販売している企業ごとに載せている情報が異なることで商品の比較を行うには複数ウィンドウを立ち上げる必要があり、必ずしも商品のスペックが見比べやすい形式で存在している訳ではない。またそのスペックもサイトや商品によって記述項目が共通していなかったり、レビューやユーザーレーティングなどの主観的なデータと入り混じって提示されたりする。サイト間で記述項目ばかりでなく、ページレイアウトや文字修飾も異なるため、見比べるのに手間のかかる場合が多い。これらの問題は、(1) 情報源が複数のウェブページやウェブサイトにもたがるというデータの分散性の問題、(2) 比較対象データ間の項目の不一致による対比の困難さの問題の2点に集約できる。

そこで本研究では、複数の事象に関するデータをひとつの比較表にまとめることで、これらの問題を解決することを狙う。

複数情報を比較表形式に統合し提供するに当たって考えられる課題として、最初の情報収集の手間

と、一度集まった情報の鮮度ならびに精度の維持管理が挙げられる。この課題に対して我々は、集合知を活用するアプローチを取る。

## 2.2 共同編集データソース

現在比較サイトとしては「価格.com」、「比較.com」や、「ベストゲート」等があげられるが、これらのサイトでは製品をピックアップして価格やスペックを比較することしかできない。欲しい情報が存在していない場合や間違っただけの製品データを発見した時に、ユーザは情報の追加や修正を行うことができないため、ウェブサイトを管理者が情報を更新するのを待つか、他のサイトを利用する他にないというのが現状である。また、掲載される情報に関してはパソコンやテレビ、カメラなどの家電量販店で取り扱うような製品のカテゴリは充実しているが、エナジードリンクのようなニッチなカテゴリや、関連性の少ない製品同士を比較したい場合には不向きである。

多数の編集ユーザが興味のある情報を持ち寄って集めることで、多くの閲覧ユーザが欲する情報を網羅すること、誤った情報や古くなった情報を適時修正、更新していくことが期待できる。多くのCGMサイトに共通するアプローチであるが、オンライン百科事典における Wikipedia のように、公共性の高い情報の集約において一定の質の担保を狙う。

Eto [2012] は、ユーザが共同でデータを追加、編集する集合知データベース Wedata を構築した。Wedata ではユーザがウェブアプリケーションで利用するデータスキーマを定義し、データの追加、更新、ならびに取得を行うことができる。共同編集データベースの構築、運用という観点で本研究と共通するが、より汎用的なデータソースとなる Wedata に対し、Wikiparison は情報の比較に特化する点、特にそのためのユーザインタフェースを提供する点で目的を異にする。

## 2.3 ウェブ上のデータの構造化

Wikiparison のもうひとつの狙いは、ユーザの手によるウェブ上のデータの構造化である。ユーザが分散する複数の情報源を参照して Wikiparison 上にまとめることで、多様なデータを検索性や再利用性の高い状態で蓄積することができる。

Wikiparison のデータ構造を図2に示す。Item (アイテム)、Key (キー)、Value (バリュー) の3つの値を組にして使用する。アイテムとは比較される値のことをいい、iPhone、ライオン、りんごなど

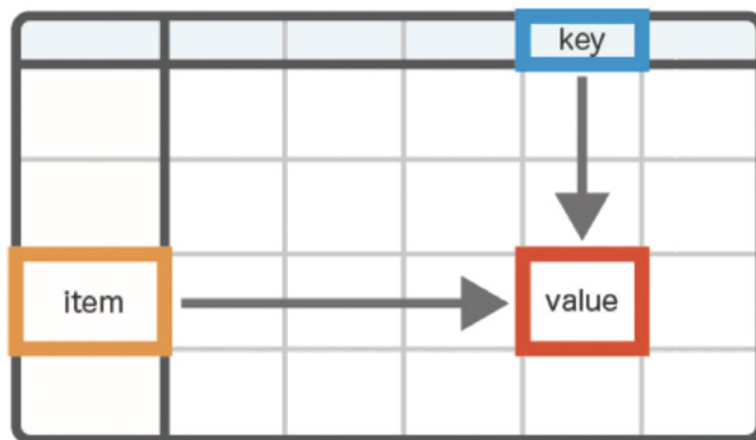


図2 Wikiparison における比較表のデータ構造

といった具体的な比較する対象を表している。それに対しキーは、重さ、寿命、高さなどのようにアイテム同士を比較するときの属性、比較項目を表している。そしてアイテムが各キーに持つ実際の値をバリューと呼ぶ。例えばアイテム「ライオン」に対する「平均寿命」というキーのバリューは「10～15年」となる。Wikiparisonにおける比較表とはアイテムとそのアイテムの持つデータ (Item, Key, Value) を複数、表形式でまとめたものことである。

これらのデータは、Item を主語、Key を述語、Value を目的語と読むことで RDF のトリプル (W3C [1999]) に変換して扱うことができるため、提案システムにおいて蓄積されるデータをリンクトデータ (Heath [2011]) として共有可能である。

## 2.4 トランスクルード

Wikiparison において、情報はアイテム単位で蓄積され、複数の比較表に同一のアイテムを埋め込むことが可能である。過去に蓄積されたデータを再利用することで新規の比較表作成時に情報入力を支援するとともに、アイテムに対する情報更新を複数比較表に渡って反映させることができる。

1960年にテッド・ネルソンによって提唱された Xanadu 計画では複数に分かれたページの一部を双方向に同期する機能が提案され、これはトランスクルードと名づけられた (Nelson [1981])。本機能はそれに倣いトランスクルード機能と呼ぶ。

トランスクルード機能により情報を分断させることなくサービス全体での更新を反映させる。このことは Wikiparison 内部でのデータの管理における利点に加え、同一の Item を一元的に扱うため、同一の主語に一意の URI を振ってリソースとして扱う Linked Data との親和性を高めている。

## 3. Wikiparison の実装

Wikiparison では Wikipedia などに見られる共同編集型でのデータ収集法を採用することで古いもしくは間違っただけの情報をリアルタイムに更新・削除する。個人のブログやサイトに掲載されるデータはそのコンテンツを運用する管理人にのみ編集権限が与えられているため、たとえ古くなっていることや間違っていることに気づいた閲覧ユーザがいても情報を更新・削除することができなかった。Wikiparison においては管理者のみならず不特定多数のユーザに編集権限を与えることで常にデータを最新状態に保つことを可能とする。

共同編集では編集者独自の解釈で記述される方式であることから、情報更新が行われたとしてもページによって比較要素 (特にアイテム・キー) の記述形式の差異が原因で発生する情報の分断化が問題視される。この差異をなくすため、まず既存の比較要素をそのまま使うよう編集者に促しユーザの入力支援 (サジェスト機能) を行う。内部処理としては、比較要素は文字列で保存するため、入力された文字列が MySQL 上に保存される比較要素と完全一致するかどうかで引用の有無を決定している。引用有と判定されれば、新しくデータ追加を行わず、その MySQL 上のデータと作成された比較表ページを紐づける。その結果、引用された比較要素が編集された場合はその比較要素が使われている全ページに更新内容が適応される。引用はあくまで文字列の完全一致の場合のみであるため、表記や表現の違いを判定することはできないが、これにより双方向引用 (トランスクルード) が可能となる。

Wikiparison は、主に比較表の作成・表示・編集の3つの機能を備える。作成機能では、比較表のタイトル及びその表の説明文、アイテムを行に、キーを列とした表を作成することが可能である。また行となる各アイテムには画像を添付することも可能である。さらに他の表計算アプリケーション等で

作成した表を流用できるよう、CSV形式のファイルをインポートする機能も設ける。表示機能では、作成機能によって作られた表を表示して閲覧することが可能である（図3）。また、表示機能と平行してIn-place Editorにより各アイテム・キー・バリューを編集する機能を備える（図4）。表内の特定の



プロ野球パリーグ所属球団の比較表です

名前	球団創立	所在地	日本一回数	リーグ優勝回数	交流戦優勝回数	試合（12年まで）	勝利（12年まで）	敗戦（12年まで）	引分（12年まで）
 北海道日本ハムファイターズ	1945年11月6日	北海道札幌市豊平区	2回 (2006)	6回 (2012)	1回	8905	4098	4471	336
 埼玉西武ライオンズ	1949年11月26日	埼玉県所沢市	13回 (2008)	21回 (2008)	0回	8441	4307	3791	343
 福岡ソフトバンクホークス	1938年3月29日	福岡県福岡市中央区	5回 (2011)	17回 (2011)	3回	9466	4760	4359	347
 東北楽天ゴールデンイーグルス	2004年9月24日	宮城県仙台市宮城野区	0回	0回	0回	1136	489	616	31

図3 比較表画面



ネコ科の主にネコ属の動物たちです。

名前	分類	体長
 チーター	哺乳綱ネコ目（食肉目）ネコ科チーター属	110-140センチメートル
 カラカル	哺乳綱食肉目ネコ科カラカル属	55-90センチメートル
 ボルネオヤマネコ	動物界脊索動物門哺乳綱ネコ目（食肉目）ネコ科	50-68センチメートル

図4 編集画面

アイテムを比較したい場合は検索機能によるフィルタリングや、キーによる比較順序を切り替えるソート機能が比較表表示画面内で利用可能である。

## 4. 運用結果

### 4.1 基礎データ

構築した Wikiparison システムを 2013 年 4 月 12 日からおよそ 1 年間に渡り運用を行った。このうち 2013 年 12 月 22 日までのアクセスは、総ページビュー数 (PV) が 53,407 件、訪問者数は 15,327 名、このうち編集を行ったユーザは 151 名であった。

作成された比較表は 195 件、含まれる総アイテム数は 3,658 件であった。キーの種類は 998 件あり、バリューのデータ数は 50,631 件である。

### 4.2 記述されたデータの特徴

ひとつの比較表に含まれるアイテム数に基づいて比較表を計数した分布を図 5 に示す。比較表中にアイテムを 1 つだけ含むものが 1 件、2 件のアイテムを含む比較表が 21 件、3 件のアイテムを含む比較表は 16 件というように、1 アイテムのものを例外としながらもアイテム数が少ない比較表の方が多い傾向にある。最もアイテム数が多い比較表には 842 個のアイテムが含まれていた。

また、アイテム当たりのキーの数は、1~61 個であり、最頻値は 22 の 842 件であった。上記の最多アイテム数の比較表に含まれるアイテムのキーの数が共通で 22 件であり、これがこのまま最頻となった。

バリューは自由記述であり、データ上は text として保存されるが、実際にはこの欄に数値データが記述されることが散見された。バリューの内容を簡易なパターンマッチにより分析を行ったところ、全 50,631 件中の 15,943 件 (31.5%) が数値データであった。ここで数値データには、数値のみでなく、単位付きの数値や数値の範囲 (「-」や「~」で結ばれた 2 つの数値) を含む。

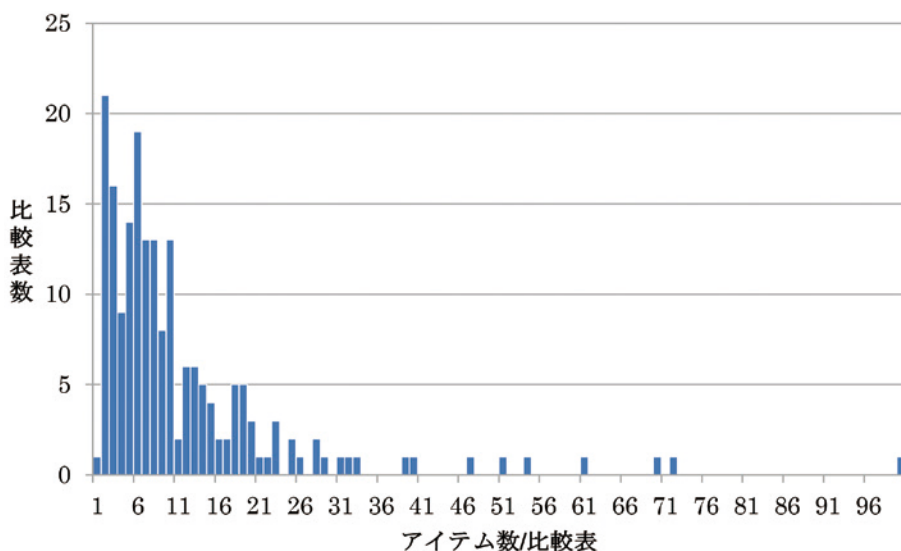


図 5 比較表当たりのアイテム数の分布

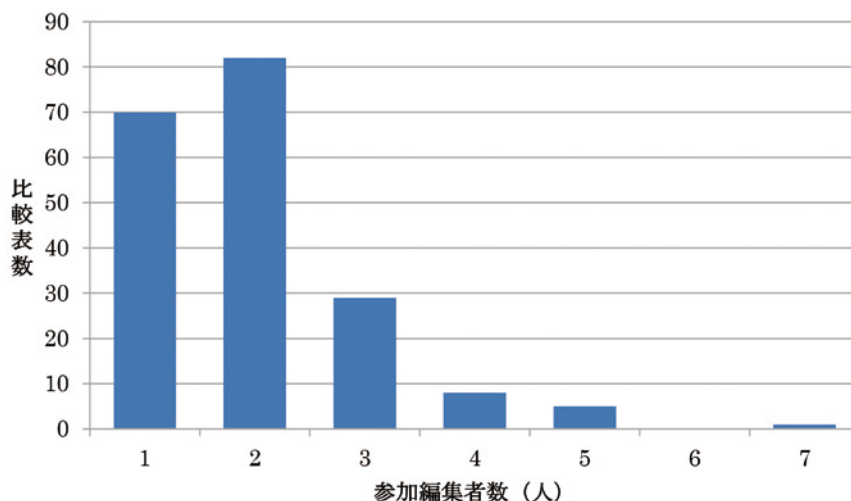


図6 比較表当たりの編集人数の分布

### 4.3 共同編集状況

次に、ひとつの比較表の編集に携わった編集者の人数を計数した。参加編集者の分布を図6に示す。1名の編集者ですべてを記述した比較表が70件、2名の編集者によって共同編集された比較表が82件であった。単独編集の比較表は全体の35.9%である。

### 4.4 比較表間での共通アイテム

全3,658アイテムの内、複数の比較表に用いられたアイテムは318件であった。

## 5. 議 論

運用中に作成された比較表の6割超が複数名のユーザによる共同編集によるものであった。このことから共同編集というスタイルが一定程度ユーザのデータ作成を支援したものと推測できる。

データの再利用を通じデータ作成を助ける目的であったトランスクルード機能は、利用頻度が低いように考えられる。これは、まだWikiparisonがサービスの初期段階にあり、全体のデータ量が多くないことにより、新しい表に使い回せる情報も多くないという点に理由があると考えられる。

本研究では比較表構築のための共同編集機能、ならびにこれを助けるトランスクルード機能の予備的分析を行ったが、比較表という形式によって比較を行うこと自体の効果は確かめられていない。これらは今後の課題である。

提案システムでは、Item, Key, Valueの3つ組による情報表現によりデータを収集した。この構造は、リソースIDとしてItemのURIを定義すると、比較表内のデータをItemを主語、Keyを述語、Valueを目的語に対応づけることで、データ構造を維持してRDF形式に変換することが可能である。本稿の扱う範囲においてはこのような実装は行っていないが、外部データソースとの接続やその応用にも提案システムの発展性があることを示している。

Linked Dataの拡充を目的とした取り組みは数多く、表形式の統計資料などをRDF化して共有する

Linked Open Data もそのアプローチのひとつである。本サービスは行政などによるトップダウン的なデータ構築とは一線を画し、ユーザ主導によって数値データを整理し、蓄積、共有することで Linked Data にも貢献できると考える。

ただしこのためにはいくつかの課題がある。

まず Value に記述されるデータが現在の実装においては書式フリーのテキスト形式であるため、同じ数値データでも編集者によって記載方法が異なる場合がある。これらを統一するような処理系、もしくはユーザの入力に制約をかけるような UI が求められる。しかしこのことは同時に、ユーザが気軽に自身の手により比較表を作成できるというサービスの思想に反する。適切なバランスでのルール作りが必要である。

第二に、Linked Data の文脈におけるデータ記述の前提には、無矛盾であることが暗黙に求められている。RDF での記載により、提案サービス内のみならず外部のデータソースとも Key と Value が共有されるが、これらが相互に矛盾する値を持つ場合、データが破綻する。Wikiparison の編集者には、Wikipedia 同様の客観性を求める編集指針を提示することも考えられるが、一方で推論時に文脈を考慮することで矛盾を許容するようなアプローチもありうると考えている。提案システム上においては、3つ組は RDF としてではなく表として記述される。このことは、表自体が文脈を形成することを意味する。比較表のタイトルや更新日時、編集者などのプロパティがこのリソースのコンテキストとなるだろう。これを考慮した推論機構は今後の研究課題である。

## 6. おわりに

本研究では、分散したウェブ情報を、集合知を用いて構造化し収集することで、あらゆるウェブデータの比較をやすくすることを目的に、共同編集型比較表システム Wikiparison を提案した。本稿では Wikiparison システムを構築し、運用した結果を報告した。またこのシステムの Linked Data への貢献の可能性を検討した。

## 参考文献

- Editage Japan, “研究成果の発表に効果的な図表の使い方, <http://pt.slideshare.net/EditageJapan/ss-12320402>, 2012.
- Eto, K., Hamasaki, M. and Takeda, H. “Wedata: a wiki system for service oriented tiny code sharing”, *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, ACM, 2012.
- Hawk, P.P. “Using graphic organizers to increase achievement in middle school life science”, *Science Education*, Vol. 70, Issue 1, pp. 81-87, 1986.
- Heath, T. and Bizer, C. “Linked Data: Evolving the Web into a Global Data Space, ser. Synthesis Lectures on the Semantic Web: Theory and Technology”, Morgan & Claypool, 2011.
- Nelson, T.H. “Literary machines”, Swarthmore, Pennsylvania: self-published, 1981. (『リテラリーマシーン—ハイパーテキスト原論』, アスキー, 1994.)
- W3C Recommendation, “Resource Description Framework (RDF) Model and Syntax Specification”, <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>, 1999.