

機関横断型文献情報 Wiki による著者情報対応付けとその編集支援機能の提案

Metadata Mapping of the Same Author by Cross-agency Bibliographic Information System using MediaWiki and Proposal of its Editing Support Function

日向野達郎*¹
Tatsuro Higano

中村早紀*¹
Saki Nakamura

増田英孝*¹
Hidetaka Masuda

山田剛一*¹
Koichi Yamada

清田陽司*^{1,2}
Yoji Kiyota

中川裕志*³
Hiroshi Nakagawa

*¹東京電機大学
Tokyo Denki University

*²株式会社ネクスト
NEXT Co.,Ltd

*³東京大学
University of Tokyo

Bibliographic databases are provided as web services from various agencies. Since entities (e.g., persons, books, and papers) are not linked across other databases, users need to repeat searches at each database. This paper propose a method for mapping entities across databases using MediaWiki, and then propose a method to support editing by micro-volunteer.

1. はじめに

現在 Web 上には、「CiNii」や「J-GLOBAL」等の、文献検索サイトと呼ばれるポータルサイトが存在する。これらのサイトは論文や書籍等の文献を集積している機関によって運営され、過去から最新の研究成果や、関連研究の調査の際に非常に便利なサービスとして多くの研究者に利用されている。

しかし、それぞれ別々の機関によってメタデータが管理されているため、複数のサイト上の文献を横断的に検索することができない。そのため、網羅的に文献を探しているユーザにとっては、複数のサイトで検索を繰り返す必要があり、非常に手間がかかってしまうというのが現状である。

本研究では、機関の枠を超えて文献情報を横断的に検索することを可能にするサービスの開発を目的としている。このためには各機関がメタデータに対して、それぞれ割り当てている固有の識別番号（論文 ID、著者 ID 等）を互いに対応付ける必要がある。そこで MediaWiki を用いることによって、複数の文献検索サイトのメタデータを容易に対応付けるための枠組みを提案している。

本研究では、主に人物情報を対象として、MediaWiki 上で対応付けを行う Wiki システムを構築している。そして編集者の負担を減らすことに加え、編集に参加する動機をもつてもらうために、facebook における友達関係を利用して、編集者と友達関係にあり、Wiki 上で対応付けの編集を必要としている人物を編集者に知らせる機能や、Wiki の記法を知らなくても編集に参加できるように編集を依頼する機能といった編集支援の機能を提案する。

2. 文献検索サイトの統合

国立情報学研究所の「CiNii」や、科学技術振興機構の「J-GLOBAL」等の文献検索サービスが様々な機関から提供されているが、横断検索等のサービスの統合はなされていない。例として「CiNii」の著者情報のページから他機関のサービスである「J-GLOBAL」へのリンクというものが存在しているが、あくまで「J-GLOBAL」でその著者の名前を検索した結果のページへのリンクであり、直接「J-GLOBAL」の著者情報ページへリンクされているわけではない。これは国立情報学研究所

と科学技術振興機構がそれぞれ所有している情報に対して独自に割り当てている識別番号（論文 ID や著者 ID 等）が、互いに対応付けられていないために起こる問題である。

さらに著者データベースでは、著者の所属の変更や結婚などの理由による姓名の変更によって発生する重複レコードや、同名同姓の複数の人物のレコードを機械的に判別し修正することは難しい [1]。そこで人手による修正が必要になってくるがここでもいくつかの問題がある。情報に間違いがあった場合に機関へ報告しても、機関のスタッフにより確認が行われた後、修正が行われるというように、間違いの発見から修正までに時間がかかってしまう。また他のサイトにおいては、自分でない他の著者の情報に間違いを発見しても修正することができないので、著者本人が間違いに気づくまで情報が間違っただまとなってしまうのが現状である。

そこで本研究では、各機関を間接的につなぎ機関の情報を横断的に検索するサービスの開発を目的とする。さらに、MediaWiki を利用することでアカウントを作成すれば誰でも編集に参加できるので、ユーザが情報に間違いを発見した場合、そのユーザの手で即座に編集をすることも可能になるということが本研究の特徴的な点である。

この目的のために、それぞれの機関のメタデータに割り当てられている ID に着目し、同一の情報同士の ID を各サイトから収集し、対応付けを行うことでサイトを統合するシステムを MediaWiki を利用して構築する。図 1 に ID 統合の概念図を示す。



図 1: ID 統合の概念図

図 1 に示すように、従来の文献検索サイトは「CiNii」の「田中一郎」の著者情報ページから、「J-GLOBAL」の「田中一郎」の著者情報ページへは直接リンクすることができなかった。本

システムで各サイト間の同一情報同士の ID を対応付けることで、各サイトを間接的につなぐことを可能にしている。

本研究では、機関の所有するメタデータの内、まず手始めに著者情報を対象として MediaWiki 上で人手によって対応付けを行う。

3. 関連研究

様々な機関の所有するデータをつなぐ研究として Linked Data と呼ばれる取り組みが近年行われ始めている [2]. Linked Data の実例として美術館や博物館の所有するデータを集めて関係づけた LODACMuseum[3] というサービスが公開されている。美術館や博物館が持つ情報に加え、Wikipedia の情報とも関連付けられており、様々な側面から情報を得ることができサービスとなっている。Linked Data の研究では一つの実体に対して一つ ID があることが前提であるが、現状では一つの実体に対して様々な機関がそれぞれの ID を割り当てているという文献検索サイトの現状と同様の問題を抱えている。Linked Open Data の研究の中に実体に対して統一的な ID を与えるというアプローチがある [4]. 本研究ではそれぞれの ID 同士を対応関係で結ぶことでつなぐというアプローチをとる。

MediaWiki を利用した Wiki システムである Wikipedia と同じように対応付けの編集は、人手によって無償で行われるため、マイクロボランティアのコンセプトを利用する。マイクロボランティアとは、短時間で参加可能なインターネット上のボランティアであり、多数の人の知や力の利用によって問題解決を行うクラウドソーシングの一種である。マイクロボランティアにおいて重要なのは、ひとりひとりの労力を少なくすることで、多くの人に参加してもらうという点である。マイクロボランティアの具体的な例として、2011 年 3 月に発生した東日本大震災直後に提供された臨時サービス「GooglePersonFinder」が挙げられる。このサービスは、人の安否を知りたいユーザが相手を特定できる情報を登録しておき、その人物の消息を知る人が現在の状況を投稿することで、消息を伝えることができる仕組みとなっている。臨時で提供されたにもかかわらず、当時 67 万件以上の安否情報が登録された。マイクロボランティアにおいて重要なのは、ひとりひとりの労力を少なくすることで、多くの人に参加してもらうという点である。

4. 機関横断型文献情報 Wiki の構築

ここでは、MediaWiki 上でどのように複数の文献検索サイト上のメタデータを対応付けていくかについて述べる。

4.1 メタデータの収集

本システムでは、各サイトの所有するメタデータを、サイト内の人物情報ページをスクレイピングすることで機械的に収集する。

各サイトから、メタデータに割り当てられている「ID」と、「名前」、「所属機関」といった人物に関する基本的な情報に加え、「研究分野」や「論文の一覧」等のように人物を特定するために参考となる情報を収集する。今回は検証のために東京電機大学の教員計 335 名の人物情報を対象とした。

4.2 Wiki への登録

収集したメタデータの Wiki への登録は編集 bot により機械的に行う。ページタイトルはそのサイトにおける ID とする。登録された人物情報の例を図 2 に示す。

図 2 は著者「増田英孝」の「J-GLOBAL」でのメタデータページである。「J-GLOBAL」における「増田英孝」の

J-GLOBAL:200901009424739052	
目次 (非表示)	
1 著者名	
2 所属	
3 ホームページアドレス	
4 研究分野	
5 論文一覧	
著者名	<ul style="list-style-type: none"> 増田 英孝(マスタ ヒダタカ)
所属	東京電機大学 未来科学部 情報メディア学科
ホームページアドレス	
研究分野	
その他	
論文一覧	<ul style="list-style-type: none"> Proposal for a Social-MVRC Social Consensus Formation System Concerning IT Risk Countermeasures

図 2: 登録されたメタデータページの例

ID は「200901009424739052」なのでページタイトルは「J-GLOBAL:200901009424739052」となる。ページ内の項目は、サイトから収集したその人物の基本情報を記載する。このようにしてページを作成していき、同名の著者のページをその人物名をページタイトルとした人物名ページに一覧としてまとめる。図 3 に人物名ページの例を示す。

増田英孝	
<ul style="list-style-type: none"> J-GLOBAL:200901009424739052 CiNii:900001651605 CiNii:900001588952 	

図 3: 作成された人物名ページの例

この例の場合、「J-GLOBAL」に一件、「CiNii」に二件レコードが存在するので、図 3 に示す通り「増田英孝の人物名ページ」には三つのページがまとめられることになる。

4.3 各サイトのメタデータの対応付け

各サイトから収集したメタデータが Wiki に登録された段階では、各サイト間の対応付けがなされていないので、同一著者のページを対応付けるという作業を行う必要がある。現状ではこの作業は人手で行っている。

編集者は、人物名ページにまとめられたメタデータページ内の、論文の一覧や、所属、研究分野、研究キーワード等の情報をもとに同一人物であるか否かの判断を行い、同一人物であった場合にはページ同士をリダイレクト関係にする。リダイレクト先は最初に Wiki に登録されたページとする。リダイレクトをさせるためにはリダイレクト元のページをリダイレクトページに変更する必要がある。ページの編集画面に「#redirect[リダイレクト先のページタイトル]」という記述を加えるだけでそのページをリダイレクトページに変更することができる。このリダイレクト先のページに対して「リダイレクトしているページのタイトルの一覧」を取り出す。タイトルは各機関における ID なので、このリダイレクト関係が ID の対応表として機能する。

4.4 人物名の曖昧さ回避

このような対応付けを行なっていくことで最終的に人物名ページは、同名の異なる人物の区別をするための曖昧さ回避ページとして機能する。図4に曖昧さ回避ページとして機能している様子を示す。

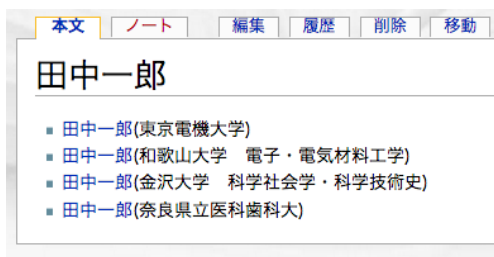


図4: 曖昧さ回避ページの例

同姓同名の人物が存在しなかった場合、人物名ページも J-GLOBAL のメタデータページへのリダイレクトページにする。こうすることで名前を検索すると直接 J-GLOBAL のメタデータページへリンクすることができる。

5. 編集支援機能の提案

2009年に行われた Wikimedia 財団による Wikipedia 参加者へのアンケート [5] に「書き込みをしない人が書き込みをしてくれるようにするためにはどうすればよいと思うか」という問いかけがある。この回答として「自分の書き込みが必要な特定のトピックがあれば書き込む」、「どうやっているか見せてくれたら」というものがあつた。このことから、本 Wiki システムの編集に参加してもらうためには、「この編集を行うためには自分の知識が必要である」という意識を持ってもらうということ、そして、「どう編集すれば良いかが分かりやすい」ことが重要であると考えられる。さらに、関連研究でも述べた通り、本システムはマイクロボランティアのコンセプトを利用するため、編集に参加するユーザの負担を出来るだけ少なくするということが重要である。この点を踏まえて、多くの人々に編集に参加してもらうための編集支援の機能を提案する。

5.1 facebook の友達関係を利用した編集支援

対応付けの編集を行うためには、メタデータページ内の論文の一覧や、所属などの情報を元に同一人物かどうかの判定を行うが、所属情報が記載されていないなど人物を特定するための情報が極端に少ないメタデータも存在する。編集者の専門分野から離れている分野の人物で所属などの情報がなかった場合、同一人物であるか判断することは難しく、間違つた編集をしてしまう可能性も高い。

このようなケースに対応するため、facebook の友達関係を利用して、編集の支援をする機能を用意する。図5に機能の概要図を示す。

図5に示すように、編集者用のページに facebook のアカウントでログインすることで、その編集者の facebook の友達一覧から登録情報である氏名、学歴、職歴の三つの情報を取得する。学歴と、職歴を所属情報として利用し、Wiki に登録されているメタデータページ内の情報と比較する。氏名と所属が一致するメタデータページがあつた場合、そのメタデータページがまとめられている人物名ページを編集者に返す。つまり編

集者用ページで facebook アカウントを入力すると、facebook 上で友達関係にあるユーザの中から、Wiki に登録されていて、対応付けが行われていない人物の「人物名ページ」の一覧を確認することができる機能である。編集者はこの一覧を見ることで、自身の知識を必要とする編集はどれなのかということが分かりやすくなる。facebook で友達関係にあるということは、編集者は対応付けが必要な人物についてある程度詳しいと考えられるため、対応付けの編集も比較的容易であり、編集の信頼性が向上すると考えられる。

5.2 編集依頼の機能

アンケート結果の「どうやっているか見せてくれたら」という回答から、Wiki のページを編集することは、HTML の構文規則よりも単純であるとはいえ、分かりづらいという意識を持っているユーザも存在しているということが分かる。そこで、Wiki の構文規則を知らなくても編集に参加できる仕組みを取り入れる必要がある。図6に編集依頼のイメージ図を示す。

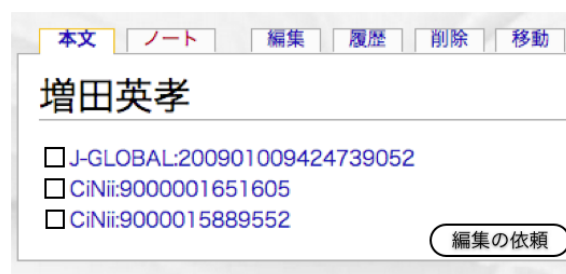


図6: 編集依頼機能のイメージ

図6のように、Wiki を編集することができないユーザが、Wiki 内の情報に間違いを発見した場合、Wiki を編集することができる編集者に対して「編集を依頼する」ことで知らせる機能を用意する。チェックボックスをチェックし、編集の依頼ボタンをクリックすることで対応づけが必要なメタデータページを編集者に知らせることができる。編集依頼の状況は編集者用ページにまとめられるので、編集者は現在編集を必要としているページをひと目で確認できる。

5.3 リアルタイムでの編集状況の提示

Wiki における全ての編集には履歴が残る。この編集履歴を RSS 等の形式で確認できるようにすることで、どの著者情報ページが編集されたかリアルタイムに知ることができるようになる。また著者情報ページ同士の対応付けも著者情報ページ同士のリダイレクト関係に編集することで行なっているので、機関ごとの著者情報の対応付けの結果も同様にリアルタイムで知ることができる。この対応付けの結果を、著者情報の重複を解決したいと考えている機関側が、一つの参考情報として活用することも可能となる。

6. 考察

MediaWiki を利用したサービスであるオンライン百科事典「Wikipedia」でも問題視されている通り、サービス内の情報の信頼性という点では、多くの課題がある。アカウントさえあれば誰でも自由に編集できるという MediaWiki の特徴から、悪意ある編集者が容易にでたらめな情報を追加するという可能性がある。しかしでたらめな編集が行われていることに他の

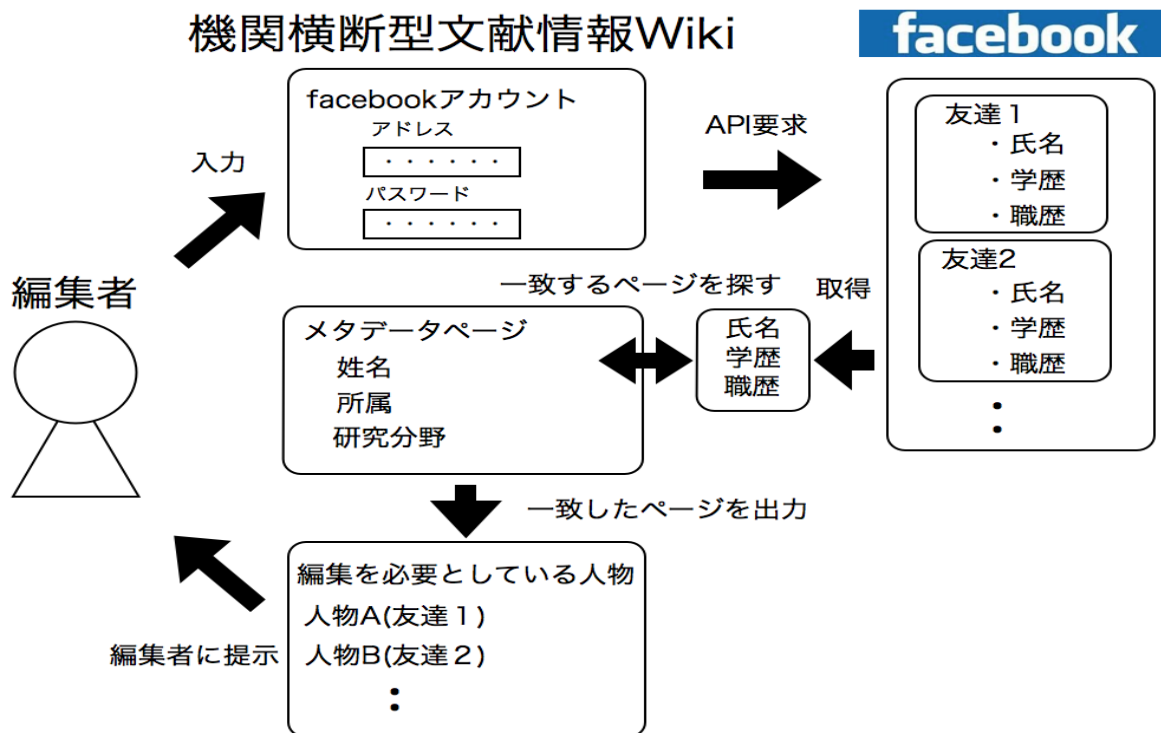


図 5: facebook の友達関係を利用した編集支援機能の概要

ユーザが気づけば、編集履歴を参照することで即座に元の状態に戻すことが可能である。そのため多くのユーザが利用してくれるようになれば情報の修正も多く行われるようになるため、ある程度の情報の信頼性は確保できるものと考えられる。

今回対応付けの編集を支援するための機能を提案した。目的としては、編集者の負担を減らすことに加え、編集に参加する動機を与えるためである。「編集を依頼する」機能によって Wiki の記法を知らなくても、編集に参加することができるようになる。そして facebook における友達関係を利用することで、編集者に近い関係にある人物が、対応付けの編集を必要としているということを明確になるため、編集者の編集への動機となり、活発に編集が行われることが期待できる。さらに編集状況を RSS で確認できるようにすることにより、編集の結果が機関の持つ情報修正に反映される可能性もある。これも Wiki に参加するための動機の一つとなることが期待できる。

7. おわりに

既存の文献検索サイトは、それぞれが別々の機関により提供されているので、メタデータの対応付けがなされておらず、横断的に文献を探しているユーザはそれぞれのサイトで検索を繰り返す必要があった。

そこで我々は、各機関が情報に対して割り当てている固有 ID の対応付けを行い各サイトの情報を間接的につなぐための仕組みである機関横断型の文献情報統合システムを MediaWiki を利用することで構築した。このシステムを ID の対応表として利用することで、機関を横断して情報を収集するということが可能になる。そして編集を支援する機能を提案した。

今後は、提案した編集支援機能を実装し、編集者に使用して

もらうことで、本機能が編集者の負担を減らし、編集に参加する動機を与えることができるかどうかを評価する予定である。

参考文献

- [1] 相澤 彰子 他, レコード同定問題に関する研究の課題と現状, 電子情報通信学会論文誌 J88-D-I, No.3, pp.576-589 (2005).
- [2] 神崎 正英, リンクするデータ、未来へのリンク, 第 19 回 Web インテリジェンスとインタラクション研究会, <http://www.kanzaki.com/works/2011/pub/0307wi2.html>
- [3] LODACMuseum, <http://lod.ac/>
- [4] 神崎 正英, 連携するデータ、リンクするデータ, デジタルアーカイブフォーラム研究会, <http://www.kanzaki.com/works/2007/pub/1129keio.html>
- [5] R. Glott, P. Schmidt, R. Ghosh, "Wikipedia Survey ? First Results" , http://upload.wikimedia.org/wikipedia/foundation/a/a7/Wikipedia_General_Survey-Overview_0.3.9.pdf