

解説

知識の横断検索 I-Discover

IEICE Knowledge Discovery (I-Discover)

山里敬也 千村保文 西野文人 伊加田恵志
秋山豊和 中野美由紀 井上友二

Abstract

電子情報通信学会の新しい文献検索システムである I-Discover (正式名称: IEICE Knowledge Discovery, 呼称: アイスカバー) は, 単に本会文献の横断検索を実現するシステムではない. 文献の関連情報も文献と有機的に結び付けることで, インターネット上で顕在化させ, 広く参照できるようにして, 本会文献の価値を最大化することを目指すシステムである. また, 従来の検索システムでは困難な, いちご摘み探索による新たな発見も可能である. 本稿では I-Discover 開発の経緯と狙い, 機能と特徴を紹介し, 普及推進に向けての活動計画と今後の開発予定について述べる.

キーワード: 文献検索システム, メタデータ, Linked Data, いちご摘み探索

1. はじめに

科学者であり文筆家でもあった寺田寅彦は「科学というものの内容も, よく考えてみるとやはり結局は「言葉」である」と書いた⁽¹⁾. 寺田は, 研究成果を論文にすることの重要性を述べるとともに「言葉」の重要性にも触れている. 適切な「言葉」でつづられた論文は記録になると同時に予言となる, と説く.

インターネット上で急激に増加し続ける膨大な数の論文の中からある論文が閲覧されるかどうか, この「言葉」の選び方に左右される. これは, 検索する者からすれば, 適切な検索ワードの選択, そして絞り込み技術にたけていなければ, 意図する論文にたどり着けないことを意味する. たとえ優れた論文であっても, 見つからなければ埋もれてしまい, 結果として評価が低くなる.

さて, 検索の対象となるデータを要約したデータのことをメタデータと呼ぶ. 例えば論文の場合, 論文タイトル, 著者名, キーワード, 抄録, 出版物名, 発行年月日などの書誌情報がメタデータである. 検索システムでは, このメタデータを手掛かりに検索を行う. 検索結果として閲覧されて初めて論文として認められるのであれば, 寺田の説く論文における「言葉」の重要性は, 現在では「メタデータ」の重要性となるであろうか.

電子情報通信学会(以下, 本会)の新しい文献検索システムである I-Discover (IEICE Knowledge Discovery)^(注1) は, 単に本会の文献(和英論文誌, 技術研究報告, 大会論文, 国際会議論文)の横断検索を実現するシステムではなく, 検索対象となる文献に付与されているメタデータを統合的に整理して提示するシステムでもある⁽²⁾. 更に, そのメタデータに関連する著者, キーワードなどもそれぞれ別のメタデータとして持ち, これら関連するメタデータ同士を互いに意味を有するリンクで結び付けて

山里敬也 正員:フェロー 名古屋大学教養教育院教養教育推進室
E-mail yamazato@m.ieice.org
千村保文 正員 沖電気工業株式会社研究開発センター
E-mail chimura730@oki.com
西野文人 正員 (株)富士通研究所ソーシャルイノベーション研究所
E-mail nishino@jp.fujitsu.com
伊加田恵志 正員 沖電気工業株式会社研究開発センター
E-mail ikada333@oki.com
秋山豊和 正員:シニア会員 京都産業大学コンピュータ理工学部ネットワークメディア学科
E-mail akiyama@cse.kyoto-su.ac.jp
中野美由紀 正員:シニア会員 芝浦工業大学教育イノベーション推進センター
E-mail miyuki@shibaura-it.ac.jp
井上友二 正員:フェロー (株)トヨタ IT 開発センター
E-mail yuji@m.ieice.org
Takaya YAMAZATO, Fellow (Institute of Liberal Arts and Sciences, Nagoya University, Nagoya-shi, 464-8603 Japan), Yasubumi CHIMURA, Member (Corporate Research & Development Center, Oki Electric Industry Co., Ltd., Warabi-shi, 335-8510 Japan), Fumihito NISHINO, Member (Social Innovation Laboratories, Fujitsu Laboratories Ltd., Kawasaki-shi, 211-8588 Japan), Satoshi IKADA, Member (Corporate Research & Development Center, Oki Electric Industry Co., Ltd., Warabi-shi, 335-8510 Japan), Toyokazu AKIYAMA, Senior Member (Faculty of Computer Science and Engineering, Kyoto Sangyo University, Kyoto-shi, 603-8555 Japan), Miyuki NAKANO Senior Member (Center for Promotion of Educational Innovation, Shibaura Institute of Technology, Tokyo, 135-8548 Japan), and Yuji INOUE, Fellow (Toyota Info Technology Center Co., Ltd., Tokyo, 107-0052 Japan).
電子情報通信学会誌 Vol.97 No.3 pp.240-246 2014年3月
©電子情報通信学会 2014

(注1) I-Discover 検索サイトの URL: <http://i-scover.ieice.org/>

いる（これを Linked Data 化と呼ぶ⁽³⁾）。このように、I-Discover は、本会の文献に関連情報も含めて有機的に結び付けることで、インターネット上で顕在化させ、広く参照できるようにして、本会文献の価値を最大化することを目指している。

メタデータとメタデータを意味も踏まえてリンクで結ぶ仕組みは、文献探索としても機能する。詳しくは後述するが、リンクをたどっていくうちに思わぬ発見も期待できる。つまり、I-Discover はセレンディピティ（何かを探しているときに、探しているものとは別の価値あるものを見つける才能、能力）につながる使い方も可能である。

本稿では、まず I-Discover 開発の経緯と狙いについて述べ、I-Discover の特徴、機能の概要と I-Discover でできることを述べる。最後に普及推進に向けての活動計画と今後の開発予定について述べる。

2. 知識の横断検索を目指す I-Discover

2.1 I-Discover 開発の経緯

会員数の長期的な減少、とりわけ企業に在籍する会員の減少は顕著であり、本会の活性化が急務となる中、伊澤会長（平成 18～19 年）の私的諮問「学会の活性化（青山主査）」の中で、IEEE（Institute of Electrical and

Electronics Engineers）の文献検索システムである IEEE Xplore と同様のシステムを構築すべきことが主要活性化策の一つとして答申された。海外においては IEEE や ACM（Association for Computing Machinery）の拡大が著しく、とりわけ IEEE Xplore は IBM（International Business Machines Corporation）ジャーナル、IET（The Institution of Engineering and Technology）ジャーナルも掲載するようになり存在感が増していた⁽⁴⁾。そのような状況で、IEEE から本会英文論文誌を IEEE Xplore に取り込みたいとオファーがあった（平成 23 年 6 月編集連絡会）。これを通信ソサイエティ（以下通ソ）執行委員会で御議論頂いたところ、最終的には電子情報通信学会版の Xplore の本格検討を開始することが承認された。また、本会全体としてどう進めていくかについて、編集連絡会、企画室で議論した結果、通ソが先行する形で具体化を進め、それをたたき台にして本会全体のシステムとして構築することが合意された。

平成 24 年 1 月に RFP（Request for Proposal）を発出し、システム構築業者からの提案を評価した結果、第 1 期システムの構築業者及び構築範囲を通ソとして決定、理事会の承認を得て、同年 4 月に I-Discover の構築が正式に決定された。その後、毎回 6 時間以上に及ぶシステム構築業者との会合を週 1 回のペースで続け、7 月に基本設計書、11 月に詳細設計書の完成、製造、システム

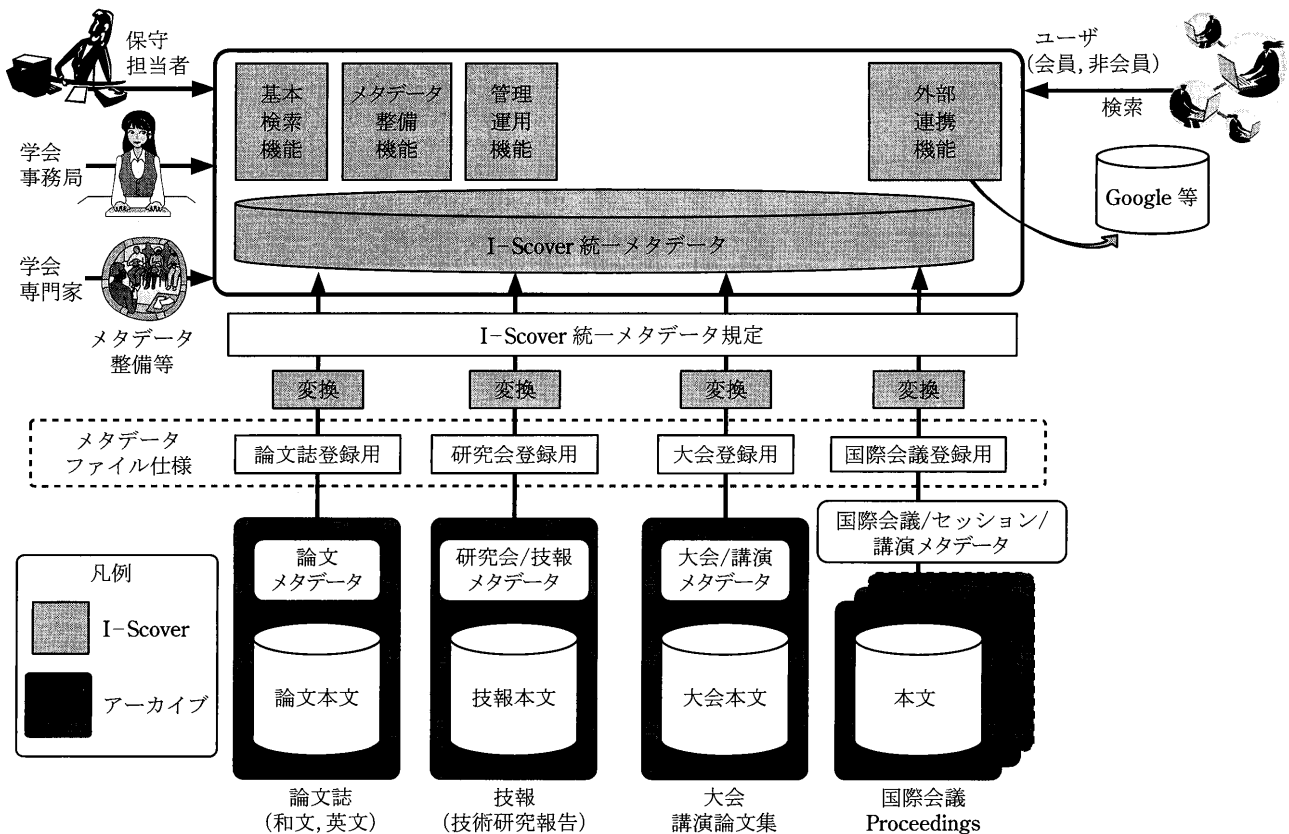


図 1 I-Discover の概要

試験, 納入試験を経て, 平成 25 年 4 月 3 日に第 1 期システムをリリースしている. 更に, 使い勝手等のマイナーバージョンアップを経て, 現在に至っている (第 1.1 期).

通ソ執行委員会での主な議論

今後ますます国際化していく中で IEEE Xplore に英文論文誌が掲載されることを希望する会員も多いであろう. 論文投稿や国際会議での発表を積極的に行っているアクティブ会員にとってのメリットは理解できる. 一方で, 会誌, 和英論文誌, 技報等の閲読を中心にしたパッシブ会員も多数いる. この者たちのメリットは何か. 慎重に検討する必要がある.

アクティブ会員にとっては IEEE 会員になれば済むことだがパッシブ会員のメリットは少ない. 本会が IEEE の配下に入ったとみなされたら会員減少に拍車がかかる. ましてや, 海外会員を引きとめる理由はなく, 国際学会を標ぼうする本会の理念とは逆行する. 更に, 産業界の会員減少に歯止めがかかっていない傾向は日本の産業界の現状を彷彿させるところがあり, 本会がどこを目指すのか明確にすべき時期に来ている.

原点に立ち戻り, 本会が, 標準化や人材育成, 最終的には学術はもとより産業が, 弱肉強食の国際社会で生き残るための切磋琢磨の場となり, コンセンサスの場となる必要がある. そのためには, IEEE とは異なる国際学会としての特色を出すべきであり, その第一歩として IEEE Xplore への対抗があるであろう. IEEE は世界中にサービスを提供するための ICT (Information and Communication Technology) 装備と, 社会への発信や連携のための広告・販促費用に巨額を投入している. 竹やりでは戦闘機と戦えない. 本会のアイデンティティの主張を行うには IT システムへの投資が不可欠である.

2.2 I-Discover の狙い

図 1 に I-Discover の概要を示す. I-Discover は, 和英論文誌, 技術研究報告, 大会論文などの既存個別アーカイブを活用しつつ, これらを包含するような仕組みをとる. 各出版物のメタデータを個別アーカイブから抽出し, これを統一メタデータとして整理することで横断検索を実現している.

また, I-Discover はアーカイブ機能を持たない (システム内に文献は持たない) ため全文検索はできない. このため, Google や IEEE Xplore 等の従来の検索エンジンとは補完関係となる新たな知識集約型のシステムを目指している. 具体的には, 近年, 注目されつつある Linked Data を採用し, メタデータの一つである技術用語の解説文を専門家が作成することにより, メタデータと有機的につながる文献自体の価値を上げることができるシステムを構築している. いわば, 研究活動に対する「目利き」機能を Linked Data として組み込むことで, 自分たちで自らの文献の価値の向上を図ることができるシステムを目指している.

2.3 I-Discover の機能と特徴

一般的な文献検索サイトでは, 書誌情報をデータベース化し, 検索結果として見つかった文献に関する情報を一覧化して表示している. これに対し, I-Discover では, 文献に関する書誌情報 (文献メタデータ) に加え, 著者, (所属) 機関, イベント, 出版物, キーワードもメタデータとして保持している. これにより, ある研究者が執筆した文献リスト, ある技術用語をキーワードとして設定している文献リストを得ることができる.

例えば, 図 2 に示すように, ある文献の関連情報である著者, 所属, キーワードなどのメタデータ同士は互い

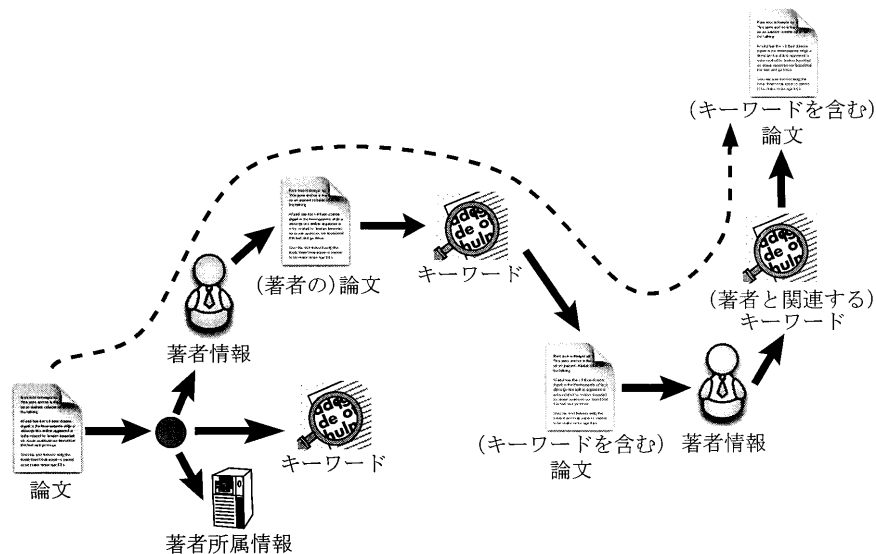


図 2 いちご摘み探索

に意味を有するリンクで結び付けている (Linked Data 化)。このため、以下のような検索が可能である。

- ・ ある研究者が執筆した文献に付与されているキーワードは？
- ・ あるキーワードが付与されている文献の著者は？
- ・ ある文献と同じキーワードを持つ他の文献に付与されているキーワードは？

このように I-Scover では、あるメタデータ (例えば著者メタデータ) から別のメタデータ (例えばキーワードメタデータ) へと次々と視点を変えて、情報を拾い集めながら文献探索を行うことができ (これをいちご摘み探索⁵⁾ という)、単に文献や研究者を探すといった一般的な検索にとどまらず、セレンディピティにつながる使い方も可能である。

更に、Google, IEEE Xplore, ACM Digital Library, CiNii など、ほかの (文献) 検索サイトによる検索結果を I-Scover 自体の検索結果と併せて表示する機能も用意している。これは、電子情報通信分野の研究者・技術者にとっての一種のポータルサイトとしての利用も可能となる。

ところで、I-Scover でデータベース化されている各種メタ情報のうち、技術用語のメタデータであるキーワードメタデータについては、その技術用語の意味、英語との対応付けなどを、本会会員である各技術分野の専門家が整備していく仕組みを構築している。このキーワードメタデータ整備活動を通じて文献の価値を高めることができる。もっとも、この作業は一朝一夕には行え

ないが、時間がたつにつれ、単に検索できる文献数が増えるだけでなく、キーワードメタデータ整備活動を通じて、データの質が高まり、より一層使いやすいサイトになるものと期待している。各研究専門委員会の多大な御尽力により、約 1,000 語のキーワード解説文を作成頂いており、JSPS 科研費を活用して作成している英訳文と併せて、I-Scover にデータ投入予定である。

3. I-Scover の普及に向けて

普及活動概要、普及ツールほかについて説明する。なお、本活動の一部については、平成 25 年度科学研究費助成事業 (科学研究費補助金：研究成果公開促進費) により行っている。

3.1 普及活動概要

平成 25 年 4 月の I-Scover のサービス開始に伴い、I-Scover の利用メリット訴求のために、会誌やソサイエティ誌に紹介記事を掲載するほか、パンフレットの配布、ホームページの充実、Twitter での情報発信や質問への回答などを実施している。また、これらの媒体による普及活動以外に、研究会への宣伝活動、海外展開のための訪問といった対面での活動も必要と考え、紹介活動も行っている。更に、I-Scover 及び文献メタデータの利活用を促進するため、後述する「I-Scover チャレンジ 2013」というコンテストを実施している。また、検索対象のコンテンツの拡大要望に基づき、会誌及び FIT (情報科学技術フォーラム：Forum on Information Technology) 論文なども検索対象とする拡張を行って

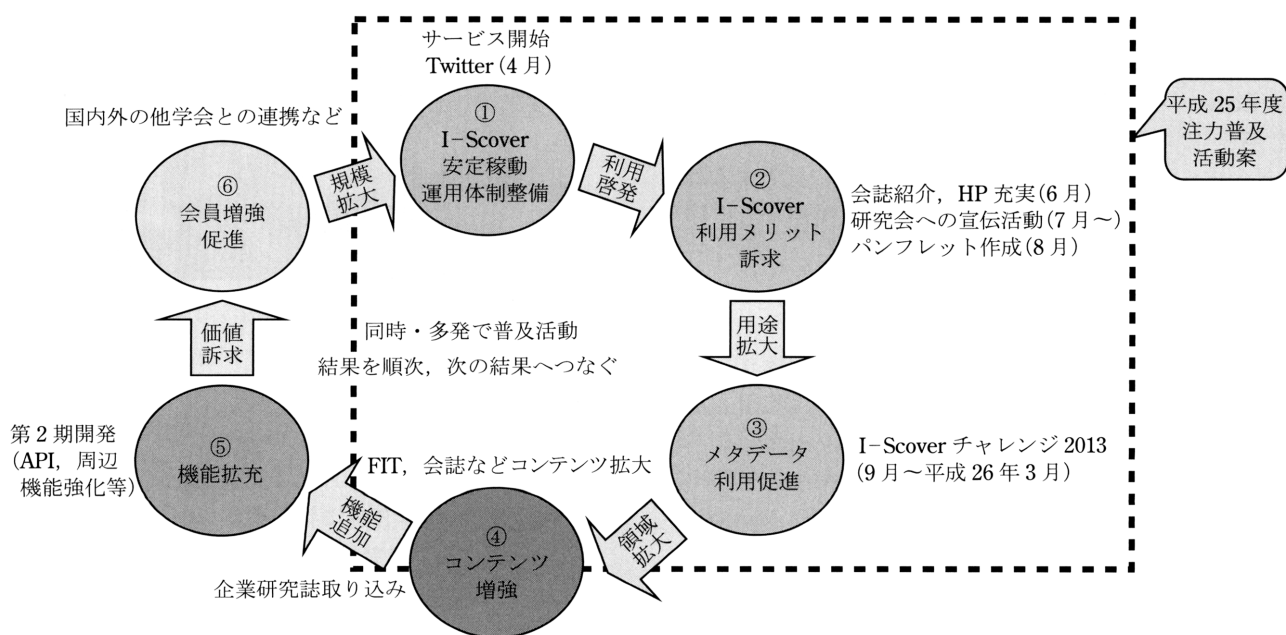


図 3 I-Scover 普及活動



図4 I-Scoverパンフレット（日本語版）

いる。ここまでを平成25年度の注力活動として実施している。今後は4. に述べる開発予定を踏まえ、更なる普及活動を行っていく（図3）。

3.2 普及ツールの拡充

I-Scoverの普及のために普及ツールとしては、日本語、英語でのパンフレットを作成（図4）し、研究会や新規会員へ配布した。また、説明機会を更に拡大するために、「I-Scover紹介ビデオ」を日本語、英語ともに作成予定である。

3.3 研究会への宣伝活動、海外展開

地道な紹介活動としては、各研究会やソサイエティ大会等に参加し、I-Scoverの目的、メリット、利用方法について説明を行っている。また、これらの活動は国内だけでなく、タイなどアジア諸国や欧州といった国際セクション向けにも精力的に行っている（図5）。

3.4 I-Scoverチャレンジ2013

メタデータを活用したI-Scoverの特徴を理解頂き、更なる利活用法を多くの方から創出して頂くために「I-



図5 海外展開活動例（タイ・チュラロンコン大学訪問）

Scoverチャレンジ2013」というコンテストを平成25年9月から平成26年3月まで開催した。コンテストの結果は、平成26年3月に新潟大学で開催される総合大会にて発表する予定である。

4. 今後の開発予定

現在、FIT と会誌を I-Scover に取り込むべき作業を進めており、平成 26 年 4 月にはこれらが検索対象になる。また、企業各社が発行する論文誌を I-Scover で検索可能とするために各社と交渉を行っており、第 2 期開発を完了後には、I-Scover から各社の企業論文誌の検索が可能となる予定である。更に、I-Scover 第 2 期システムでは、和英論文誌及び英文レター誌の参考文献を新たにメタデータとして取り込むことを考えている。

ところで、I-Scover では、メタデータを Linked Data としてデータベース化しているが、外部 API 提供による他システムから利用可能な形として公開、すなわち LOD (Linked Open Data) 化は行っていない。外部 API 提供にはサーバの負荷増加や 5. で述べる著作権の問題を解決する必要があるため、これらを解決した上で、第 2 期システムで実装する予定である。その前段階として、3. で述べた I-Scover チャレンジ 2013 におい

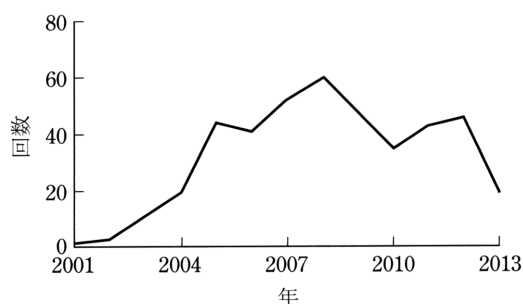


図 6 キーワード「MIMO」の年ごとの出現回数

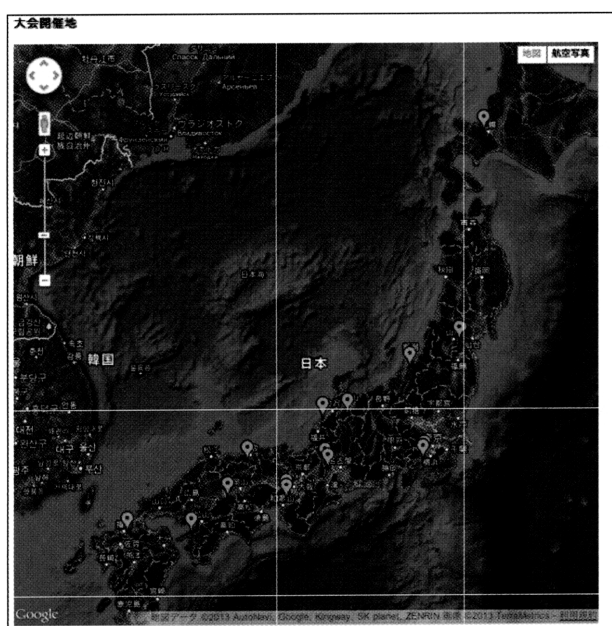


図 7 大会開催地を Google Map 上に表示

て、国立情報学研究所の御協力を得て、外部 API 提供 (LOD 化) を試行している。具体的には、I-Scover に格納された Linked Data を検索対象とする (SPARQL Protocol and RDF Query Language) エンドポイントを構築頂いている⁶⁾。SPARQL の詳細な説明は割愛するが、SPARQL を活用することで、ソサイエティ、あるいは研究会単位で独自の検索・集計・分析サイトを構築するなど、アイデア次第で多彩なシステム構築も可能となる。SPARQL による検索結果をグラフ表示するツールはオープンソースをはじめ幾つか存在し、それらを使うことにより、どのようなキーワードが文献に多く出現するかといった分析、図 6 に示すようなキーワード「MIMO (Multiple-Input and Multiple-Output)」の年ごとの出現をグラフ化し傾向をつかむなどが容易に行える。更に、外部サイトが提供する地図情報データと文献のイベント (開催地情報) メタデータを関連付けることで得られるイベント表示 (図 7)、検索結果に含まれるキーワードやイベントなどのメタデータと関連するニュースや特許の表示などもできる。このように、SPARQL を利用して作成できる簡単なデモの有用性を確認し、第 2 期システムでは外部 API として SPARQL を提供する予定である。

5. おわりに

I-Scover は本会会員の意見、要望に基づき、自らの技術、知見を結集して構築している知識横断検索のための戦略的なツールである。今後、電子情報通信分野の研究は、電子情報の有効活用が欠かせない。そのような時代に、本会会員の皆様にとって少しでも有益なツールとなるよう、開発、普及活動を行っていく所存である。

I-Scover の特徴を最大限に生かすためには、外部 API 提供を通じたメタデータの公開 (LOD 化) が不可欠である。一方で、本会の著作権規程は本会文献の著作権を守ることを前提にしているため、幅広く使って頂くことを前提にしたメタデータに当てはめることが適切かどうか検討すべき段階になっている。今後、企業誌等、本会が著作権を保持しない文献のメタデータも I-Scover に取り込む予定であり、著作権管理委員会の皆様の御理解・御協力を賜りたい事項が顕在化してきている。

また、I-Scover を本会のビジネスに活用する検討は緒についたばかりであり、現時点で確度の高いビジネスモデルを具体化しているわけではないが、現時点での見通しを以下に述べる。

I-Scover が保持するデータ (メタデータ) は検索サイトでも利用されている基本的にはオープンな情報である。I-Scover 第 2 期システムで予定している LOD 化 (外部 API 提供) により第三者が当該データを加工・再利用してビジネスに供することも可能となるため、I-

Scover 単体で新しい収入源を見いだすのは難しいと考えている。

IEEE では、Xplore を通じてメタデータを無償で公開する一方で、論文本文をオンライン閲覧する権利を企業や大学等にサイトライセンスとして販売し、また Pay Per View によるオンライン販売を行うことで大きい収入源としている。本会でも、I-Scover を核として論文本文のオンライン閲覧権の収入を増やすことも一案と思われる。

しかしながら、上記の実現のためには、紙媒体出版物の収入に頼っている現状のビジネス構造を大きく変更することが必要となる。現在の論文や技報など学術コンテンツによる収入は学会収入の5割を超えている。しかしながら、そのうちオンラインによる収入は数%にすぎない。また、CiNii からの還付金は1論文当り数円程度にすぎない。10年後の閲覧がほとんど電子化されているという想定を立てれば、本会のオンライン収入を飛躍的に増加させることが必須である。

本会の現状のシステムはアーカイブが一本化されていないことに加え、出版物ごとに閲覧権が独自に設定されており、統一した管理を行えないなど、技術的な課題も多い。いずれも一朝一夕に解決される問題ではないが、将来の財政面を考えると、検討が急がれる課題と思われる。I-Scover を契機に、これらの検討が進展することを期待している。

最後に、本会の査読の方針が「粒よりの論文をそろえるというよりは、玉石混在でもよい、石を捨てることを恐れる余りに宝石となるものを逃すことがないようにする」とあるように、I-Scover で検索できる文献には、執筆時には著者あるいは査読者も気付かなかった予想外の価値が埋もれている。寺田が述べた優れた論文の持つ「予言」を I-Scover で発見できることを願ってやまない。

文 献

- (1) 寺田寅彦, “科学と文学,” 寺田寅彦随筆集, 青空文庫, 2010.
- (2) “16 万件以上の文献を誰でも無料で検索可能” 電子情報通信学会, [I-Scover] を公開～文献, 研究者, 技術用語を探すなら [I-Scover] ～, プレスリリース, 2013.
- (3) 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 (冬号), pp. 240-244, Dec. 2012.
- (4) 井上友二, “会長就任にあたって—社会や世界との新たな関わりを求めて—,” 信学誌, vol. 96, no. 7, pp. 488-494, July 2013.
- (5) M.J. Bates, “The design of browsing and berry picking techniques for the online search interface,” Online Review, vol. 13, no. 5, pp. 407-424, 1989.
- (6) E. Prud'hommeaux and A. Seaborne, “SPARQL query language for RDF,” W3C recommendation, 2008.

(平成 25 年 12 月 11 日受付 平成 25 年 12 月 15 日最終受付)



山里 敬也 (正員:フェロー)

平 5 慶大大学院博士課程了。工博。現在、名大教養教育院教授。平 24 本会通信ソサイエティ功労顕彰状、IEEE Communications Society 2006 Best Tutorial Paper Award を受賞。可視光通信、ITS、eラーニングなどの研究に従事。IEEE 会員。



千村 保文 (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、電子交換システム及び VoIP 機器のソフトウェア開発、標準化活動に従事。現在、研究開発センタ・スマート社会ビジネスイノベーション推進部エグゼクティブ・スペシャリスト。著書「SIP 教科書」など。



西野 文人 (正員)

昭 54 東工大・理・情報科学卒。昭 56 同大学院博士前期課程了。同年(株)富士通研究所入社。以来、機械翻訳、情報検索、ナレッジマネジメントの研究に従事。現在、同社ソーシャルイノベーション研究所専任研究員。著書「コンピュータで翻訳する」(共著)、「自然言語処理—基礎と応用」(共著)など。



伊加田 恵志 (正員)

平 13 奈良先端情報科学研究科修士課程了。同年沖電気工業株式会社入社。現在、データ分析による最適化システムの開発に従事。



秋山 豊和 (正員:シニア会員)

平 11 阪大大学院工学研究科修士課程了。平 12 同博士課程中退後、同大学サイバーメディアセンター助手を経て、平 17 から同センター講師。平 20 から京都産業大・コンピュータ理工・講師を経て、平 23 同大学准教授。分散システム、インターネットアプリケーション、キャンパス情報システムなどに興味を持つ。博士(工学)(平 15-09, 阪大)。情報処理学会、IEEE CS 各会員。



中野 美由紀 (正員:シニア会員)

東大・理・情報科学卒。博士(情報理工学)。富士通株式会社勤務。昭 60-07 東大生産技術研究所助手(平 16 助教)。平 20-07 特任准教授。平 25-11 芝浦工大教育イノベーション推進センター教授。データベースシステム、ストレージシステム、データ工学の研究に従事。IEEE、情報処理学会、ACM、日本データベース学会各会員。



井上 友二 (正員:フェロー)

昭 23 福岡生まれ。昭 48 九大大学院工学研究科電子工学修士課程了。同年日本電信電話公社(現 NTT)に入社。デジタルネットワークの研究開発と国際標準化。平 12NTT データ取締役、平 14NTT 取締役で R & D 責任者。平 19(社)TTC 理事長。平 22(株)トヨタ IT 開発センター代表取締役会長(現職)。工博、IEICE 及び IEEE フェロー。本会会長。