

情報の収集方法、情報の提供方法、 情報提供を持続する方法の課題解決 に向けた取り組み。



社会情報学科 木村ゼミ

佐藤 栄作 吉岡 晃汰 齊藤 竜成 奥山 和樹

制作:齊藤竜成



三姉妹という
キャラ付け

肉声からの卒業
提供の持続



制作:武市和真



制作:奥山和樹



制作:伊賀世杜奈



Taxi 奥山和樹

商大生の伝統を引き継ぐタクシー
相乗りアプリ

Cafemap 新井田瑠璃

商大生がおすすしあうカフェマップ

Timetable 佐藤栄作

商大生の生活スタイルに合わせた
乗り換えカウントダウン

Weather Report 藤原里紗

地震版どうなっているかな天気予報

Q&A 佐久間千佳

暖かい教室から、冷たい先生まで
聞けるQ&A

Skill 内山莉々菜

小遣い稼ぎができる Skill sharing

情報の提供



取得



表形式データ

担当
佐藤栄作

有価証券報告書

▶機械判読しにくい

情報の収集

情報発信には、「収集」「提供」「持続」の3段階が存在する

- 社会情報学科である木村ゼミは、**情報を扱う取り組み**を行っている
- 情報発信の過程には、「**収集**」「**提供**」「**持続**」という3つの段階が存在する



情報の「収集」「提供」「持続」という3つのテーマ

それぞれのテーマに関する問い

→ それに答える形で、異なる3つの取り組みについての発表を行う

収集: データの収集にはどのような課題があるのか

提供: どのように安定した情報提供の基盤を構築するのか

持続: どのように動画による情報提供を持続させるのか

情報の「収集」「提供」「持続」という3つのテーマ

第一に、「収集」をテーマとする問いに答える

収集: データの収集にはどのような課題があるのか

提供: どのように安定した情報提供の基盤を構築するのか

持続: どのように動画による情報提供を持続させるのか

収集: データの収集にはどのような課題があるのか

表形式データの収集

～有価証券報告書の表を対象として～

社会情報学科 3年
佐藤 栄作

本研究の背景とは

オープンデータ化が進み、Web上で多くのデータが公開されている

オープンデータ: 国や地方公共団体が公開している、
誰もが利用可能なデータ。

- ① **オープンライセンス**: 二次利用の条件が緩い
- ② **オープンフォーマット**: 特定のソフトウェアが不要
- ③ **機械判読可能**: 一般に**表形式データ**

➡ **表形式データに注目**

ウェブサイトに含まれる表の特徴とは

- 表には**様々な種類がある**
 - 各セルは、**エンティティ(固有名詞)**、**属性**、**値**に分類される
 - **項目名(1行1列目)**のセルの種類の組み合わせにより、表の種類が変化
- 各セルの種類を判別できなければ、**表の比較ができない**

関係データベース	属性と値のペア	行列の項目が固有名詞																								
(a) <table border="1"><thead><tr><th>Name</th><th>Age</th><th>Sex</th></tr></thead><tbody><tr><td>Alice</td><td>21</td><td>Female</td></tr><tr><td>Bob</td><td>35</td><td>Male</td></tr></tbody></table>	Name	Age	Sex	Alice	21	Female	Bob	35	Male	(c) <table border="1"><thead><tr><th colspan="2">Alice</th></tr></thead><tbody><tr><td>Age</td><td>21</td></tr><tr><td>Height</td><td>5'4"</td></tr></tbody></table>	Alice		Age	21	Height	5'4"	(e) Ranks <table border="1"><thead><tr><th></th><th>2015</th><th>2016</th></tr></thead><tbody><tr><td>Alice</td><td>2</td><td>4</td></tr><tr><td>Bob</td><td>1</td><td>8</td></tr></tbody></table>		2015	2016	Alice	2	4	Bob	1	8
Name	Age	Sex																								
Alice	21	Female																								
Bob	35	Male																								
Alice																										
Age	21																									
Height	5'4"																									
	2015	2016																								
Alice	2	4																								
Bob	1	8																								
(b) Alice <table border="1"><thead><tr><th>Year</th><th>Rank</th><th>Score</th></tr></thead><tbody><tr><td>2015</td><td>2</td><td>82.5</td></tr><tr><td>2016</td><td>4</td><td>76.2</td></tr></tbody></table>	Year	Rank	Score	2015	2	82.5	2016	4	76.2	(d) Bob <table border="1"><tbody><tr><td>Age</td><td>35</td></tr><tr><td>Height</td><td>5'10"</td></tr></tbody></table>	Age	35	Height	5'10"	(f) Bob's scores <table border="1"><thead><tr><th></th><th>Sep.</th><th>Dec.</th></tr></thead><tbody><tr><td>2015</td><td>80.3</td><td>84.1</td></tr><tr><td>2016</td><td>70.4</td><td>76.2</td></tr></tbody></table>		Sep.	Dec.	2015	80.3	84.1	2016	70.4	76.2		
Year	Rank	Score																								
2015	2	82.5																								
2016	4	76.2																								
Age	35																									
Height	5'10"																									
	Sep.	Dec.																								
2015	80.3	84.1																								
2016	70.4	76.2																								

表の理解の難しさとは

KDD 2021 Tutorial: From Tables to Knowledge: Recent Advances in **Table Understanding**

KDD 2021と呼ばれる国際会議のチュートリアルにおいても、**表の理解の難しさ**が示された

表の理解のためには、
「各セルの意味の理解」が必要である

項目名を比較する際の難しさ

- 表記ゆれ：組織名、会社名、企業名 ...
- 曖昧性：(前年度の)合計、(今年度の)合計 ...
- 正規化：100万円、1,000,000円 ...

Parts of a table in dataset annotations

Metadata

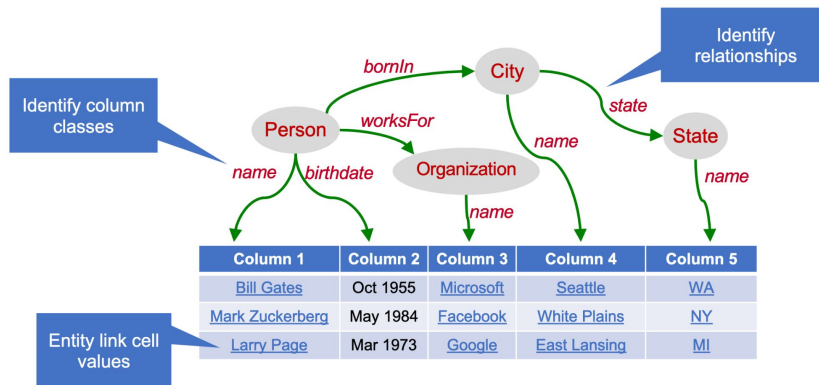
Table 1: Number of acceptances by category of work in conferences ¹						
Year	Conference	Papers		Workshops	Tutorials	Total
		Research	Industry			
2021	KDD	800	500	25	40	1365
2021	WWW	700	400	25	40	1165
2020	KDD	600	300	25	35	960
2020	WWW	500	200	25	35	760

¹The source of this data is Jay's made-up data generation

Chen, Z., Cafarella, M.: Integrating spreadsheet data via accurate and low-effort extraction. In: KDD 2014.

Koci, E., Thiele, M., Romero Moral, O., Lehner, W.: A machine learning approach for layout inference in spreadsheets. In: IC3K 2016.

Subtasks of semantic modeling



表の理解が望まれている場面は？

- **有価証券報告書**: 上場企業等の事業・経理報告書
 - **XBRL**の導入により、表の理解に取り組んでいる
- **XBRL**: XMLベースの言語。**タクソミ**と**インスタンス**を用いて、**各セルに意味付け**をする。
 - **タクソミ**: **要素の種類や属性**(要素名)を設定
 - **インスタンス**: 値に対しての期間、通貨単位などを設定

➡ **タクソミに注目する**

タクソノミがあるとどうなる？

「各セルの意味」を定義し、表の理解への手掛かりとなる

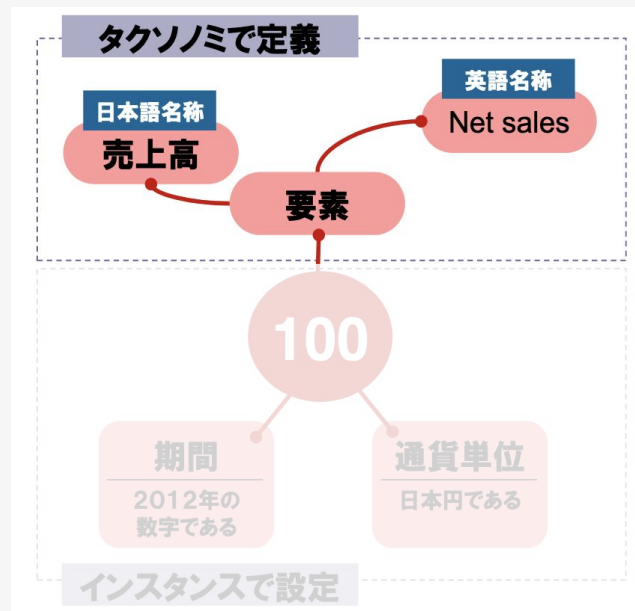
要素の定義

数値データ or 文字列データ

名称の定義

どのようなデータか
ex) 売上高、純利益、企業名

➡ 十分に付与されているのか？



XBRLにおけるタクソノミ及びインスタンスのイメージ

TOPIX100の有価証券報告書に含まれる 表形式データの収集の問題点を明らかにする

収集対象とする有価証券報告書

- **TOPIX100**: TOPIX(東証株価指数)を構成する銘柄のうち、**流動性と時価総額の高い100銘柄**を算出対象として選定した株価指数
- 対象年: **2016(平成28年) ~ 2021(令和3)年**の5年間
- 対象報告書数: **482**

以下の問いに答える

- Q-1** 表形式データの数はいくつであるか
- Q-2** タクソノミはどの程度付与されているか
- Q-3** 年度間や会社間でタクソノミに差異はあるのか

Q-1 表形式データはいくつであるか

表形式データの総数は、「107,034」である

- 総数: **107,034**
- 最大値: 526
- 平均: **221.9**
- 最小値: 103

企業名 \ 開始年	平成 28 年 2016 年	平成 29 年 2017 年	平成 30 年 2018 年	平成 31 年 2019 年	令和元年 2019 年	令和 2 年 2020 年	令和 3 年 2021 年	合計
1 ANA ホールディングス	175	182	179	180	0	178	0	894
2 HOYA	230	212	216	220	0	217	0	1,095
3 住友金属鉱山	204	215	291	255	0	249	0	1,214
4 JXTG ホールディングス	252	250	204	198	0	196	0	1,100
...
95 資生堂	0	231	218	202	0	207	224	1,082
96 野村ホールディングス	346	257	267	264	0	273	0	1,407
97 Zホールディングス	227	208	218	218	0	223	0	1,094
合計	18,556	20,881	21,286	21,407	343	21,850	2,711	107,034

Q-2 タクソミはどの程度付与されているのか

タクソミの付与率は30%以下である

- 1セクションあたり**最高でも約31%**しか付与されていない
- 3つのセクション以外、**ほとんど付与されていない**

セクション	表の数	タグを含む表の数	タグを含む割合
企業の概況	6,358	1,967	30.94 %
事業の状況	7,764	21	0.27 %
設備の状況	3,488	114	3.27 %
提出会社の状況	26,767	6,782	25.34 %
経理の状況	38,866	6,388	16.44 %
提出会社の株式事務の概要	624	0	0.00 %
提出会社の参考情報	599	0	0.00 %
企業情報	5	0	0.00 %
保証会社情報	247	0	0.00 %

Q-3 年度間や会社間でタクソノミに差異はあるのか

年度間でタクソノミに大きく差がある

およそ1.7倍に増加している

※法律の改正によるものと思われる

平成28年	平成29年	平成30年	平成31年/令和元年	令和2年	令和3年
99163	107978	184024	198229	196890	23041

年度毎によるXBRLタグの数

現時点で提出されている報告書数自体が少ない

主に2つの問題点が示された

- タクソミの付与率が**30%以下**である
 - タクソミが付与されていないと、**所望のデータの収集や比較**ができない
- **70%以上**の表を活用するには、**セルタイプの自動識別**が必要
 - 年度や会社間で**増減するタクソミ**に対応する必要がある

情報の「収集」「提供」「持続」という3つのテーマ

「収集」をテーマとする問いに対し、以下の答えが得られた

収集: データの収集にはどのような課題があるのか

表、有価証券報告書では

- ・タクソノミの付与率が**30%以下**である
- ・70%以上の表を比較するためには、**セルタイプの自動識別**が必要

提供: どのように安定した情報提供の基盤を構築するのか

持続: どのように動画による情報提供を持続させるのか

情報の「収集」「提供」「持続」という3つのテーマ

第二に、「提供」をテーマとする問いに答える

収集: データの収集にはどのような課題があるのか

表、有価証券報告書では

- ・タクソノミの付与率が**30%以下**である

- ・70%以上の表を比較するためには、**セルタイプの自動識別**が必要

提供: どのように安定した情報提供の基盤を構築するのか

持続: どのように動画による情報提供を持続させるのか

提供:どのように安定した情報提供の基盤を構築するのか

Flutter、ミニアプリ構造による 保守性の高い情報提供アプリの開発

社会情報学科 3年
吉岡 晃汰

現代における情報提供のあり方とは

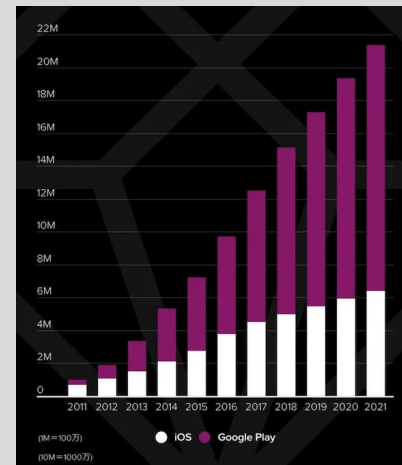
スマホの普及に伴うアプリ市場の隆盛

一人当たり 月間利用平均アプリ数

40個

一人当たり 日間平均アプリ利用時間

4.8時間



↑ アプリの世界累計リリース数 推移

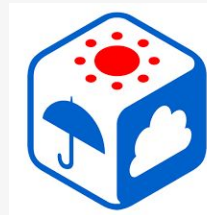
情報提供の手法としてのアプリ



Gunosy



SmartNews



NHK
ニュース
防災



アプリ開発に存在する2つの課題

- 開発の負担が大きい

現在は**様々な角度からの情報提供、双方向的な情報提供**など多機能、機能が増えることに比例して負担も大きくなる

- 環境の違い

iOS、AndroidやWebなどの様々なプラットフォームがあり、それぞれ**使用言語やコードを変化させる必要**がある

OS	開発環境	言語
 Android	Android Studio	java, C#, Kotlin
 iOS	Xcode	Objective-C, Swift

Android、iOS両方に対応させようとする**と二種類の言語を学び、コードも二つ必要!**

解決策となる技術「Flutter」



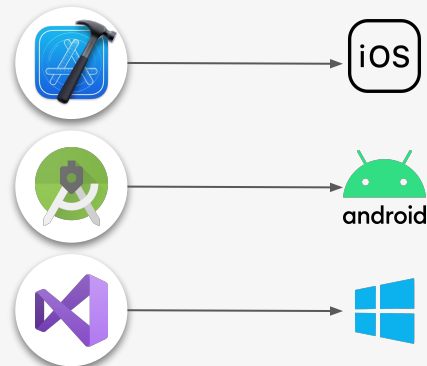
Flutter

Googleによって開発されたフレームワーク
マルチプラットフォーム開発に対応している

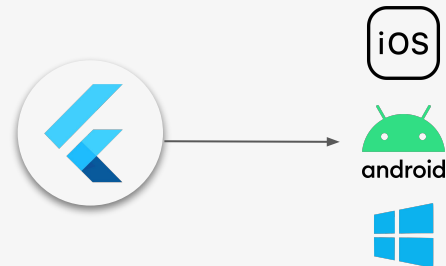


「環境の違い」への対処

<一般的な開発>



<Flutterでの開発>

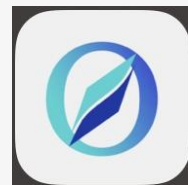


App
└ App

ミニアプリ構造

アプリの中にアプリを作成すること
機能別に開発することができる

「開発の負担」の軽減



Tap!



OUCnavi: Flutterで作られたアプリ



OUCnavi: ミニアプリ構造が用いられたホーム画面

Q&A Box

佐久間千佳

暖かい教室から、
冷たい先生まで聞けるQ&A

Weather Report

藤原里紗

地獄坂どうなっているかな天気予報

Cafemap

新井田瑠璃

商大生がおすすめしあうカフェマップ



Skill Sharing

内山莉々菜

個々の学生のスキルを活かした
お小遣い稼ぎ

Timetable

佐藤栄作

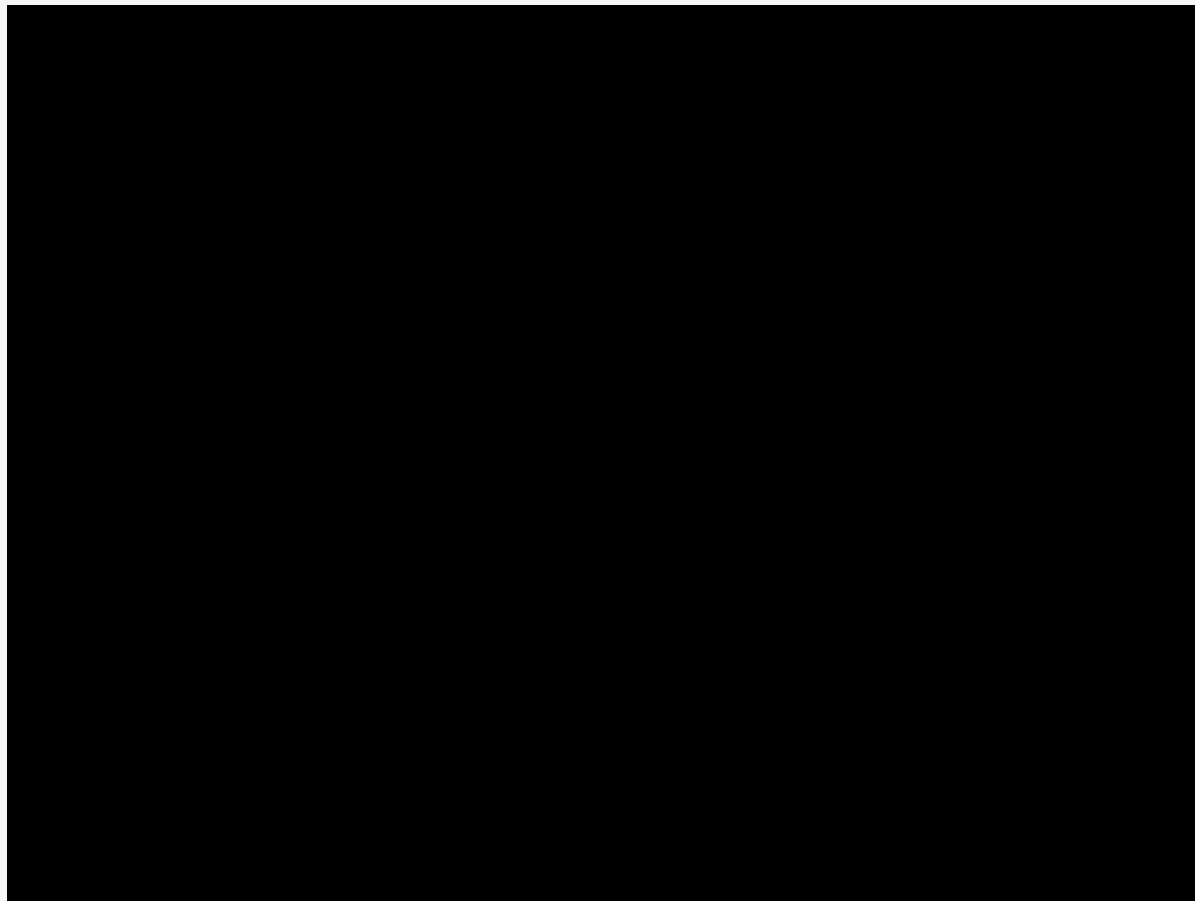
商大生の生活スタイルに合わせた
カウントダウン式時刻表

Taxi

奥山和樹

商大生の伝統を引き継ぐ
タクシー相乗りアプリ

OUCnavi: ミニアプリ構造で開発された機能



どのように安定した情報提供の基盤を構築するのか

Flutter、**ミニアプリ構造**を用いて開発を行うことで
安定した情報提供の基盤を構築できる



Flutter

Google開発のフレームワーク
マルチプラットフォーム開発に対応



「環境の違い」への対処

App
└ **App**

ミニアプリ構造

アプリの中にアプリを作成すること
機能別に開発することができる



「開発の負担」の軽減

情報の「収集」「提供」「持続」という3つのテーマ

「提供」をテーマとする問いに対し、以下の答えが得られた

収集: データの収集にはどのような課題があるのか

表、有価証券報告書では

- ・タクソノミの付与率が**30%以下**である
- ・70%以上の表を比較するためには、**セルタイプの自動識別**が必要

提供: どのように安定した情報提供の基盤を構築するのか

- ・**Flutter**を用いて、多様な環境に対応させる
- ・**ミニアプリ構造**で設計することで、チーム開発を行いやすくし、開発者の負担を軽減する

持続: どのように動画による情報提供を持続させるのか

第三に、「**持続**」をテーマとする問いに答える

収集: データの収集にはどのような課題があるのか

表、有価証券報告書では

- ・タクソノミの付与率が**30%以下**である
- ・70%以上の表を比較するためには、**セルタイプの自動識別**が必要

提供: どのように安定した情報提供の基盤を構築するのか

- ・**Flutter**を用いて、多様な環境に対応させる
- ・**ミニアプリ構造**で設計することで、チーム開発を行いやすくし、開発者の負担を軽減する

持続: どのように動画による情報提供を持続させるのか

維持:どのように動画による情報提供を持続させるのか

合成音声技術を用いた 引き継ぎ問題の解決とその環境の持続

社会情報学科 3年
齊藤 竜成
奥山 和樹

背景

大学生による部活/サークル/ゼミなどの活動において、
頻繁に**引き継ぎ作業**が発生

▶例えば動画による情報発信などがある

課題

個人に依存しすぎた作業が存在し、引き継ぎが困難に

▶メンバーが動画の公開終了を希望する、
情報発信のクオリティが変わってしまうなどすることも

容易な引継ぎに必要な条件とは

● 個人に対する依存の排除

- 誰が作成しても一定のクオリティが担保される

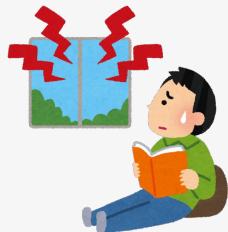


スキル



使いやすい

- 個人の環境に影響されない



場所



機材

個人に依存しやすい作業とは

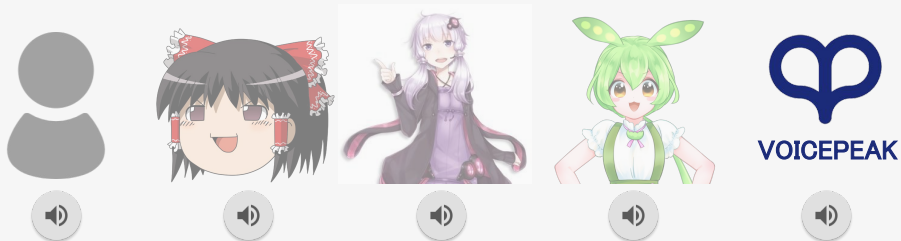
動画制作において個人に強く依存する作業

➡ 個人やその環境の影響によってクオリティが変化するもの

➡ 動画撮影、ナレーション



条件を満たす制作様式とは



	肉声	AquesTalk	VOICEROID	VOICEVOX	VOICEPEAK
音声の精度	○	×	△	△	◎
操作性	○	○	△	△	◎
商用フリー	○	△	△	○	◎
非属人性	×	○	○	○	○
キャラクター性 (identity)	○	○	○	○	×

誰が作成しても
一定のクオリティが担保される
AI音声合成による高性能な音声出力
容易に調整が可能なUIデザイン

個人の環境に影響されない
声質、機材の有無、
声を出せる場所かに影響しない

しかし
identityを1から作る必要がある

目的、何を明らかにするのか、どのような問題を解決するのか

目的

本研究では、動画作成をする際に、キャラクターの Identity を保ちつつ、**引き継ぎ問題**を解決できる方法を提案し、有効であることを明らかにする。

- 個人に依存せず引き継ぎが容易な方法とはどのようなものか

引き継ぎ問題 → VOICEPEAKを利用した動画作成

- キャラクター性がない合成音声を使う場合、どのようにしてアイデンティティを保つのか

Identity を保つ → 3姉妹 twitter による性格、特徴の継続 (Voicepeakだと、identityが十分でない)

キャラクターの引き継ぎ問題を解決する方法とは

■VoicePeakの導入により動画制作効率化して引き継ぎ問題を解決

導入前

制作開始～
動画撮影・ロケ

8日目～
構成・台本制作

14日目～
音声収録

21日目～
動画編集

導入後

制作開始～
動画撮影・ロケ

8日目～
構成・VoicePeak出力

14日目～
動画編集

音声収録とセリフを同時に行え
動画制作の効率化に貢献

■VoicePeakは視聴者を減らさずに引き継ぎ問題を解決

導入前の動画

通常のイントロを下回る動画



【小樽商大】 空から商大を見てみよう！ドローンで空中散歩してみた！

34%



【ミスおたると対談！？】 ミスおたる山川穂乃果さんと仲良くなっちゃおう...

32%

導入後の動画

通常のイントロを上回る動画

0:30での視聴者保持率



【春から商大のキミへ！】 商大発の便利コンテンツを使ってみてくださいませんか...

60%



小樽のスイーツが美味すぎてとまらない！

52%

VoicePeakを導入して動画はどのように変化したか

導入前



導入後



キャラクターのアイデンティティを確立、維持させるには

■声色の調整

キャラクターの
イメージにあった
ナレーションを実現

無感情

NOI



感情

カスタム設定

幸せ

楽しみ

怒り

悲しみ

さらに...

■キャラクターごとに異なるタイプの情報をNSで提供

EMAによる学生向け情報提供

 EMA   @EMA_sainoma · 2022/10/03

本日10/3(月)から10/7(金)まで後期科目履修登録変更期間となっておりますのでキャンパススクエアの確認をお忘れなく！



SAIによる小樽の紹介



最終日の花火も最高でしたね🔥
屋台のごはん食べ足りないです...
#潮まつり #花火 #小樽



情報の「収集」「提供」「持続」という3つのテーマ

「持続」をテーマとする問いに対し、以下の答えが得られた

収集: データの収集にはどのような課題があるのか

表、有価証券報告書では

- ・タクソノミの付与率が**30%以下**である
- ・70%以上の表を比較するためには、**セルタイプの自動識別**が必要

提供: どのように安定した情報提供の基盤を構築するのか

- ・**Flutter**を用いて、多様な環境に対応させる
- ・**ミニアプリ構造**で設計することで、チーム開発を行いやすくし、開発者の負担を軽減する

持続: どのように動画による情報提供を持続させるのか

- ・**VoicePeak**を用いて動画制作を効率化し、引き継ぎ問題を解決した
- ・**キャラクターにあわせたSNS運用**でアイデンティティを確立・維持し、一貫した情報提供を持続化させた

まとめ

- 「収集」「提供」「持続」という3つのテーマに関する問い

→ それに答える形で、異なる3つの取り組みについての発表を行った

収集: データの収集にはどのような課題があるのか

表、有価証券報告書では

- ・タクソミの付与率が**30%以下**である

- ・70%以上の表を比較するためには、**セルタイプの自動識別**が必要

提供: どのように安定した情報提供の基盤を構築するのか

- ・**Flutter**を用いて、多様な環境に対応させる

- ・**ミニアプリ構造**で設計することで、チーム開発を行いやすくし、開発者の負担を軽減する

持続: どのように動画による情報提供を持続させるのか

- ・**VoicePeak**を用いて動画制作を効率化し、引き継ぎ問題を解決した

- ・**キャラクターにあわせたSNS運用**でアイデンティティを確立・維持し、一貫した情報提供を持続化させた