I-Scoverを使用した研究，アプリケーションについて

Study and Applications Using I-Scover

I-Scoverプロジェクト普及推進チームリーダー　千村保文

Abstract

2015年3月に文献検索システムI-Scoverのデータを用いたアプリ等のコンテスト「I-Scoverチャレンジ2014」の応募作品発表会を実施した．本発表時に最優秀賞，優秀賞を受賞した作品について，その目的，概要，特徴を受賞者から紹介する．I-Scoverは第2期システムからアプリケーションインタフェースを公開する．本稿で紹介する作品を今後のI-Scover活用の参考にして頂きたい．

キーワード：I-Scover，マイニング手法，Wikipedia Link API，時系列分析，技術要因分析，Expert finding，Expert profiling，Linked data，Web information extraction

1. はじめに

本会では，論文誌掲載論文・技術研究報告・大会講演論文等の知的リソースを横断的に検索するシステム「I-Scover」（アイスカバー，http://i-scover.ieice.org/）を構築した．I-Scoverでは，単に論文検索を行うことができるだけでなく，論文やその著者とその所属，出版物，イベント，キーワードなどの情報をデータとして再利用しやすい形式（Linked Data）で整理してある．これにより研究動向の分析や著者の分析などがしやすくなっており，I-Scoverデータを活用した様々な応用が期待されている．I-Scoverの普及促進活動を行っているI-Scoverプロジェクトでは，2013年からI-Scoverの積極的な利活用を促すため，これらの蓄積したデータを提供し，研究動向分析技術などI-Scoverデータの活用の可能性を示すアプリケーションや研究事例の投稿を募集する「I-Scoverチャレンジ」を実施している．本稿では，「I-Scoverチャレンジ2014」にて最優秀賞，優秀賞を受賞した作品について，受賞者からの紹介記事を掲載する．

2. I-Scoverチャレンジ2014概要

I-Scoverチャレンジ2014では，I-Scoverに蓄積されている論文，研究者，技術用語等の構造化データ（以降，I-Scoverデータセット）の分析／視覚化結果，I-Scoverデータセットを利用したアプリケーション（プログラム），及びI-Scoverの活用事例を2014年10月から2015年1月まで募集した．その結果，17件の応募があり，2015年3月の総合大会にて発表，審査会を実施した．

その結果，表1の作品を優秀作品として表彰した．

3. 優秀作品紹介

最優秀作品，優秀作品2件について，受賞者からの紹介記事を以下に示す．

千村保文　正員　沖電気工業株式会社経済・政策調査部
E-mail　chimura730@oki.com
Yasubumi CHIMURA, Member (Government & External Relations Division, Oki Electric Industry Co., Ltd., Tokyo, 105-8460 Japan).
電子情報通信学会誌　Vol.98 No.12 pp.1106-1118 2015年12月

表1　I-Scover チャレンジ 2014 優秀作品

		作品タイトル	応募者氏名
最優秀賞		研究会への参加による研究者のコミュニティ解析	宇野毅明（国立情報学研究所）
優秀賞（2作品）		I-Scover データベースを用いた時系列・技術要因分析モジュールの開発	槇　俊孝，若原俊彦（福岡工業大学）
		Topic-sensitive expert finding and profiling for I-Scover	Ruiyu Fang, Lu Fang, Qingliang Miao, Yao Meng（Fujitsu R&D Center in Beijing）
特別賞（4作品）	プレゼン賞	I-Scover オープンデータによる検索で見つけてもらえる論文のための正しいキーワードの付け方～サルでもわかる！ やさしい論文キーワードの付け方～	五味 弘（沖電気工業株式会社）
	技術賞	論文著者のつながりからイノベーションを生み出す「場」の形成を知る	岸本康成，飯田恭弘（日本電信電話株式会社），岸田吉弘，鬼塚　真（大阪大学大学院）
	スポンサー賞（OKI 賞）	論文紹介 Twitterbot	鈴木秀友
	学生賞	卒業論文タイトルジェネレーター SRTG v. 1.0	馬目慎太郎，佐々木祐三，朝香卓也（首都大学東京）

◎最優秀作品
「研究会への参加による研究者のコミュニティ解析」

宇野毅明（国立情報学研究所）

何らかの研究を行うとき，あるいは技術を探し出すときなど，研究者が作るコミュニティを把握することが重要である機会は多く存在する．新しい知り合いができたときに，その知り合いが所属するコミュニティを知ることで，その人のバックグラウンドや研究における文化のようなものを知ることができる．隣の分野で研究を始めたい，あるいはあるデータやプラットホームを軸に新しいビジネスを起こしていきたいときなど，関わりのある研究分野に存在するコミュニティを把握することで，どの人とコンタクトを取るべきか，どの分野を抑えておくべきか，といったことを俯瞰することができる．学会，あるいは日本全体の研究者コミュニティ構造を調べれば，分野による活性度の偏り，伸びている分野と停滞している分野，国際的に競争力を持つ，あるいはこれから持ち得る分野，横断的な協力体制を持つ領域などを知ることができ，それを基に将来的な重点目標を立てることもできる．

実際，私のところを尋ねていらっしゃる研究者，企業の方も，人を探すことに大きな難しさを感じている．研究や開発の最初のステップは，状況や問題点の整理を行い，それに現場や分野の問題意識を合わせ，そこに既存技術や研究分野の方向性を重ねた上で良い課題設定を模索し，実行可能な研究開発プランを議論する．そのステップにおいては，特定の専門性だけが高い必要性を持つわけではなく，いろいろな知識や考え方を俯瞰しつつ議論することが重要となる．このような状況で議論を行う，あるいは議論をしてもらう人を探すときには，研究者と技術を俯瞰した地図のようなものを眺めながら考え，探す作業が重要になる．私を訪ねる方々も，多くは「自分の持っているデータをどう使っていいか分からない」というような抽象的な問題意識を持ち込む方が多く，近年研究開発におけるこのステップの重要度が急激に高くなっていると感じさせられる．トピックだけでなく，コミュニケーションの取られ方と今後の方向性を意識した研究者データ解析が重要になるのである．

従来，研究者コミュニティは論文などの共著関係を軸にしてマイニングされることが多い．共著論文があるからにはある程度深いつながりがあるはずであり，その共著関係が作るネットワーク上で密な構造を見つけ出せば，それは確かにつながりのある人々のコミュニティであると考えられる．しかし，学会の論文誌などでは，多くの場合著者は学生と指導教員であることが多く，同一の研究室に所属しているという関係を表していることが多い．現実のコミュニティは共著関係から発生するものではなく，むしろほかの要因，学会や研究会における face to face のコミュニケーションなどを基に形作られていると考える方が自然であろう．

また，論文のキーワードを用いて研究トピック，あるいは研究分野を自動発見する試みも存在する．多くの論文で共通して用いられているキーワードは関連が深いと

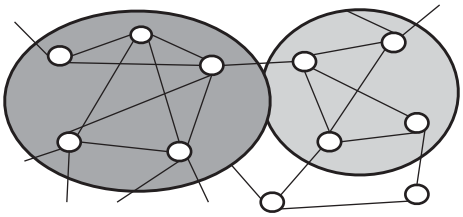


図1　ネットワーク上のコミュニティ

考えられ，関連が深いキーワード群が作るグループ（クラスタ）は研究のトピックを表していると考えられるからである．また，似たキーワードを用いている論文のクラスタも，同様にトピックを表していると考えられ，コミュニティマイニングなどの手法によりトピックマイニングとして研究が盛んに行われている．しかし，これらについても上記同様，研究分野は研究者の属人的なつながりにより形成されることが多いと考えられ，キーワードだけでは情報が不十分であると考えられる．例えばAというトピックとBというトピックがあったとして，それらを共通して研究している研究者が多数存在すれば，両者のトピックのつながりは深くなり，両者を関連付け，間を埋めるような志向を持つ研究も増えると考えられる．研究者を軸にクラスタリングをすれば，このようなトピック間の関連性も考慮した研究分野のマイニングが可能となる．

そこで，今回は研究会への参加，及び研究者とキーワード，という関係性に注目することで，上記の問題意識を解決するようなマイニング手法を提案する．研究者コミュニティに関しては，同一の研究会に投稿した（異なる回の研究会は異なるものとして扱う），という関係性に着目し，論文を投稿した研究会の集合が似ている研究者を似ているとみなすことで face to face のつながりを作り得る関係を捉えた研究者クラスタの発見を行う．研究トピックに関しては，それぞれの研究者に対して，その人が投稿した論文のキーワードをその研究者のキーワードだと解釈し，キーワード集合が似ている研究者のクラスタを作ることで関連する研究をしている研究者のクラスタを作成するとともに，キーワードを，そのキーワードを持つ研究者の集合が似ているという関係を軸にクラスタリングすることで，研究者コミュニティを軸とした研究トピックのマイニングを行う．

一般に，このような類似性を軸としたクラスタリングには，グラフクラスタリングをはじめとするクラスタリング，及びコミュニティマイニングが用いられる．図1のように，人やものの関わりを示す関連ネットワークや似ているものを枝で結んだ類似度グラフから，丸で囲まれた部分のような密構造を取り出すことでコミュニティを発見する．しかし，既存のクラスタリング手法のほぼ全てが，多数の多様なクラスタが存在するような状況で精度高くクラスタを発見することができず，多くのものが混在している巨大なクラスタを見つける，小さすぎる（大きさ2や3）クラスタを大量に見つける，一つのものが複数のクラスタに属するソフトクラスタリングが苦手，といった本質的に解決の難しい弱点を多く持つ．また，コミュニティマイニングの手法は重なりの大きな似たクラスタを取り扱い不可能なほど大量に発見する，計算コストが大きい，クラスタの粒度が小さすぎるといった，これまた本質的な改善が難しい弱点を多く持つ．そこで今回は，これら本質的な弱点を持たない，データ研磨手法によるクラスタ発見手法を用いる．データ研磨は，データの持つ雑音や欠損などによる揺らぎを消すことでクラスタの境界を明確にし，それによりコミュニティマイニングの精度を高める手法である．詳しくは添付の技術研究報告（データ研磨によるクリーク列挙クラスタリング http://ci.nii.ac.jp/naid/110009659424）を参照されたい．

データ研磨によるクラスタリングは，一部の研究者が作るコミュニティの芯の部分を従来法よりも大きめに抽出する．重なりが小さいが一部は重なりを持ち，粒度が一定で大きさの大小が余りなく，内部的につながりが強いようなコミュニティ（クラスタ）を類似度グラフから抽出する．類似度グラフとは，この場合，研究者（キーワード）が頂点であり，二人の研究者（キーワード）が，何らかの意味で似ている，この場合は二人の属性の類似度が，与えられたしきい値以上である場合枝で結ばれているグラフである．研究者，キーワードの属性として，Event，author を用い，類似度としてここでは，Jaccard 係数を用いる．著者 A の参加した研究会の集合を Event(A) とすると，「著者 A と B が類似する」⇔「Event(A) と Event(B) の Jaccard 係数がしきい値 θ 以上である」とモデル化できる．Jaccard 係数は $|\mathrm{Event}(A) \cap \mathrm{Event}(B)| / |\mathrm{Event}(A) \cup \mathrm{Event}(B)|$ で定義され，一般的によく利用される集合類似度の一つである．

今回は，I-Scover のデータから研究会投稿数が 5 回以上の研究者を抽出して類似度を計算した．類似度グラフの枝数は，しきい値 θ を 0.1 にした場合 14,000 程度，0.2 にした場合 11,000 程度であった．得られたクラスタは，大きさ 5 以上のものが，しきい値を 0.1 にした場合 536 個，しきい値を 0.2 にした場合 541 個であった．このほかのしきい値でも似たような値が観測され，しきい値の変化に対する変動が小さいことから，コミュニティの構造は比較的安定しているとみなしていいだろう．クラスタの平均の大きさはどちらも 10 程度であり，コミュニティの核の部分の大きさとしては受け入れやすいサイズである．現在，本会通信ソサイエティの研究会数（発表時）は，常設の研究会が 21，時限の研究会が 7，過去に終了し，継続していない研究会が 24 となっている．これらの数と比べると，今回発見したコミュニティ数は非常に大きい．しきい値が 0.1 の場合のクラスタに

ついて各人が所属しているクラスタの数を調べると，所属クラスタ数が 0 である人が 20 万程度，1 である人が 3,500 人程度で非常に多いが，2，3，4 である人の数も 407 人，44 人，3 人と，少なからず存在している．しきい値を 0.2 にした場合もほぼ同じであり，これらのことから，研究会をまたいだコミュニティがあること，一つの研究会に属するコミュニティでも，時間的に離れたもの，トピックが異なるもの，などが存在するようなことが，実際のクラスタの目視からも見て取れた．ある程度の割合の研究者は複数のコミュニティに所属し，コミュニティ構造は非常に多様性に富んでいることが考察される．

幾つかのコミュニティについて，その著者の 50% 以上が参加した研究会を調べてみたものが下記になる．

コミュニティ 1：
ICM2012-16，IT2006-1，ICM2011-52，ICM2011-51，ICM2012-45，ICM2012-25
コミュニティ 2：
IT2006-16，CS2009-54，IE2009-141，ITS2010-9，IE2010-114，IE2009-140，IE2010-116
コミュニティ 3：
SCE2010-20，IT2006-16，SCE2008-31，SCE2010-33，SCE2011-18，SCE2010-22，SCE2007-32，SCE2007-31
コミュニティ 4：
IT2006-16，RECONF2006-2，RECONF2005-14，RECONF2005-36，CPSY2004-67
コミュニティ 5：
IT2006-16，MI2006-84，MI2007-19，MI2006-83，MI2006-85，MI2006-54，MI2006-55，MI2007-18，MI2007-121

（ICM：情報通信マネジメント，IT：情報理論，IE：画像工学，ITS：高度交通システム，CS：通信方式，SCE：超伝導エレクトロニクス，CPSY：コンピュータシステム，RECONF：リコンフィギャラブルシステム，MI：医用画像）

例えば 2 番目のコミュニティは画像工学が中心であるが，通信や交通システムにも関わりを持っていることが分かるなど，示唆に富んだ情報が得られている．

同様の解析をキーワード：著者の関係でも行ってみたが，非常に大量のクラスタが生成されてしまった．キーワードを基にした研究者の類似性は，コミュニティ構造を反映しにくいであろうことが観察された．この方向から研究トピックやコミュニティをマイニングするためには，トピックに関連し得るキーワードの重要度を算定する，語のつながりをほかの方法で補完して強調する，といった解析が必要になってくるであろう．これらに関しては，将来的な課題である．

今回の研究では，研究分野におけるコミュニティ発見を研究会での発表を同時に行っている，という情報を基に発見する方法について提案した．元来このような研究会などを軸とした人脈やコミュニティは，研究や技術開発を行う上で非常に重要なファクタであると考えられ，それらに関わる情報収集がある種重要な意味を持つと言われてきたが，その人脈やコミュニティの発見は容易ではなく，実際に研究会に所属してその場の雰囲気から研究分野を構成する人員とそのつながりを調べていく，その研究会に所属する人物に直接中の状況を教えてもらう，というような形で俯瞰を試みることが多かったが，どちらも容易にできるものではなく，かつ誰でもできるというものでもない．論文共著関係から導くコミュニティは，この状況に一石を投じることはできたと考えるが，上述のように face to face を軸としたコミュニケーションの要因を考慮できないため，非常に強い結び付きの部分しか抽出できない．今回の分析の結果，研究会数よりも非常に多いコミュニティが発見されたことはある種の驚きである．研究会をまたいだコミュニティがあること，一つの研究会でも，中には複数のコミュニティ，時間的に離れたもの，分野的に離れたもの，などが存在することが示唆されている．

今回得られたクラスタを検索することで，ある人物がどのような研究コミュニティに参加しているのかを容易に得ることができるようになり，ある研究会に含まれるコミュニティにはどのようなものがあるか，また分野をまたがるコミュニティによりつながっている研究分野はどこであるか，といった情報が簡単に得られるようになった．近年は google scholar や SPYSEE などにより，個人の業績や個人と個人のつながりを調査することは容易になった．しかし，それらの業績や人的なつながりから全体を把握する作業はいまだ困難である．当研究のような人的交流を基にしたコミュニティ構造の解析により，このような人的情報の分析と利用に対して新たな方向性が加わり，全体と個人の両面を踏まえた情報収集が可能となっていくだろう．

宇野　毅明（正員）

1998-03 東工大大学院総合理工学研究科博士課程了，博士（理学）を取得．1998-04 東工大大学院社会理工学研究科経営工学専攻助手着任，2001-02 国立情報学研究所助教授着任．2014-04 同教授着任．2005-05 から 2006-08 までスイス連邦工科大に滞在．現在，情報学プリンシプル研究系教授．日本オペレーションズリサーチ学会，情報処理学会に所属．専門はアルゴリズムの理論と応用，特に離散アルゴリズム，列挙アルゴリズム，計算量理論，組合せ最適化など．データマイニング・データ解析・ゲノム情報学では，クラスタリングや類似性などの基礎計算を大規模データで高速に行う手法を研究．2010 文部科学大臣表彰科学技術部門若手科学者賞受賞．

◯優秀作品 1
「I-Scover データベースを用いた時系列・技術要因分析モジュールの開発」

槇　俊孝，若原俊彦（福岡工業大学）

現代社会は，人・もの・情報の流れが活発であり，価値観が多様化している．これにより，時間の経過に伴って常用語や技術用語の発生，変化，消滅が度々起こっている．例えば，記憶障害や判断力の低下で知られる「認知症」という用語は 2005 年頃から使われており，これより以前は「痴呆」と表現されていた．このような現象は様々な分野で生じていることが確認されており[(1)]，その時代の価値観によって表現が変化することがある．

計算機の処理能力や記憶能力は日々進化しているが，蓄積されている膨大なデータはその時々の状態を保持している．計算機は基本的に「認知症」と「痴呆」を別の用語として処理するため，「痴呆」のメタデータが付与されたデータを「認知症」では検索できない．この問題は計算機に「認知症」と「痴呆」が同義関係にあることを教示して解決できるが，手動で全用語に同義関係を教示することは非現実的である．同義関係による問題は数多くあり，近年注目されているビッグデータ分析も例外ではない．ビッグデータの時系列分析や特徴自動抽出等の研究[(2),(3)]は世界的に行われているが，日本語や英語等の自然言語を含むデータを正しく判読できなければ分析結果に大きな影響を及ぼす可能性がある．科学技術振興機構（JST）が提供している文献・特許データの分析ツール[(4)]の場合，「MANET」，「Mobile Ad Hoc Network」，「モバイルアドホックネットワーク」，「無線アドホックネットワーク」をそれぞれ時系列分析すると抽出できないデータが多々あり，同義語にもかかわらず異なった分析結果を出力することも度々あった．また，本会の I-Scover や国立情報学研究所の CiNii Articles，情報処理学会の電子図書館，IEEE の IEEE Xplore 等も意味解析相当の処理が施されていないので検索できない文献もあった．そこで応募者らは，従来システムやビッグデータ分析に容易に適用可能な Linked Data[(5)]形式の辞書 API（Wikipedia Link API）を新たに開発し，検索性能と分析性能の向上を目指すこととした．

・I-Scover チャレンジへの挑戦

我々は，本会情報・システムソサイエティのライフインテリジェンスとオフィス情報システム研究専門委員会（LOIS 研究専門委員会）に関わっており，2013 年の春頃，この研究会の動向を分析するために時系列分析システムの開発を検討していた．LOIS 研究専門委員会は，当初，オフィスシステム（OS）研究専門委員会として昭和 61 年（1986 年）に発足し，2009 年にオフィス業務だけでなく一般ユーザのライフログを含む身近な情報も対象とすることになり，研究会の名称にライフインテリジェンスが追加された．文献題目や著者名，キーワード等のメタデータを用いて LOIS 研究専門委員会の研究動向を分析したところ，名称が変更される 1〜2 年程前からライフログ関連の研究発表が増加していることが分かり，また発表件数も毎年 70 件程度に増加していることを確認した[(6)]．このように研究動向を分析することで研究トレンドが把握できることはもとより，組織活動の変遷もたどることができる．我々は，更に本会における研究会の位置付けを分析するため，文献メタデータの収集方法を検討していた．ちょうどこの頃，第 1 回目の I-Scover チャレンジが開催されることを知り，I-Scover に登録されている文献メタデータを入手できるという好

表 1　各文献検索システムにおける表記揺れの影響（調査日：2015 年 6 月 18 日）　検索結果数（件）

検索キーワード	I-Scover	CiNii	電子図書館
MANET	777	980	542
Mobile Ad Hoc Network	779	1,506	1,313
モバイルアドホックネットワーク	535	960	217
Smartphone	374	1,293	687
スマートフォン	1,033	4,776	3,652
スマートホン	5	46	44
スマホ	22	1,572	214

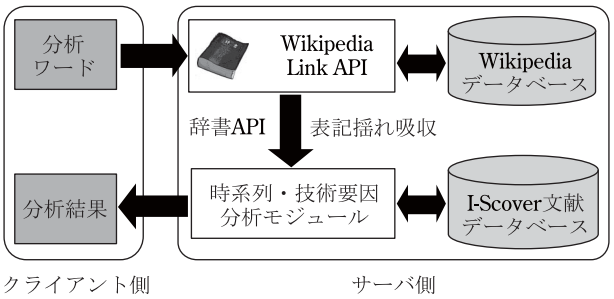


図 1　提案システムの構成図

表 2　Wikipedia Link API（WLA）の機能

type	機能	例
0	表記揺れ訂正 ほかの表記を取得	ようつべ→YouTube
1	日英・英日辞書	セキュリティ→Security
2	上位概念の取得	テキストマイニング∈｛データマイニング，自然言語処理｝
3	下位概念の取得	｛福岡，祭り｝∋｛小倉祇園太鼓，博多祇園山笠，博多どんたく，…｝
4	関連用語の取得	形態素解析→｛ChaSen, JUMAN, …｝

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:owl="http://www.w3.org/2002/07/owl#">
  <rdf:Description rdf:about="http://ictlab.ce.fit.ac.jp/wikilink.php?type=0&q=MANET">
    <rdfs:label>モバイルアドホックネットワーク</rdfs:label>
    <skos:closeMatch rdf:resource="http://ictlab.ce.fit.ac.jp/wikilink.php?type=0&q=MANET"/>
    <skos:closeMatch rdf:resource="http://ictlab.ce.fit.ac.jp/wikilink.php?type=0&q=モバイルアドホックネットワーク"/>
    <skos:closeMatch rdf:resource="http://ictlab.ce.fit.ac.jp/wikilink.php?type=0&q=Mobile ad hoc network"/>
    <owl:sameAs rdf:resource="http://ja.dbpedia.org/page/モバイルアドホックネットワーク"/>
    <owl:sameAs rdf:resource="http://ja.wikipedia.org/wiki/モバイルアドホックネットワーク"/>
    <owl:sameAs rdf:resource="http://www.wikipediaontology.org/instance/モバイルアドホックネットワーク"/>
  </rdf:Description>
</rdf:RDF>
```

図 2　WLA による「MANET」の表記一覧取得

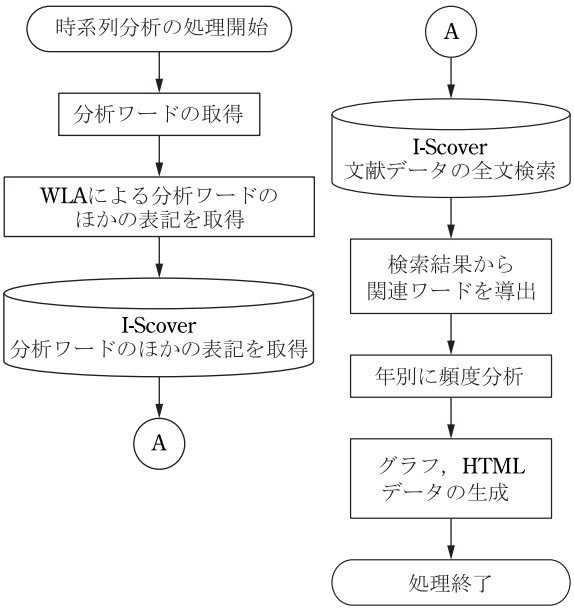
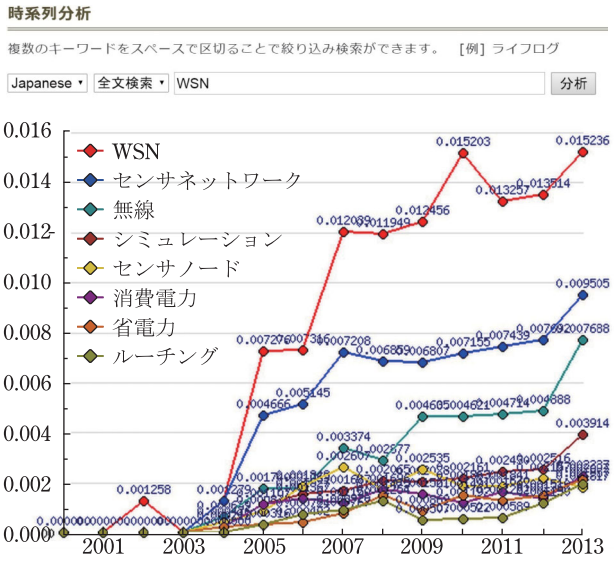
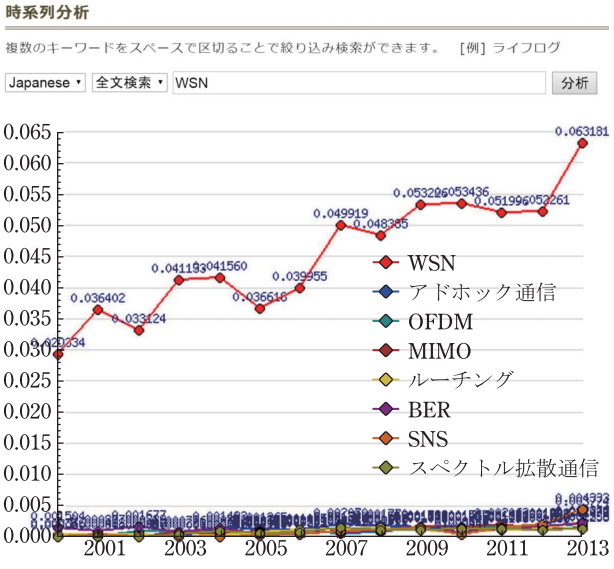


図 3　時系列分析モジュールの処理フロー



(a) 従来の時系列分析を実施した結果



(b) 意味関係を考慮して時系列分析を実施した結果

図 4　WSN（センサネットワーク）の時系列分析結果の比較

機をつかんだ．この I-Scover の文献メタデータは Linked Data 形式で約 16 万タイトルの文献が整理されているため，機械判読が容易であり信頼性の高い分析が可能である．CiNii Articles も文献メタデータを利活用できるように OpenSearch[7] という Web API を公開しており，1 回のクエリで最大 200 タイトルのメタデータが得られる．しかし，これを用いて信頼性及びリアルタイム性を追及することは難しい．一方，I-Scover では，約 16 万タイトルの文献メタデータを一度に取得でき，専用のデータベースを構築できるため高信頼性及びリアルタイム性を比較的容易に実現できる．我々は，I-Scover に大きな可能性を感じ，I-Scover チャレンジを通して学会論文や文献の各種分析モジュールの開発に取り組むことを決めた．

・提案システム

常用語や技術用語の表現は，時代の価値観や技術者の思想により表現が変化することがある．表 1 は，I-Scover，CiNii Articles（CiNii），電子図書館の各文献検索システムを用い，検索キーワードの表記揺れの影響を調査した結果である．同表から，検索アルゴリズム上の相違はあるが各文献検索システムは表記揺れによって検索結果数が異なることが分かる．MANET を例に挙げて I-Scover の検索性能を評価したところ，表記によって適合率は 0.52～1.00，再現率は 0.57～0.82，F 値は 0.60～0.90 のように変化することを確認した．この問題を解決するため，我々ははじめに辞書 API を開発し，次に時系列分析モジュール及び時系列分析結果の技術要因を

分析できるモジュールを開発することとした．図1に提案システムの構成を示す．

（1） Wikipedia Link API

Wikipedia データベースを用い，メタデータの自動補完を目的として Wikipedia Link API（WLA）を開発した[8]．Wikipedia は，語彙の網羅性と更新性が優れており，またブリタニカ百科事典と同程度の正確性[9]があるため，辞書 API の開発に適したデータセットと言える．WLA は，表2に示すように表記揺れ一覧の取得（type＝0）や日英・英日辞書（type＝1），上位概念の取得（type＝2），下位概念の取得（type＝3），関連語の取得（type＝4）の計五つの機能を有している．DBpedia[10]や Wikipedia オントロジー等の関連サービスが既に存在するが，WLA はメタデータの自動補完を目的としており，リアルタイム性を求めるために機能をタイプ分けし，処理コストを最小化している．

例として，下記URIを用いてサーバにアクセスすると図2に示すように「MANET」の表記一覧を取得できる．

http://ictlab.ce.fit.ac.jp/wikilink.php?type=0&q=MANET

現在は非公開であるが I-Scover データベースを導入した辞書 API を開発している．Wikipedia だけでは難しい専門性の高い技術用語にも対応できるため，これを I-Scover チャレンジ2014の応募作品にも利用している．

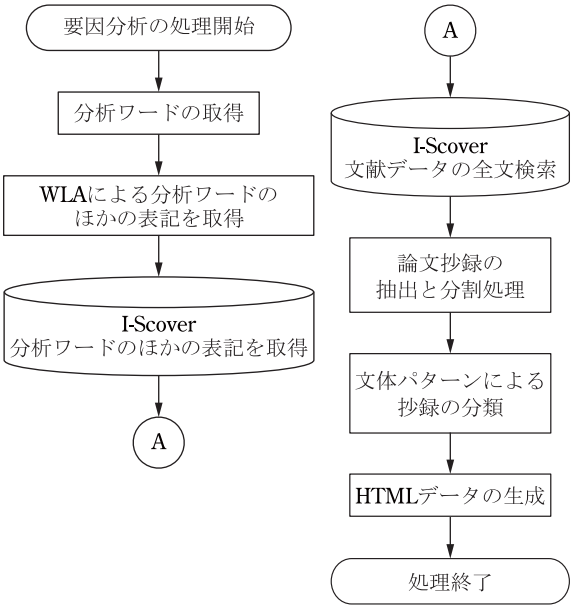


図5 技術要因分析モジュールの処理フロー

（2） 時系列分析モジュール

時系列的に研究動向を分析できるモジュールであり，I-Scover の論文メタデータを活用している．このモジュールは Wikipedia Link API を用いており，入力されたキーワードの意味関係を考慮した分析ができる．図3に時系列分析モジュールの処理フローを示す．図4は

図6 「センサネットワーク」の技術要因分析結果

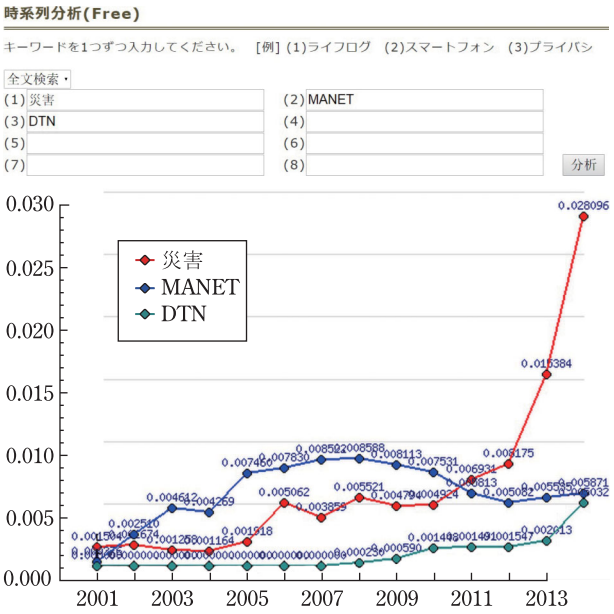


図 7 「災害」,「MANET」,「DTN」の時系列分析結果

「WSN（センサネットワーク）」を時系列分析した結果である．横軸は年代を表しており，縦軸は年別の全論文数に対する抽出された論文数の比率を表している．従来の時系列分析による結果を表した図 4 (a) から，「WSN」,「センサ」,「ワイヤレスセンサネットワーク」などの研究が 2005 年辺りから急激に増加している．これに対して意味関係を考慮して時系列分析した結果である図 4 (b) から上記の「センサネットワーク」関連の研究は「WSN」にまとめられて緩やかに活性化しているように見受けられる．このように意味関係を考慮することで正確なデータ分析を実現できることが分かる．

（3） 技術要因分析モジュール

時系列分析結果の技術要因を特定するためのモジュールであり，同様に I-Scover の文献（論文）メタデータを活用している．また，本モジュールにも辞書 API を導入しており，意味関係を考慮して分析する．一般的に論文は，「本研究では～である」,「～という課題がある」,「～が期待されている」等のように一定の文体が存在しており，比較的容易に特徴抽出が可能である．ただし，文体の種類は多くあり，表記違いの文も数多くあるため辞書 API を用いることで定義を簡略化している．図 5 に技術要因分析モジュールの処理フローを示す．図 6 は，「センサネットワーク」の技術要因を分析した結果である．本モジュールは，論文メタデータに含まれる概要を文単位に分割し，文体ルールに従って意味関係を考慮して一般，動向，注目，課題，例示の計 5 種類に自動的に分類する．

本作品により技術要因分析及び技術要因分析をリアルタイムに行えるようになり，動向調査の効率化が期待できる．本作品の研究開発はこれからも継続する予定であり，実用化して近日公開予定である．

・自然災害とネットワークに関する動向調査

近年，東日本大震災や浅間山噴火等を含む自然災害，及びこれに基づく二次災害に関するニュースが連日報道されている．これに対して，MANET や DTN（Delay/Disruption Tolerant Networking）等を含む通信ネットワークの研究が世界的に行われている．図 7 は，「災害」,「MANET」,「DTN」をそれぞれ時系列分析した結果である．同図から，「災害」に関する研究が 2012 年頃から急激にトレンド化していることが分かる．

また，同時期頃から「DTN」に関する研究がトレンド化しているように見受けられる．このため，「災害」と「DTN」には何らかの関係性があることを推察できる．この関係性を明らかにするために「DTN」の技術要因分析を行う．これにより図 8 に示すような分析結果が得られ，動向に「災害時に DTN が有効」という旨の内容が提示されている．このように本作品は，複数のキーワードを用いて研究動向を調査することも可能である．

常用語や技術用語は表記揺れが存在しており，データ

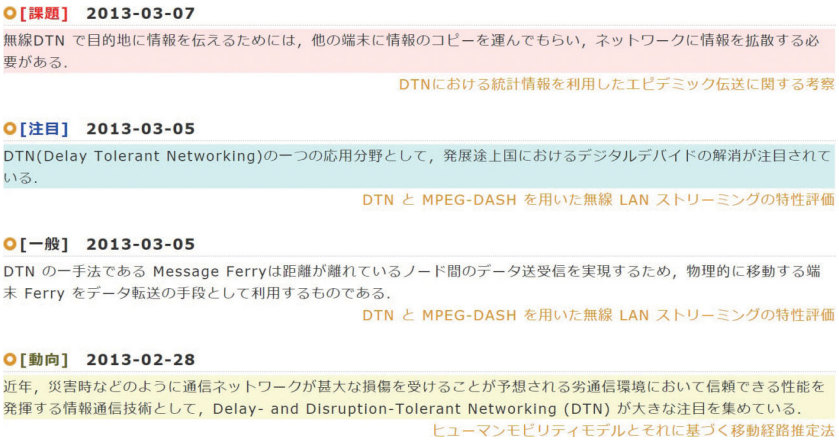
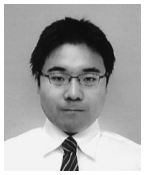


図 8 「DTN」の技術要因分析結果

の検索や分析を行う上で見過ごせない問題となる．このため我々は，メタデータの自動補完を目的としたWikipedia Link APIを開発し，これを導入した時系列・技術要因分析モジュールで有効性を確認した．今後は，各モジュールの実用化を目指して開発研究に取り組み，近日一般公開する予定である．



槙　俊孝（学生員）

2013福岡工大・工卒．2015同大学院修士課程了．現在，同大学院博士後期課程在学中．Webアプリケーション，自然言語処理及びコミュニティシステムの研究に従事．2014本会学術奨励賞，2014本会情報・システムソサイエティ優秀ポスター賞，2014 I-Scoverチャレンジ2013優秀賞，2014 LOIS若手研究者賞，2015 I-Scoverチャレンジ2014優秀賞ほか各受賞．

若原　俊彦（正員：シニア会員）

1970東工大・工・電子物理卒．1972同大学院修士課程了．同年日本電信電話公社（現NTT）電気通信研究所入所．電気通信システムの研究実用化に従事．博士（工学）．1999早大国際情報通信研究センター客員教授．2003から福岡工大・情報工・情報通信・教授．2012-2013本会ライフインテリジェンスとオフィス情報システム（LOIS）研究専門委員会委員長．2014 I-Scoverチャレンジ2013優秀賞，2015 I-Scoverチャレンジ2014優秀賞，2015本会情報・システムソサイエティ活動功労賞各受賞．IEEE，情報処理学会，画像電子学会各会員．

◯優秀作品2

“Topic-sensitive expert finding and profiling for I-Scover”（英語）

Ruiyu Fang, Lu Fang, Qingliang Miao, Yao Meng (Fujitsu R & D Center in Beijing)

・Exploring I-Scover Data

I-Scover, officially known as IEICE Knowledge Discovery, is an academic search engine. I-Scover provides information about publications, authors, and keywords, these data is organized as linked data by a uniform schema. Although linkage characteristic of linked data has already brought convenience for bibliographic information discovery through exploratory search [11]. We believe more interesting functions can be explored based on linked data to satisfy user's demands.

In this paper, we assume that users shall hold requirements to find out experts within specific research fields. Additionally, they would like to acquire more information about experts besides current I-Scover data. To address above issues, we propose an expert finding and profiling tools as illustrated in Fig. 1 Expert finding focuses on identifying persons with relevant expertise or experience for the given technical keyword. We adopt a topic-sensitive ranking approach that uses I-Scover data : entities and relationships among entities (e. g. authors, publications, keywords and heterogeneous relationships between authors and authors, authors and publications, publications and keywords) to rank authors. And expert profiling aims at integrating I-

Fig. 1　Main page of expert finding and profiling tools（http://59.108.112.7/Challenge/ExpertFindingProfiling.jsp）

Fig. 2 Topic-sensitive expert finding (top 3 experts) Left region is a topic navigation, selecting the topic "ライフインテリジェンスとオフィス情報システム" leads to ranking experts related to this topic.

Scover data with relevant data extracted from the web to build comprehensive pictures of authors, we use web information extraction methods to complement authors' attributes, e. g. homepages, experiences and awards. User study results show our expert finding and profiling tools are helpful for academic search.

Section 2 describes our expert finding system and includes some use cases, section 3 describes how we do expert profiling for authors in I-Scover and presents screenshots for illustration. In Section 4, we summarize our system and give future works.

・Expert Finding : Find Top Experts in I-Scover

On I-Scover website, we assume that a user with a specific demand to find the most authoritative authors or experts within certain domains, and authoritative authors refer to those who have scholarly influences in their research fields. And we call this demand expert finding, which is also important for academic search and mining. In our expert finding system, a search query can be the "technical keyword" entity in I-Scover [12], but different from what I-Scover retrieves [13], we merely focus on finding out and ranking authoritative authors related to the given keyword. We first find out all the authors who have mentioned the keyword in one of their publications, and then we do ranking based on several influence indexes. Notice that many technologies, specified by the query keywords, may be used in different research fields, e. g. "SVM" technology can relate to various research fields such as "画像," "音声" and "情報". In addition, an expert always dabbled in various research fields simultaneously or at different time period. Due to this fact, our searching result should also be research fields or topics related.

In a word, Expert finding finds out the leading authors over different research fields with given query keyword, and authors are ranked separately with respect to different topics as well. Fig. 2 shows a screen of the search results when we want to find experts who used "I-Scover" in their research fields. The retrieve experts are categorized into several topics concerning their related research fields. For topic definition, we pre-defined 89 topic categories according to IEICE 2013 general conference's publication taxonomy (†1). We simply display top three authors for each index.

According to our experiences, we usually concern the scholarly life, paper number and co-authors and other indexes to rank an expert. We define five indicators (indexes) to reflect the authorities of authors as follows, and by manipulating the weights of these indicators with our own criteria, we get different ranking results. E. g., by setting Longevity's weight to 1, article number' weight to 2 and others' to zero. We get a ranking list considering only the academic life and publications of authors, and value the article number index twice more important than his academic life, as Fig. 3 reveals.

(†1) http://www.ieice.org/jpn/event/program/2012G/Settings/html/info/group.html#iss

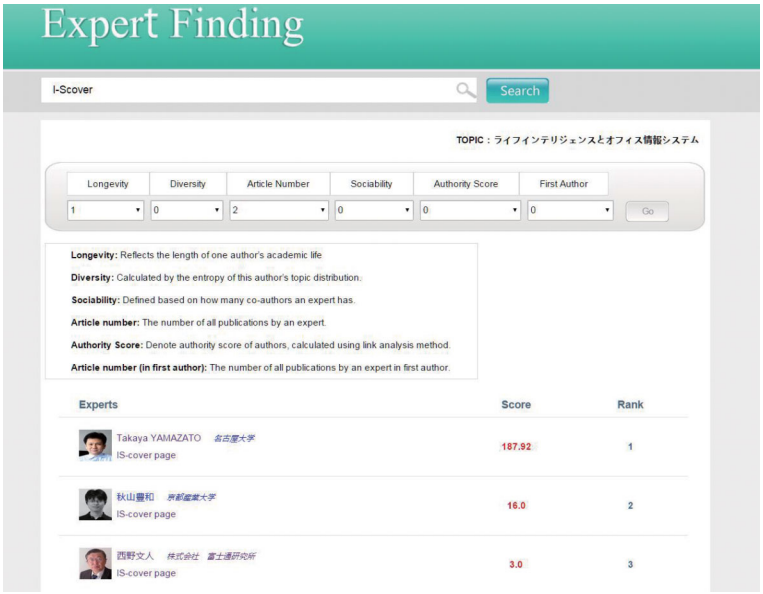


Fig. 3　Expert ranking with weight regulation

Longevity : Longevity reflects the length of one author's academic life. We consider the year when one author published his/her first paper as the beginning year of his academic life and the last paper as the end year.

Sociability : basically defined based on how many co-authors an expert has.

Diversity : Generally, an expert's research may include several different research fields. Diversity is defined to quantitatively reflect the degree, and is then calculated by the entropy of this distribution.

Article number : the number of all publications by an expert.

Authority Score : denote authority score of authors, similar to PageRank score on webpages, and is calculated based on link analysis on I-Scover data network like Nie proposed in [14].

Article number (in first author) : the number of all publications by an expert in first author.

・Expert Profiling : Integrating I-Scover Data with Web Data

Detailed information about a person is widely distributed on the web, in the forms of semi-structured and unstructured data. Person information may come from various types of website sources like homepages (e.g. working place homepages, organization homepages and personal homepages), news releases (e.g. news about his research achievements and projects) and third-party scholarly websites (like research-map, kanken and so on). In current version of I-Scover, a small fraction of information about an author is covered relative to his detailed information on the web, the I-Scover author entity only holds metadata which represent academic facets, like author name, publication and affiliation, a lack of information that reflects more details of other aspects of authors makes deep understanding of them impossible. However, according to our preliminary experiment results, about 42.77% authors in I-Scover have external homepages, and utilizing these information will enormously complement I-Scover author data. To this end, we propose our expert profiling tools, which integrates I-Scover data with external data extracted from the web to build a comprehensive picture of an author.

In practice, we use web information extraction method to acquire complementary information for authors in I-Scover, such as portrait, homepage, affiliation, experience and award. Firstly, we collect necessary attributes like authors' names and affiliations from I-Scover to locate and gather relevant webpages via global search engines (†2), and I-Scover information can also play an important role in webpage disambiguation, like we use author's affiliation to disambiguate those same name homepages, and then more fine-grained attributes are extracted using information extraction methods.

By this way, we finally generate profile for every

(†2)　http://search.yahoo.co.jp

Fig. 4　Expert profiling page　Clicking on homepage hypertext brings you to his homepages.

Fig. 5　Topic area of expert profiling page, which reveals that“無線通信システム”is his major research area during 2000 to 2013, and his last research interest is“ライフインテリジェンスとオフィス情報システム”

author in I-Scover, and by clicking author's name in Fig. 2, 3, we can view his homepages, research area, experiences and award information, etc., like Fig. 4 presents. Among them, Research areas, experiences and awards are extracted from his homepages, and homepages are identified using classification methods like[15] illustrated. Keywords region displays keywords used in his publications with word frequency considered.

Topic area (in Fig. 5) are generated using topic identification for publications. Topic area displays the temporal trend of his research interests and leads to full knowledge of which research interest (research area) he was focusing on at different times.

・Conclusion and Future Works

In this paper, we describe the achievements that I-Scover already made using linked data, and more than this, to promote the system usability and meet user's demands of knowledge discovery and mining on I-Scover data, we propose our expert finding and profiling

tools. We adopt data mining methods on I-Scover data and analyze author impacts in each research area, which is helpful for users to know the leading experts in target research fields, and also, we acquire complementary information from word wide web to build I-Scover authors' profiles. We believe our tools will bring convenience and practicability for I-Scover.

As future work, we will extract complementary information for entities like events and organizations in I-Scover, more than just authors. And events in I-Scover can also be ranked with their impacts on different research areas considered.

Ruiyu Fang

Ruiyu Fang is a Researcher in Fujitsu Research & Development Center. He received his master degree from Xiamen University in 2013. He has been engaged in research and development of natural language processing including machine translation, information retrieval, and knowledge Graph.

Lu Fang

Lu Fang is a Researcher in Fujitsu Research & Development Center. She received her master degree from Soochow University, China, 2011. Her primary research interests include linked open data, natural language processing and Web information extraction.

Qingliang Miao

Qingliang Miao is a associate research manager in Fujitsu Research & Development Center. He received his Ph.D. degree in pattern recognition and intelligent systems from Chinese Academy of Sciences, China, 2011. His primary research interests including linked open data, knowledge extraction and management and natural language processing.

Yao Meng

Yao Meng is a research manager in Fujitsu Research & Development Center. She received her Ph.D. degree in computer science from Harbin Institute of Technology, China. Her research focuses on natural language processing, with broad applications on Web information processing and machine translation.

4. おわりに

本会の内外において，I-Scover は着実に認知度を向上させている．I-Scover を活用することにより，本会の文献へのアクセスが増えている．また，I-Scover 第 2 期システムのアプリケーションインタフェースを活用することで，論文の傾向分析などをシステマティックに実現可能となることが期待される．今後は，I-Scover やその他の学会活動を通して，本会のコンテンツの価値がより向上し，学会の収益にも貢献していけるよう努力していく所存である．

文　献

（1） P.M. Greenfield, "The changing psychology of culture from 1800 through 2000," SAGE journals, Psychological Science, Aug. 2013.
（2） 松原靖子，櫻井保志，C. Faloutsos，"大規模時系列データからの特徴自動抽出，"日本データベース学会，DEIM2014, D4-2, March 2014.
（3） S. Gupta and C.D. Manning, "Analyzing the dynamics of research by extracted key aspects of scientific papers," Proceeding IJCNLP, pp. 1-9, 2011.
（4） 科学技術振興機構，"J-GLOBAL 分析ツールβ版，" June 2015, http://foresight.jst.go.jp/analyzer/
（5） C. Bizer, T. Heath, K. Idehen, and T.B.-Lee, "Linked data-The story so far," IJSWIS, vol. 5, no. 3, pp. 1-22, Aug. 2009.
（6） 若原俊彦，槇　俊孝，岡本　学，山元規靖，茂木　学，小舘亮之，"LOIS 研究の動向分析(3)～文献検索 I-Scover とその応用システムを利用した分析～，"信学技報，LOIS2014-9, pp. 97-102, May 2014.
（7） CiNii Articles，"メタデータ・API-CiNii Articles 論文検索の OpenSearch，" June 2015, http://support.nii.ac.jp/ja/cia/api/a_opensearch
（8） 槇　俊孝，髙橋和生，若原俊彦，"メタデータの自動付与のための Wikipedia リンク API を用いた論文データの類似度評価の一検討，"信学論（D），vol. J97-D, no. 12, pp. 1705-1708, Dec. 2014.
（9） J. Giles, "Internet encyclopedias go head to head," Nature, vol. 438, pp. 900-901, Dec. 2005.
（10） J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P.N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "DBpedia-A large-scale, multilingual knowledge base extracted from Wikipedia," Semantic Web Journal 1-5, pp. 1-29, 2012.
（11） 西野文人，"Linked Data～つながるデータ，広がるサービス～" 信学通誌，no. 23, pp. 240-244, Dec. 2012.
（12） T. Yamazato, Y. Chimura, F. Nishino, and S. Ikada, "IEICE Knowledge Discovery（I-Scover)," 2014.
（13） 西野文人 "I-Scover-Linked Data に基づく電子情報通信学会文献検索システム，" 信学通誌，no. 25, pp. 49-53, June 2013.
（14） Z. Nie, Y. Zhang, J. Wen, and W. Ma, "Object-level ranking : Bringing order to Web objects," Proc. WWW, pp. 567-574, 2005.
（15） S.D. Gollapalli, C. Caragea, P. Mitra, and C.L. Giles, "Researcher homepage classification using unlabeled data," Proc. WWW, pp. 471-482, 2013.

（平成 27 年 6 月 30 日受付　平成 27 年 7 月 22 日最終受付）

千村（ちむら）　保文（やすぶみ）（正員）

昭 56 日大・理工・電気卒．同年沖電気工業株式会社入社．以来，データ通信機器，VoIP の研究開発，標準化活動に従事．現在，経済・政策調査部上席主幹．平 26 年度 TTC 会長賞受賞．著書「SIP 教科書」など．