

オルタナティブデータ活用の 黎明期を超えるために： AI経済センシングセンター構想

和泉 潔

東京大学大学院 工学系研究科

世界最初のウェブサイト info.cern.ch

World Wide Web

The WorldWideWeb (W3) is a wide-area [hypermedia](#) information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an [executive summary](#) of the project, [Mailing lists](#), [Policy](#)

[What's out there?](#)

Pointers to the world's online information, [subjects](#), [W3 servers](#), etc.

[Help](#)

on the browser you are using

[Software Products](#)

A list of W3 project components and their current state. (e.g. [Line Mode](#), [X11 Viola](#), [NeXTStep](#), [Servers](#), [Tools](#), [Mail robot](#), [Library](#))

[Technical](#)

Details of protocols, formats, program internals etc

[Bibliography](#)

Paper documentation on W3 and references.

[People](#)

A list of some people involved in the project.

[History](#)

A summary of the history of the project.

[How can I help?](#)

If you would like to support the web..

[Getting code](#)

Getting the code by [anonymous FTP](#), etc.

World Wide Web プロジェクト (1992)

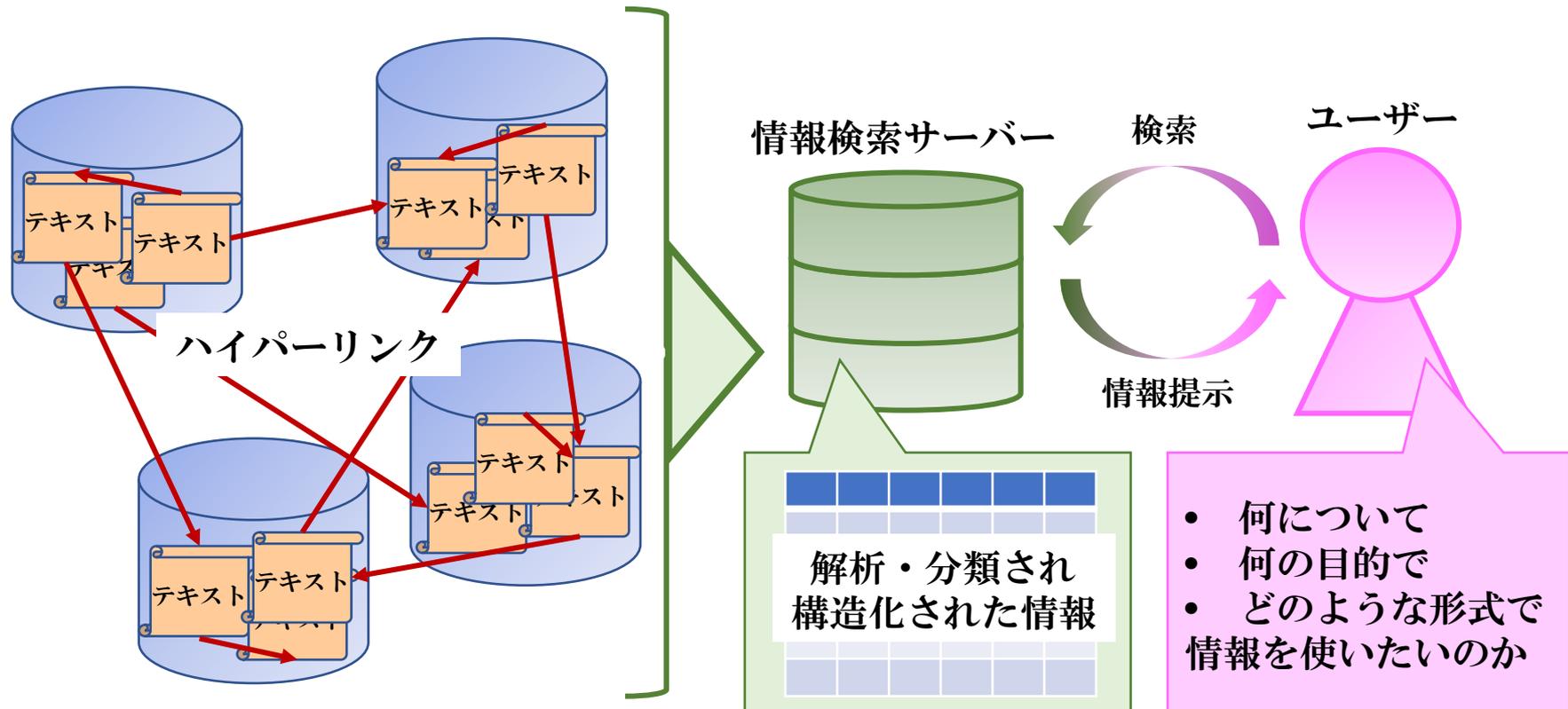
Webサービスはどうして普及したのか?

コンテンツ技術の側面から見た要因

ハイパーテキストによる
テキスト間の相互関連付け

+

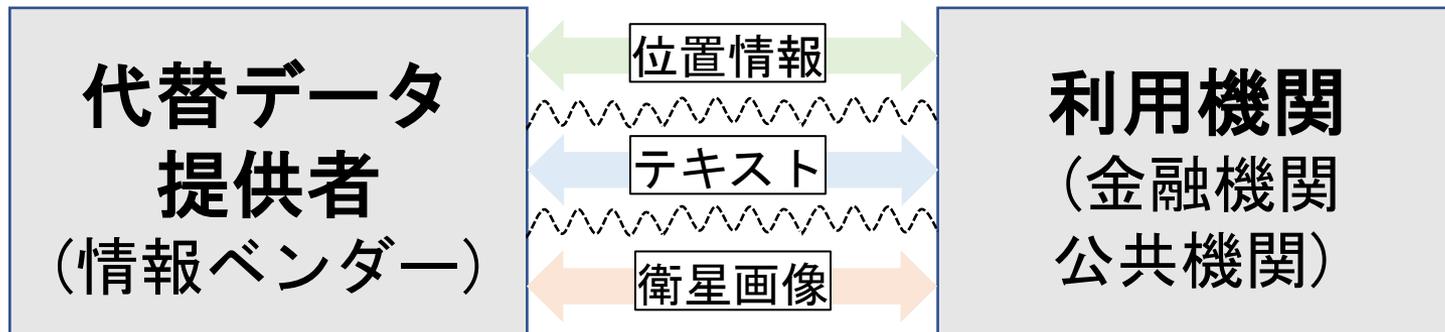
ロボット型検索エンジン
による情報検索の高度化



代替データ分析の現状

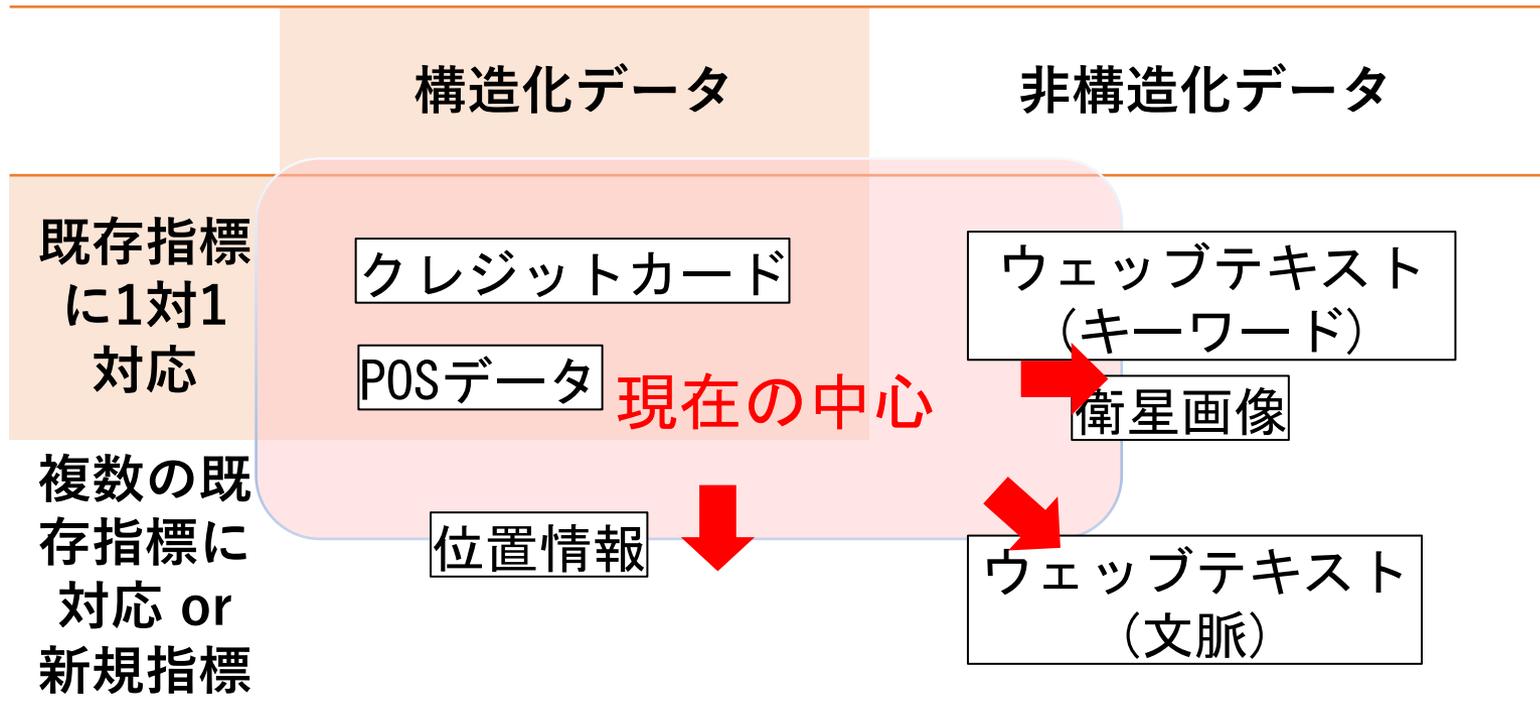
まぜるな危険

各代替データごとの一発芸的な分析



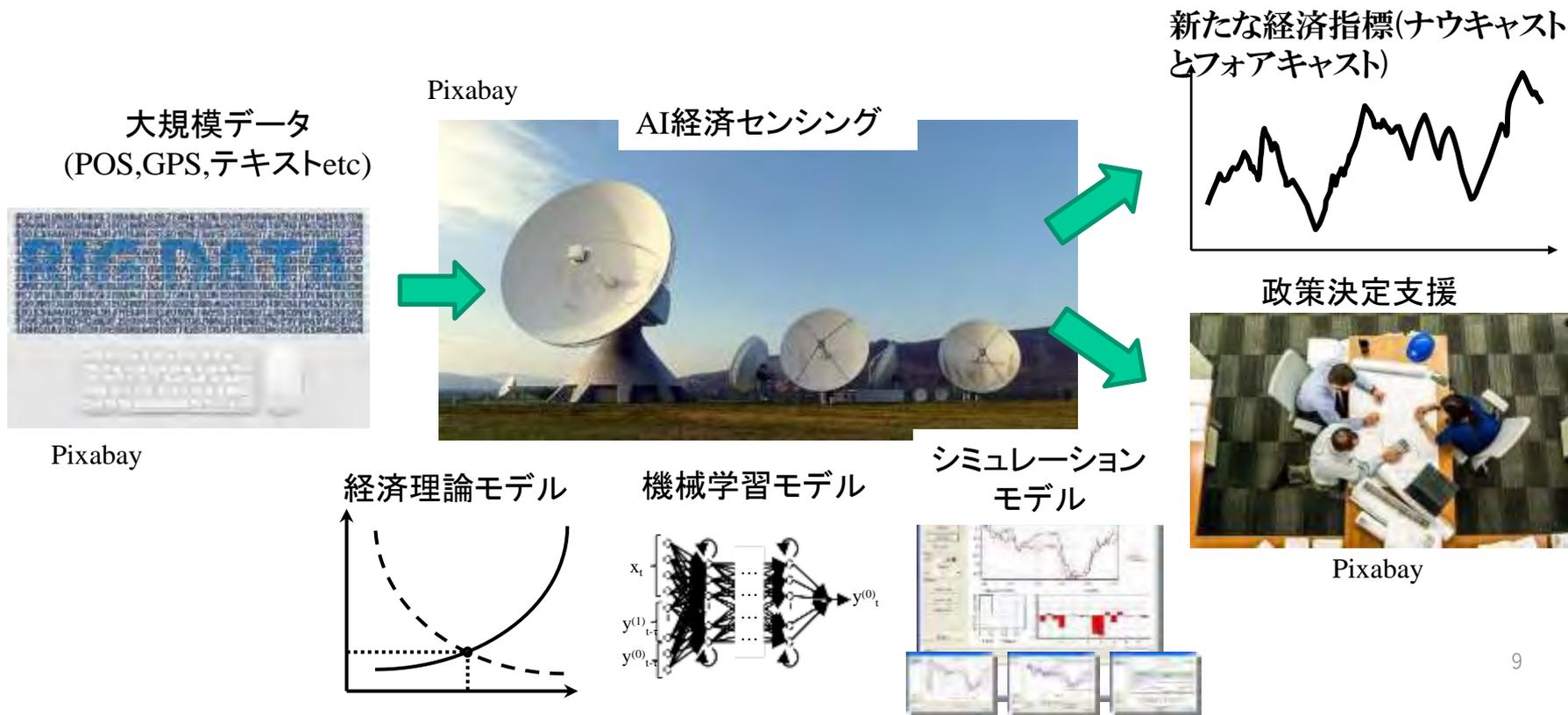
代替データ活用の方向性

- 構造化データ から 非構造化データ へ
- 既存指標の補完 から 新規指標 へ
- ナウキャスト (現状把握) から フォアキャスト (予測) へ



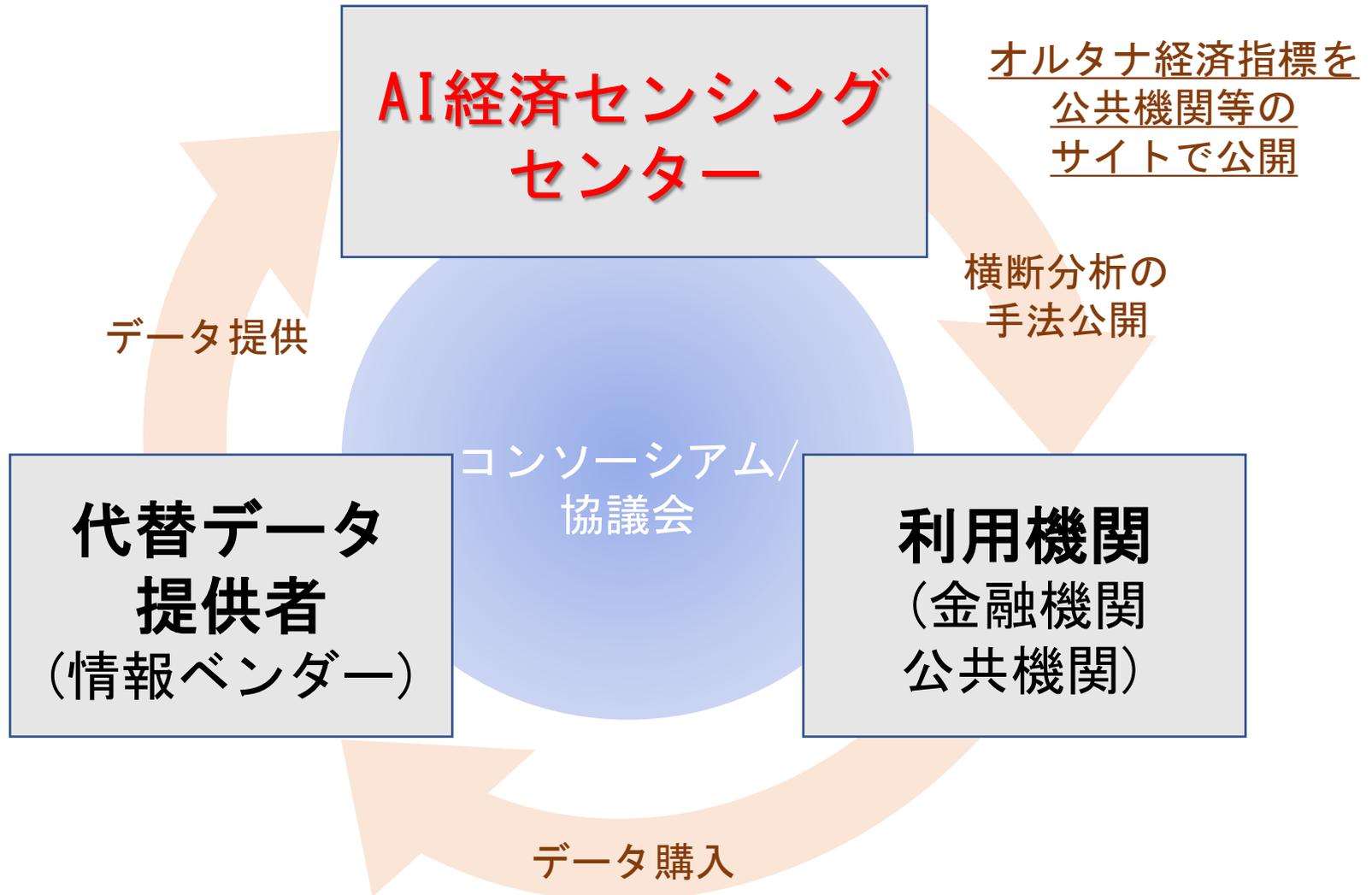
【最終目標】 AI経済センシングセンター 構想

- ・ **オルタナティブデータ**解析に横串を！
 - 基本的な解析方法のベースラインを公開して共有
- ・ **経済理論モデル+データマイニング**によるフォアキャストと政策決定支援



「三方一両得」の社会実装

- 分析手法公開による代替データの高付加価値化



JST未来社会創造事業 (2020年11月-)

「代替データと理論モデルの融合による新たな経済観測」

東大グループ

和泉 潔(PL)
坂地 泰紀
久野 遼平/東京大学

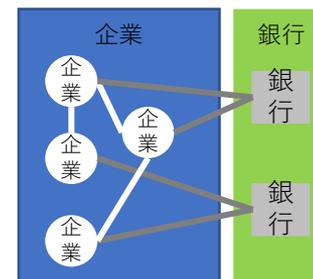
金融市場 シミュレーション



一橋大グループ

宮川 大介/一橋大学
滝澤 美帆/学習院大学

大規模パネルデータ/
企業間・金融取引
ネットワーク



慶應グループ

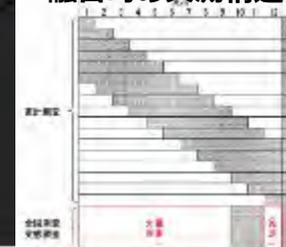
星野 崇宏/慶應義塾大学
加藤 諒/神戸大学
森脇大輔/サイバーエージェント

異種データ融合
(位置情報と
小売POSなど)/
政府統計個票の利用

GPS位置情報



家計調査と全消の 融合時の欠測構造



+ 金融機関、データ提供会社

オルタナティブ経済指標の公表

各種の代替データから経済指標を作成する。

- **テキストデータから業種別景況感**の月次指標
 - 開発技術：機械学習＋自然言語処理＋時系列の融合
 - 従来景気動向指数よりも頻度高、詳細化
- **企業情報とPOSデータから月次の業種別企業活動・売上高**指標
 - 開発技術：機械学習＋時系列予測＋クラスタリング
 - 従来指標よりも頻度高、幅広い業種をカバー
- **会計データから業種別の財務**指標
 - 開発技術：時系列予測＋欠測データ解析
 - 従来指標よりも頻度高、幅広い業種をカバー

上記以外に 移動情報、衛星画像、クラウドファンディング、不動産情報、リース情報、医療情報 等のデータ提供を交渉中。

ご関心のある機関(利用者・提供者)は、ぜひご連絡ください！

代替データをまぜる際の技術的課題

代替データ間の構造 (関係性) をどうやって作るのか？

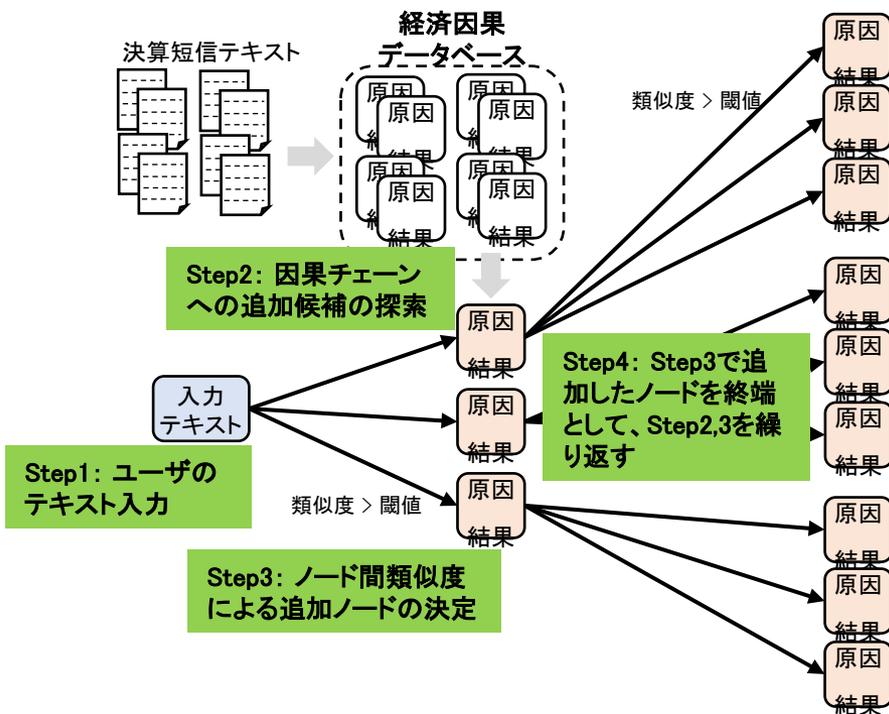
- 経済的な因子間の関係は、人間の活動に基づく
→ 自然科学のような第一原理はない
- 自然科学的な現象の場合とは異なるデータ分析手法が必要となる
- **経済現象に関する人間の認識**を軸とした、新たな代替データ分析手法を構築する
 - 統計的因果推論と自然言語処理、社会シミュレーションの融合

因果チェーンで代替データに横串を！

- 経済理論や因果関係などに基づく代替データ間の関係性の構造化

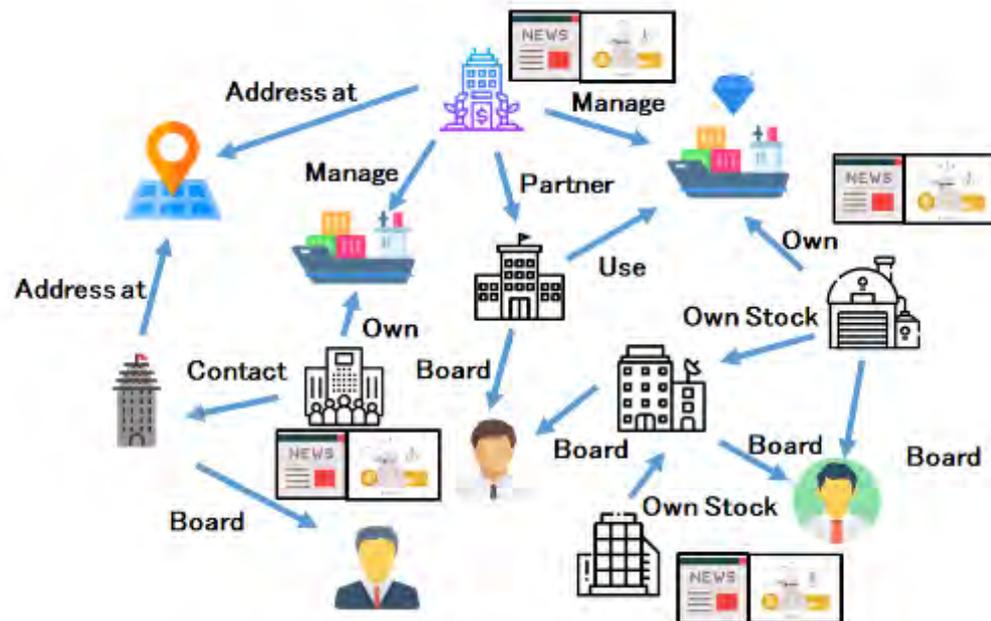
経済因果チェーン構築

テキストデータを基に人々が認識している経済的な因果関係の連鎖を検索する



経済ナレッジグラフ

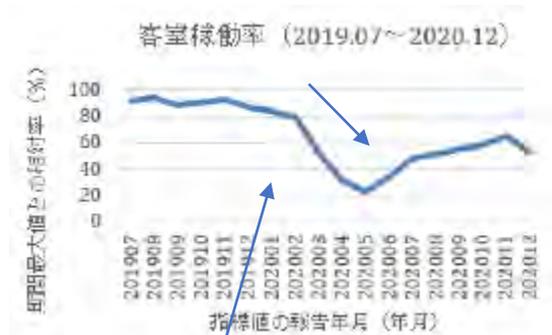
経済取引など裏がとれているファクトのデータを統合的なデータベースとして融合



クイズ

1. **感染症**が拡大すると、影響を受ける企業指標は？
2. **米国大統領選挙**があるときに、影響を受けるマクロ経済指標は？

クイズの答え

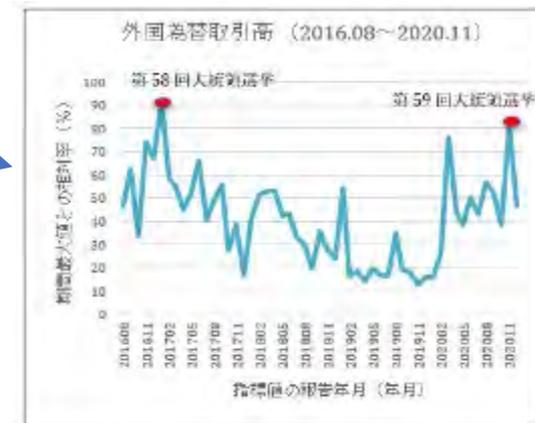


1. **感染症**が拡大すると、影響を受ける企業指標は？

→ 特定産業の売上高・利益率、**客室稼働率** etc.

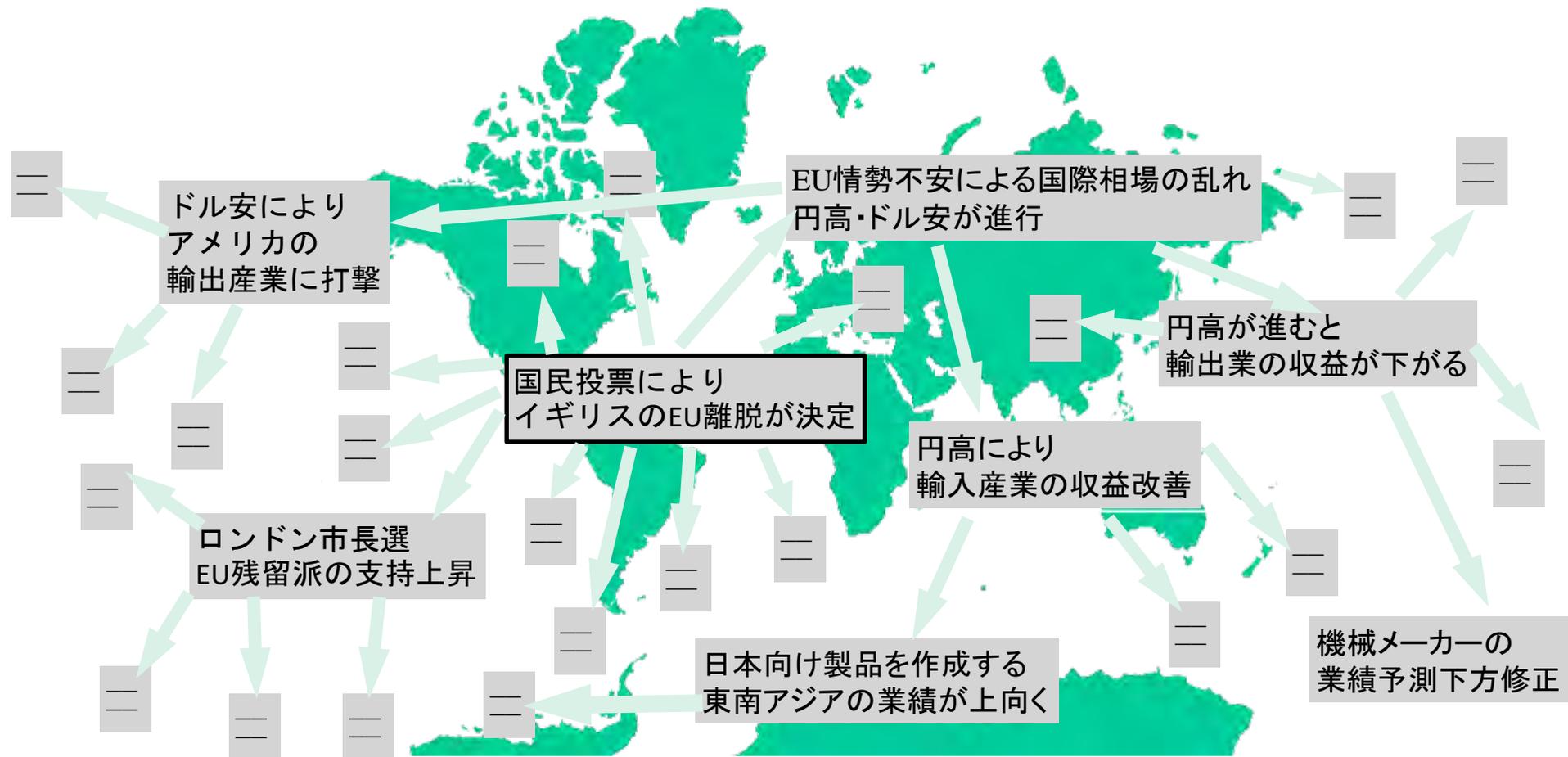
2. **米国大統領選挙**があるときに、影響を受けるマクロ経済指標は？

→ **外国為替取引高** etc.



経済金融に関わる全ての因果関係のデータベースを構築したい！

- 因果関係のネットワークを構築することで、どのような影響があるかを知ることができるようになる



因果チェーン検索サービス @ Symphony Market

(和泉研究室×電通国際情報サービス)

和泉研究室にて公開している**決算短信テキスト**より抽出した**因果情報**から**因果チェーン**を検索するサービスを金融機関向けコラボレーションプラットフォーム「Symphony」上にリリースを致します

The screenshot shows the 'Causal Chain Search' interface. On the left, a sidebar lists 'Symphony Market' and 'Causal-Chain Search'. The main area displays a search for '消費機 増損' (Consumer Machine Increase/Decrease) for the period 2012/10 ~ 2018/9. A flowchart shows the causal chain from '消費機 増損' to '消費機 増損' (via '増損'), then to '消費機 増損' (via '増損'), and finally to '消費機 増損' (via '増損').

Annotations on the left side of the interface:

- 「感染症」
- 「オリンピック」
- 「小麦価格の高騰」

Annotations on the right side of the interface:

- 旅行業界への波及効果
- ゴルフ業界への波及効果
- 製粉機械メーカーへの波及効果

Annotations on the far right side of the interface:

- 特定の社会現象に関する因果の要素候補の抽出
- 特定の社会現象 = $X \text{ or } Y$
- 原因候補 = $\{X_1, X_2, \dots, X_n\}$
- 結果候補 = $\{Y_1, Y_2, \dots, Y_m\}$
- 交絡要因候補 = $\{Z_1, Z_2, \dots, Z_l\}$



因果チェーン構築：決算短信に含まれる主な情報

下記に関する定性的情報が含まれる

- 業績予想

- 経営成績

- 経営成績に関する分析

- 財政状態に関する分析

- 事業等のリスク

- 継続企業の前提に関する重要事象等

- 経営方針

- 会社の経営の基本方針

- 目標とする経営指標

- 中長期的な会社の経営戦略

- 会社の対処すべき課題

業態を越えた販売競争の激化及び人件費の高騰、物流コストの上昇等により引続き厳しい経営環境が続いております

天候不良は、これらの事業における売上の低迷をもたらし、業績と財政状況に悪影響を及ぼす可能性があります

因果関係文の判別

使用したテキストデータ：
2012年10月から2018年5月の
決算短信テキスト

Step 1

Company X, publishing date

業態を越えた販売競争の激化及び人件費の高騰、物流コストの上昇等により引続き厳しい経営環境が続いております

Step 2

Company Y, publishing date

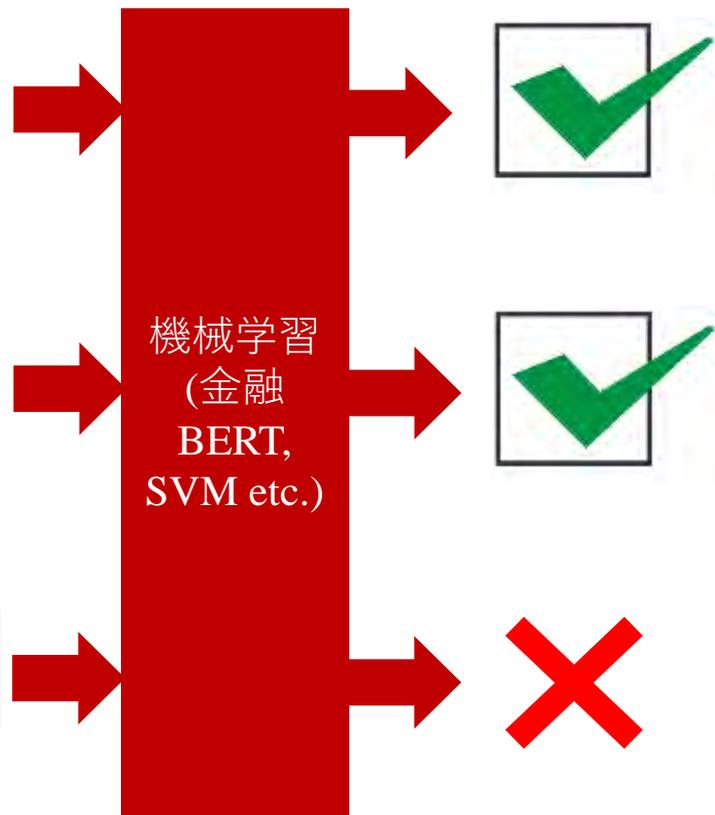
天候不良は、これらの事業における売上の低迷をもたらし、業績と財政状況に悪影響を及ぼす可能性があります

Step 3

Company Z, publishing date

総資産は、前連結会計年度末に比べてXXX百万円増加し、YYY百万円となりました。

⋮



因果表現 (原因/結果情報) の抽出

Step 1

Step 2

Step 3

業態を越えた販売競争の激化及び人件費の高騰、物流コストの上昇等により引続き厳しい経営環境が続いております

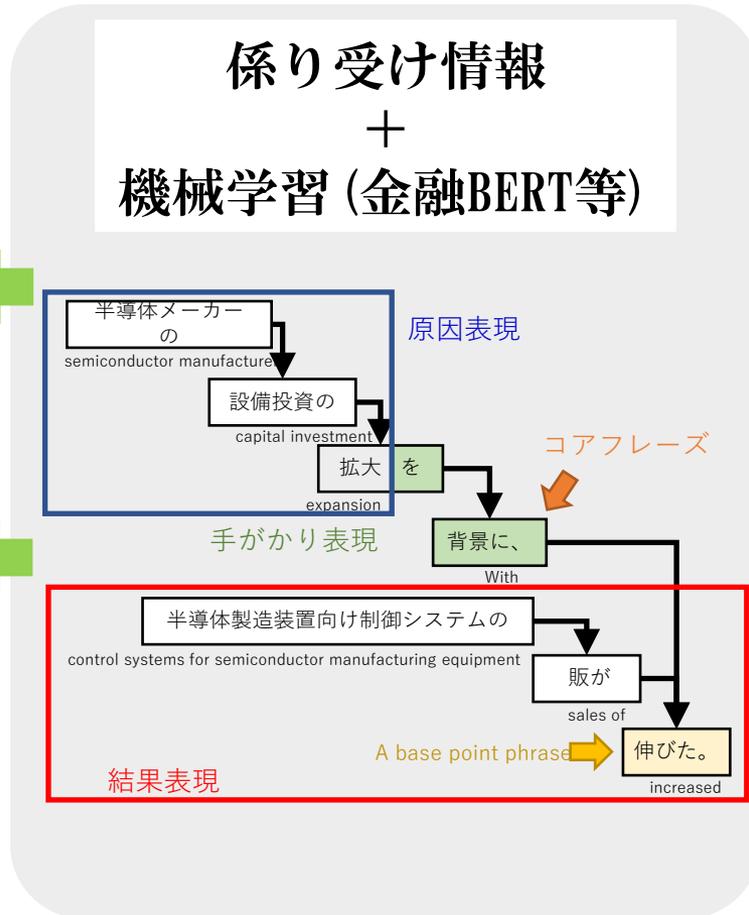
天候不良は、これらの事業における売上の低迷をもたらし、業績と財政状況に悪影響を及ぼす可能性があります

原因表現

結果表現

経済的因果データベース

1,078,542 組の因果



決算短信から抽出した因果関係のデータベース

因果関係のデータベースを構築する

原因表現

結果表現

銘柄:XXXX ○○年○月○日

業態を越えた販売競争の激化及び人件費の高騰、物流コストの上昇等



引続き厳しい経営環境が続いております

銘柄:YYYY ○○年○月○日

天候不良

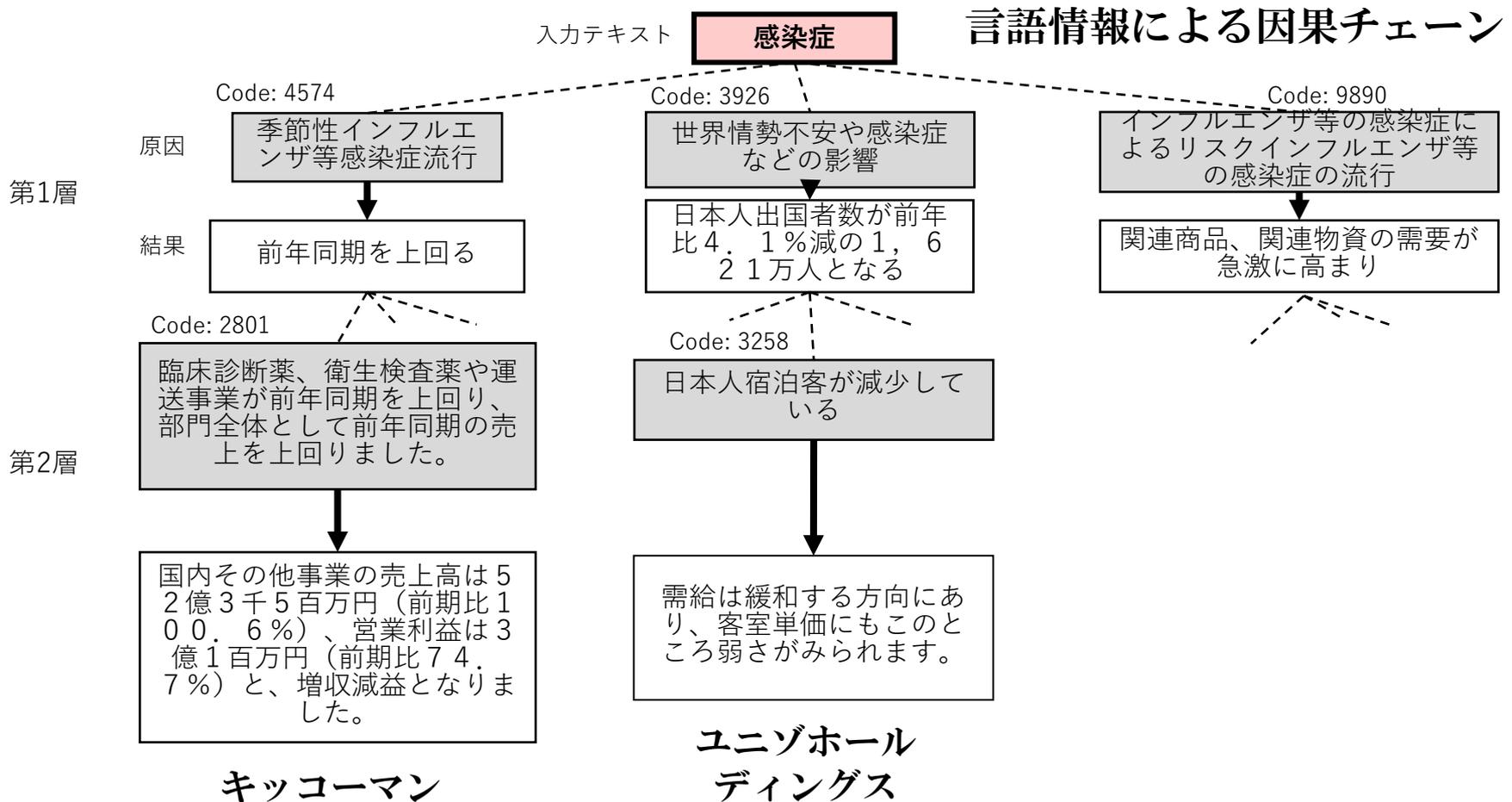


売上の低迷

使用したテキストデータ:

2012年10月から2018年5月の決算短信テキスト
抽出した因果関係: 1,078,542個

1. 因果チェーンの構築

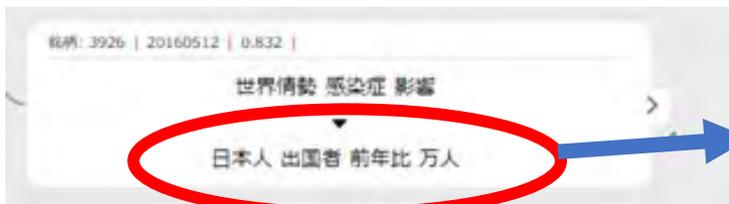


因果チェーンから関連経済指標の提示

■経済指標に関連するテキストデータ (Deep Data Research社 提供)

上場企業が自社サイトからIRやPRとして開示するマンスリーレポートからデータを取得。元のPDFを前処理してデータセットに整理。

- 全体で上場企業約500社程度が毎月PDFや該当ウェブページを開示。
- データ期間：2019年1月～2021年2月



A	B	C	D
1	KPI	コメント	
2	売上高	前年比 前年比 売上	
3	売上高	日本人 出国者 取引	
4	出国者数 出国者 取引	
5	出国者数	日本人 出国者 取引	
6	店舗数	
7	店舗数	
8	売上高	
9	売上高	
10	店舗数	

キーワードの種類が2以上の場合、いずれか2個一致したメモに該当するKPIを抽出。
 <例>対象キーワード： 日本人、出国者、前年比
 抽出されたKPI： 売上高、出国者数

1. 因果チェーンの各BOX内・結果の欄の名詞をコメント欄に**2つ**以上含む指標を抽出。

2. 抽出された各指標の修正オッズ比(関連度)で順位付け

関連数値指標 (上位5位)	
数値指標	関連度
1 平均賃料坪単価	8.093
2 NOI	8.093
3 ADR	6.641
4 RevPAR	6.636
5 客室稼働率	3.994

「感染症」からの波及効果

前向き検索 (波及効果) の結果一部

入力テキスト

感染症

Code: 4574

季節性インフルエンザ等感染症流行

原因

前年同期を上回る

結果

Code: 3926

世界情勢不安や感染症などの影響

日本人出国者数が前年比4.1%減の1,621万人となる

Code: 3258

日本人宿泊客が減少している

Code: 9890

インフルエンザ等の感染症によるリスクインフルエンザ等の感染症の流行

関連商品、関連物資の需要が急激に高まり

Code: 2801

臨床診断薬、衛生検査薬や運送事業が前年同期を上回り、部門全体として前年同期の売上を上回りました。

国内その他事業の**売上高**は52億3千5百万円（前期比100.6%）、**営業利益**は3億1百万円（前期比74.7%）と、増収減益となりました。

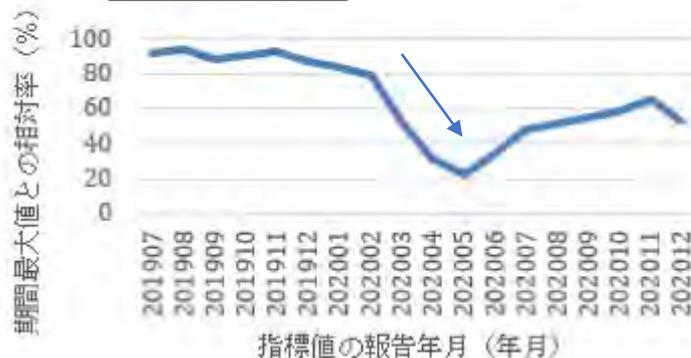
キッコーマン

(株価 2020.07-2021.06)



(Google Financeより)

客室稼働率 (2019.07~2020.12)



ユニゾホール
ディングス

需給は緩和する方向にあり、客室単価にもこのところ弱さがみられます。

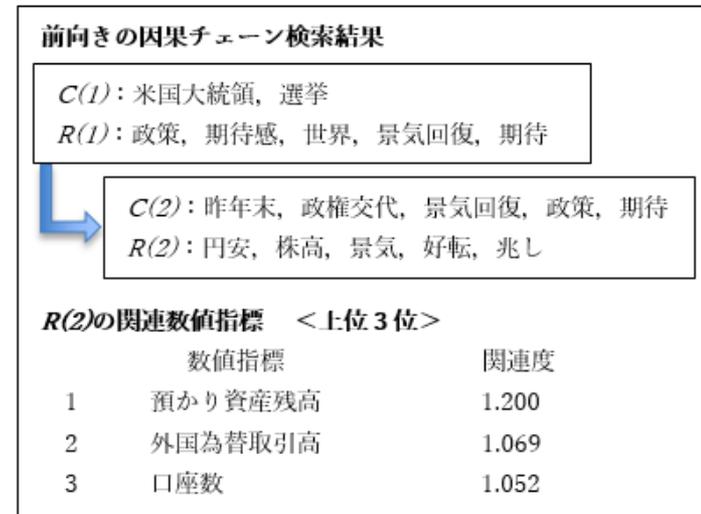
特願2021- 67163

統計的因果推論と自然言語処理の融合による言語表現と代替指標の因果による構造化

「感染症」からの波及効果予測



「米国大統領選挙」からの波及効果予測



【最終目標】 AI経済センシングセンター 構想

- ・ **オルタナティブデータ**解析に横串を！
 - 基本的な解析方法のベースラインを公開して共有
- ・ **経済理論モデル+データマイニング**によるフォアキャストと政策決定支援

