

令和4年度 メディア芸術連携基盤等整備推進事業
分野別強化事業

「クレジット情報テキスト識別の更なる高精度化
とメタデータ処理の汎用化」 実施報告書

一般社団法人日本アニメーター・演出協会

令和5年2月

目次

第1章 事業概要	3
1.1 目的（目的・背景）	3
1.2 実施内容（概要）	3
1.3 実施体制	4
1.4 実施スケジュール	5
第2章 成果・課題	6
2.1 実施結果	6
2.1.1 自動化／高精度化／処理量削減／ユニット化	6
2.1.2 メタデータ処理	8
2.1.3 周知活動	9
2.1.4 （仮称）アニメ作品情報連絡調整会	11
2.2 結果からの推測、課題	13
2.2.1 OCR 選定及び自動化	13
2.2.2 学習データ作成	14
2.2.3 災害等に強い録画体制	14
2.2.4 配信など多種映像メディアへの対応	15
2.3 利活用事例	16
2.4 その他特記事項（アニメデータベースを構築する意義の再確認）	18
第3章 実施内容	21
3.1 会議開催	21
3.1.1 定例会議	21
3.1.2 周知活動	21
3.1.3 （仮称）アニメ作品情報連絡調整会	21
3.1.4 その他会議	22

目次

3.2	高精度化・メタデータ処理	22
3.2.1	機能要件、動作概要	22
3.2.2	成果・効果・制約事項	30
3.3	再配布可能なシステム	31
3.3.1	機能要件、動作概要	31
3.3.2	将来への展望、課題	37
3.4	周知活動	39
3.4.1	「マチ★アソビ」（令和4年（2022年）10月15日）	39
3.4.2	JAniCA 定例配信（令和4年（2022年）10月29日）	43
3.4.3	Japanese Visual Media Graph（2023年1月27日、28日）	44
3.5	採録状況	47
3.5.1	全録サーバ（杉並）実績	48
3.5.2	全録サーバ（千代田）実績	49
3.6	メディア芸術データベースへのデータ提供	51

第1章 事業概要

1.1 目的（目的・背景）

本事業は、アニメーション（以下、アニメ）のスタッフ情報を収集してデータベース（以下、DB）に記録し、どのような人物や組織がアニメの制作・製作に携わっていたのかを、将来において判別・確認できるようにすることを目的としている。日本が世界に誇る文化であり、同時に重要な産業でもあるアニメがどのような人材により、どのような環境で作られてきたかを知る上で、こうしたスタッフ名は最重要の情報となるからである。

このため、平成 29 年度では「リスト制作委員会における採録手法の研究」、平成 31 年／令和元年度では「アニメ番組の全録画」、令和 2 年度では「クレジット表示された画面を含む全画面のキャプチャ画像生成」「クレジット表示されたキャプチャ画像を使用して OCR を利用したテキスト変換」、令和 3 年度では「全録サーバから生成したクレジット情報テキスト識別の高精度化とメタデータ処理の自動化の試行」を行った。

過去の事業を積み重ねた結果、録画したアニメ番組から文字情報が出ている場面を選び出し、キャプチャを行った上で OCR に投入し、テキスト情報を抽出した上で個々の文字列の属性を判断し、メタデータとして記録する作業を、現実的な作業負荷で集約可能にするシステムの構築・運用について、一定のめどが付いた。

しかし、OCR 技術は日進月歩であり、継続的な調査研究の必要性があった。また、令和 3 年度事業ではメタデータ処理を試行したが、集約した作品／スタッフ情報については、将来の利活用を見据えた場合には万全な状態でのメタデータ処理ができていない状況にあった。

そこで、令和 4 年度は「クレジット情報テキスト識別の更なる高精度化（高識別精度化、高速化、システム化等）」と「メタデータ処理の汎用化（クレジット識別のためのシステム化）」を試行した。

1.2 実施内容（概要）

令和 4 年度の本事業では、「クレジット情報テキスト識別の更なる高精度化」として、「最新状況での OCR サービス群の機能評価」「令和 3 年度で試行した機能のシステム（ユニット）化」「更なる機能開発」などに取り組んだ。また、「メタデータ処理の汎用化」として「教師データの調整」「AI 等の機能活用」「令和 3 年度で検討した機能開発」などに取り組んだ。

放送されるアニメ作品に掲出されるスタッフやキャスト、企業などの名称をキャプチャによって画面ごと切り出し、OCR によって記録しやすいテキストへと変換する手法に取り組んでいる本事業においては、OCR の読み取り性能が採録した情報の精度に大きく関わってくる。一方で、大量の情報を処理する必要があるため、OCR 処理に要するコストも問題となってくる。そのため高精度でかつ安価な OCR を選定する作業に取り組み、クラウド上で提供されているサービスの中から現時点において適切なものを選定した。これにより、処理にかかる時間やサービスの使用料、及び携わる人的コストがどの程度に達するかも判明した。

加えて、全録サーバに録画されたアニメ映像からキャプチャ画面を抽出し、OCR へと投入して取

第1章 事業概要

り出せたテキストを記録していく流れを、それぞれのソフトウェアやサービスを結合させてユニット化し、作業手順の省略にも一定のめどを立てた。こうしたシステム構成の公開によって、映像からのテキスト抽出作業を誰でも行えるようになり、アニメ以外でテレビ放送されるドラマやバラエティなど娯楽番組として消費され顧みられる機会に乏しい制作スタッフや出演者・キャストの情報を、抽出してデータベース化する可能性も膨らんだ。

メタデータ処理は、抽出されたテキストの属性を AI によって判別し、データベースへの格納の際に属性に合致した項目に格納し、以後の検索・利用しやすいようにする上で、欠かせないプロセスとなっている。その際には、キャプチャ画面上に表示される文字列から役職名を判断し、これに続く文字列をスタッフ名と判断するような処理を行う運びとなる。人間の作業員なら目で見ても容易に判別できる内容でも、データベースへの自動登録化には役職名への理解、人名としての判別などが必要となり、この際に AI を活用して判定できれば人手を削減できる。

これについて、全録サーバの稼働や、リスト制作委員会によって収集・構築された「リスト DB」の膨大なデータの教師データとしての活用により、メタデータ処理を可能とする道筋は見えた。一方で、教師データそのものの汎用的なものとしての配布には著作権的、コスト的な課題があることも見えて、解消に向けて取り組む必要性が浮上した。これが、アニメ関連で共通に利用できる情報ソースの確保の重要性を改めて高める結果を呼んだ。

このこともあり本事業では、あわせて、アニメ関連データベースのより一層の充実につながり、利活用の促進にも寄与する施策として、アニメ業界側からのビジネス情報（宣材やクレジットデータなど）を集約するための組織を組成する必要があると判断した。（仮称）アニメ関連情報連絡調整会を立ち上げるために、アニメ関係団体、アニメ制作会社や製作委員会会社、アニメ関連情報を採録している個人・団体、アニメ情報を発信している雑誌やウェブといった関係者への働きかけを行った。

こうした開発作業と並行して、アニメ関連の情報が網羅されたデータベースが存在していることの意義を改めて確認しつつ、データ収集に関わる労力の多さを周知するべく、徳島県徳島市で定期的で開催されているアニメ関連イベント「マチ★アソビ」や、アニメ関連団体の一般社団法人日本アニメーター・演出協会（JAniCA）会員向けに配信しているウェブ番組、同人誌即売会のコミックマーケットに並行して実施される研究会のコミケットフリンジで講演などを実施。シーズンごとに数十本単位で増えていく新作アニメから、スタッフ・キャストなどの情報を採録するためにどれくらいの労力が必要なのかを話した。採録されデータベース化された情報を元にスタッフ・キャストが自身のキャリアに信頼性を与えられること、過去情報の分析により、作品の傾向の把握や携わったスタッフへの再評価を行えることも説明して、アニメ関連データベースの存在に理解を求めた。

1.3 実施体制

本事業の実施体制は以下のとおりとなる。

主催：文化庁

一般社団法人 日本アニメーター・演出協会（以下、JAniCA）

第1章 事業概要

コーディネーター：高橋 望（特定非営利活動法人アニメ特撮アーカイブ機構、以下 ATAC）

オブザーバー：大坪 英之（JAniCA）

関係者：三好 寛（ATAC）

須山 大介（合同会社 Prod.）

原口 正宏（リスト制作委員会）

磯部 正義（リスト制作委員会）

谷口 隆一（ライター）

1.4 実施スケジュール

本事業の実施スケジュールは以下のとおりである。

令和4年4月28日 記録作成のための方針会議

令和4年5月31日 内部会議（第1回）

令和4年6月23日 内部会議（第2回）

令和4年7月29日 内部会議（第3回）

令和4年8月30日 内部会議（第4回）

令和4年9月30日 内部会議（第5回）

令和4年10月13日 中間報告会

令和4年10月15日 徳島「マチ★アソビ」でのアニメ DB 関連講演登壇

令和4年10月29日 JAniCA チャンネル「アニメのデータベース」配信

令和4年10月31日 内部会議（第6回）

令和4年11月25日 内部会議（第7回）

令和4年12月23日 内部会議（第8回）

令和4年12月30日 コミケットフリンジ登壇

令和5年1月13日 （仮称）アニメ関連情報連絡調整会顔合わせ

令和5年1月20日 内部会議（第9回）

令和5年1月28日 ドイツ・シュトゥットガルト「JVMG」登壇

令和5年2月17日 内部会議（第10回）

令和5年2月20日 最終報告会

第2章 成果・課題

2.1 実施結果

テレビ放送などのアニメ作品を全録サーバに録画し、スタッフやキャストが表示されている画面をキャプチャしてOCRに投入し、テキストデータとして抽出する作業を続ける中、令和3年度の事業では、「全録サーバから生成したクレジット情報テキスト識別の高精度化とメタデータ処理の自動化の試行」に取り組んだ。過去の事業を積み重ねた結果、焼き込まれた文字やアニメ番組の管理情報を、現実的な作業負荷で集約可能な運用について一定のめどが付いた。

しかし、OCR技術は日進月歩であり継続的な調査研究の必要性がある。また、令和3年度事業ではメタデータ処理の試行を行ったが、集約した作品／スタッフ情報については、将来の利活用を見据えた場合、万全な状態でのメタデータ処理をできているとは言えない状況にあった。そこで今年度は、「クレジット情報テキスト識別の更なる高精度化」として、「最新状況でのOCRサービス群の機能評価」及び「令和3年度で試行した機能のシステム（ユニット）化」「更なる機能開発」に取り組んだ。また、「メタデータ処理の汎用化」として「教師データの調整」「AI等の機能活用」「令和3年度で検討した機能開発」なども行った。ここでは、全録サーバについて、「メタデータへの高精度な変換」機能の向上を指向した。

周知活動としては、アニメ情報の採録活動とデータベースの構築に対する理解促進を目的に、アニメ関係のイベントや業界関係者向けのウェブ配信、海外の大学での講演を行った。さらに、より正確な情報を迅速に入手できる窓口を作り、全録サーバからの収集に止まらない情報の採録により多面的な判断を下せる環境を整える目的で、アニメ業界側からのビジネス情報（宣材やクレジットデータなど）を集約し、同様に採録を行っている団体・個人との連携を図れるようにする組織となる「(仮称)アニメ作品情報連絡調整会」の組成を目指し、関係者への働きかけを行った。

2.1.1 自動化／高精度化／処理量削減／ユニット化

「全録サーバ」を、現在の高度な人材が手間暇をかけて運用・メンテナンスを行う試験機という位置付けから、ある程度の専門知識は必要なものの普及・再頒布可能なシステムとしての公開を目指した。クレジット情報テキスト識別の更なる高精度化作業フローの自動化により文字部分を一举に抽出してのテキスト化を検討した。

アニメ作品のクレジットをテキスト化するシステムで、最重要となるプロセスはOCRとなる。クレジットを読み取りテキスト変換する精度に結果が左右されるからで、どのサービスが運用している変換エンジンを選ぶべきかを、幾つかのサービスから比較検討した。

OCRは、サービスの提供形態によって、①商用パッケージソフト、②PaaS (Platform as a Service) クラウドAPIサービス、③オープンソースがある。このうちパッケージソフトは、パナソニック株式会社提供しているものや、株式会社NTTデータNJKが提供する「メディアドライブ」などがある。パッケージソフトは性能面に関して高い精度を実現できるが、利用に当たってはコストが問題と

第2章 成果・課題

なった。具体的な費用を確認しようにも、商談を経た上で見積りを出さなければ見えてこないところもあり、具体的な費用算定に至るまでに時間と手間がかかるため、選択から除外した。

結論として、本事業では PaaS (Platform as a Service) によるクラウド API サービスを活用した。こちらにも幾つかのサービスが存在している。主なものとしては、①Google Cloud Vision、②マイクロソフト提供の Azure Computer Vision、③LINE 提供の CLOVA OCR API、④Amazon Textract (日本語未対応) などがある。本事業では、この中からコスト面や利用に当たっての契約面でハードルが低かった Google Cloud Vision を採用した。

CLOVA OCR API も検討したが、料金体系が月額利用料金+従量料金で、月額利用料金部分だけでも 22,000 円となり除外した。CLOVA OCR API は、国立国会図書館 (以下、NDL) が行っている書籍のスキャンにおいて利用されており、性能がある意味で担保されている。Google Cloud Vision に勝る面もあるが、書籍をスキャンするという NDL の事業主体と事業目的を考えると、アニメ画像のスキャンに適しているかを判断できず除外した。

OCR のコストについても算定した。1 話分の画像が重複などを含めて 50 枚と仮定すると 1,000 回の OCR で 20 話分を処理できる。2022 年秋季の TV レギュラー番組は 67 タイトルあり、これらを集計すると 1 シーズン当たりの OCR 処理枚数は 50 枚×67 タイトルで 3,350 枚に及ぶ。Google Cloud Vision については、最初の 1,000 ユニットについては無料となる。以後 1,001~5,000,000 ユニットで、1,000 ユニットについて 1.5 ドル、5,000,001 以上のユニットで 1,000 ユニット当たり 0.60 ドルとなる。

これらを元に 1 話当たりで 53 回分を OCR にかけた場合の料金を推計すると、1,001 ユニット以上で 10.48234 円、5,000,001 ユニット以上からは 4.19283 円となる。1 話当たり 10 円程度で利用できるとして、それを 1 シーズンに放送される 67 タイトルに当てはめた額が、1 シーズン分の OCR コストとなる。

こうした作業を、より効率的に行えるようにするモジュール作りにも取り組んだ。以前は、映像からいったん切り分けた静止画を OCR に投入し、テキストデータを抽出していた。今回は、処理の流れをモジュール化し、動画ファイルを読み込めば自動的にキャプチャが行われ、OCR によるテキスト化処理を行えるようにした。

モジュール化に当たっては、利用者への提供のしやすさを検討し、Google による機械学習・深層学習の環境を整備できる無料サービス「Google Colaboratory」を使用した。これ以外の OCR サービスでは、料金面などで利用者の負担が増えてしまう。設備などを何も持っていない状況から OCR を使い、映像を OCR に渡して焼き込まれた文字をテキスト化する流れを作る上では、この仕組みが最適と考えた。

作業では、全録サーバで録画し作品を、1 シーズン分ひとまとめにした映像ファイルを Google Drive にアップする。セットアップまでの時間は 1 分程度。続いて画面上で映像ファイルを指定して

第2章 成果・課題

OCRに投入すると、キャプチャされた画面が抽出される。処理にかかる時間は、1シーズンの13本でOCRを含めて50分ほどだった。

試行はシングルスレッドで行った。これをマルチスレッドで行えば処理にかかる時間は数分の1に短縮できる。動画から静止画像を切り出す部分の処理が重くなりがちで、これをより速く処理する手法が見つければ、更なる時間短縮も可能になる。動画ファイルから静止画に書き出さないで処理すれば更に速度は上がるが、現段階では想定していない。

処理では、同じ文字が出ている部分が重複して残されたり、文字のように見えてしまう部分を持った画像も残されたりする事象が起こる。ここで、Google Colaboratory 版ではない OCR へと読み込んで、同じ文字列が見える重複画像を人間が目で見確認し、間引いていけば、枚数を減らしてシステムの負荷やコストを削減できる。ただし、目で見確認して重複を削除する作業が必要となるため、金額的にも、また目視によって本来選ぶべき画像を削除してしまう可能性からも、処理対象となる画面を選ばずに OCR に投入する方が、自動化の面でも省力化の面でも良いといった見解も出た。

これは、画面から切り出した画像がテキストとセットになって存在している状態を重視しているため、切り出された画像さえ存在していれば、誰でも“原典”として参照し、変換ミスや記録ミスを探して訂正できる。時短を急ぐ余りにクレジットが出ている画面を見落としては本末転倒となる。後世に照合可能なデータを残しておく意味も考慮して、システムを構築する必要があると確認した。

2.1.2 メタデータ処理

文字が出ている画像を切り出し、OCRに投入してテキスト化を行う作業では、OCRの性能もあって1回の投入で完全なテキスト化ができる訳ではない。幾度か動作させるとより正解に近い文字が出てくるが、その分余計にコストがかかる。定額制ではなく従量制の料金体系を採用しているクラウドサービスのOCRならではのコストと言える。

こうした事情から、カスタマイズできない部分のメタ情報処理を行うためのデータセットを構築することにした。データセットは人海戦術による作成も可能だが、今回はこのデータセット生成を効率良く行うための自動生成システムの構築に取り組んだ。具体的には、キャプチャした画面に対し、文字を認識するソフトを再利用し、画面上の文字を確定した。この作業はローカルのPC上で行っているため、クラウドサービスを使う必要がなく、余計なコスト負担がない。

続けて、確定した文字に対して解析を行い、属性を把握して役職名か人名かで色分けを行った。一つのタイトルについて、オンエアされたエピソードに掲出される人名などの文字列に属性を付け、データセットとして活用してAIのモデルを作成する。この作業を行えば、シリーズが終わっていても、続けてオンエアされた新規の話数に対して、自動的にメタデータ、各項目の属性がついていくことになる。

アニメ作品の場合、クレジットのレイアウトについてタイトルごとに一定のフォーマットがある。第1話の放送話と、シーズン途中のエピソードが盛り上がった話とでフォーマットが変わる場合もあ

第2章 成果・課題

るが、何パターンか学習させておくことで、新しくオンエアされる話でも高い精度で属性を把握できるようになった。

ここまでは、クラウドを必要としなくても行える作業だが、これにクラウドの OCR を使うと、テキストデータに更に情報を載せられるようになる。加えて、間違っただけの修正がツールを使えばできるようになっている。このツールで認識した情報は構造化されているため、いつでもクレジットデータの詳細な情報を位置も含めて保存できる。

こうした作業を推し進めていった結果、最終的には、クレジット画面に対して専用の OCR が作れるようになった。こうした作業のために、画面のレイアウトを認識するためのデータセットも生成した。OCR 用のデータセットを生成するためのソフトもできており、追って文字に対する OCR のモデル作成も行う予定となっている。

2.1.3 周知活動

10月15日に徳島市で開催のイベント「マチ★アソビ」でトークイベント「あなたの知らないアニメデータベースの世界」を実施した。

「マチ★アソビ」はアニメへの関心をよりどころにして、地方にファンを誘導して観光や経済を盛り上げる狙いで平成21年(2009年)にスタートした。現在は特定非営利活動法人マチ★アソビを中心に徳島県も運営に加わり、アニメ制作会社やパッケージメーカー、グッズ会社、出版社などが出展し年に2回、春と秋にイベントを開催している(現在はコロナ禍対応のため流動的な開催)。

トークイベントを行った「マチ★アソビ Vol.25」は、10月6日から11月6日までを開催期間として各種イベントを開催。このうち10月15日、16日はクライマックスラインとして、多くのイベントが繰り広げられ、全国から参加者が集まった。「あなたの知らないアニメデータベースの世界」は、こうした参加者や出展のために徳島市を訪れたアニメ業界関係者に、アニメ関連データベースの存在について知ってもらい、意義を周知する目的で行った。

一般向けに開放されている「マチ★アソビ」に来場するアニメファンは男性女性を問わず年齢層が幅広い。「あなたの知らないアニメデータベースの世界」の参加者も同様で、それぞれが捉えるアニメの範囲も多岐に及んだ。新作を中心にしているファンもいれば、旧作の知識が豊富なファンもいて、それぞれに向かう関心に差異がある。こうした来場者に、アニメのジャンルが一般向けやキッズ向け、深夜アニメ、アート指向の高いアニメなどに分類できること、その収入源も玩具の売上げやパッケージ収入、非営利といった具合に違っていることなどを示し、認識を広げることに取り組んだ。

分類が多岐に及ぶ実態に加え、放送されているタイトルも多い状況を説明。半年間で約16,000話を全録サーバで録画し、そこからクレジットデータを収集している現状を説明すると、会場から多さに対して驚きの反響があった。アニメファンは、最終的な成果物としてデータベースのクレジットデータを見るが、それらがどのようなプロセスを経て集められているかを意識していない。トークイベントを通してクレジットデータを収集する手順を見てもらい、かかる手間や携わる人の多さを知って

第2章 成果・課題

もらえた。そこから、人力による収集は難しく、システム化の必要性への関心を惹起 [じゃっき] できた。

「マチ★アソビ」ではほかに、メディア芸術データベース（以下、MADB）の構築・充実によってどのような利便性が得られるのかも示した。具体的には、より正確なデータを利用可能になる点、様々なデータベースが統合されたジャパンサーチと連携して利用できるようになるメリットなどを挙げた。そうした将来に向けて、データセットの構築や API の確保など、自分たちで何をすべきかを改めて確認した。

10月29日には、JAniCA がオンラインで開催した「第13回 JAniCA チャンネル」で、主に会員向けにアニメのデータベースとは何かを説明した。JAniCA はアニメーターや演出が所属する団体として、その地位向上や技術継承などに資する活動の実施を目的に平成17年（2005年）に発足した。ウェブ配信では、会員登録画面を示しつつ、人が主となり手掛けた作品が従となる登録の仕組みを伝え、クリエイターに紐 [ひも] 付く形で作品名が積み重ねられていき、キャリアを証明するものとして働く状況を示した。ここで、掲載された情報が正確なものかを確認する場合、作品マスターなり参加した各話のマスターがあつて、それらを参照できるようにしておくのが確実と指摘し、採録活動への理解を促した。

JAniCA への申請は、現状は自己申告制となっているため、本人が作品名を誤認していて正しく登録されない場合がある。作画作業等に参加していた時点と、実際に放送された時点でタイトルが変わる場合もあるため、申告だけでは正式な作品名が出されない事態も起こりえる。有志によって運営されている「作画@wiki」(<https://w.atwiki.jp/sakuga/>) の情報を添えて申請してくるクリエイターもいる。クリエイター本人が見て正しいものとオーソライズした情報を言えるが、本人が認めた正確な情報だといったことを「作画@wiki」に戻す手段がないため、採録団体等が連携できるような「(仮称)アニメ作品情報連絡調整会」の必要性が浮かび上がった。

「マチ★アソビ」と同様、こちらでもアニメの定義について触れた。「アニメ」として、深夜アニメや子供向けアニメしか見ていない層も存在するが、「アニメ」には、NHKのEテレなども含め幅広く含める必要がある。また、過去のアダルト作品には今では有名になった錚々 [そうそう] たるクリエイター関わった作品があり、テレビ作品ではできない表現などに挑んで腕を磨いて出てきた人もいる。こうしたキャリアをどうやって掘り起こし、フォローするかが課題である。

本事業を始めてから初となる海外での周知活動にも取り組んだ。令和5年（2023年）1月27日、28日にドイツ・シュトゥットガルトにある Stuttgart Media University で開催された「Japanese Visual Media Graph」(以下、JVMG) に招聘 [しょうへい] され、日本におけるアニメ関連のデータベース構築に向けたプロジェクトについて1月28日に発表を行った。

この「Japanese Visual Media Graph」(以下、JVMG) プロジェクト (<https://jvmg.iuk.hdm-stuttgart.de/japanese-visual-media-graph/>) とは、Stuttgart Media University の Magnus Pfeffer

第2章 成果・課題

教授らがメンバーとなって、日本の視覚メディア（アニメ、マンガ、コンピュータゲーム、ビジュアルノベルなど）に関する研究用データベース構築を目的としている。単なるデータベースではなく、Google のナレッジグラフに似た、グラフベースの高度に相互接続されたデータベース構造と、柔軟な検索インターフェース、分析ツールの組合せを想定したものとなっている。欧州におけるメディア芸術関連データのプラットフォームとして、各所にデータを提供できるようにしたいとの意向を持っている。

研究用データベースは、ウェブ上の多くの愛好家コミュニティによって作成・管理されている日本の視覚メディアに関するデータ利用を意図している。幾つかの大規模なコミュニティのウェブサイトを中心に調査したところ、情報の深さ、ソース資料への深い理解、ボランティア貢献者の細部への高いこだわりを持ったコミュニティを発見できた。「JVMG」では、これらのコミュニティと接触し、彼らのニーズや動機について学ぶことが重要と位置付け、これらのコミュニティサイトの代表者や管理者と有意義な話し合いを行い、双方に利益をもたらす長期的な協力関係を築いていきたいと考えている。

日本のアニメに関する DB には世界中に様々なものが存在しているが、そこに採録されているデータが正しいか、間違っているのかを判断する基準を彼らは持っていない。様々な DB を比較検討し、ほぼ同じだから信用できるものだろうといった判断はできても、本当に正確かは分からない。そうした検討の中で、2019 年度に ATAC が実施した『アニメスタッフデータベースの持続的な構築・整備に向けての調査・検証』の報告書を読み、リスト制作委員会の原口氏の活動に関心を得られた。リスト制作委員会や原口氏が採録したデータや、全録サーバによって採録したデータは正しいものといった判断を下せ、それらが MADB を通じて使用できる可能性があると分かった。

その結果、日本のアニメに関する DB の現状についてヒアリングを行いたいとの要請があり、このプロジェクトが構築を検討している研究用 DB と関係できるとなれば、日本国内よりも先に海外で評価を得る結果にはなるが、広く多くに知ってもらえると考えた。そのため、海外であっても出向く意味があると判断し、参加して発表を行った。

発表では、まず現状整理として日本国のアニメに関する状況を紹介しつつ、「マチ★アソビ」「JAniCA チャンネル」で行ったようなアニメ作品の定義を示した。その上で、MADB への採録状況や、信頼できる情報ソース例を示して、本事業におけるテレビ作品からの採録事例が持つ特徴も知ってもらった。この後、アニメ作品情報にどのような利活用の可能性があるかも話し、作品情報がクリエイターの情報も含めて世界的に知られることの価値、動的ネットワークが構築されて最新の情報をやりとりできるようになる意義などを訴えた。

2.1.4 (仮称) アニメ作品情報連絡調整会

今年度の事業では、アニメのスタッフクレジット収集に当たり、製作委員会の主幹事社、制作会社からリスト収集を行っている公的・私的団体までが参加した緩やかな組織を作り、情報を補完し合うような関係を築くための「(仮称) アニメ作品情報連絡調整会」を組成する準備を進めた。

前提として、昨年度の実業でアニメ関連のクレジットデータを放送作品から採録する以外に、アニメ制作会社や製作委員会主幹事会社から情報を入手する道について検討を行い、アニメ制作会社、パ

第2章 成果・課題

パッケージメーカー、テレビ局等のプロデューサーにヒアリングを行った。結果として、より現場に近い情報は所有しているが、データの拠出には製作委員会の判断が必要になり、制作現場レベルの判断では自由には取り扱えない実態が分かった。

一方で、受皿となる窓口なり団体等が存在していれば、製作委員会参加会社なり宣伝・PR 会社から情報を拠出できるようになる可能性も示唆された。こうした調査結果を鑑 [かんが] み、窓口となる組織として「(仮称) アニメ作品情報連絡調整会」の組成を検討し、その活動方針や参加してほしい企業・団体・個人などについても検討を行った。

組織の狙いとして、オフィシャルのキービジュアルや各話スチールなどを私的なリスト採録・発信組織が使う際、公認とまでは行かずとも黙認をしてもらえる関係を作り、連絡が取り合えるようにすることを挙げた。情報が滞るなどの緊急時に相互扶助を可能にする課題も検討。それらを通して相互繁栄を遂げる目標などを挙げた。声かけを行う候補としては、①メーカーや産業側 以前にヒアリングをお願いした法人、②ウェブや紙の媒体、③アニメ業界団体 JAniCA、ATAC、一般社団法人日本動画協会 (AJA)、④採録している個人・団体等をリストアップした。

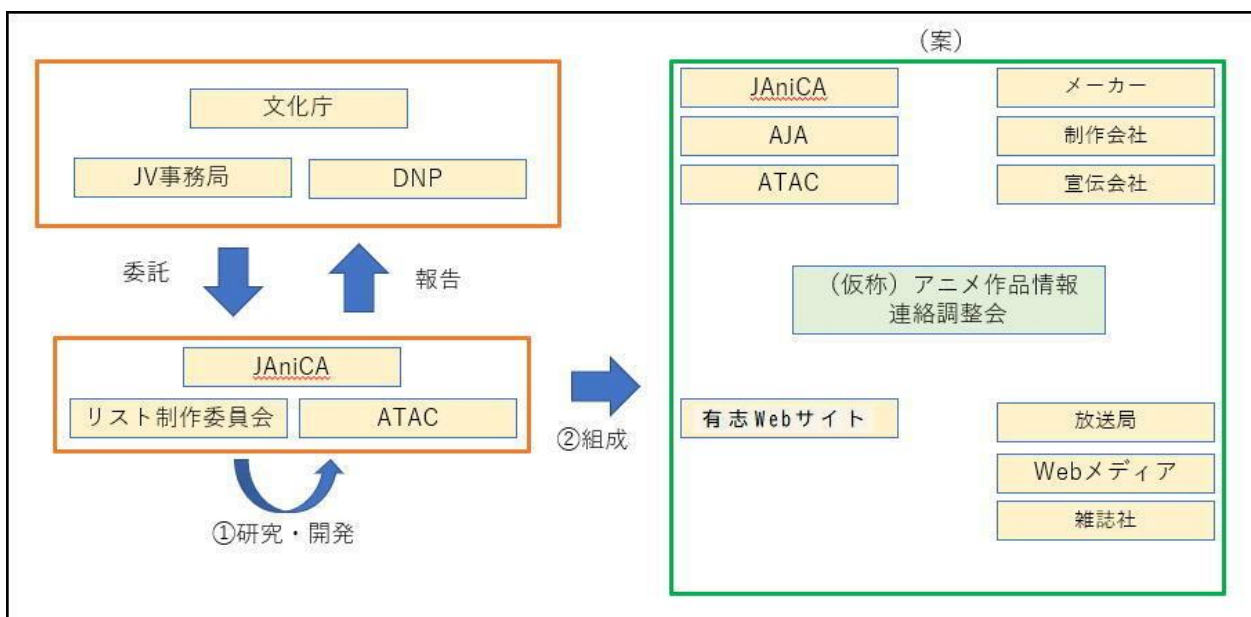


図 2-1 (仮称) アニメ作品情報連絡調整会の構成イメージ

全録サーバの場合、居住地の理由から録画できる範囲が首都圏に限られる。ここから外れてしまうような場合、連絡を取り合えるような経路があるだけで状況が変わってくる。また、集合知によって作成されているアニメスタッフのサイトの橋渡しを行えば、競争ではなく共栄へと発展させられる。それぞれのサイトでは少しずつ異なるデータが欠落している実状もあり、間を取り持つことで相互補完し合うような状況も作り出せると考えられる。

こうした問題意識を前提に、「(仮称) アニメ作品情報連絡調整会」の発足に向けた初会合を、令和 5 年 (2023 年) 1 月 13 日に実施。一定数の参加者を得て、将来の組成に向けた話し合いが進展した。

第2章 成果・課題

参加者はアニメ関連団体、アニメ放送事業者、アニメ制作会社、採録している個人・団体、アニメ雑誌編集者、アニメ関連サイト編集者、データベース研究者、アニメライターなど多岐に及んだ。

会合では冒頭で、JAniCAの入江泰浩代表理事が挨拶し、「(仮称)アニメ作品情報連絡調整会が実現すれば、アニメ制作会社を初め業界に関わる人たちにとってメリットが大きいと感じている。今後お願いする機会も増えると思うので、よろしくお願ひしたい」と協力を呼びかけた。ATACからは研究員の高橋望が「ネットを通じて情報は収集できるが、統合された信頼性の高いスタッフリストが要求されている。古いデータも新しいデータも大切だ。業界全体でスタッフのデータベースが持つ価値を認識し、作っていったら良いと考えている」とアニメ関連データベースの必要性を訴え、その活動に役立つ調整会の必要性を示唆した。

会合では、アニメ関連データベースの現状と、調整会の活動内容に関するプレゼンテーションが行われた。参加者からは「システムが稼働して追加キャストなどが分からないときなど、データ作成が楽になれば良いと考えている」(アニメ関連サイト編集者)、「次世代を視野に入れてどのように効率良く合理的にデータなり知識を蓄積していけば良いのか、などの問題に貢献するコミュニティの成立に、アカデミアに関わっている人間として興味を持っている」(研究者)との意見や期待の声が出た。

参加者数の多さ、参加職種の広さからアニメ業界にこうした調整会に対する関心の高さが見て取れたが、アニメ制作会社などは参加を誘う際に法人名なり部署名で案内状を送付した場合、先方で誰を参加させるかで迷うケースが発生した模様。活動が軌道に乗るまでは、権限を持つ責任者や関心を持つ個人名で参加を呼びかける必要性が、次回以降の課題として判明した。

2.2 結果からの推測、課題

2.2.1 OCR 選定及び自動化

本事業では、OCRに料金や性能の観点からGoogle Cloud Visionを利用したが、マイクロソフト提供のAzure Computer Visionも53回の利用で1,000トランザクションまでが7.356125円、1,000から1万トランザクションまでは4.781501円、1万から10万トランザクションが4.413681円、10万トランザクション以上が2.942454円となる。

67タイトルについて1回50枚を処理したとして、1週でGoogle Cloud Visionが663円、Azure Computer Visionが302円、1年分ではGoogle Cloud Visionが31,824円、Azure Computer Visionが13,496円となって、Azure Compute VisonがGoogleの半分程度のコストとなる。Azure Computer Visionは性能面についても上昇中となっており、今後の利用について検証を準備してみたいと考えている。この場合、マイクロソフトからの検証用アカウントを提供してもらう必要があるため、現時点で配布可能なシステムを構築するなら、やはりGoogle Colaboratoyの上で動くものが良いと言える。

いずれにしても、利用すればするほど料金が上がっていく従量制の場合、使用状況によって請求金額が高くなる事態が起こりえると改めて確認できた。また、こうしたクラウドAPIの場合、利用結果から文字情報などを追加しての学習がほぼできないため、ローカルに学習データを用意する必要性も見えてきた。

第2章 成果・課題

OCR からのスキャンでは、高解像度のデータなら極めて精緻に文字を拾い、テキスト化できるめどが立った。今後の課題は、低解像度の映像についても同様に OCR 化を行えるかの検証となる。リスト制作委員会の原口正宏氏宅に保管されている、古い時代の VHS テープなどには貴重な映像も多いと考えられる。こうしたデータを発掘し、デジタル化して DB に採録していく上で、低解像度のデータでも対応可能な OCR 化の技術が必要となる。

2.2.2 学習データ作成

OCR は、使用すれば使用するほど性能が上がっていく特徴がある。利用した結果をリファレンス（参照）データとしつつ、間違った部分を修正した結果を取り込ん（再帰する）で変換精度を洗練させていき、これを利活用すれば、コストも抑えつつ性能の高い専用の OCR を得られる。これを推し進めれば、スタッフ情報のメタ化専用 OCR を作成できるのではないかと考えられる。

こうした仕組みはモジュール上で動いている。今後はこのモジュールを洗練化させ、シーズンごとのアニメ作品に当たる分量を効率良く変換できる方向に持っていける見込みが立った。

利用に当たっては、1 回データセットを作ってモデルを作れば、スマホでも動作できるようになる。処理能力が高い機材でもなくても OCR は稼働するため、スマホなどで写真を撮るだけで文字を読み取って日本語に変換し、映し出すシステムも稼働し始めた。

ただし、初動の部分でデータセットを構築する作業などで、人の手がかかりハードウェアの負担も重くなる。後続の人が同じようなシステムを独自に構築する場合、最初のデータセットをどのように手に入れるか、先行する者が渡すのかなどの点が課題となる。しっかりとしたデータセットを生成できるかも一種のノウハウとして継承する必要がある。これは知的財産とも言え、権利的に出せない事態も起こりえる。機械的に使うものであるため、著作権の適用範囲外に当たる可能性もあるが、やはり道義的な部分で、機械学習用だからという理由で、実際に放送されたクレジット画面の使用や、機械学習用途ではない目的で連綿と蓄積された役職や人名データを無対価にて共有・公開することはためられる。こうした課題を解消し、誰でも使いやすいデータセットを構築し、汎用的に使える環境を作る必要がある。

構築したシステムのインターフェースを一般の人なり、リスト制作委員会のような団体が使いこなせるかも課題となる。現状では、録画した映像をシステムに投入すれば、テロップが出ている部分だけが全て抽出され、OCR で処理されてテキスト化される。課題はその先で、テキストデータを属性ごとに振り分けてデータベースに収納するには、機械的な OCR 処理とは異なり、柔軟な判断が可能な機能開発という大きな壁が明らかになった。その壁を乗り越えることが今後の課題となる。

2.2.3 災害等に強い録画体制

録画体制については、①リスト制作委員会の原口氏宅にて従来方式でのテレビ番組の録画実施と並行し、本事業で一昨年度に構築した②全録サーバ（通称：杉並サーバ）と、予備系として運用を行っ

第2章 成果・課題

ている③全録サーバ（通称：千代田サーバ）の3拠点において、引き続きそれぞれ異なる録画手法を用いてテレビアニメ作品の全録画を行い、耐障害性を高めている。異なる地域で実施している理由は主に災害対策である。これは、純粋な荒天・火災だけでなく、地域的な停電や受信障害も想定して、同じ原因で全ての録画が止まることのないよう注意を払っている。

さらに、それぞれの録画現場において①目視、②EPG、③集合知を用いたアニメ番組の認識により、アニメ番組の録画漏れの低減にも取り組んでいる。一昨年度事業に始まり、昨年度事業、今年度事業にて全録サーバを稼働させたため、1年間（令和4年（2022年）1月1日から12月31日）で約10,000番組（再放送や関連番組を含む）を録画する必要があり、HDD容量として約15TBが必要と分かった。最終的には相互に録画データは補完できたものの、それぞれの拠点において、録画機器の不調や受信障害などが発生、また、風水害やニュース速報、大規模スポーツイベントによる放送時間の移動、コロナ禍やハッキングによる制作遅延、編成変更（遅延振替、総集編、再放送、最終話2段積み、一挙再放送）といった理由から、1%程度の頻度で録画に失敗することが今年度も確認できた。どの録画手法においても録画の失敗数をゼロにできなかったため、単一の録画方式の採用は望ましくなく、複数の拠点と手法での運用が、耐障害性を高める上で有益であると再確認できた。

災害については、施設の切替えなどによる停電なら事前に情報が得られる場合が多く、対応は可能だが、落雷などによる急な広域停電の回避は難しい。水害も発生してからでは機材の運び出しなどの対応は難しいと考えられる。火災による消失も避けなければならない。原口氏宅で録画された番組のテープ類などの中には、他所〔よそ〕に保管されていないものも相当数あると見られる。これらが焼失等によって永遠に失われないためにも、早急のデータ化及びデータ類のバックアップが必要との認識に達した。

NHKが、令和4年（2022年）9月14日に放送した番組『ザ・バックヤード 知の迷宮の裏側探訪』の「国立映画アーカイブ お宝映像の宝庫」回によれば、アーカイブ処理が進められても、その価値が認められて“お宝”と言われるようになるまでには、時間がかかるという。アニメなりポップカルチャーの分野も、将来に向けて保存施設を形にしていく必要がある。施設建設には批判もあるが、現実問題として施設がなければ安定的な長期保存はできないため、様々な災害に備えた施設の設置・運用を急ぐなど、今後の動きが大切になってくる。

2.2.4 配信など多種映像メディアへの対応

本事業では、年間に生成・公開される作品の本数換算で約8割を占めると推測しているテレビアニメ作品について全録サーバにて録画を行い、蓄積された映像に対して省力的で効率的なクレジットデータのテキスト化やメタ化処理を実施している。しかし、アニメ作品は劇場作品やパッケージでの新作映像などの形でも公開されており、近年はこれに加えて配信サービスでの提供も増えてきた。新規サービスも続々と立ち上がり、契約者獲得に向けて作品の独占配信を行うところも少なからず出てき

第2章 成果・課題

ている。このようなメディア形状で公開されている作品についても、法的・技術的・道義的な観点に配慮した上で、全録サーバに映像データを格納すれば、テレビアニメ作品と同様にクレジット情報の抽出が可能と考えられる。全録サーバで録画された映像からクレジット掲出部分をキャプチャして静止画として記録し OCR へと投入しているが、技術的には放送・配信中の映像をビデオボードで PC に取り込んでリアルタイムで OCR にかける方法も視野に入れる必要がある。

劇場作品やオリジナルビデオアニメ（OVA）などパッケージによって提供される作品について、従来は原口正宏氏とリスト制作委員会が目視と手作業でスタッフ・キャスト名を抽出し、スタッフデータベースを作成してきた。こうした作業の全録サーバと OCR の組合せへの置換に加え、配信作品からのクレジットデータ収集が可能となれば、メディア芸術データベースの利用者の調査・研究や、事業展開を行う上で有意義に働くと期待できる。

これは同時に、リスト制作委員会が負担してきた作業の軽減にもつながり、原口氏の持つアニメスタッフやアニメスタジオに関する知見の、学術的な分野での活用にもつながる。日本におけるアニメ産業の変遷を、より精緻に掘り起こして歴史として定着させられるようになる。長く懸案となっている、原口氏宅に所蔵されたテープ類の映像の把握・検証・分類にも本格的に着手して、データ量を増やすこともできる。

本事業を通して、「映像から文字情報を効率的に取り出す」という技術が確立されれば、その適用範囲はアニメにとどまらず、あらゆる映像作品に対してクレジットデータなり、掲出されている様々な文字情報をテキスト化して保存し、容易に利活用できるようになる。非常に汎用的な可能性を秘めた技術と言える。

2.3 利活用事例

氷川竜介・明治大学大学院特任教授から「リスト DB」に記録されているデータを参照し、調査したいとの依頼があり、対応した。明治大学大学院生の令和4年（2022年）度修士論文（非公開）「SF・ロボットアニメにおけるモニターグラフィックス」の執筆指導に当たり「モニターグラフィックス」という役職名がアニメのスタッフクレジットにいつから出現したかを、リスト制作委員会の原口正宏に尋ねてもらいたいというものであった。ATACを経由して依頼を受け、原口氏から預かっている「リスト DB」の記録から「モニターグラフィックス」と付けられている役職名、更に関連していると思われる役職名を一式、検索して抽出した。

クレジット表記の抜粋(各話役職、役職ヨミに"モニタ"を含む)

DISTYEAR	TITLE	NUMBER	WORKNM	
2001	ゾイド 新世紀/ZEF	1	モニターエフェクト	藤本さとる
2001	ゾイド 新世紀/ZEF	1	モニターエフェクト	加藤榮
2001	ゾイド 新世紀/ZEF	1	モニターエフェクト	ソニーFCL
2001	ゾイド 新世紀/ZEF	1	モニターエフェクト	KangDonghyuk
2001	ゾイド 新世紀/ZEF	1	モニターエフェクト	森本聖子
2001	ゾイド 新世紀/ZEF	1	モニターエフェクト	ラスオハウス
2001	ゾイド 新世紀/ZEF	1	モニターエフェクト	おおのやよい
2001	ゾイド 新世紀/ZEF	1	モニターエフェクト	中野剛
2003	宇宙のステルヴィア	1	モニターCG	加藤榮
2003	宇宙のステルヴィア	1	モニターCG	ソニーFCL
2003	宇宙のステルヴィア	1	モニターCG	KangDonghyuk
2003	宇宙のステルヴィア	1	モニターCG	中野剛
2003	宇宙のステルヴィア	1	モニターCG	藤本さとる
2003	宇宙のステルヴィア	1	モニターCG	ラスオハウス
2003	宇宙のステルヴィア	1	モニターCG	おおのやよい
2003	宇宙のステルヴィア	1	モニターCG	森本聖子

図 2-2 役職読みに「モニター」を含む例 1

クレジット表記の抜粋(シリーズ共通役職、役職ヨミに"モニタ"を含む)

DISTYEAR	TITLE	NUMBER	WORKNM	NAME
1998	青の6号	2	モニターグラフィックス	滝老川謙武
1998	青の6号	2	モニターグラフィックス	rock[ロゴ]
1998	青の6号	2	モニターグラフィックス	歌由紀さおり・安田祥子
1998	青の6号	2	モニターグラフィックス	宮尾佳和
1998	青の6号	2	モニターグラフィックス	「KZUNA〜絆」
1998	青の6号	2	モニターグラフィックス	金山
2004	爆裂天使	1	モニターグラフィックス	KDOI
2004	SAMURAI 7	5	モニターグラフィックス	佐藤菜津子
2004	SAMURAI 7	5	モニターグラフィックス	作[?]鈴木まなか
2004	爆裂天使	2	モニターグラフィックス	佐藤菜津子
2004	爆裂天使	2	モニターグラフィックス	作[?]鈴木まなか
2004	爆裂天使	2	モニターグラフィックス	滝老川謙武
2004	爆裂天使	2	モニターグラフィックス	KDOI
2004	爆裂天使	1	モニターグラフィックス	佐藤菜津子
2004	爆裂天使	1	モニターグラフィックス	作[?]鈴木まなか
2004	爆裂天使	1	モニターグラフィックス	滝老川謙武
2004	爆裂天使	1	モニターグラフィックス	サイトロン・デジタルコンテンツ株式会社
2004	爆裂天使	2	モニターグラフィックス	サイトロン・デジタルコンテンツ株式会社

図 2-3 役職読みに「モニター」を含む例 2

メディア芸術データベース(以下、MADB)に格納されているデータに「モニターグラフィックス」という役職が搭載されていても、検索結果として抽出できるのは作品名や話数となってしまう、役職名を単品で取り出すことは難しい。また、「いつからか」という条件に関して、精密にトレースすることも難しい。今案件は原口氏から預かった「リストDB」のデータがあったため、全データに当たることが可能となって、適切に対応できた。

第2章 成果・課題

依頼に応じ、揺らぎを含む「モニターグラフィックス」の役職名が初めてクレジットに登場した作品や、その後に登場した作品、担当した人名又は会社名をリストアップした。作業時間は15分くらいだった。氷川氏からは「当初、これほどの短時間で作業を終えられるとは予測しておらず、実証研究に十分に使える」といったコメントを得られた。「リストDB」がMADBにつながれば、こうした使い方ができるメリットを先行的に確認できた一例と言える。

調査の結果、小沢さとる氏の漫画原作をGONZOがアニメ化した『青の6号』（1997年）に、「モニターグラフィックス」として海老川武氏の名前が、最初の登場例としてクレジット表記されている状況が判明した。これは事前に氷川特任教授が予想していた状況と合致する（同作は、2D、3Dをハイブリッドしたフルデジタル制作のため）。また、研究テーマの副題は「一デジタル制作以後の役割変化について」であるため、実証的に裏付けられた成果は、研究価値を高めるものである。検索結果には、会社名や歌手の名前も並んでいるが、その中では、『SAMURAI7』の佐藤菜津子氏も、氷川特任教授が求めた「モニターグラフィックス」に該当する人名と見なせると判明した。

「モニターグラフィックス」という名称を調べると、同じ作品で複数の話数が出てくる。各話のスタッフリストを見ていくと役職名と人名が見つかる。こうして2段階3段階と進めていけば作業を完了できる。このように、ある目的を持って探す行為により、データの信頼性を高められる。調査結果の提供とは別に、信頼性確保のためにこうした副次的な対応を行う必要性も明らかになった。

「アニメ全作品年間パーフェクトデータ」でも、テキストで入稿されたものについて、「リストDB」に全量が反映されている訳ではない状況も分かった。原口氏が保有する作業用のテキストファイルを串検索した方が、より網羅的に調べられた可能性が浮上した。予備調査だったが、その中でもデータの不整合やつながり方のミス、データの欠落に気づけた。

2.4 その他特記事項（アニメデータベースを構築する意義の再確認）

本事業の中間報告会が令和4年10月13日に開催され、有識者検討委員会の委員から本事業が最終的な目的とするアニメ関連情報収集の意義について質問があった。平成29年度から継続的に取り組んできた事業であるが、採録のための全録サーバ構築やキャプチャ画面の作成・OCR等のシステム構築、及びOCRなどの技術的成果を採求する傍らで、本目的とするアニメデータベース構築の意義について、認識が薄れてきた状況を鑑み、有識者検討委員会委員からの疑義に対する答えを改めて提示して、本事業そのものへの関心喚起にも取り組んだ。

クレジットを読み込み収集する作業に、どのような必然性があるかとの問いかけについてだが、アニメを含むクリエイティブの成果としての作品は、「人」が主（しゅ）となって制作に携わり、結果として「作品」という従（じゅう）が出来上がる。ここで主となる人が作ったものとして作品の記録を残していく行為が、そうした人たちが携わった作品群としての記録と伝承につながると考えている。

第2章 成果・課題

現在、アニメに携わる人たちの労働環境の不安定性が問題とされるケースがある。スタッフ名が作品に紐付けられて正確に記録されていけば、キャリアをリアルタイムに把握できるようになり、労働環境の向上につながるものと見ている。

アニメ制作会社や製作委員会構成会社にとっても、こうしたスタッフリストの存在は、新規の企画やスタッフ構成の策定に役立つものと考えられる。あるキャラクターデザインを手掛けたデザイナーが過去にどのような作品を手掛けてきたか、監督として起用を予定している人の業歴はどのようなものだったかを確認する際に、本事業で収集したデータを利活用すれば、より正確な情報を簡単に得られて、求める作品のクオリティを達成できる。

同種の研究を、例えばNHKの放送技術研究所などが行っている可能性もあって、重複を避けるために調査したが、放送技術研究所において将来的に公開される研究は行われていない実態が分かった。また、NDLが書籍を対象にOCR処理を行い、膨大なデータを蓄積したと発表されたが、主に白地に黒で筆記される書籍と、背景が常に変化する映像とでは同じOCRでも性質が大いに異なる点があった。NDLの場合はまた、教師データの蓄積に長い年月を費やしており、全く異なる教師データが必要なアニメ関連情報のOCRに技術を直接流用するのは困難なことも分かった。

そもそも本事業は、「網羅的なスタッフデータベース」のニーズがアニメ関連業界団体にあって、その目的に沿うように構築を始めたものではなく、現状では原口正宏氏を主宰とするリスト制作委員会の活動成果となる「リストDB」において存在していない、より正確でより作品に近い形でスタッフ・キャスト情報を採録する仕組みを、代替できるようなシステム作りを目指したことを発端としている。シーズとしてより正確でかつ網羅的なデータの存在している状況を第一義とし、それらが将来的に保存・運用されるとどのようなメリットが生じるのか、例えば研究の促進なり新規事業の拡大に資するなどのニーズを考えた利活用により、重要性の再確認を目指したものと言える。

過去において営々と積み上げられてきた正確なスタッフ・キャスト情報のリストが、令和3年(2021年)から開催されている『アニメージュとジブリ展』において、著名なアニメーターであった大塚康生氏の経歴作成などに役立つ事例もあり、改めてリスト制作の意義が確認され、本事業が持つ役割も裏打ちされた。本事業に限らず、一般社団法人日本動画協会が「アニメ_NEXT100」プロジェクトとして構築を目指したデータベース「アニメ大全」が公開となり、株式会社スカパーJSATが事業を通して収集してきたものに、事業者から提供されるデータを加えて「コンテンツデータベース」の構築を目指すと発表した事実も、アニメデータベースの必要性を裏付けている。

こうした一連の動きの中でも、本事業がよりどころとする放送番組からのスタッフ・キャスト情報採録が持つ正確性と迅速性はほかにはない特色であり、ほかのデータベースの併用により比較検証も可能になる。ここに至るまでに、原口氏個人を中心としたリスト制作委員会メンバーの活動が果たしてきた役割に敬意を払いつつ、昨今のアニメ制作本数と制作スタッフ数の増加によって人力では不可能となっているデータ採録を、技術によって省力化し作業の一部なり全部を代替できるようにする目

第2章 成果・課題

的もあって、本事業は立ち上がり、ここまで継続してきた経緯を改めて指摘しておきたい。

第3章 実施内容

3.1 会議開催

3.1.1 定例会議

今年度はコロナ禍における感染防止の観点から、実際に集まったの会合は避け、ネット会議ソフトの Zoom を利用して参加メンバーによる月次の定例報告会議を開催した。

令和4年5月31日 内部会議（第1回）
令和4年6月23日 内部会議（第2回）
令和4年7月29日 内部会議（第3回）
令和4年8月30日 内部会議（第4回）
令和4年9月30日 内部会議（第5回）
令和4年10月31日 内部会議（第6回）
令和4年11月25日 内部会議（第7回）
令和4年12月23日 内部会議（第8回）
令和5年1月20日 内部会議（第9回）
令和5年2月17日 内部会議（第10回）

3.1.2 周知活動

アニメのデータ収集と DB 構築が持つ意義を知ってもらい、活動に理解を得るために、アニメファンやアニメ研究者が集まるイベント会場や、アニメ関係者が視聴するウェブ配信チャンネルに登壇した。海外からも日本のアニメに関連した情報の記録状況を知りたいという要望があり、ドイツのシュトゥットガルトにある Stuttgart Media University で進められている「Japanese Visual Media Graph」プロジェクトのイベントで、本事業の概要を含む日本のアニメ関係 DB の現状を講演した。

令和4年10月15日 徳島「マチ★アソビ」でのアニメ DB 関連講演登壇
令和4年10月29日 JAniCA チャンネル「アニメのデータベース」配信
令和4年12月30日 コミケットフリンジ登壇
令和5年1月28日 ドイツ・シュトゥットガルト「JVMG」登壇

3.1.3 (仮称) アニメ作品情報調連絡調整会

全録サーバを活用して放送作品からスタッフ・キャスト情報を採録することとは別に、こうした情報を持つアニメ制作会社や製作委員会会社、宣伝会社から情報を入手できるように関係を構築するとともに、同様にアニメ関連情報を発信している雑誌やウェブ媒体、独自にアニメ関連情報を採録している団体・個人とも関係性を作り、情報をやりとりしたり、緊急時の相互扶助を行えるようにしたりして、アニメに関わる者たちが相互に繁栄していきけるような仕組みとなる「(仮称) アニメ作品情報連

第3章 実施内容

絡調整会」の創設を企図。その準備に向けて関係者の顔合わせを行う場を設けた。

令和5年1月13日 (仮称) アニメ作品情報連絡調整会顔合わせ

3.1.4 その他会議

上記以外の会議については適宜行った。

令和4年4月28日 記録作成のための方針会議

令和4年10月13日 中間報告会

令和5年2月20日 最終報告会

個々の作業においては、ネットワークを接続して、千代田サーバ、杉並サーバで録画した番組を基に、OCR化の効率化や生成されるデータの精度向上に必要なプログラムの作成、それぞれに採集してきたデータの検証、付随する EPG データのテキスト化に必要なプログラムの作成、リスト DB 収録データのメディア芸術データベース向けデータへの変換プログラム作成、リスト制作委員側の要請による不足データの提供を行った。付随する相談、検討についても随時行った。

3.2 高精度化・メタデータ処理

3.2.1 機能要件、動作概要

採録するデータの高精度化には、学習データの充実が必要である。既存のクラウドサービスは、出力されたテキストに誤字などがあっても、クラウドサービスに指示しての再学習や、修正結果の反映などはできない。このため、カスタマイズできない部分のメタ情報処理を行うためのデータセットを構築することにした。

教師データを活用し、画面に表示されたクレジットを、自動的に異なる色の枠で囲んで役職や名前の属性を指摘する取組となる。

第3章 実施内容



図 3-1 クレジットへの属性添付

教師データを参照し、固有名詞に対してメタデータ処理を行う作業となる。この自動化については将来的には可能とのめどが付いたが、現実的な実用レベルとするためには、より多くの教師データが必要となると確認された。現状では、アニメ関連の情報全てに対応できるだけのデータセットが存在していない。今後、役職など数百項目のデータを作り、反映していく必要がある。

ここで使用した、教師データから属性を指摘する AI については、一つ一つをゼロの部分から作っていった。作業では、キャプチャした画面に対し、文字を認識するソフトを再利用し、画面上の文字を確定する。続けて、確定した文字に対して解析を行い、属性を把握して役職名か人名かで色分けを行う。

第3章 実施内容

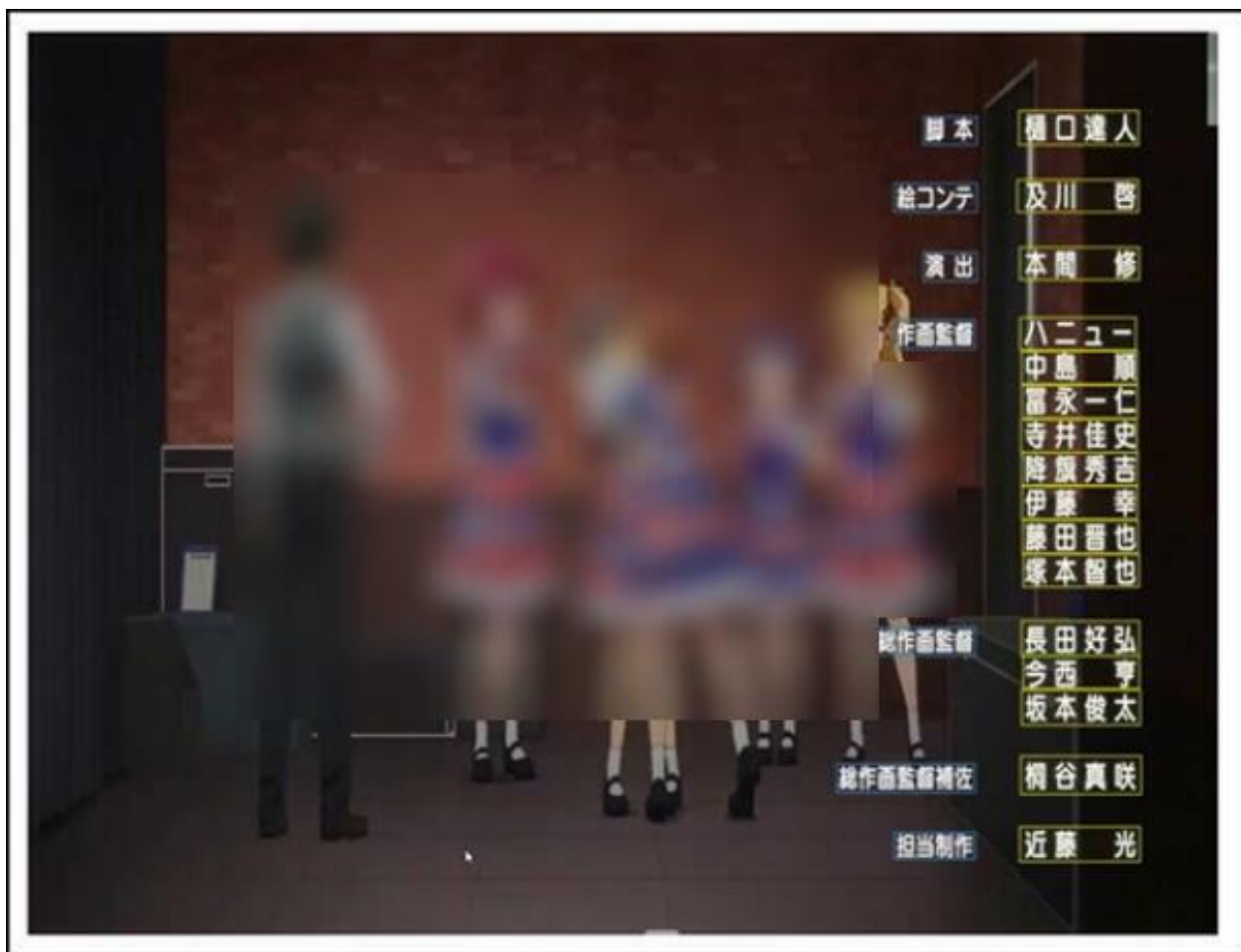


図 3-2 文字部分を把握する

第3章 実施内容

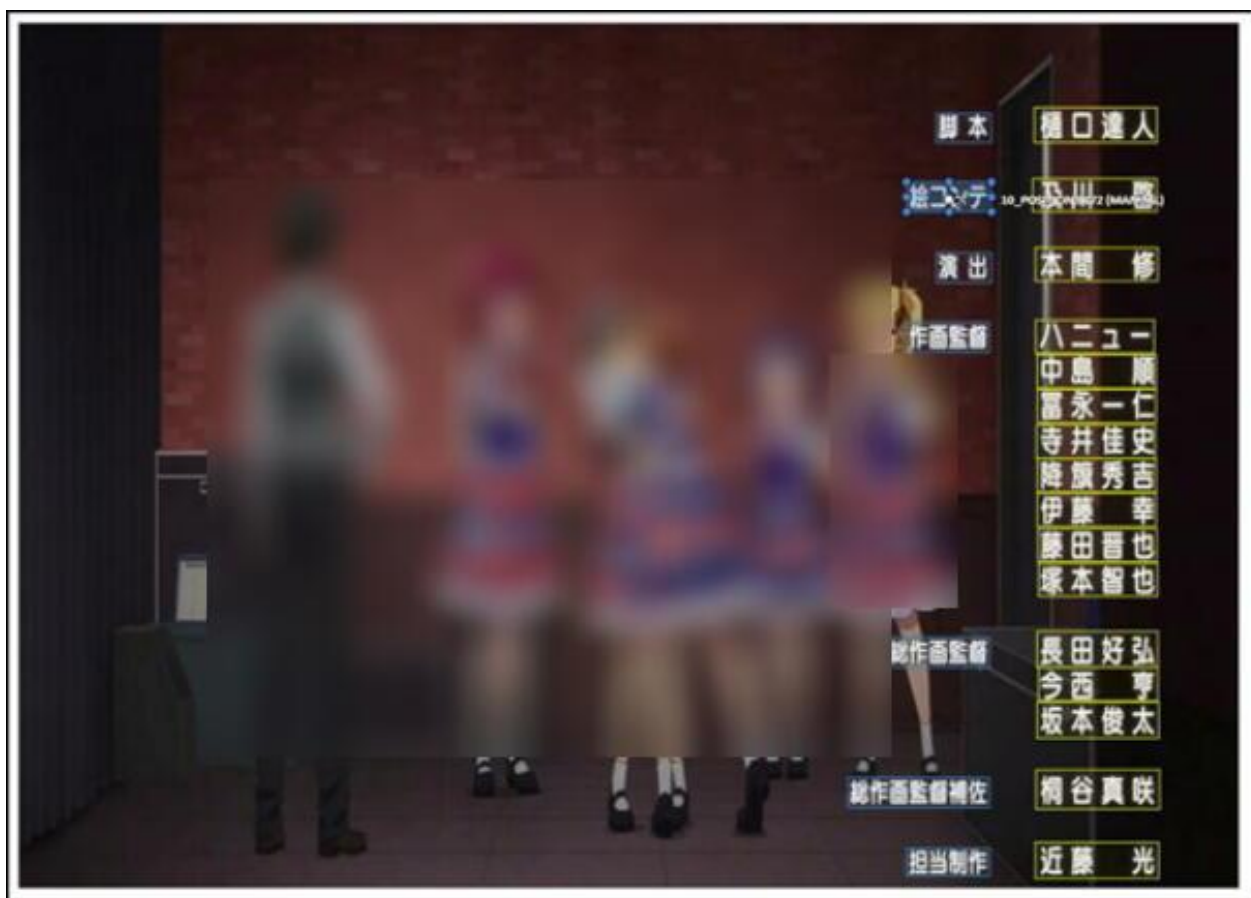


図 3-3 文字解析によって属性を把握し色分けをする

一つのタイトルについて、オンエアされた分について属性を付け、データセットとして活用してAIのモデルを作成する。話数ごとにこうした作業を行っておくと、シリーズが終わっていても、続けてオンエアされた新規の話数に対して、自動的にメタデータ、各項目の属性が付いていくようになる。

第3章 実施内容



図 3-4 別タイトルでのレイアウト把握からの文字解析による属性付与と色分けの例

アニメの場合、タイトルごとにレイアウトのフォーマットが存在している。同じ作品でも第1話の放送時と、途中でエピソードが盛り上がったときとでフォーマットが変わるケースもある。幾つものパターンを学習させておくことで、新しくオンエアされる話数でも高い精度で属性を把握できるようになる。

ここまでは、クラウドを必要としなくても行える作業だが、これにクラウドの OCR を使えば、データに情報を載せられるようになる。間違ったものの修正がツールで可能となる。

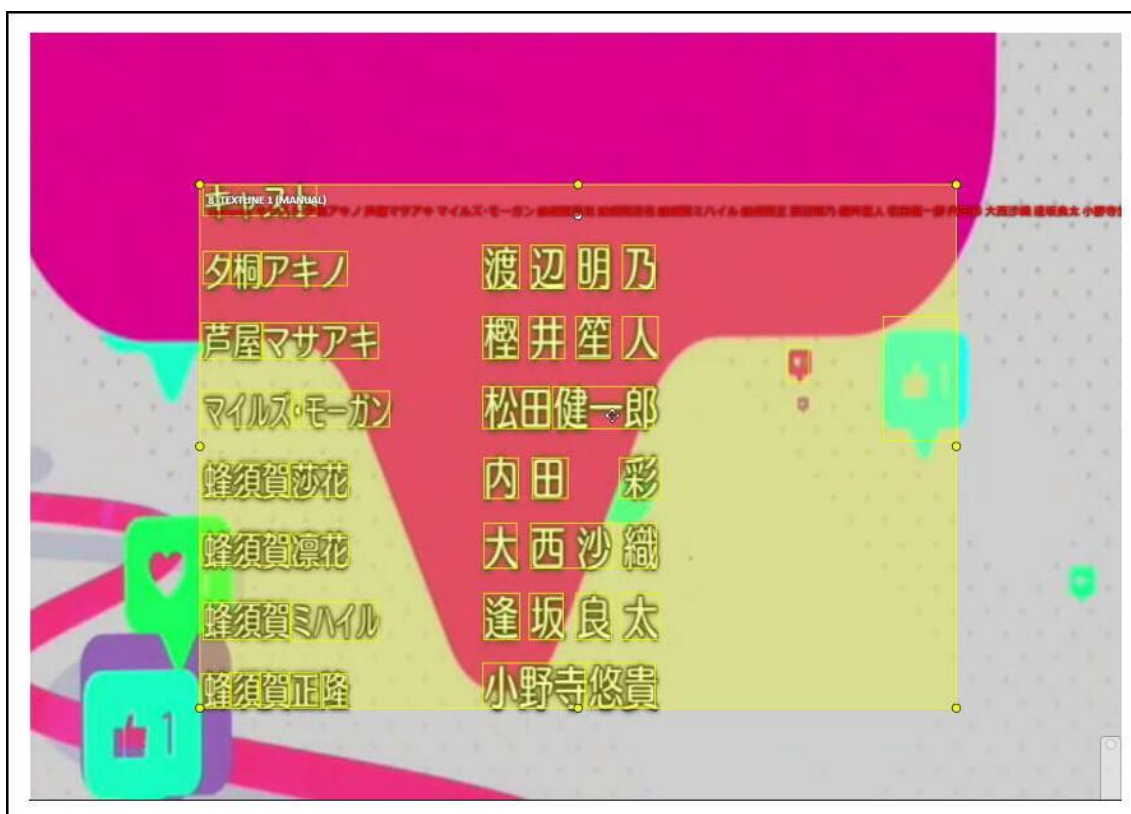


図 3-5 OCR による文字部分の把握

ツールで認識した情報は構造化されているため、いつでもクレジットデータの詳細な情報を位置も含めて保存し、必要に応じて再現できる。

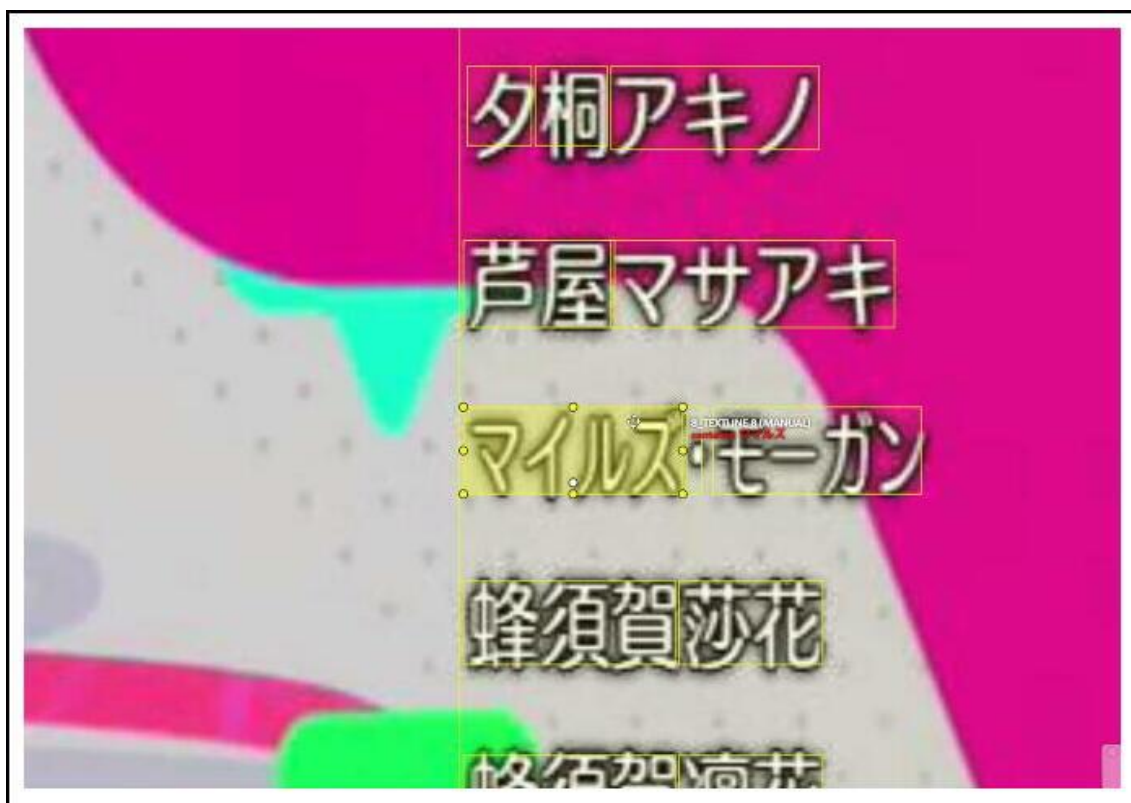


図 3-6 ツールを介在させることで位置などを把握し記録できる

こうした作業を推し進めていけば、最終的には、クレジット画面に対して変換精度の高い言わばアニメ専用の OCR を構築できるようになる。

画面のレイアウトを認識するためのデータセットも生成した。出力されたデータが実際の画面と一致しているかの検証は、画像とレイアウトされたテキストを重ねることによって行う。人間が比較検証をしてどこが間違っているかといった検証を行える。

第3章 実施内容

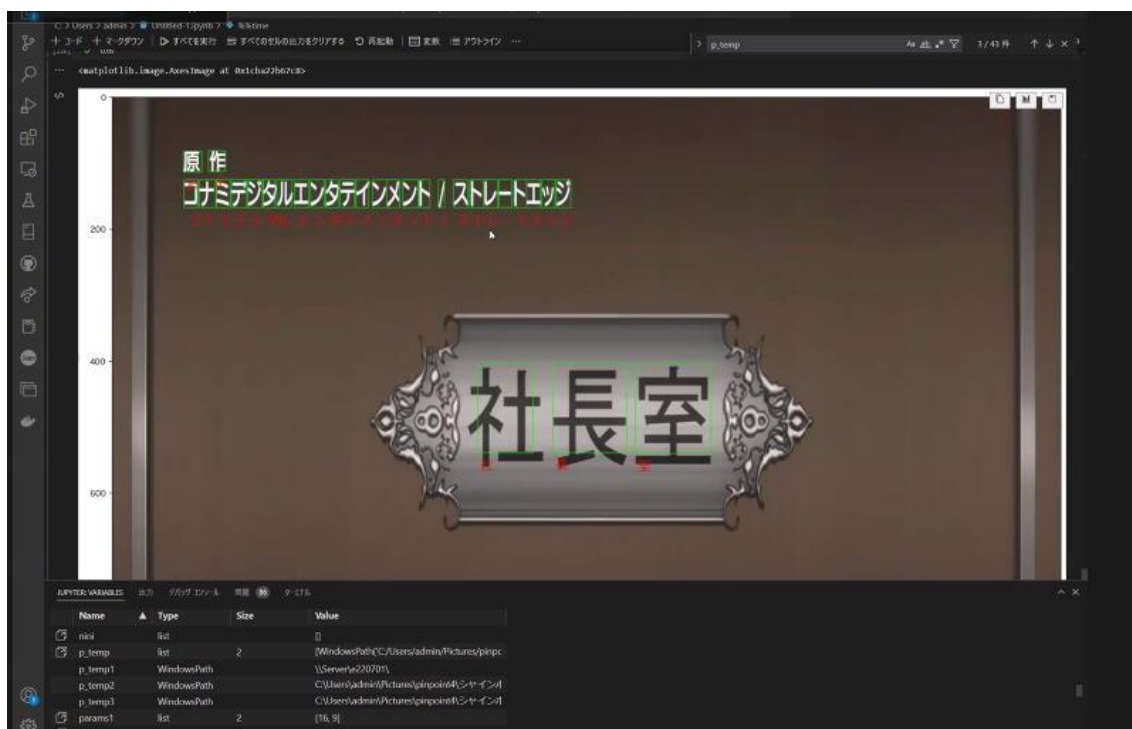


図 3-7 キャプチャ画面の文字と読み取ったテキストの照合

テキストの基となった文字が、画面上のどこの位置に掲出されていたかといった情報を持っているため、テキストだけで元の画面と同じようなレイアウトを再現できる。フォントのサイズも付加すれば、ほぼ同じ見た目での再現が可能になる。

現行でクレジットデータのキャプチャ画面からの採録を行っているリスト制作委員会では、テレビ画面をデジタルカメラで撮影し、テキストと照合して正誤の確認をしている。レイアウト情報を保存しつつキャプチャ画面も保存しておけば、作業に慣れていない初心者でも容易に差異を確認できるようになる。

レイアウトに関しては、画面上に掲出されたクレジットの並びに近い形でのテキスト化にも取り組んだ。画面の片端からスキャンを行う OCR の機能上、検知された順に文字がテキスト化されるため、投入する画面によっては先に役職名が並び、その後に人名が並ぶ事象が発生する。これを役職名と人名の並びを再現する形でテキスト表示できるようにして、見ただけで属性を把握しやすいようにした。

第3章 実施内容

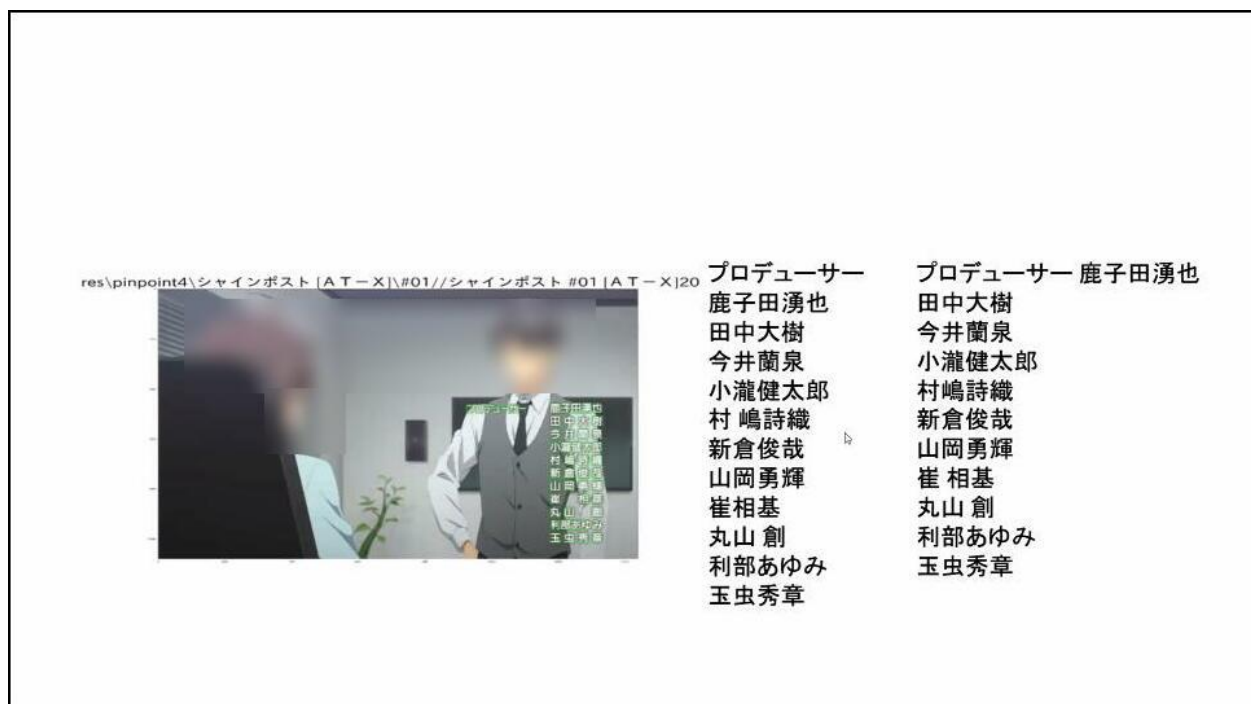


図 3-8 レイアウト把握から並び順を再現しテキスト化する例 1



図 3-9 レイアウト把握から並び順を再現しテキスト化する例 2

3.2.2 成果・効果・制約事項

キャプチャ画面から文字を認識するためには、AI モデルが必要となる。ここで高性能のクラウド AI を利用するとなると、料金が高額なものになってしまうため、独自に作る必要がある。ただし、AI を含む文字判定機能にどれだけの費用をかけられるのか、どれくらいのデータを準備できるのか、そこで教え込むための教師用データの質に関する課題が生じる。

第3章 実施内容

再配布可能なシステムを作ろうとした場合、ハードウェアや OCR といったハードやソフトウェアなどのシステム自体はコピーして構築できても、判断を行う“頭脳”とも言えるノウハウ部分は、独自に構築した教師データや、操作する側の知見に依拠することになる。システム自体の再配布は可能でも、こうした AI や知見が及ばず、物足りないものになる可能性がある。これが今年度の作業を通して得られた課題と言える。

そのため、各人で構築したシステムが実用的なものとなるかならないかといった分かれ目は、やはり AI の基礎学力となる。AI のテンプレートを基盤にいろいろなものが組み込まれていれば、より高度な結果を得られる。本事業でも、数年来の作業で積み重ねてきた基礎データに、新たなデータを加えるアプローチを行っているが、レベルはまだ低い。

また、OCR によって文字をしっかりと認識することができても、その文字が何と書かれているかを認識するには、圧倒的なリソースと経験値が必要となる。それを所持しているクラウドサービスは識別精度も費用も高くなる。加えて、アニメのリソースを所持しているサービスも現時点では存在していない。ただ、これに特化させることは学習によって可能であり、将来のためにアニメ特化のためのベース作りを積み重ねておく必要がある。

3.3 再配布可能なシステム

3.3.1 機能要件、動作概要

全録サーバからクレジットデータを抽出し、テキスト化する作業を行う上で必要なソフトウェア群をモジュール化するプラットフォームは「Google Colaboratory」を選択した。このうち OCR 部分については、Google の Google Cloud Vision API を活用した。

作業では、Google Colaboratory 上に構築したモジュールにアニメ作品の映像ファイルを選んでセットアップし、OCR 部分を稼働させてアニメ作品からテキストを抽出する。この作業で、Google ドライブに全話数分の映像ファイルをセットアップするまでにかかる時間は1分程度となる。

第3章 実施内容

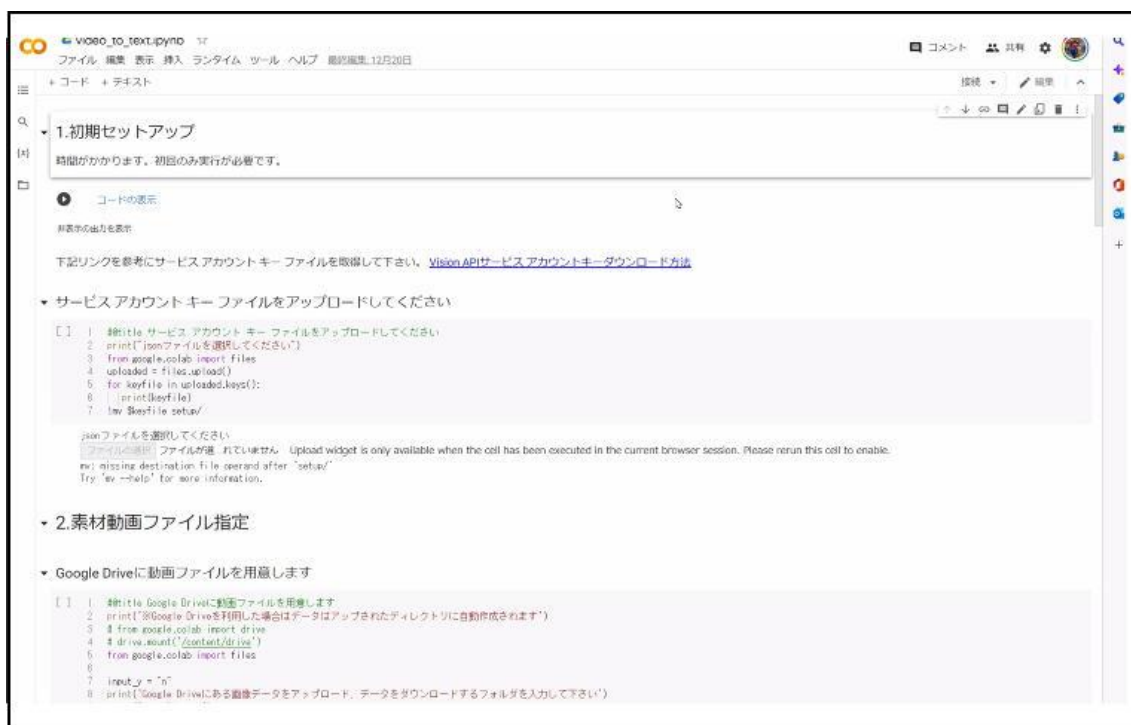


図 3-10 Google Colaboratory への映像ファイルのセットアップ



図 3-11 素材動画ファイルの指定

第3章 実施内容

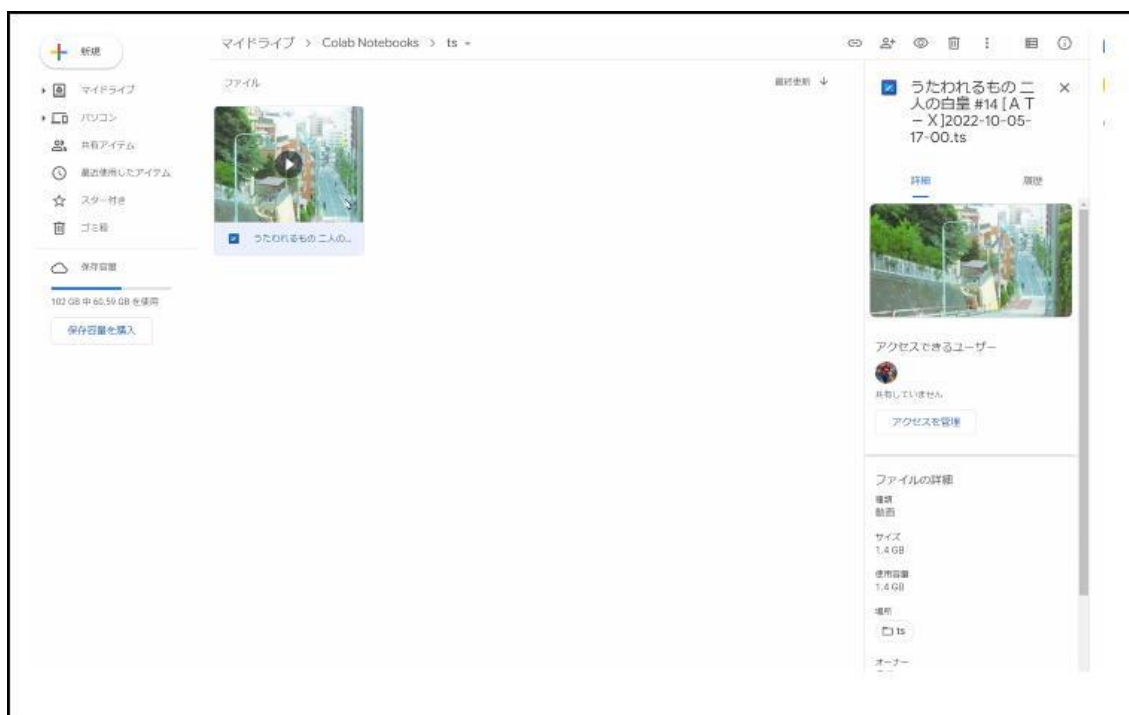


図 3-12 Google Colaboratory への映像ファイルのアップロード

映像ファイルを指定すると、キャプチャされた画面が抽出される。ここからテキストが掲出された画像を選 [え] り分ける。

第3章 実施内容

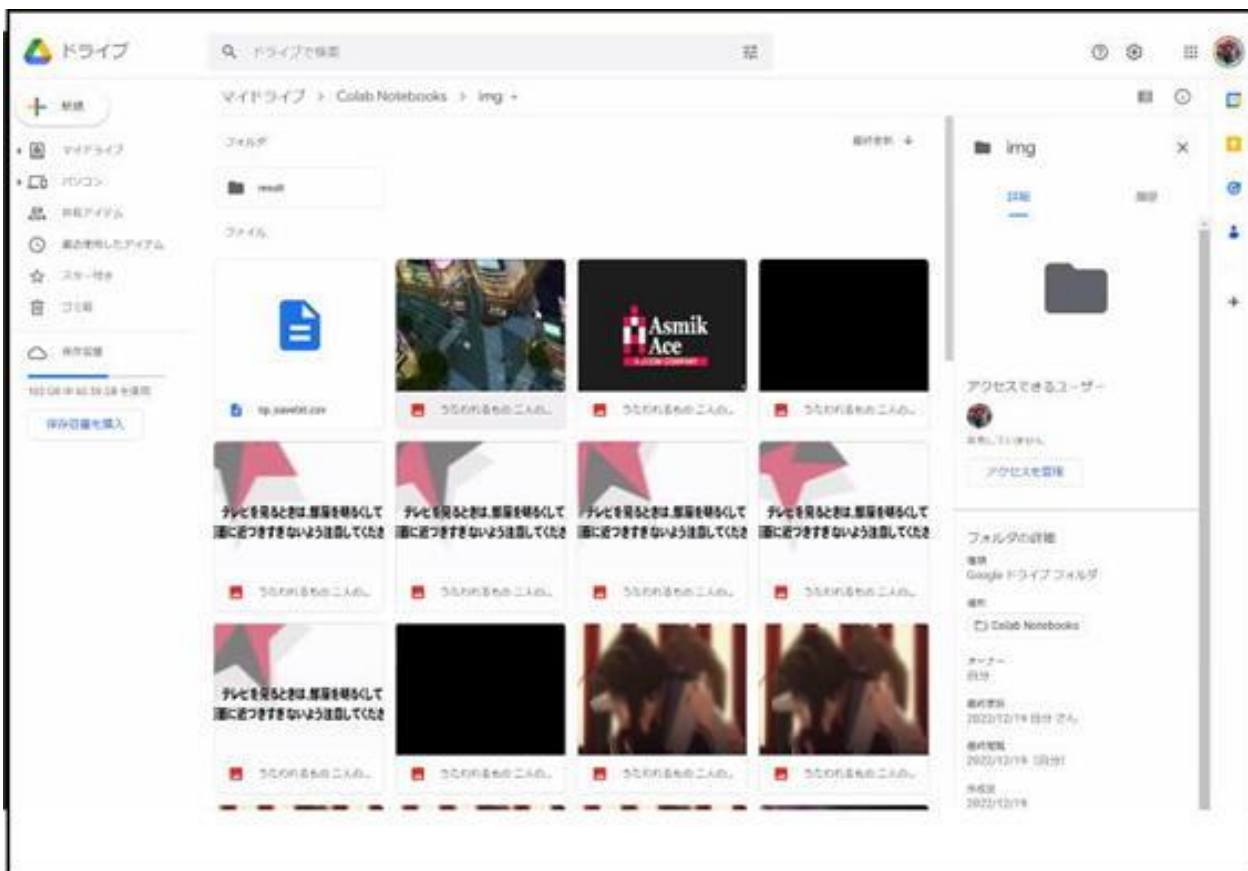


図 3-13 映像よりキャプチャされた画面

第3章 実施内容

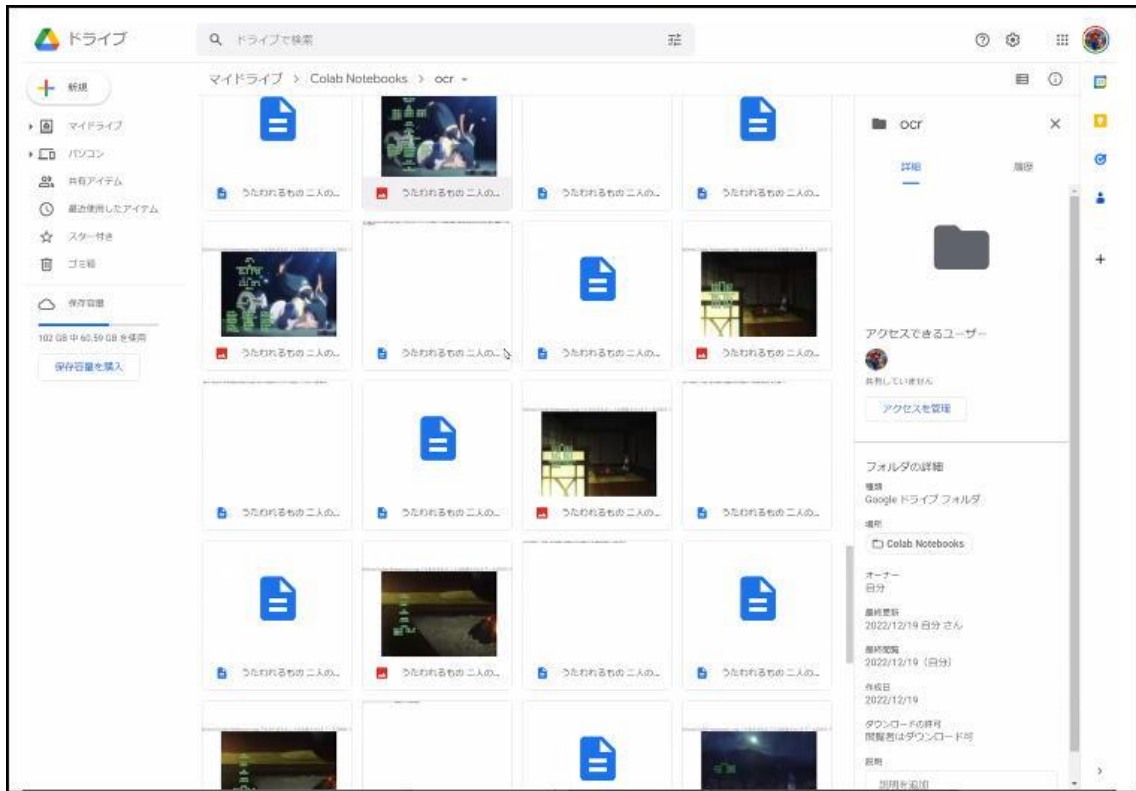


図 3-14 テキスト部分を持った画像の抽出

これらの画像を OCR へと取り込みテキストを抽出する。

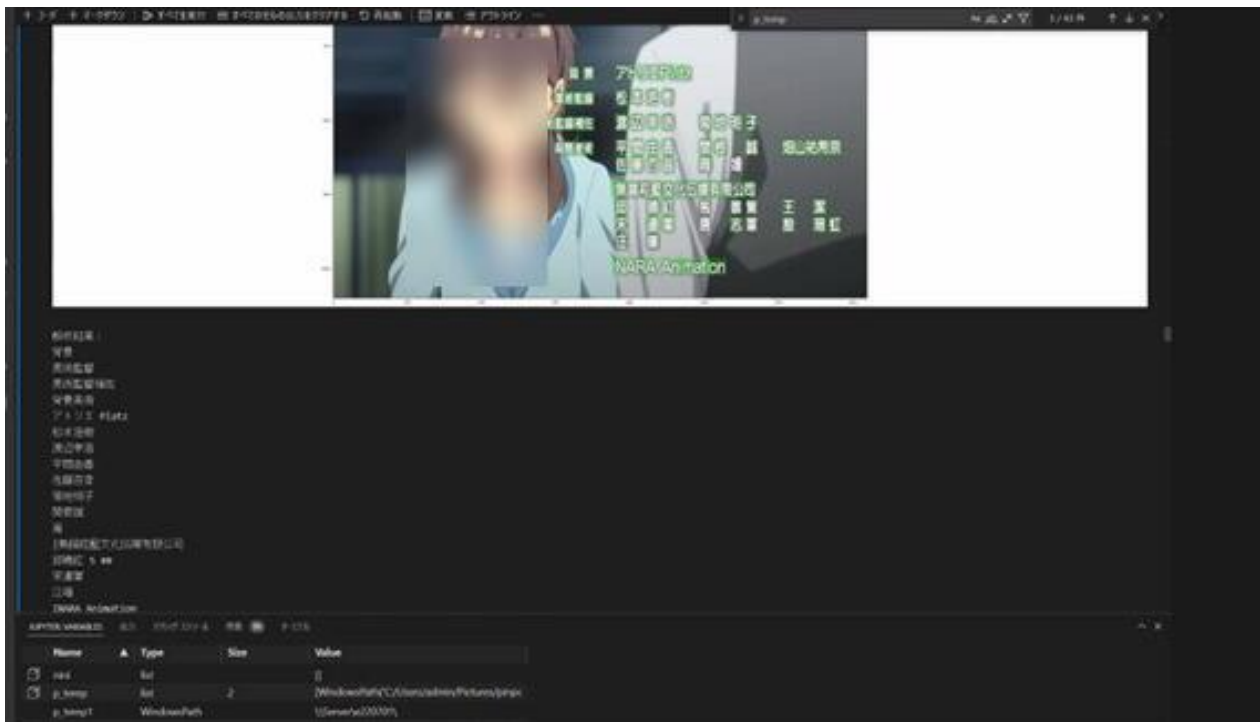


図 3-15 抽出された文字列

第3章 実施内容

アニメ作品の1クール分13本について処理を行って時間を計測すると、OCRを含めた処理時間は50分ほどとなった。これはシングルスレッドで処理した時間で、複数を同時に処理するマルチスレッドで行えば、時間は数分の1に短縮できる。

システムの要となるOCRの性能が日進月歩で進化しているため、誤字率が減ってきている。システムに再投入する手間が省けて全体の処理時間も縮まっている。ただし、OCRの高度な機能を利用しようとするれば、その分の費用がかかるためコストと性能との見合いになる。現時点で出力された文字でも十分な精度を得られているのは確かで、これらをコピー&ペーストしていくだけでも、簡易的なリストなら作成できる状況にある。

キャプチャ画像から、文字が出ている部分を選び分け、同じ文字が出ている画面を削除すれば、OCRへと投入する枚数を減らしコストや処理時間を削減できる。

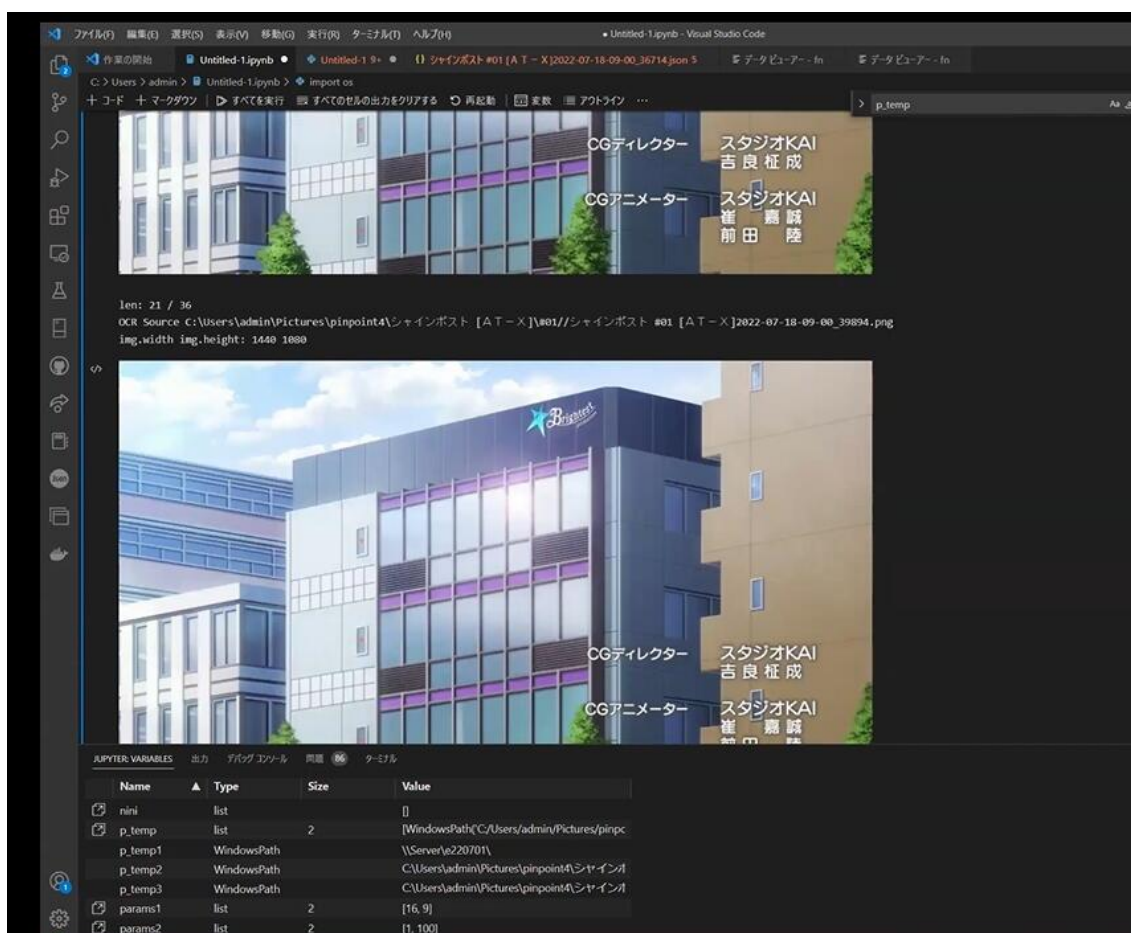


図 3-16 重複した画面の一方を手動で削除し処理枚数を減らす

人間が目で見えて重複画像を選び分けるため、見誤って重複しておらずOCRが必要な画像を削除してしまう可能性もある。コスト負担もそれほど多くはないため、全ての画面をOCR処理する方が結

第3章 実施内容

果的にミスが減らし、スピードも高められる。

そのスピードについては、作業を行う人数を増やす人海戦術を取っての生産性向上も可能だが、同種のシステムがない現状では、そうした機能向上を図るよりも、多くの人々が映像の画面をキャプチャして文字データを抽出し、データベースを作る作業の一助になるツールを早急にまとめるべきだろう。人間が監修に当たる前提で、高い精度でテキスト化を行えるようにしておく方が先決と言える。

こうした作業を、日々のタスクの中にどのように組み込んでいくかを考える必要もある。メディア芸術データベースへの即応を考慮するなら、毎日の放送直後にテキスト化すべきなのかもしれないが、シーズンが終了してから1作品についてまとめて処理する方が、学習データを利用して効率的との考え方もある。

1日当たりで7本ほどの放送があると考えそれらを毎日処理するとなると、作業時間は1時間程度となる。週で5時間なら、時間給5,000円としても25,000円程度に収まる。1週間のうちの1日を1週間分の処理を行う作業日とし、ほかの日を別の作業をするようにしても良い。この作業量を人月で計算すると4~5人月となる。1人月を10万円とすれば月間50万円ほど、1年で600万円ほどとなる。全ての処理を人が行った場合、労働時間数の積算だけで1,000万円超程度かかることを考えると、相当に削減できる。

3.3.2 将来への展望、課題

膨大なクレジットからデータセットを作りモデル化を手掛ける人材は、現実にはなかなかおらず、既に取り組んでいるチームでより効率良く進めることが重要となってくる。ある程度モデルが作られれば、自動判別されるようになると考えられる。

現在は、文字が出る部分の特定をリアルタイムで行えるようになってきている。この技術を使い、画面をキャプチャしながらリアルタイムでOCR処理からのテキスト化を行えるようにする準備を進めた。実現すれば、ネット配信の作品などを取り込みながら、クレジットデータをダイレクトにテキストとして抽出できるようになる。テレビ番組でも、放送中にリアルタイムでOCR処理からのテキスト化処理を行えるため、後の作業時間の短縮につながる。

これまでのAIのモデルでは、画面上に映し出されるイラストにはアプローチしていなかった。イラストを察知するデータセットを搭載していなかったからだが、最近になって、対応できるAIが登場してきた。イラストのようなものでもモデルを作って学習させられるため、学習データを構築していけば、より広範囲で映像から様々な情報を読み取れるようになる。

第3章 実施内容



図 3-17 イラスト枠の指定

将来的には、画面に出たものを全て認識してラベル付けできるくらいに発達するだろう。現状では、絵で描かれたものを自然物の何なのかを判断しても、それが絵で描かれたものだと正確に認識できる訳ではない。画像の読み込み技術を進化させれば、アニメの中に描かれる小道具や建物やプロップなどを実物と関連付けられるよう、情報を分解できるようになる。人間が目で見れば容易に判断できることでも、AIが行うとなると課題が多いが、こうした技術がキャラクターに適用されれば、誰か特定のキャラクターが出ているシーンを検索できるようになり、記録を取る上で大いに役立つ。

アニメに特化したデータセットの構築も急務となっている。NDLが書籍や雑誌のスキャンで使用したもののように、文字部分の読み取りについて鍛え上げられたOCRはあるが、用途が偏っておりオープンな利用を前提にしたものでもない。全ての文字を読み取れるモデルは容易には手に入れない。これを活用するなら、システムに関わる機能を分解し、クラウドのサービスを利用する部分は最小限にして、文字判定に利用できるような機能だけを取り出して提供してもらえようとした方が良いでしょう。

教師データを構築できたとしても、その汎用化に向けては一定期間、例えば2年分のデータを学習用として共有する仕組み作りも必要となる。アニメの制作者や製作委員会など権利者からは、映像使用に疑念を持たれる可能性があるため、何らかの制度的な道筋を付けておくべきだろう。「(仮称)アニメ作品情報連絡調整会」のような場を通し、必要な権利処理の手続きを行える仕組みや土壌の醸成をする必要があるだろう。

第3章 実施内容

3.4 周知活動

本事業では、アニメ関係のイベントやアニメ関係者が視聴しているウェブチャンネル等を通してアニメデータベースが持つ意義と、そのためのデータ収集の大変さを知ってもらう試みを行った。

3.4.1 「マチ★アソビ」(令和4年(2022年)10月15日)

10月15日の16:30～17:30に、徳島市で開催の「マチ★アソビ」でトークイベント「あなたの知らないアニメデータベースの世界」を実施した。出席者はJAniCA事務局長の大坪英之で、本事業ではオブザーバーを務めている。



図 3-18 「マチ★アソビ」登壇

トークイベントでは、アニメを「動かす面の集合体」と定義し、例として手描き、CG、切り絵、クレイなどを挙げつつ、それらを含んだ放送リストの一部を見せて作品数の多さを示した。

第3章 実施内容

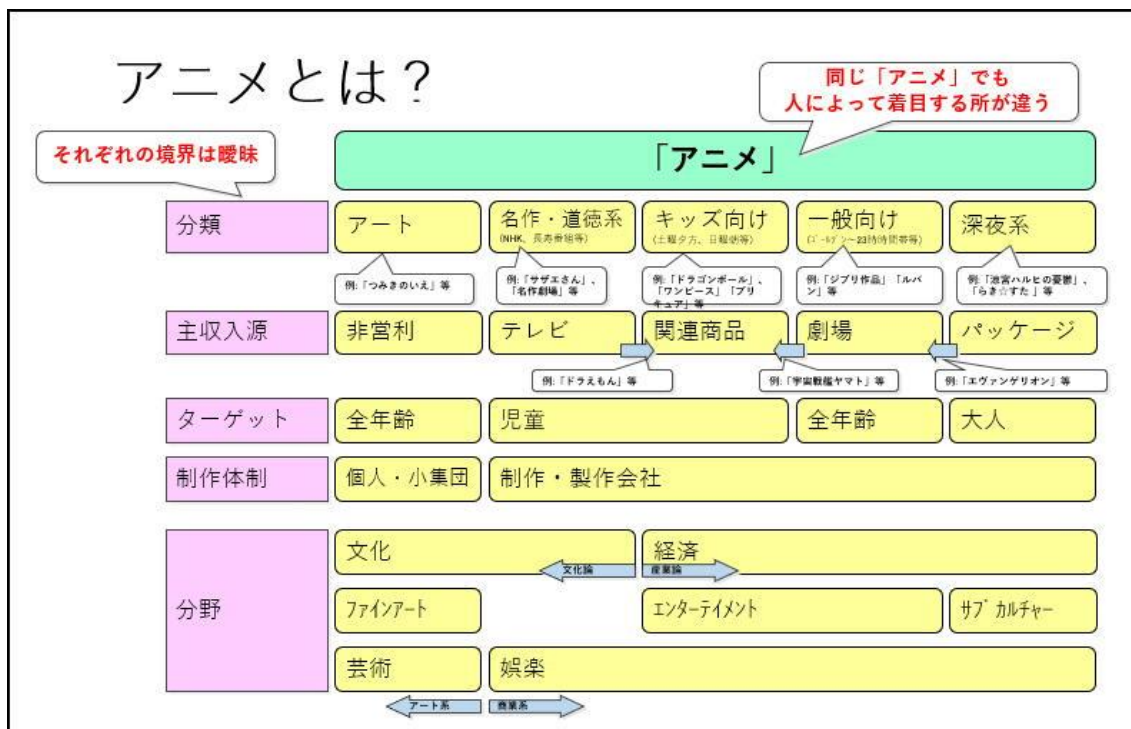


図 3-19 アニメの範囲

ヨミ	作品タイトル	放送局	放送種	放送開始日	時刻	備考	制作会社	種	監督	シリーズ
デュエル	デュエル・マスターズWIN	地上	FX	09/04(日)	09:50~		アニメ・コース・MS		石塚 豪	加藤勝一 飯 典久
子キップ	子キップダンス(第2期)	地上	Eテレ	09/27(水)	08:45~		ファンワークス		ラレコ	田辺 茂徳、ラレコ
うものし	うものし(第2期)	地上	FGVY MX	09/20(金)	22:30~		ライダフィルム		山本貴博	横田家子 山内 遼
ぼくのひ	僕のヒーローアカデミア(第6期)	地上	日本テレビ	10/01(土)	17:30~		ボンス		長崎健司	黒田洋介 馬場高志
のうみん	農林園遊のスキルばっか上げてたら神域が強く	地上	FGVY MX	10/01(土)	22:00~		studio 8+CAF		なみはま	横田家子 玉岡正美
あまく	あまく(第2期)	地上	FGVY MX	10/01(土)	22:30~		MAHO FILM		羽原久美子	橋本 祐内らら
すばい	SPYxFAMILY(第2期)	地上	FX	10/01(土)	23:00~		STUDIO CUBO		斎藤 一浩	橋田知規
こころ	こころ(第2期)	地上	FGVY MX	10/01(土)	23:50~		BN Pictures		冨田 洋二	冨田 洋二 太田 隆雄 竹内 浩二
ベルセル	ベルセルク 黄金時代篇 MEMORIAL EDITION	地上	FGVY MX	10/01(土)	24:50~		STUDIO 4℃		佐野 浩二	佐野 浩二
ボブデビ	ボブデビビック TVアニメーション作品第二シリーズ	地上	FGVY MX	10/01(土)	25:30~				清水 純	清水 純 ×
ぶにるん	ぶにるん	地上	FX	10/02(日)	09:45~				遠山 裕隆	市川十徳 一石 小百合
きどろ	機動戦士ガンダム 水星の魔女	地上	FBS	10/02(日)	19:00~		サンライズ		小林 寛	大河内 一博
ムサシ	ムサシ(第2期)	地上	FGVY MX	10/02(日)	19:00~					
ユウシャ	勇者パーティを道連れしたビーストタイム	地上	FGVY MX	10/02(日)	22:00~		EMスクエアード		黒川 敦	あおしま 山本 園千
アイドリ	アイドルリジェンセン Third 超次元	地上	FGVY MX	10/02(日)	22:30~		TROYCA		朝所 誠人	隈根 アユ 深川 可純
いじら	伊集院 ちんぽんは遊びたい。(おぶる)	地上	FGVY MX	10/02(日)	24:00~		AF-Xが1日早い		ENGU	三浦 知也 あおしま 藤原 亨
ハニム	ハニム(第2期)	地上	FGVY MX	10/02(日)	25:00~					
かみは	かみは(第2期)	地上	日本テレビ	10/03(月)	10:25~					
まのき	まのき(第2期)	地上	フジ	10/03(月)	11:45~					
しんま	しんま(第2期)	地上	FGVY MX	10/03(月)	22:30~					
ゴール	ゴールデンカムイ(第4期)	地上	FGVY MX	10/03(月)	23:00~					
めいた	めいた(第2期)	地上	FGVY MX	10/03(月)	25:05~					
ライビ	ライビ(第2期)	地上	チバテレビ	10/03(月)	26:00~					
しんが	しんが(第2期)	地上	FGVY MX	10/04(火)	23:00~					
ヤマス	ヤマス(第2期)	地上	FGVY MX	10/04(火)	24:30~					
おぼけ	おぼけ(第2期)	地上	FX	10/05(水)	07:05~					
ボス・	ボス・パイビー シーズン2	地上	Eテレ	10/05(水)	19:00~					
ヒュー	ヒューマンパワーズ 不死軍団不浄軍団	地上	FGVY MX	10/05(水)	22:00~					
パスロ	PASSLOCK THE ANIMATION	地上	FGVY MX	10/05(水)	22:30~					
でんせい	でんせい(第2期)	地上	FGVY MX	10/05(水)	23:00~					
かけの	かけの(第2期)	地上	FGVY MX	10/05(水)	23:50~					
ふとく	ふとく(第2期)	地上	AT-X	10/05(水)	23:50~					
どろー	どろー(第2期)	地上	FX	10/05(水)	24:00~					
モブ	モブサイコ100 第2期	地上	FGVY MX	10/05(水)	24:00~					
ふとく	ふとく(第2期)	地上	FGVY MX	10/05(水)	25:35~					
マブラ	マブラヴ オルタネイティヴ 第二期	地上	フジテレビ	10/05(水)	24:55~					
シルバ	シルバニアファミリー フレアのハッピーダイア	地上	FGVY MX	10/06(木)	19:22~					
わか	わか(第2期)	地上	FGVY MX	10/06(木)	23:50~					
あき	あき(第2期)	地上	FGVY MX	10/06(木)	24:00~					

図 3-20 アニメのリスト例

ここから、録画やDVD・BDといった所有の形式、テレビ、劇場、パッケージといった公開形式、娯楽としての映像という表現形式をアニメという概念が含んでいる点を指摘。一方で興味の対象とし

第3章 実施内容

て作品があり、その着目点として物語や世界観、キャラクター、キャストといった項目がある状況を示して、個々のファンがアニメに対して抱いているイメージにも差異がある実態を示した。

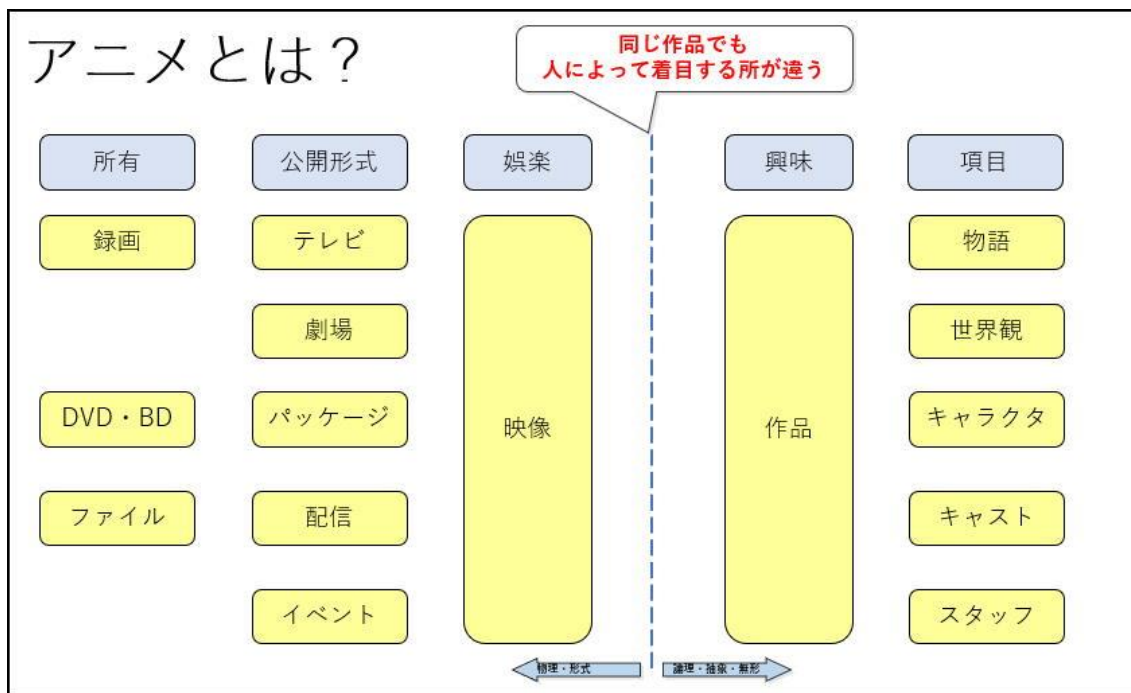
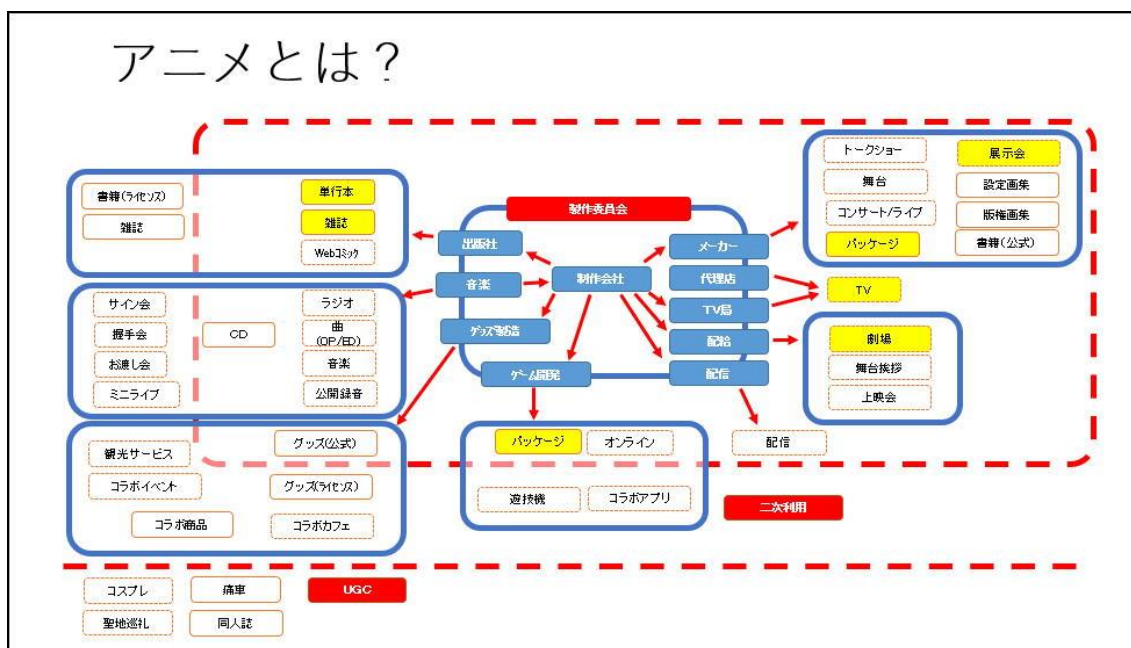


図 3-21 「アニメ」という概念の多相性

また、アニメに携わる企業・組織も、アニメを直接作る制作会社をはじめとしてテレビ局や配給会社、グッズメーカー、音楽会社、出版社などが製作委員会構成会社として存在し、更にその周囲に各種イベントや施設などがあって広く文化であり、産業を構成している実状も示した。



第3章 実施内容

図 3-22 「アニメ」が関わる範囲

これらを集計して、テレビで年間に 3,000 件、映画で 200 件、パッケージで 1,500 件程度の「アニメ」が送り出されている現状、ほかに配信によってテレビや劇場、近年では配信オリジナルの作品もあって、総量を把握できないほどの多くの「アニメ」が送り出されている実態を指摘して、それらを記録する作業の大変さを示した。

続いて、記録に当たって構築した録画環境について説明。「foltia ANIME LOCKER」を導入してチューナーを追加（現在：地上 8ch+衛星 8ch）し、改編時期に録画台帳を作成して録画対象番組を把握しつつ、毎朝に特番などの確認も行なって漏れを防いでいる現状、さらに、実際の録画作業として、毎朝、正常録画の確認を行い緊急特番等の補完も実施していることを伝えた。

こうして録画した番組から、どのような項目を採録しているかについても説明した。テレビ 1 話当たりおおむね 200 人くらいがクレジットに表記されること、本編についてはシリーズ共通のスタッフ（職名、人名、会社名）、各話スタッフ（職名、人名、会社名）、キャスト（人名、役名）、挿入歌（曲名、歌手名、作曲者名、編曲者名、作詞者名）などを記録する必要があること、オープニングやエンディングについてもスタッフ（職名、人名、会社名）、楽曲（曲名、歌手名、作曲者名、編曲者名、作詞者名、レーベル等）を記録していく必要があることを指摘した。

スタッフ・キャスト

テレビ 1 話あたり 200 人くらいが
クレジットに表記される。

- ・ 本編
 - ・ シリーズ共通のスタッフ {職名、人名、会社名}
 - ・ 各話スタッフ {職名、人名、会社名}
 - ・ キャスト {人名、役名}
 - ・ 挿入歌 {曲名、歌手名、作曲者名、編曲者名、作詞者名}
- ・ OP/ED
 - ・ スタッフ {職名、人名、会社名}
 - ・ 楽曲 {曲名、歌手名、作曲者名、編曲者名、
作詞者名、レーベル等}

図 3-23 アニメ作品から記録する項目

第3章 実施内容

トークイベントでは、こうした項目を、膨大な作品数について実施する大変さを伝え、人力で書き写す作業の困難さも示唆した。全録サーバから自 OCR を経てテキスト化する作業の有用性を訴えて、本事業の活動への理解を促した。あわせて、稼働を始めているメディア芸術データベース（以下、MADB）の存在を告知し、アニメデータベースが持つ有用性を知ってもらった。

3.4.2 JAniCA 定例配信（令和4年（2022年）10月29日）

10月29日には、JAniCA がオンラインで開催した「第13回 JAniCA チャンネル」で、主に会員向けにアニメのデータベースとは何かを説明した。JAniCA はアニメーターやアニメ演出家らの地位向上と技術継承を目的に設立された団体で、令和4年（2022年）10月23日時点で正会員363人、無料正会員932人、業界会員62人、準会員430人、賛助個人25人、三条法人1法人等が所属している。個人／法人の応援団も329人を得ている。

入会に当たって、携わった作品名を申請してもらっており、クリエイター個人にそうした作品名や担当話数が附属する形になっている。ここに、年間3,000件にも及ぶテレビアニメや200件に及ぶ映画作品などが新規に加わっていくことになる。量だけでも膨大な上に、クレジットが掲出される OP（オープニング）や ED（エンディング）の位置も話数によって変化するため、採録を人力だけに頼っていては難しいことを指摘。テキスト化に当たっての OCR の使用、掲出位置を判断する AI の活用に有用性があることを示唆した。

そもそも、アニメのデータを蓄積することにどのようなメリットがあるのかも説明した。JAniCA 会員にとっては、業歴が正確に把握・表現されるようになる。社会にとっては、アニメに関する範囲が明確化され、現物収集についてのめど（費用や人員等）が立てられるようになること、収集のための制度設計ができるようになること、定量的なデータを用いて分析を行えるようになること、データモデルについて研究できるようになることがある。さらに、クリエイターやスタジオの着実なキャリアマップが可視化できるようになること、メディア展開の裏側を理解してビジネス展開に役立てられること、データを活用して1995年頃のインターネット化以前の作品について再評価を可能にすること等のメリットがある。

データ蓄積のメリット

- JAniCA
 - 会員の業歴が正確に表現される
- 社会
 - 第1段階・「全て」範囲が明確化される
 - ・現物収集についての目処（費用や人員等）が立てられるようになる
 - ・収集のための制度設計ができるようになる
 - ・定量的なデータを用いて分析を行うことができる
 - ・データモデルについて研究することができる
 - 第2段階・クリエイターやスタジオの着実なキャリアマップが可視化できる
 - ・表出された作品を通じての各種分析ができる
 - ・埋もれた作品や下支えするスタッフ「再評価」につながる
 - 第3段階・メディア展開の裏側がわかればビジネス展開に役立つ
 - ・隠された（抹消された）事項について気付くことができる
 - ・裏面調査ができる
 - 第4段階・データが存在することでInternet化以前(1995年以前)の作品の再評価
 - ・他分野（隣接）との比較検証

図 3-24 データ蓄積のメリット

こうした理由から、アニメクレジットのデータ化が高い意義を持つことを知ってもらい、MADBで採録からデータベース化が行われていること、及び本事業でもデータベース構築に取り組んでいる状況を説明して、活動の周知を図った。

3.4.3 Japanese Visual Media Graph (2023年1月27日、28日)

2023年1月27日、28日にドイツ・シュトゥットガルトにある Stuttgart Media University で開催された、日本におけるアニメ関連のデータベース構築に向けた「Japanese Visual Media Graph」プロジェクト (<https://jvmg.iuk.hdm-stuttgart.de/japanese-visual-media-graph/>) のワークショップに参加し、28日13時に「Further Data sources」と題して講演を行った。

第3章 実施内容



図 3-25 「Japanese Visual Media Graph」プロジェクト (<https://jvmg.iuk.hdm-stuttgart.de/japanese-visual-media-graph/>) サイトトップ




図 3-26 JVMG でのワークショップにおける講演の様様

第3章 実施内容

講演では、現状整理として MADB が 2001 年の文化芸術振興基本法の公布を経て事業として構築されてきた歴史を説明し、次いで日本における「アニメ作品」の定義を説明。「マチ★アソビ」や JAniCA チャンネルでも示した、日本におけるアニメの多彩さと受容の深さを紹介した。

Accumulation and Development of Information on Animation Works in Japan

History of the Media Arts Database



December 7, 2001.
“Basic Act for the Promotion of Culture and the Arts” was promulgated.

From 2010 to 2014
“Project to Improve Networking and Archival Infrastructure of Media Arts”
<Mori Building>

From 2015 to 2016
“Project for the Promotion of Cooperation of Media Arts” <Kyoto Seika>

From 2017 to the present
“Project for the Promotion of Cooperation of Media Arts” <DNP>

<https://mediaarts-db.bunka.go.jp/>

図 3-27 JVMG で紹介した MADB と連携事業の歴史（英語版）

この後、アニメ作品のクレジット・キャスト情報を採録する活動として、MADB がどのような取組を行っているかを説明。「リスト DB」などを例に挙げ、信頼できる情報ソースとは何かを指摘し、総体としての MADB が持つ信頼性も説明した。加えて、本事業でのテレビ作品の採録事例を紹介。どのような作品を対象に、どのような手法でクレジットデータを採録しているかを説明した。アニメ作品に付随するデータにどのような種類のものがあるかも示して、採録が必要なデータが多岐にわたる状況を知ってもらった。

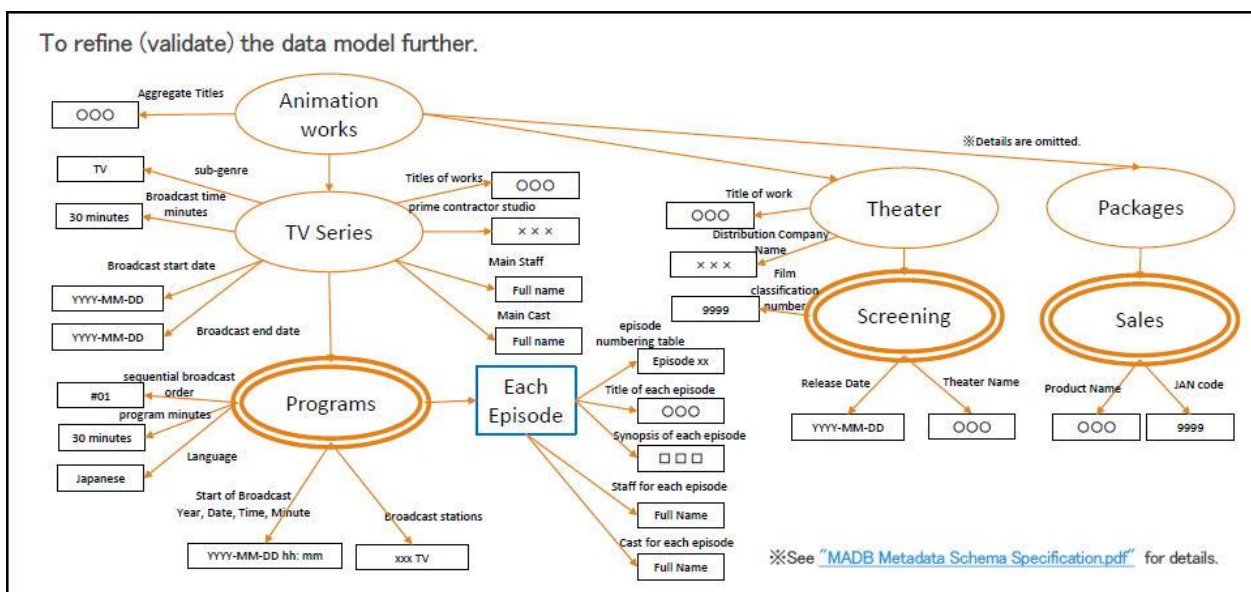


図 3-28 アニメ作品に付随するデータの種類（英訳版）

このほか、採録されたアニメ作品情報が、クリエイターのキャリアマップ作成や、文化・産業としてのアニメの地位確立に役立つこと、様々なデータベースの集合体となるジャパンサーチとの接続を通してほかの分野からアニメ情報への接続、及びアニメ情報からほかのデータへの接続といった相互運用が可能となり、それが相対的にアニメの存在感の向上に繋 [つな] がっていきと説明した。

今回の JVMG に参加した印象として、アニメーションのデータ等に向き合うスタンスとして、①現物を集める、②現物をデータ化する、③使いやすい形式に変換して提供する、④データを使って分析・研究する—といった分類ができると分かった。リスト制作委員会の原口正宏氏や全録サーバのスタンスは①と②に含まれ、MADB を含む本連携事業は①、②、③が当てはまる。学術研究機関の研究者は③を目的としており、その他の多くの人は④を目的として見ているようだった。

3.5 採録状況

本事業では、全録サーバを設置してアニメに関連した番組を全て録画してクレジットデータを抽出したり、スタッフリストを採録・研究している原口正宏氏が主宰するリスト制作委員会と情報共有したりする業務を行っている。全録サーバは杉並と千代田に設置。それぞれに録画作業を行っている。

第3章 実施内容

アニメ全録 受信状況								
No	放送局	区分	原口(1階)	原口(2階)	杉並	千代田	大坪	須山
1	地上波局(関東広域圏)	無料	○	○	○	○	○	○
2	BSデジタル局	無料	○	○	○	○	○	○
3	TOKYO MX	無料	○	○	○	○	○	○
4	tvk	無料	○	○	×	×	○	×
5	チバテレ	無料	×	×	×	○	○	×
6	テレビ埼玉	無料	×	×	×	×	×	×
7	AT-X	有料	×	○	×	×	×	○
8	アニマックス	有料	?	?	×	×	○	?
9	WOWOW(3局)	有料	?	?	×	×	○	?

図 3-29 録画体制

3.5.1 全録サーバ(杉並)実績

杉並サーバについては、地上波、BS、CS をカバーして録画している。アニメ専門チャンネルも AT-X のように最新の作品を放送しているチャンネルをチェックして録画している。単発で放送される番組もあるため、個別にチェックしてアニメ作品を録画している。

録画は TS ファイルで行っている。番組名も放送時のものをそのまま採録している。30 分の番組で 1 番組当たり 2GB から 3GB といったファイルサイズとなり、そのままでは HDD を圧迫するため、不要なファイルを都度削除して容量を確保した。今年 1 月 1 日から 12 月 31 日までの重複削除後の杉並サーバの録画ファイル数は 16,200 件余りとなった

杉並サーバの録画リストについては、全てが 1 番組で 1 ファイルになっていけば後々の取扱いが容易になるが、番組によってはリスト上に 2 段で表示されるものもある。

杉並サーバ					
リーマンスクラブ #9	2022/3/27	2:30:00	0:30:00	テレ朝チャンネル1	2022/3/27 2:30:00
リーマンスクラブ #9	2022/3/27	22:30:00	0:30:00	テレ朝チャンネル1	2022/3/27 22:30:00
リーマンスクラブ #9	2022/6/19	22:00:00	0:30:00	テレ朝チャンネル1	2022/6/19 22:00:00
リーマンスクラブ #9	2022/6/30	20:00:00	0:30:00	テレ朝チャンネル1	2022/6/30 20:00:00
リーマンスクラブ #9【スマニメーション】[宇]	2022/3/27	1:30:00	0:30:00	テレビ朝日	2022/3/27 1:30:00
リーマンスクラブ #9-10	2022/6/18	5:00:00	1:00:00	テレ朝チャンネル1	2022/6/18 5:00:00
リーマンスクラブ【CSオリジナルボイスドラマ付DX一挙放送】	2022/4/24	12:00:00	5:30:00	テレ朝チャンネル1	2022/4/24 12:00:00
リコリス・リコイル #02 [宇]	2022/7/11	22:00:00	0:30:00	AT-X	2022/7/11 22:00:00
リコリス・リコイル #02 [宇]	2022/7/13	10:00:00	0:30:00	AT-X	2022/7/13 10:00:00
リコリス・リコイル #02 [宇]	2022/7/15	16:00:00	0:30:00	AT-X	2022/7/15 16:00:00
リコリス・リコイル #03 [宇]	2022/7/18	22:00:00	0:30:00	AT-X	2022/7/18 22:00:00
リコリス・リコイル #03 [宇]	2022/7/20	10:00:00	0:30:00	AT-X	2022/7/20 10:00:00
リコリス・リコイル #03 [宇]	2022/7/22	16:00:00	0:30:00	AT-X	2022/7/22 16:00:00
リコリス・リコイル #2 [The more the merrier]	2022/7/9	23:30:00	0:30:00	TOKYO MX1	2022/7/9 23:30:00
リコリス・リコイル #2 [The more, the merrier]	2022/7/9	23:30:00	0:30:00	BS11イレブン	2022/7/9 23:30:00
リコリス・リコイル #3 [More haste, less speed]	2022/7/16	23:30:00	0:30:00	BS11イレブン	2022/7/16 23:30:00
リコリス・リコイル #3 [More haste, less speed]	2022/7/16	23:30:00	0:30:00	TOKYO MX1	2022/7/16 23:30:00
リコリス・リコイル #4 [Nothing seek, nothing find]	2022/7/23	23:30:00	0:30:00	BS11イレブン	2022/7/23 23:30:00
リコリス・リコイル #4 [Nothing seek, nothing find]	2022/7/23	23:30:00	0:30:00	TOKYO MX1	2022/7/23 23:30:00
リスアニ！LIVE 2022	2022/4/4	22:00:00	3:00:00	エムオン!	2022/4/4 22:00:00

図 3-30 杉並サーバ録画作品リスト

第3章 実施内容

『リコリス・リコイル』のように、放送局によって書き方が違ってくる場合がある。TOKYO MX は英文のサブタイトルが「The more the merrier」と全角で記録されるが、BS11 は「The more, the merrier」と半角で記録される。AT-X についてはサブタイトル自体が入っていない。こうした記載方式の違いについて、データを取りまとめる段階で、手作業によって修正をかけて揃 [そろ] える必要があるため、手間がかかる実態が分かった。

杉並サーバでは、新番組でも再放送でも、基本的に放送されている番組は全て録画した上で必要なデータを残し、不必要なデータを削除する方針を採っている。1 週間に同じ作品を何度も放送するチャンネルの場合は、一度だけ録画する。再放送については録画していない。

録画をするかしないかは、人間が恣意的に判断している。大量に録画しているため、不要なデータを自動的に判断して削除できれば良いが、現状ではそうした機能は実装されていない。このため、不要なデータを削除する作業は、録画する番組を設定して録画していく作業よりも重たいものとなっている。削除作業を簡略化するためのツールを駆使しても、半期で 8,000 件に達する番組全てで削除を行うのは大変な手間となる。

同じ番組名で予約しておけば、地上波、BS、CS といったいろいろなチャンネルで放送されたものをカバーして録画することができる。地上波で放送されても BS や CS で放送予定がない場合でも、タイトルで予約録画をかければ録画ミスを防げる保険となる。

AT-X で放送される番組では、残り数分で独自にミニ枠が付いて地上波とは形態が違う場合がある。こうした場合は残り数分を切らずに録画し続ける必要がある。

AT-X も BS も映像ファイルサイズが大きいものとなるため、30 分番組でもそのまま全て録画する作業は記録領域の理由から本来なら避けたい。以前も BS では、厳密に番組放送時間を見計らい、関係がなくなる残り 2 分を削って容量を確保していた。ただ、杉並サーバでは録画ミスが発生した事態もあり、録画時間を削らないように方針を改めた。

放送の尺が変わったり、ミニ番組が入ったりする事態の事前察知は難しい。後日、容量を確保するために番組の冒頭や結尾を確認しようとして録画ミスが起こったことを発見するよりも、始めから全て録画しておく方がミスを防げると判断した。

AT-X は番組の終わりにエンドカットが入るため、番組と宣伝の境目が分かる。この境目を検出する技術を動作させれば、少しでも録画時間を削って映像ファイルサイズを減らせるだろう。完全に稼働させるには、録画プログラムを作り込む必要があるため今後の課題としたい。

3.5.2 全録サーバ（千代田）実績

千代田サーバも、1 年で 17,100 件以上を録画した。録画対象は杉並サーバと同じだが、新作としてのアニメが中心で、再放送の録画は少ない。代わりに放送前や放送後、シーズン終了後、シーズン途中に放送される特番を追って録画していた。

第3章 実施内容

17127	HD	0109	2230	薔薇王の葬列	1	Wars of the Roses		
17128	HD	0116	2230	薔薇王の葬列	2	Father Is My Light		
17129	HD	0123	2230	薔薇王の葬列	3	I Am the Shadow, the Darkness Behind the Light		
17130	HD	0130	2230	薔薇王の葬列	4	I'm Afraid I'm Going to Lose the Answers I've Gotten and... Everything Else		
17131	HD	0206	2230	薔薇王の葬列	5	If I Don't Grab Your Hand, I Won't Have to Let Go		
17132	HD	0213	2230	薔薇王の葬列	6			
17133	HD	0220	2230	薔薇王の葬列	7	Even if I sin..		
17134	HD	0227	2230	薔薇王の葬列	8	Even alone, it wasn't frightening. Rather, the same (mess) has been kind ever since leaving Father		
17135	HD	0306	2230	薔薇王の葬列	9	I remember when Christmas was approaching. The time when I lost Father		
17136	HD	0313	2230	薔薇王の葬列	10	My name is Richard Plantagenet		
17137	HD	0320	2230	薔薇王の葬列	11	That is my... my punishment..		
17138	HD	0327	2230	薔薇王の葬列	12	Richard is dead, died alone in the woods that day		
17139	HD	0403	2230	薔薇王の葬列		第1クールダイジェスト～出逢いと別離～		
17140	HD	0410	2230	薔薇王の葬列	13	After all, this body is just a prison of the soul..		
17141	HD	0417	2230	薔薇王の葬列	14	Good night. Brother		
17142	HD	0424	2230	薔薇王の葬列	15	If you hide it, you won't feel sadness		
17143	HD	0501	2230	薔薇王の葬列	16	The devil has the devil's way		
17144	HD	0508	2230	薔薇王の葬列	17	Because God has chosen this bloody demon		
17145	HD	0515	2230	薔薇王の葬列	18	Your name		
17146	HD	0522	2230	薔薇王の葬列	19	I will protect Edward. That child is my son		
17147	HD	0529	2230	薔薇王の葬列	20	The light is overhead		
17148	HD	0605	2230	薔薇王の葬列	21	It should still be in time now		
17149	HD	0612	2230	薔薇王の葬列	22	Use your head more, damn brat. Don't die even if you are weak		
17150	HD	0619	2230	薔薇王の葬列	23	Mother Goodbye		
17151	HD	0626	2230	薔薇王の葬列	24	Requiem of the Rose King		

図 3-31 千代田サーバ録画作品リスト

コロナ禍の流行時に番組制作環境が混乱し、特別番組が増えたケースがあった。終息に伴い急に特番に差し替わるようなことは少なくなったが、時期によって感染拡大が見られるため、今後の変動にも注意を向ける必要がある。

千代田サーバでは、NHKの短尺アニメや、番組内で放送されるミニアニメなどを録画した。原口氏から指摘のあった作品を録画する機会が多いが、全録サーバ側でも適宜調べて録画していった。一例では、「スッキリ!」で『がんばれ!長州くん』という番組が放送されると分かり、録画対象番組に追加した。

特撮番組については、『ウルトラマン』や『仮面ライダー』の特番は、メディア芸術という枠組みの中で録画した。音楽番組については、アニメソングを数多く作曲した渡辺宙明氏に関連した番組、声優が出演する音楽番組は録画するように努めた。このため、千代田サーバの方が録画する作品の対象が広がった。杉並サーバとはタイトルは8割ほどが重なっているが、2割ほどは独自のものになっている。

千代田サーバでも、HDDの容量確保と障害への対策を行った。HDDは破損が常に懸念されるため、対策として毎月の冒頭に前月の1日から月末までの映像を別HDDにコピーして、録画リストを関係者と共有した。PC上で番組録画を行うためのソフト「foltia ANIME LOCKER」はデータの復旧が比較的容易だが、完全にデータが破損してしまう場合もある(2019年9月に全損)。このため、おおむね10日置きにデータをコピーして被害の低減に努めた。

録画に当たっては、原口氏の知見に本事業メンバーの知見を出し合ってミスがないように努めた。原口氏は民生機を使い二重三重の録画体制を整えているが、それでも録画ミスが発生するため、全録

第3章 実施内容

サーバ事業にバックアップとしての信頼を置いてくれている様子だ。

3.6 メディア芸術データベースへのデータ提供

原口正宏氏が主宰するリスト制作委員会が作成してきた「リスト DB」から、メディア芸術データベース（以下、MADB）へのデータ投入を検討・着手した。取り扱った作品数は 5,094 件、話数は 59,647 件に及んだ。共通スタッフ数（延べ）は 570,354 件、各話スタッフ数（延べ）は 877,283 件、各話声優数（延べ）は 121,580 件に達した。これらの中身の調査を進めて、MADB へとデータを提出した。

この作業において、作品リストについては、MADB と「リスト DB」のリストの突き合わせを行っていた。

#N/A	ボク	ボール・ポジション	ボクボク	A	1989	12.1
#N/A	ボク	ボンキッキ むかしぼなし	ボンキッキ	A	1989	10.2
C8587	ボク	まじかる ハット	マジカルハット	A	1989	10.18
#N/A	ボク	魔動王【キング】 グランゾート	マドウオウ	A	1989	4.7
C8567	ボク	魔法使いサリー【新】	マジックサリー	A	1989	10.9
C8035	ボク	まんが はじめて面白塾	マンガが面白い塾	A	1989	2.4
C8041	ボク	ミラクル・ジャイアンツ 童夢くん	ミラクルジャイアンツ	A	1989	4.2
C8561	ボク	桃太郎伝説	モモタロウ伝説	A	1989	10.2
#N/A	ボク	YAWARA! a fashionable judo girl!	ヤワラ	A	1989	10.16
C8055	ボク	らんま 1/2*	ランマ1/2	A	1989	4.15
C8595	ボク	らんま 1/2* 熱闘編	ランマ1/2	A	1989	10.20
#N/A	ボク	ルパン三世 バイバイリパティール危機一発	ルパン三世	A	1989	4.1
C8589	ボク	レスラー軍団（銀河編） 聖戦士ロビン1号	レスラー軍団	A	1989	10.19

図 3-32 「リスト DB」と MADB の作品 ID

図の最左に位置する A 列が MADB の作品 ID で、B 列が「リスト DB」の作品 ID となる。同じ作品がうまくつながるか、それともつながらないかをある程度事前に調べておき、実際に投入する際の作業負担を軽減できるようにした。ある程度日付や数字を綺麗 [きれい] に揃えておくことで、膨大な作業量の負担を軽減できた。

各話については、更にデータの一致性を検討する必要があった。一例では、『はりもぐハーリー』という作品について、「56」「57」「58」といった話数が「リスト DB」と MADB で本来は一致しており、容易に紐付くと考えられるが、本当に第 56 話が「さがし物は何ですか?」だったのかという確証はない。放送日が違っていたり、57 話として放送されていたりといったケースが考えられる。ほかには「?」が半角と全角で異なって記録されていて、うまく重ならないケースも想定できた。

こうした長大な作品の場合は、起点を揃えると以下がそれぞれに対応し、しっかりと紐付くと考えられる。今回の 5,000 件については、コレクション ID を揃えアイテム ID を揃えて、合うか合わないかを調べた。

第3章 実施内容

このほか、『オオカミ長者』という番組で、「WORKNW」という役職を示すIDに「CG」が登場してスタッフ名が書かれてしまっていた。『オオカミ長者』にはCGは使われておらず、どのような理由から入ってしまったのかを調査する必要があった。

NUMBI	ED	SUBTITL	Year	Date	DCI	WOR	WORKNM	STU	NAME
0000	A	オオカミ長者【11回A(再)】	017	924	DC0	015	文芸	999	久光正
0000	A	オオカミ長者【11回A(再)】	017	924	DC1	11F	演出	999	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC1	11L	絵コンテ	999	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC3	31G	CG	-99	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC3	31E	作画コーディネ	-99	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC3	31U	動画	999	瀬田みのり
0000	A	オオカミ長者【11回A(再)】	017	924	DC3	31E	作画	-99	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC3	31u	キャラクターデ	-99	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC3	31u	【動画】	-99	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC3	31U	OP・END・DI	999	瀬田みのり
0000	A	オオカミ長者【11回A(再)】	017	924	DC3	31G	原画	-99	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC4	41B	美術	999	瀬田龍
0000	A	オオカミ長者【11回A(再)】	017	924	DC4	41E	背景	999	永田明枝
0000	A	オオカミ長者【11回A(再)】	017	924	DC5	511	色彩設定	999	竹村菜月
0000	A	オオカミ長者【11回A(再)】	017	924	DC5	515	仕上	999	福江南菜美
0000	A	オオカミ長者【11回A(再)】	017	924	DC6	616	撮影	999	青映社
0000	A	オオカミ長者【11回A(再)】	017	924	DC6	616	撮影	999	密本麻衣
0000	A	オオカミ長者【11回A(再)】	017	924	DC6	616	撮影	999	加藤あみ

図 3-33 スタッフ情報

スタッフや声優、楽曲のデータの接続は、作品名で紐付けが果たされて後の段階となる。まずは作品と各話のリストについて渡せる分を渡すための作業を行った。

本報告書は、文化庁の委託業務として、大日本印刷株式会社が実施した令和4年度「メディア芸術連携基盤等整備推進事業 分野別強化事業」の成果をとりまとめたものであり、第三者による著作物が含まれています。
転載複製等に関する問い合わせは、文化庁にご連絡ください。