「東アジア DH ポータル」の構築と課題：デジタルヒューマニティーズの研究ノウハウのオープンな知識基盤を目指して

菊池 信彦・宮川 創・二ノ宮聡（関西大学 アジア・オープン・リサーチセンター）

本稿では，関西大学アジア・オープン・リサーチセンターが運営している「東アジア DH ポータル」について論じる．「東アジア DH ポータル」は，広く人文学研究に携わる者がデジタルヒューマニティーズ（DH）を学ぶ機会を拡大させることを目的として，既存の DH 研究ノウハウを公開・共有するためのポータルサイトとして開設したものである．本稿では，このポータルサイト構築の背景や提供コンテンツ，そして公開後の課題等について論じている．

Construction and Problems of the “DH Portal for East Asian Studies”: Toward an Open Knowledge Base of the Digital Humanities Research Know-How

Nobuhiko Kikuchi / So Miyagawa / Satoru Ninomiya (Open Research Center for Asian Studies, Kansai University)

This paper discusses the "DH Portal for East Asian Studies" operated by the Kansai University Open Research Center for Asian Studies. The "DH Portal for East Asian Studies" has been established as a portal site to share existing DH research know-how with the aim of expanding opportunities for those who are engaged in humanities research to learn about Digital Humanities (DH). This paper explains the background, its contents, and the challenges of the portal site after its opening to the public.

１．はじめに

「デジタルヒューマニティーズ」（以下，DH と略）あるいは DH の前身から数えれば，誕生してからすでに 70 年ほどが経過している．その間，DH は情報学の進展ともあいまって，着実にその方法論を深化させていった．しかし，国内の人文学研究者に対する DH の浸透，あるいは，人文学研究者による DH への参入は，海外ほどは盛んではないように思われる．たとえば，試みに国内の研究者の多くが登録している Researchmap を調べてみると，研究分野のうち人文学分野である「哲学，倫理学」等の小分類 33 件のうちどれか 1 つでも登録がある 35,685 人のうち[1]，「Digital Humanities」や「デジタルヒューマニティーズ」「人文情報学」，「デジタル人文学」のいずれかを研究キーワードに挙げているのはわずかに 63 人である．一方で，世界中の人文学研究者用 SNS である Humanities Commons[2]では，ユーザ数 24,622 人に対し，“Digital Humanities”をキーワード検索すると 687 人がヒットする．このことからも，国内の研究者による DH への参入率の低さが際立っていることが分かるだろう．

この理由には様々な要因が考えられるが，特に本研究が着目するのは，DH の研究手法の共有という問題である．本報告は，その課題解決を目指して関西大学アジア・オープン・リサーチセンター（以下，KU-ORCAS と略）が開発したポータルサイト「東アジア DH ポータル」[3]の取り組みについて論じるものである．

２．先行研究・事例の整理とその課題

DH の研究手法を新たに学ぼうとする場合，大学での講義への参加を除けば，①DH サマースクール等の長期の講習会に参加する，②特定のトピックに絞った短期のワークショップに参加する，③セルフラーニングのサイト等で学習する，のいずれかに分けられる[4].

国内においては，①は現在のところ公的な実施事例はなく[5]，②に関しては JADH で行われているワークショップや ToDH を挙げることができる．一方で，③に該当するものが少ない．例えば，『歴史情報学の教科書』（文学通信，2019）や『文化情報学事典』（勉誠出版，2019）といった入門者用の「教科書」等があるが，研究をどのように進めればよいのかという点で直接参考になるものとは言い難い．唯一特筆すべきは，「UTDH/東京大学人文情報学」の YouTube チャンネルの動画であろう[6]. ここでは永﨑研宣や小風尚樹といった DH 研究者らによるデータベースの解説や正規表現の紹介が行われている．しかし，総じてこのようなセルフラーニングを支援する取り組みは，国内ではまだ事例が少ないという現状にある．

©2020 Information Processing Society of Japan

一方で，③に該当する海外の事例は豊富に見いだせる．例えば#dariahTeach は，EU7 カ国による DH 教育のためのセルフ学習用リソースを提供するプロジェクトとして始められたもので，様々な機関・専門領域での教育・学習の場面に転用可能なリソースとなることを目指されている[7]. また，別の例に The Programming Historian がある[8]．これは，2008 年にカナダで始まったプロジェクトで，当初は歴史研究者をはじめとする人文学研究者向けの Python チュートリアルサイトであった．その後，2012 年に査読付きオープンアクセスジャーナルの体制に拡大し，英語を中心に，スペイン語・フランス語でプログラミング技術やデジタルツールのノウハウの提供を行っている．The Programming Historian については，後で改めて言及する．

しかし，#dariahTeach や The Programming Historian 等の欧米のリソースサイトは，TEI（Text Encoding Initiative）や Python 等の汎用的な DH の知識・技術の学習リソースを提供することに成功しているものの，当然，そこで使用される言語は英語を中心とした EU 圏域の言語である．また，扱う内容も欧米を対象にしたものとなっている[9]. そのため，英語とはいえ日本人研究者にとっては習得にハードルを感じたり，また，東アジアを研究する者にとっては，自身の研究に転用できるかどうかを判断したりすることが難しい．

したがって，先行研究・事例における課題は，国内での日本語による，あるいは，東アジア研究に関する DH 研究ノウハウを学ぶオープンな知識基盤の未発達にあると指摘できる．とりわけ昨今は，新型コロナウイルス感染症の流行によってオンライン講義の導入が進められたが，オンラインで利用可能な学習リソースの乏しさも大きな課題となったことは記憶に新しい．欧米の具体的なノウハウを日本語話者に提供すること，それと同時に，日本および東アジア研究での DH 研究ノウハウを学ぶことができるようにすること，そのためのオンライン環境の開発が課題である．

３．東アジア DH ポータルの構築と各カテゴリの内容

東アジア DH ポータルは，前章で挙げた課題解決を目指し，2020 年 4 月 29 日の「DH の日」[10]に，KU-ORCAS が公開したポータルサイトである（図 1）．

KU-ORCAS とは，2017 年に採択された文部科学省私立大学研究ブランディング事業[11]を推進するために，関西大学が設置したセンターである．デジタルアーカイブを通じて研究資料をオープンライセンスで提供する「研究リソースのオープン化」，研究資料やデータのノウハウを広く共有する「研究ノウハウのオープン化」，研究活動を大学外へと開いていく「研究グループのオープン化」，そして，学会報告や論文等の成果物を機関リポジトリで公開する「研究成果のオープン化」という 4 つのオープン化を活動の柱として掲げている．本稿で述べる東アジア DH ポータルは，このなかの「研究ノウハウのオープン化」に該当するもので，「デジタルアーカイブで提供されるデジタル資料のデータを研究利用するための意見交換の場を作ること，そして，教育・学習用リソースの提供機能」[12]として実施しているものである．なお，ポータルサイトは，Wordpress を採用し，KU-ORCAS のサーバで運用している．

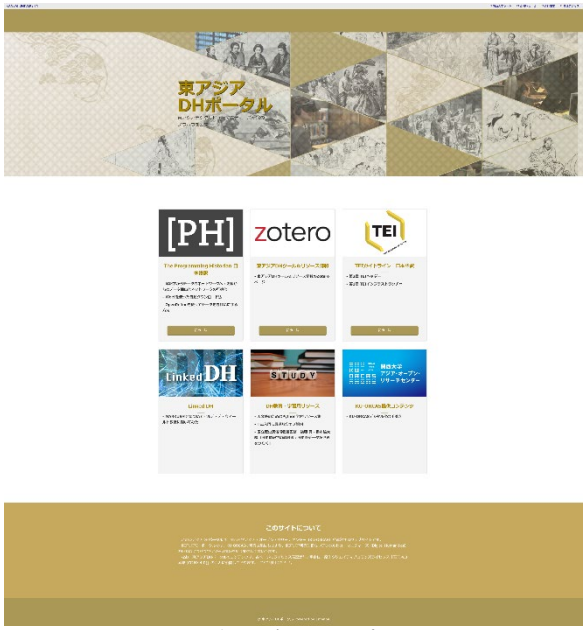

図 1　東アジア DH ポータル

東アジア DH ポータルで現在提供している記事カテゴリは，次の 6 つである．

① The Programming Historian 日本語訳
② 東アジア DH ツール＆リソース情報
③ TEI ガイドライン日本語訳
④ Linked DH
⑤ DH 教育・学習用リソース
⑥ KU-ORCAS 提供コンテンツ

各記事には，それぞれタグ付けをしており，それによりカテゴリをまたいだ検索を可能としている．また，「オープン化」への対応として，サイトコンテンツは，各ページにライセンス表記がない場合は，原則クリエイティブコモンズライセンス「表示 4.0 国際 (CC BY 4.0)」のもとに公開している．以下，各節では，東アジア DH ポータルの 6 つのカテゴリについてそれぞれ詳述していく．

©2020 Information Processing Society of Japan

3.1. The Programming Historian 日本語訳

「The Programming Historian 日本語訳」のページは，後述する TEI ガイドライン日本語訳と同様に，前章で課題として指摘した「欧米のノウハウを日本語話者に提供すること」を目的としている．このカテゴリでは，The Programming Historian の英語ページから次の 5 つのレッスン，「正規表現を利用した OCR テキストのクリーニング手法」「SPARQL を使ってリンクト・オープン・データ（LOD）へアクセスする」「Python を使って電子図書館から挿絵のページを抽出する」「OpenRefine を使ってデータをきれいにする方法」「Wget を使った自動ダウンロード法」を翻訳公開している．2020 年 10 月現在，The Programming Historian の英語ページでは全 82 レッスンが公開されており，その中からまずは研究の初期段階で必要となる，データ収集に関するレッスンを選択した．

3.2. TEI ガイドライン日本語訳

「TEI ガイドライン日本語訳」のページでは，TEI ガイドライン P5 を翻訳公開している．

TEI とは，Text Encoding Initiative（テキスト符号化イニチアチブ）の略で，このイニシアチブが標準化した XML (eXtensive Markup Language) 形式は，DH において，文献テキストを機械可読な形でマークアップする際のデファクト・スタンダードになっている．欧米では，写本などの文献のメタデータ，コーデクス学的・古文字学的情報，装飾やマージンなどのパラテクスト情報，校勘情報，人名・地名などのオントロジー，度量衡に関する情報など，そのテキストが書かれた文献にまつわる様々な情報をデジタルテキストに付与する際，TEI XML 形式に従って，テキストをマークアップする．そして，このデータを基に，デジタルエディションなどの視覚化や，ディスタントリーディングなどの様々な分析を施す．しかし，TEI による TEI XML の標準化は，欧米中心で進められたため，和書や漢籍など東アジアの文献資料のマークアップに暫く対応していなかった．しかしながら，近年，TEI コンソーシアムの SIG (Special Interest Group)である東アジア/日本語分科会の活動とともに，東アジア諸言語の文献への対応が急速に進み，日本でも TEI XML を用いることが普及しつつある．そのような潮流の元，TEI の東アジアでの普及を促進するため，本サイトでは，TEI の最も基本的なガイドラインである TEI ガイドライン P5 の日本語訳を順次公開している．現在第１章と第２章の翻訳が終了しているが，TEI-C 東アジア/日本語分科会の主催による TEI ガイドライン輪読会[13]によって，本訳がチェックされ，翻訳の精度と読みやすさを向上させている．なお，本家の TEI ガイドラインのウェブサイトでは，サイト自体が TEI/XML でマークアップされているが，東アジア DH ポータルではそこまでの対応は現時点では行っていない．

3.3. 東アジア DH ツール＆リソース情報

「東アジア DH ツール＆リソース情報」のページでは，The Programming Historian 日本語訳や TEI ガイドライン日本語訳で得た知識やノウハウを，東アジア研究に適用する際に必要となる DH リソース情報の集約と提供を目的としている．しかし，そこで扱うべきデジタルアーカイブ等のリソース情報は多岐にわたり，かつ，多言語であることから，筆者ら数名で収集することは困難である．そのため，Paula Cartis の Digital Humanities Japan グループのように収集を複数名で，そして，グループ外からの投稿も受け付けることで，効率的な情報の集約を目指した[14].

しかし，Wordpress 上でデータベースを表示するには維持管理が負担になるうえに，外部ユーザによる投稿を受け付けるには困難であった．そのため，KU-ORCAS 関係者以外の誰でも情報を追加することが可能なうえに，自前でデータベースを構築することなく検索を容易にできるようなツールとして，文献管理ツールの Zotero のグループ機能を採用した．この Zotero を採用した背景には，第一次世界大戦研究国際協会（International Society for First World War Studies）による”First World War Studies Bibliography”が，Zotero のグループ機能を活用して，多くの情報収集に成功しているという事実も後押ししている[15]．現在，KU-ORCAS の Zotero グループには 79 件の情報を掲載しており，それらは日中台韓とその他地域の 5 つに分類している．

3.4. Linked DH

「Linked DH」のページは，国内外の様々な DH プロジェクトのうち，グッドプラクティスな事例へと「つなぐ」ことを目的としている．端的に言えば，他のプロジェクトの紹介とそこへのリンクの提供である．2020 年 10 月現在，「MARKUS 開発についてヒルデ・デ・ウィールト教授に聞いてみた」と題した記事一件のみを掲載している．これは，中国語や韓国語を中心に，東アジア言語資料のマークアッププラットフォームである MARKUS について，開発者の Hilde De Weerdt に尋ねたインタビューの翻訳記事[16]である．

3.5. DH 教育・学習用リソース

「DH 教育・学習用リソース」のページは他の機関やプロジェクトによって公開されている DH

©2020 Information Processing Society of Japan

の教育や学習に役立つようなリソースをまとめ，公開している．現在掲載しているのは，「人文系のための Python 学習リソース集」と「TEI 入門に最適なウェブ資料」，そして，「国立歴史民俗博物館監修　後藤 真・橋本雄太編『歴史情報学の教科書：歴史のデータが世界をひらく』」の 3 つである．

「人文系のための Python 学習リソース集」は，人文系の学生や研究者向けに，プログラミング言語のなかでも，デジタルヒューマニティーズの研究でよく利用されている Python について，学習の助けとなるようなリソースをまとめている．近年の Python 人気を背景に，様々な Python の学習用資料が公開されており，そのなかでも学術的かつ DH に関わるようなものを選択している．

「TEI 入門に最適なウェブ資料」では，まず，ウェブ上かつ日本語で入手できる TEI の入門に最適なリソースを紹介している．例えば，3.2 でも取りあげた「TEI-C 東アジア/日本語分科会」が作成している TEI の日本語資料を用いた入門 Wiki[17]，ネット上で公開されている永崎研宣氏の TEI 概説[18]や TEI 設立メンバーによる TEI の歴史に関する概説の日本語訳[19]などを紹介している．さらに，TEI by Example[20]や#dariahTeach など，英語による入門教材も最後に紹介し，日本語と英語の入門教材を両方載せることで，東アジア文献の TEI 化特有の問題と西洋諸言語の文献との共通点を学習者に気づかせることができるような構成になっている．

最後の，「国立歴史民俗博物館監修　後藤 真・橋本雄太編『歴史情報学の教科書：歴史のデータが世界をひらく』」は，人文情報学の基盤形成を解説している同名の書籍の目次情報を公開している[21]．同著が発行元の文学通信からクリエイティブコモンズライセンス CC-BY-SA のもと，HTML，PDF，epub・Mobi の各形式でも公開されていることから，取り上げて紹介すべき資料と考えた．

3.6. KU-ORCAS 提供コンテンツ

最後のカテゴリが，「KU-ORCAS 提供コンテンツ」である．ここでは，KU-ORCAS 独自の研究ノウハウに関わるコンテンツを公開するが，現在のところ，「KU-ORCAS デジタル化の手引き」の 1 件のみを公開している．

『デジタル化の手引き』は，KU-ORCAS で書籍（和装本・洋装本）や絵画資料（軸装）等をデジタル化する[22]際の撮影マニュアルである．手引き作成の経緯はおおよそ次の通りである．KU-ORCAS は，関西大学アジア文化研究センター（CSAC）[23]を前身としており，そこでのデジタル化作業が基盤となっている．CSAC での作業は，ゼロからのスタートであり全てが手探りであった．資料の公開方法や画像に関しては国立国会図書館や早稲田大学演劇博物館を主に参考とした．また撮影機材の選定にあたっては，国立国会図書館の『デジタル化の手引き』[24]やウェブ上で紹介されている各種撮影機材の情報を参照した．その他，ハーバード大学燕京図書館などの他機関のデジタル化作業[25]を見学すると共に，定期的に撮影機材や方法に関する基準項目の見直しを続けた結果，ようやく現在の『デジタル化の手引き』が完成した．

デジタル化作業の大部分は，関西大学の大学院生がアルバイトとして従事している．彼ら／彼女らはデジタル化作業においては全くの素人であるが，『デジタル化の手引き』を使い撮影練習をすれば，1 日の練習で基本的な技術はほぼ身につけることができる．大学院生は数年で卒業してしまうため，撮影技術の安定性・新たな作業者に対する技術指導という面から言えば大変効率が悪い．しかし，本学のデジタル化作業は，大学院生が多くの資料に触れることで専門分野以外の知見を深めて欲しいという教育的側面も有するものとなっている．

４．反響と考察

東アジア DH ポータルは 4 月 29 日に公開し，日本語だけでなく英語でも SNS で広報を行った．これにより，The Programming Historian と TEI ガイドラインの日本語訳に対して，海外の DH 研究者コミュニティから大きな反響が寄せられた[26]．一方で，Zotero によるリソース情報のまとめには，国内外のどちらからも目立った反応がない．

東アジア DH ポータルの利用に関し，現段階では掲載情報の少なさもあり目立ったものはないが，そのなかでも先述のように TEI-C 東アジア/日本語分科会が TEI ガイドラインの輪読会を行っており，そのこともあって TEI ガイドラインの翻訳ページの利用が他に比して多い傾向にある．

５．おわりに

東アジア DH ポータルはまだ公開から日が浅く，DH 研究ノウハウのオープンな知識基盤となるためには，今後のコンテンツの拡充は必須である．そのため，サイトのコンテンツを利用したワークショップ等を行うことで，そこで得られた知見をもとに，日本および東アジア研究でのオリジナルな研究ノウハウの開発や日本における DH プロジェクトの優良事例の紹介，あるいは日本語で読むことのできる DH ノウハウのサイト情報の公開を行う想定である．

©2020 Information Processing Society of Japan

また，KU-ORCAS 自身が 2021 年度で終了する事業であることから，ポータルサイトの継続に向けた取り組みも必要である．そのためにも，学内外にユーザコミュニティを形成し，コンテンツの拡充や維持管理，国内外の同種の取組みとの連携に向けた検討が必要である．

それらの実践を積み重ねていくことで，日本から，あるいは東アジア研究に直結する情報の発信と共有を目指していきたい．

参考文献

[1] Researchmap で，大分類[人文・社会]を選択し，その下位にある小分類から人文学の分野を選択し，重複を削除して人数を割り出した．データ採取日は 2020 年 10 月 23 日である．
Researchmap. https://researchmap.jp/, (accessed 2020-10-29).

[2] Humanities Commons. https://hcommons.org/, (accessed 2020-10-29).

[3] 東アジア DH ポータル．https://www.dh.ku-orcas.kansai-u.ac.jp/, (accessed 2020-10-29).

[4] “The challenges and prospects of the intersection of Humanities and Data Science: White Paper from The Alan Turing Institute”. The Alan Turing Institute. 2020-08-04. https://www.turing.ac.uk/research/publications/challenges-and-prospects-intersection-humanities-and-data-science, (accessed 2020-08-26).

[5] 国立歴史民俗博物館による「総合資料学の創生」プロジェクト内の人文情報ユニットが開発合宿を行ったことがあるが，これは新たに学ぶという趣旨とは異なるものと位置づけられる．
1) “「DH システム開発・データ構築合宿」兼人文情報ユニット研究会を開催”. https://www.metaresource.jp/8%e6%9c%8821%e6%97%a5%e3%80%9c23%e6%97%a5%e3%80%80%e3%80%8cdh%e3%82%b7%e3%82%b9%e3%83%86%e3%83%a0%e9%96%8b%e7%99%ba%e3%83%bb%e3%83%87%e3%83%bc%e3%82%bf%e6%a7%8b%e7%af%89%e5%90%88%e5%ae%bf%e3%80%8d/, (accessed 2020-08-26).

[6] “UTDH/東京大学人文情報学”. YouTube. https://www.youtube.com/channel/UC-LwTQvNdrNO-f4HinoPuAA, (accessed 2020-09-01).

[7] “#dariahTeach: online teaching, MOOCs and beyond”. DH2016. https://dh2016.adho.org/abstracts/292, (accessed 2020-09-01).

[8] The Programming Historian. https://programminghistorian.org/, (accessed 2020-09-01).

[9] Sichani, Anna-Maria, James Baker, et al. “Diversity and Inclusion in Digital Scholarship and Pedagogy: The Case of the Programming Historian”. Insights. 2019. 32 (1), p.16. http://doi.org/10.1629/uksg.465, (accessed 2020-09-01).

[10] “Day of DH 2020”. centerNet. https://dhcenternet.org/initiatives/day-of-dh/2020, (accessed 2020-08-28).

[11] 文部科学省による助成終了を受け，2020 年度からは関大研究ブランディング事業として継続している．

[12] 出典では「教育・教育用リソース」となっているが，「教育・学習用リソース」の誤りである．
菊池信彦，内田慶市，永崎研宣.”「越境する」デジタルアーカイブの機能要件を考える―KU-ORCAS が備えるべきもの―”．研究報告人文科学とコンピュータ（CH）．2018-CH-117(6), pp. 1-6. http://id.nii.ac.jp/1001/00187422/, (accessed 2020-09-02).

[13] TEI-C 東アジア/日本語分科会．https://www.facebook.com/groups/2165074817079710, (accessed 2020-10-24).

[14] Paula R. Curtis, Hoyt Long, Molly Des Jardin and Mark Ravina. Digital Humanities Japan: Building Community and Sharing Resources. The Digital Orientalist. 2020-03-27. https://digitalorientalist.com/2020/03/27/digital-humanities-japan-building-community-and-sharing-resources/, (accessed 2020-10-29).

[15] Jean-Christophe Peyssard. First World War Studies Bibliography. Aldébaran. 2017-01-02. https://journals.openedition.org/aldebaran/7266, (accessed 2020-10-29).
“First World War Studies Bibliography”. Zotero. https://www.zotero.org/groups/55813/first_world_war_studies_bibliography, (accessed 2020-10-29).

[16] ALÍZ HORVÁTH and Hilde De Weerdt. OpenMethods Spotlights #1: Interview with Hilde De Weerdt about MARKUs. OpenMethods. 2020-10-13. https://openmethods.dariah.eu/2020/10/13/openmethods-spotlights-1-interview-with-hilde-de-weerdt-about-markus/, (accessed 2020-10-29).

[17] 日本語向け TEI ガイドライン（作成中・試行版）．https://github.com/TEI-EAJ/jp_guidelines/wiki, (accessed 2020-10-24).

[18] 永崎研宣（2019）「歴史データのさまざまな応用 −Text Encoding Initiative の現在−」国立歴史民俗博物館監修，後藤真・橋本雄太編『歴史情報学の教科書』131-154．東京：文学通信．https://bungaku-report.com/blog/2019/03/chapter-8-text-encoding-initiative.html, (accessed 2020-10-24).

[19] Ide, Nancy M., C. M. Sperberg-McQueen, Lou Burnard (2018)「TEI：それはどこからきたのか．　そして，なぜ，今もなおここにあるのか？」『デジタル・ヒューマニティーズ』1, 3-28. https://www.jstage.jst.go.jp/article/jadh/1/0

©2020 Information Processing Society of Japan

/1_2/_article/-char/ja, (accessed 2020-10-24).
[20] TEI by Example. http://teibyexample.com/TBE.htm, (accessed 2020-10-24).
[21] 国立歴史民俗博物館監修　後藤 真・橋本雄太編『歴史情報学の教科書　歴史のデータが世界をひらく』全文公開サイト．文学通信．https://bungaku-report.com/metaresource.html, (accessed 2020-10-24).
[22] KU-ORCAS でデジタル化した資料は「関西大学デジタルアーカイブ」として公開している．https://www.iiif.ku-orcas.kansai-u.ac.jp/,（accessed 2020-10-22）
[23] 文部科学省「私立大学戦略的研究基盤形成支援事業」により平成 23～27 年度の 5 年間，関西大学東西学術研究所に設置された．http://www.csac.kansai-u.ac.jp/,（accessed 2020-10-22）
[24]国立国会図書館デジタル化の手引き．https://www.ndl.go.jp/jp/preservation/digitization/guide.html, (accessed 2020-10-23）．
[25] ハーバード大学の各図書館のデジタル化作業については，当時 CSAC 研究員であった氷野善寛が 2013 年に現地を視察している．
[26] Twitter. https://twitter.com/So_Miyagawa/status/1255298231736643589, (accessed 2020-09-02).

©2020 Information Processing Society of Japan