クラウドソーシングを用いた社会課題因果関係 LOD の構築

江上 周作　川村 隆浩　古崎 晃司　大須賀 昭彦

国内の自治体は交通渋滞，ポイ捨て，放置自転車など複数の社会課題を抱えており，対策に向けた議論や施策検討が行われている．より課題解決を進めるためには自治体の抱える各種社会課題の分類や因果関係，費用対効果等のデータの整理が望まれるが，現在これらのデータは整理されていない．そこで，本研究では社会課題の因果関係と関係費用を含む Linked Open Data(LOD) を構築し，社会課題の分析基盤の実現を目指す．本稿では，社会課題の因果関係と自治体事業予算のスキーマを設計し，クラウドソーシングと自然言語処理を用いて LOD を半自動的に構築した．また，構築した LOD を活用した分析・考察事例を示した．

Local governments have a number of urban problems, such as crime, traffic jam, and littering. However, various factors are socially intertwined; thus, structural management of the related data is required for solutions to such problems. Moreover, to implement the action plans, local governments need to grasp the cost-effectiveness. Therefore, this paper aims to construct Linked Open Data (LOD) that include causal relations of urban problems and the related cost information in the budget. We first designed LOD schema that represents the urban problem causality and budgets information. Next, we constructed LOD based on designed schema using NLP and crowdsourcing. Finally, we provided some use cases of our LOD to consider solutions of urban problems.

1　はじめに

現在，我が国では交通事故，放置自転車，騒音，ごみ問題など複数の社会課題を抱えており，自治体ごとに対策に向けた議論や施策検討が行われている．こうした社会課題の解決に向けたデータドリブンアプローチが注目されており，2013 年に行われた G8 サミットでの「オープンデータ憲章」への合意以降，自治体によるオープンデータの取り組みが普及している．大阪市ではオープンデータポータルサイト †1 で約 10,000 件のデータセットを公開しており，オープンデータを活用した課題解決に役立つアプリケーションの事例を紹介している．横浜市では，NPO 法人横浜コミュニティデザイン・ラボにより，横浜市オープンデータを活用した地域課題解決のビジネス創出を支援するクラウドファンディングサイトが運営されている †2

このようにオープンデータを用いた社会課題解決が各地で注目されている．しかし，社会課題は様々な原因が複雑に絡み合っているため，より課題解決を進めるためには自治体の抱える各種社会課題の分類や因果関係，解決に向けた施策や効果に関するデータの整理が望まれる．そこで，本研究ではこれらのデータのスキーマを整備し，各データの要素をリンクさせた Linked Open Data (LOD) としてデータを蓄積することで，社会課題の分析基盤の構築を目指し

---

* Constructing Urban Problem LOD with Causality using Crowdsourcing

Shusaku Egami, Akihiko Ohsuga, 電気通信大学大学院情報理工学研究科, Graduate School of Informatics and Engineering, University of Electro-Communications.

Takahiro Kawamura, 科学技術振興機構, Japan Science and Technology Agency.

Kouji Kozaki, 大阪大学産業科学研究所, Institute of Scientific and Industrial Research, Osaka University.

†1 https://data.city.osaka.lg.jp/

†2 http://yokohama.localgood.jp/

ている．LOD とはデータを主語・述語・目的語の三項構造 (トリプル) で記述する Resource Description Framework 形式で構造化したオープンデータであり，Web 標準化団体 W3C により推奨されているデータ普及方法である．

また，社会課題の本格的な解決には自治体の施策実行が望ましいが，自治体は施策実行に慎重であるため事前に費用対効果の把握を要する．これは横浜市政策局の職員にヒアリングを行った際に得た情報である．そこで，本研究では自治体事業の予算情報を収集し，社会課題の因果関係と紐づけた LOD を構築する．この LOD により，例えば社会課題の因果関係や階層関係のリンクをたどって課題の影響範囲を分析・予測することや，その結果を基に自治体の対策方針や予算削減案を検討することが可能になる．本稿では，まず社会課題の因果関係と自治体予算情報を表す LOD のスキーマを設計した．次に Web から収集した情報を基に自然言語処理とクラウドソーシングを用いて社会課題の因果関係を抽出し，設計したスキーマに基いて LOD を構築した．最後に，構築した LOD を用いて社会課題の因果関係と関連予算情報の検索例を示した．したがって本論文の貢献は以下の 4 点である．

1. 都市問題の因果関係と自治体予算情報スキーマの設計
2. 因果関係 LOD を半自動的に構築する手法の提案
3. 社会課題因果関係 LOD の公開
4. 社会課題解決や予算削減の検討に有用な検索例の提示

以下，2 章では関連研究について述べ，3 章では社会課題の因果関係と自治体事業の予算情報を含む LOD のスキーマ設計について述べる．4 章で因果関係の半自動的な抽出および LOD の構築について述べる．5 章では構築した LOD を使用した検索例を提示し考察する．最後に 6 章でむすびとする．

2 関連研究

2.1 社会課題解決に向けた Linked Data

これまでに，社会課題解決に向けた幾つかの Linked Data が公開されている．Szekely らは人身売買問題解決に向けて，Web の悪質広告から収集した情報を基に Linked Data を構築し，迷子検索システムを開発し法執行機関や NGO に導入している [14]．Szekely らが特定の社会課題解決に向けて Linked Data を構築しているのに対し，我々は複数の社会課題間の関係ををを表す Linked Data を構築し，課題解決を目指す点で異なる．

我々はこれまでに，社会課題の一つとして放置自転車問題を対象とし，放置自転車問題解決に向けた LOD を構築し，可視化することで市民の問題意識向上を行ってきた [5]．この研究で，我々は放置自転車のように日常的に生じる社会課題に関する LOD スキーマ設計の方法論を提案した．しかしながら，スキーマの設計はワーカー 1 人の知識に依存し，因果関係のカバー率が十分でないという問題点があった．また，設計されたスキーマは単一の社会課題を対象としており，社会課題間の因果関係を辿った推論や，予算情報を含めた検索までは不可能であった．そこで，本稿では社会課題の因果関係と予算情報を含む新たなスキーマを設計した．

Shiramatsu ら [13] は社会課題の解決目標を LOD 化するためのデータモデルを提案している．さらにこの LOD 活用し，課題解決者同士のマッチングをすることでシビックテックを促進させる手法を提案している．Web アプリケーションとして Goalshare を開発しており，実際に社会課題解決に向けてシビックテックイベントでの活用を行っている．この研究で構築された LOD は課題解決目標に関するものであり，社会課題の因果関係については記述されていないが，将来的に本稿で述べる LOD とのリンクにより課題解決促進が期待できる．

2.2 市の指標分析に向けた Linked Data

Santos ら [12] は市の様々な指標の値を算出するために Linked Data を構築し，可視化をサポートする quality of experience (QoE) Indicators Ontology を新たに提案している．また，Linked Data からデータを取得して可視化ウィジェットを生成できるダッシュボードを開発している．Pileggi ら [11] は首都圏の動的かつ詳細な指標を表すために，Conceptual

Framework を定義し，OWL-DL[†3] オントロジーとして実装している．データ構造の理解を容易にするために，Indicator, Data, Profiling, Computations, Geographic Context の 5 つのサブオントロジーに分割している．

OpenSpending.org では各国の自治体が公開する予算・支出情報のオープンデータのカタログサイトである．現在 75 カ国のデータが集約されている．これらのデータ LOD 化したものに LinkedSpending [8] がある．統計データの RDF データ・モデルを表現する RDF Data Cube Vocabulary[†4] に従って OpenSpending のデータを Linked Data 化している．これらの Linked Data には社会課題情報は含まれていないが，予算情報や指標については本研究で提案する LOD とリンク出来る可能性がある．

2.3 Linked Data 構築に向けたクラウドソーシングと自然言語処理

Demartini ら [4] はクラウドソーシングを用いた Linked Data のための Entitiy Linking 手法を提案している．クラウドソーシングはリンクの品質を向上させるために使用され，信頼性の低いユーザを排除し，一貫性のない結果を統合するための確率ネットワークを開発している．Celino ら [3] は，Linked Data における Point of interent (POI) データと写真をリンクするためのモバイルアプリケーションを開発している．ユーザにインセンティブを与えるために Game with a purpose (GWAP) [1] の概念を導入し，クラウドソーシングによりデータをリンクさせている．しかしながら，これらの手法は正解データが存在することを前提としており，社会課題の因果関係のように正解データが存在しない問題に対して適用することは困難である．

Nguyen ら [10] は Conditional Random Fields を使用し Twitter からユーザの行動を推定し，Linked Data を構築する手法を提案している．ユーザの行動や日時，場所，目的等をトリプルとして構築している．主に，震災時の行動分析や，ナビゲーションに応用が可能である．LODfier [2] は Wikifier [9] を用いた Named Entity-Recognition により非構造化テキストからエンティティを抽出し，DBpedia や WordNet にリンクする RDF を生成する．Linked Data 構築に向けた自然言語処理手法は数多くあるが，我々の知る限り，これらの非構造化テキストからの単語抽出手法を因果抽出に適応するには，精度の面で不十分である．したがって，自然言語処理とクラウドソーシングを組み合わせることで社会課題の因果関係を抽出する手法を提案した．

3 社会課題課題因果関係と予算情報のスキーマ設計

本研究で構築する LOD の主要な使用用途は，自治体の抱える社会課題（都市問題）の解決施策や予算削減案の導出支援である．具体的には，因果関係を基に都市問題解決策の波及効果や，各都市問題に関する自治体ごとの予算を検索可能にし，都市問題解決施策の作成や予算削減案の作成を支援する．そこで，我々は都市問題の因果関係と自治体の予算情報を記述するスキーマを設計した．図 1 に設計したスキーマを示す．図 1 の上部は都市問題の因果関係を記述する語彙である．因果関係部分ではリソースは都市問題かそうでないもののどちらかに分類される．因果関係を表すプロパティは大きく分けて `factor` と `affect` があり，推移律を持つ `TransitiveProperty` である．これら二つのプロパティは related プロパティのサブプロパティである．また，クラウドソーシングプロセスにおける一致人数に応じてサブプロパティを定義している．このようにプロパティを定義することで，用途に応じた推論が可能になる．例えば，クラウドソーシングプロセスにおける一致人数の高い因果関係を知りたい場合 `factor_level4` や `affect_level4` プロパティを使用し，一致人数に関係なく因果関係を全て知りたい場合は `factor` と `affect` プロパティを使用する．図 1 の下部は自治体の予算情報を記述する語彙である．自治体の予算情報は Microsoft Excel や Adobe PDF などの表形式データで公開されているため，RDF Data Cube Vocabulary を使用して記述する．さらに，本研究では Online analytical processing (OLAP) のデー

†3 `https://www.w3.org/TR/owl-guide/`
†4 `https://www.w3.org/TR/vocab-data-cube/`

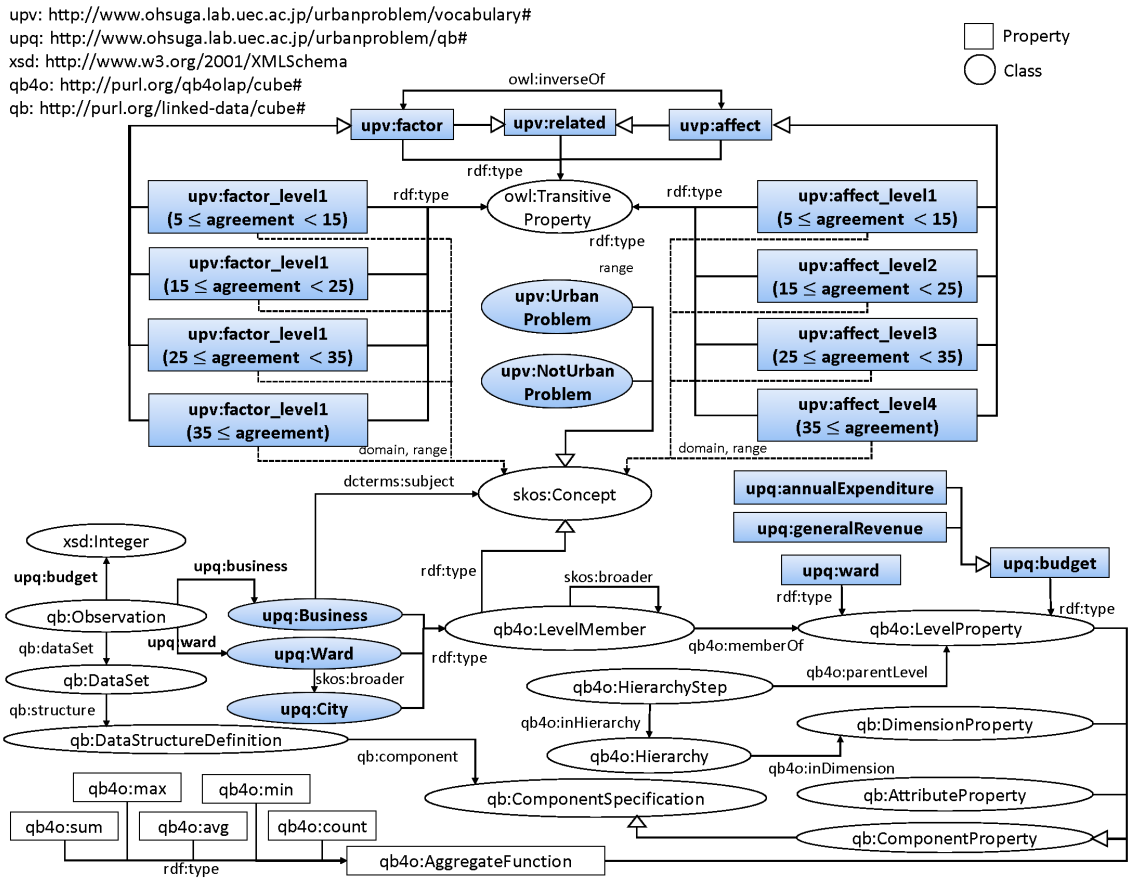

図 1　社会課題の因果関係と予算情報を表す Linked Data スキーマ

タモデルで使用され，RDF Data Cube Vocabulary の拡張である QB4OLAP [6] に基づいて予算情報を記述する．QB4OLAP により，都市問題対策事業の地域全体の予算合計や部局ごとの予算合計，最大値最小値などを検索することが可能になる．自治体の事業名プロパティの range は事業名クラスであり，事業名クラスは少なくとも一つの dcterm:subject プロパティを持ち，その値は skos:Concept クラスである．このように設計することで，都市問題の波及効果から関連予算まで取得することが可能になる．

4　社会課題因果関係 LOD の構築

4.1　社会課題の因果関係抽出

本章では社会課題の因果関係を半自動的に抽出する手法を提案する．提案手法は以下の 4 ステップで構成されている．

(1)　検索エンジンによる Web 文書の収集
(2)　係り受け解析による文書からの因果単語抽出
(3)　抽出した因果単語を基にワードクラウドを生成
(4)　クラウドソーシングによる因果単語のフィルタリング

まず，対象とする社会課題名とその同意語を一番目の検索語とし，“要因” の同意語を二番目の検索語として検索エンジンを用いた記事検索を行う．以下，郊外型犯罪の要因を例として説明する．郊外型犯罪の場合，一番目の検索語は “郊外型犯罪” を使用し，二番目の検索語は “要因” とその同意語であり，日本語 WordNet[†5] から “要素”，“原因” を取得して使用する．同意語として取得したが社会課題に関する日本

†5　http://nlpwww.nict.go.jp/wn-ja/

語文書において使用頻度の少ないと考えられる，“因子”，“素因”，“導因”，“エレメント”，“ファクター”等に関しては，これらの検索語がメインとして検索されてしまったため除外した．本研究では検索エンジンとして Google と Bing を使用し，Google Custom Search API[†6] と Bing Web Search API[†7] に「郊外型犯罪　要因」のようなキーワードパラメータを与えて結果を取得する．取得するファイルタイプとして HTML と PDF を別々に指定し，一番目の検索語と二番目の検索語の全ての組み合わせで記事を 50 件ずつ検索する．HTML と PDF を分けて検索した意図としては，社会課題の因果関係について言及している記事を，行政の調査報告や議事録等と一般人の SNS やブログ等から満遍なく収集するためである．また，記事中における社会課題名の出現回数が低い場合は，この後の要因単語抽出処理でノイズが多く混ざってしまうため，社会課題名の出現回数でフィルタリングを行う．

収集した記事の内，“要因”，“要素”，“原因” を含む文を抽出し，文ごとに形態素解析を行い名詞を抽出する．この後の処理であるクラウドソーシングによる要因単語選択の処理において，ある程度意味が通じる名詞となるように，サ変接続名詞の直前・直後に名詞がある場合はそれらを連結して一つの名詞とする．例えば “非行防止” は通常 “非行 (名詞-一般)” と “防止 (名詞-サ変接続)” に分割されてしまうが，本手法ではサ変接続を結合して “非行防止” とする．

次に日本語係り受け解析を行い，“要因”，“要素”，“原因” を含む文節と直接的に依存関係のある文節から，前の処理で抽出した名詞句に相当する部分を抽出する．間接的に依存関係のある文節まで取得するとノイズが多くなってしまうため，本研究では直接的に依存関係のある文節のみ使用した．

同様に社会課題の影響に関しての抽出も行う．“影響” の他，日本語 WordNet から “引き起こす”，“惹起”，“もたらす”，“来たす”，“招く”，“誘発”，“生み出す” などの同意語を取得し，これらを二番目の検索語として要因抽出と同様の処理を行う．

4.2　クラウドソーシングを用いた因果単語のフィルタリング

次に抽出した因果単語を基にワードクラウドを生成し，クラウドソーシングによる因果単語の選択を行う．まず，4.1 章で抽出した因果単語の内，類似度の高い単語を文字列の類似度計算結果に基いて統合した．文字列の類似度計算には Jaro-Winkler 距離 [16] を使用し，類似度が 0.8 以上であった場合に文字列の長い単語に統合した．次に因果単語の出現回数を基にワードクラウドを生成した．ワードクラウドを使用する理由は，ワーカーになるべく負担をかけず重要な単語の抽出を容易にするためである．図**??**に郊外型犯罪の要因単語のワードクラウドを示す．出現回数が多いほど中心に位置し文字サイズが大きくなる．文字の色はランダムに決定する．このワードクラウドから「郊外型犯罪の原因として考えられるものを 10 個選択する」「郊外型犯罪が引き起こすものや影響範囲として考えられるものを 10 個選択する」というタスクを設定し，クラウドソーシングによる因果単語の選択を行う．今回は金銭的報酬を与えるマイクロタスク型クラウドソーシングを行った．クラウドソーシングサービスとしてランサーズ [†8] を使用し，上記２つのタスクの遂行報酬を 50 円として，一つの社会課題につき最大 50 人に作業を依頼した．このうち作業人数の 1 割にあたる 5 人以上に選択された単語を使用する．選択された単語の中に類似の単語がある場合は，選択した人数の多い単語に統一する．図 2 内で赤枠で囲まれた単語は実際にクラウドソーシングで 5 人以上に選択された単語である．

さらに，より多くの因果関係を構築するために，クラウドソーシング後に最終的に抽出された因果単語について，提案手法の (1) から (4) のステップを繰り返し実行する．しかしながら，必ずしも全ての単語が因果関係を持つとはかぎらないため，1 回目のループで抽出された因果単語について Google 検索結果の上位

†6 https://developers.google.com/custom-search/?hl=ja

†7 https://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/

†8 http://www.lancers.jp/

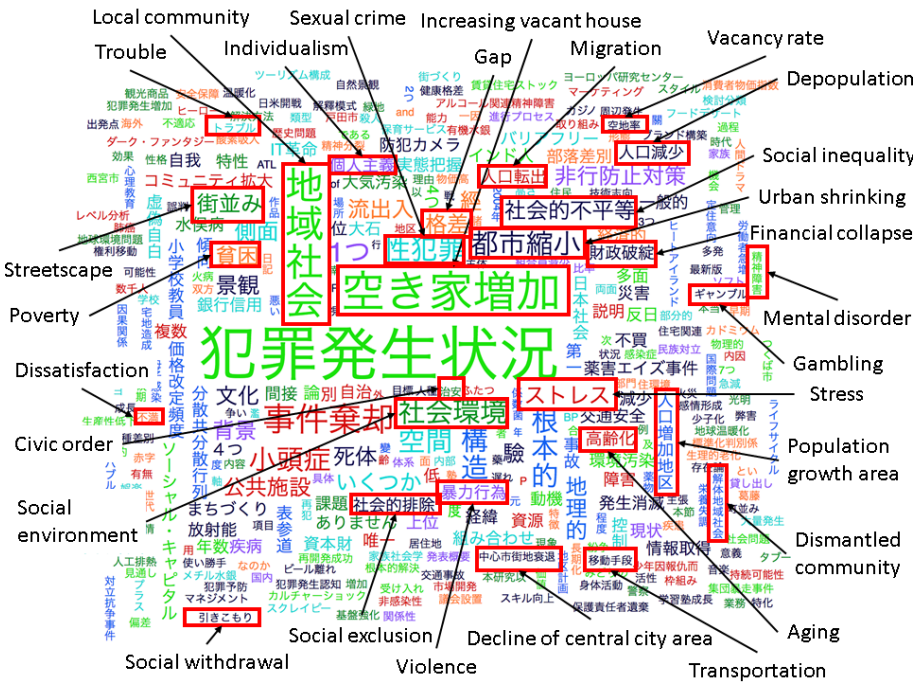

図 2　郊外型犯罪の要因のワードクラウド

50 件の記事から共起語を調査し，原因，要因，影響，効果などの単語を 20 以上持つ単語のみこのステップを実行する．

4.3　因果単語抽出結果に基づく LOD の構築

3 章で設計したスキーマに基いて，抽出した因果単語に関する LOD を構築した．クラウドソーシングの結果は TSV データで取得できるため，このデータを基に Apache Jena を使用して RDF に変換した．具体的には，都市問題は `UrbanProblem` クラスのサブクラスとして生成され，それ以外は `NotUrbanProblem` のサブクラスとして作成される．各クラス間の因果関係はクラウドソーシングにおける一致人数別に factor, affect のサブプロパティを用いて生成される．更に，各クラス名を形態素解析して名詞を抽出し，上位クラスとして作成した．全ての単語は Google Translate API[†9] で英語に自動翻訳し，リソースの一意な URI はこの英語を基に生成する．

4.4　自治体の予算情報に基づく LOD の構築

大阪市は大阪市オープンデータポータルサイト [†10] で様々なデータを CC-BY 4.0 ライセンスで公開している．本稿では大阪市をモデルとし，このポータルサイトから予算情報のデータを取得して RDF に変換する．まず，予算情報のデータは表形式の PDF データとして公開されているため，Adobe Acrobat を使用して Excel データ (XLS) に変換する．次に Apache POI[†11] を用いてセルから情報を抽出し，Apache Jena を用いて 3 章で設計したスキーマに基いて RDF を生成する．

次に，自治体の事業リソースと都市問題およびの都市問題の因果リソースとのリンク付けを行う．元のデータには事業に関する説明は記述されていないため，事業名称に含まれている単語を基に因果リソースへのリンク付けを行う．しかしながら，「路上喫煙対策事業」と「タバコ」のような関係は既存のシソーラス等を使って取得することはできない．そこで，自

†9　https://cloud.google.com/translate/docs?hl=ja

†10　https://data.city.osaka.lg.jp/

†11　https://poi.apache.org/

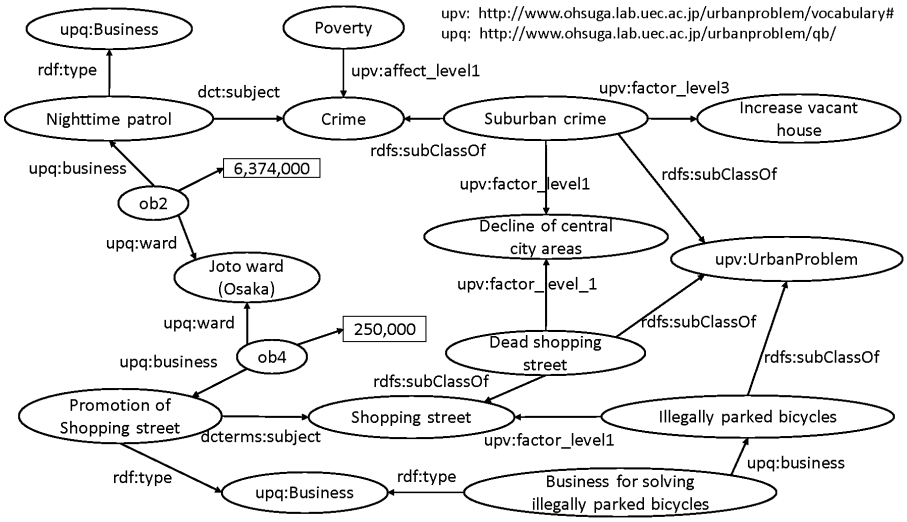

図 3　最終的に構築される LOD の一部

治体事業名からリンク付け候補の単語を増やすことで，事業リソースから因果リソースへのリンク付けを行った．

まず自治体事業名を形態素解析し名詞を抽出する．その際汎用的な後をストップワードとして削除している．次に各名詞の同義語を日本語 WordNet から取得し，リンク付けの候補を増やす．さらに，抽出した名詞の意義素を日本語 WordNet から取得する．意義素とは単語の意味を定義したものであり，その単語を使用した例文が含まれている．本研究では単語の意味を定義した文章のみを使用し，形態素解析により名詞を抽出する．このようにしてリンク付けの候補と成る単語を多く抽出し，これらの単語と同じ名称の都市問題およびその因果リソースのクラスと `dct:subject` プロパティでリンク付けを行う．

図 3 は最終的に構築される LOD の一部分である．トリプル数は 49,386 となった．構築した LOD は我々のウェブサイト †12 で，LOD に対して検索が実行可能な SPARQL Protocol and RDF Query Language (SPARQL) エンドポイントとともに公開している．さらに，ある事物により生じる影響は対象先の事物の上位概念に伝搬するというルールを新たに定義し，以下のように Semantic Web Rule Language (SWRL) で記述し，構築したデータを格納している RDF ストアに追加した．

(?x upv:affect ?y), (?y skos:broader ?z)
-> (?x upv:affect ?z)　　(式 1)

4.5　因果関係抽出結果の考察

表 1 に最終的に使用した記事数，抽出文数，抽出単語数の合計を示す．3.1 章で述べたように要因に関しては “素因”，“導因” などの検索ノイズとなる類語を多く除き，影響の方が使用した類語が多かったため，文書，文，単語の数が多くなっている．また，クラウドソーシングの一致度を計るため，Fleiss の Kappa 係数 [7] を計算した．ワーカー数は 50 人であり，要因に関する一致度の平均は 0.291，影響に関する一致度は 0.212，全体の平均は 0.256 となった．すなわち，クラウドソーシングによるワードクラウドからの因果単語選択には弱い一致度が確認でき，本手法はある程度妥当であったといえる．一致度が下がった原因として，社会課題と関連しない単語が中心に集まり，本来関係のある単語が目立たなくなり選択されづらくなったことが考えられる．これは，行政の文書は冗長的な表現が多く複雑な重文が多く含まれていたため，係り受け解析を用いた因果単語候補の抽出が上手くいかない例があったことにより生じている．また，ワー

†12 `http://www.ohsuga.lab.uec.ac.jp/urbanproblem/`

表 1　因果単語抽出結果

	# of documents including urban problem words	# of sentences including synonyms of “factor” and “influence”	# of extracted words
Factor	1,438	4,481	3,110
Influence	2,465	9,082	4,661

カーの予備知識が共通していないことも考えられ，社会課題の要因と影響の予備知識・問題意識が不足しているからこそ，社会課題が解決していないことも読み取れる．一方で，交通事故の要因に関しては 0.443 と高い一致度が確認できた．これは，警視庁，教育機関，報道機関などが一般に向けて交通事故に関して説明する機会が多いため，ワーカーの予備知識が共通していたのではないかと考える．

5　社会課題解決施策の検討に向けた LOD の検索

5.1　ホームレスの因果関係

図 4 はホームレスの因果関係とその予算情報について検索するクエリにより得られた部分グラフの例である．この図から，経済的困窮，雇用不足の他，ギャンブル依存等が原因でホームレスが増えていることがわかる．また，ホームレスは貧困ビジネスの発生を招き，貧困ビジネスは生活保護受給者の増加や地域差の増大や格差社会の助長につながることがわかる．貧困ビジネスとは経済的に困窮している貧困層を対象として利益を上げるビジネスであり，違法行為や非人道的な方法で金銭を巻き上げることが社会問題となっている．図の因果関係の具体例として，生活保護未受給の路上生活者に対して，生活保護の受給を強要してだまし取る貧困ビジネスが存在する．

2015 年の統計データによると大阪市は生活保護費が 2,907 億円を超えており，国内で最も生活保護費が多い自治体である．中でも西成区の生活保護受給費は 624 億円を超えており，二位以下に 2.5 倍以上の差をつけている．西成区ではホームレスを対象とした貧困ビジネスの発生も問題視されており，アパートを契約させて生活保護費からピンハネする事例や，携帯の名義貸しと引き換えに報酬を受け取り，後で高額な請求が来る事例が問題となっている．このようなホームレスから始まる負の連鎖を止めるために，ホームレスの原因となっている雇用不足やギャンブル依存の対策を各自治体で進めることが望ましい．しかし，現在西成区だけでなく大阪市全体でもこのような取り組みは行われていない．雇用不足やギャンブル依存の対策事業を行うことで，生活保護受給者の減少および生活保護費の削減につながると考えられる．

5.2　ポイ捨ての因果関係

図 5 はポイ捨ての原因と影響およびそれらに関連する予算情報を検索するクエリにより得られた部分グラフである．この図から，灰皿設置（不足）がポイ捨てに影響を与えており，ポイ捨ては地下水汚染を引き起こすことがわかる．また，ある事物により生じる影響は対象先の事物の上位概念に伝搬するというルール (式 1) により，ポイ捨てから地下水へ `affect` プロパティによる直接的な関係が推論できる．さらに，地下水は騒音に影響を与えることが，騒音に関する因果関係抽出結果から得られている．結果として，灰皿設置（不足）はポイ捨てを引き起こすだけでなく，連鎖的に地下水汚染および騒音に影響を与える可能性があることが `affect` プロパティの推移律により判明した．この因果関係の具体例としては，路上喫煙者がポイ捨てしたタバコが地下水に流れ込むことで地下水が汚染されたり，地下水の水質汚濁改善のため地下で配管工事が行われ，騒音被害が発生することなどが挙げられる．

都市問題とその予算の関係に関しては，大阪市住之江区は土壌汚染・水質汚濁対策事業に区・シティマネージャー自由経費予算として 336,000 円を投じていることがわかる．一方で同区において路上喫煙対策事業は 76,000 円しか予算を投じていないことがわかる．そこで，一連の都市問題のより上流に位置する路上喫煙対策事業予算を増額することでより効果的な

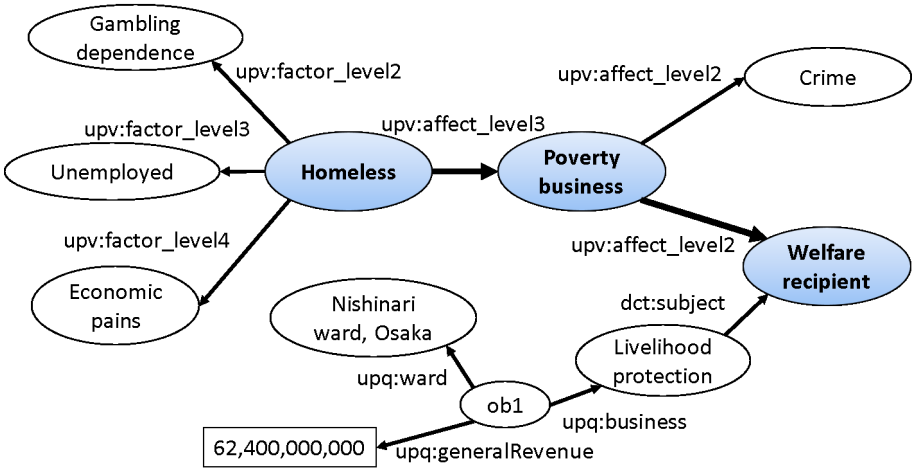

図 4　ホームレスの因果関係の部分グラフ

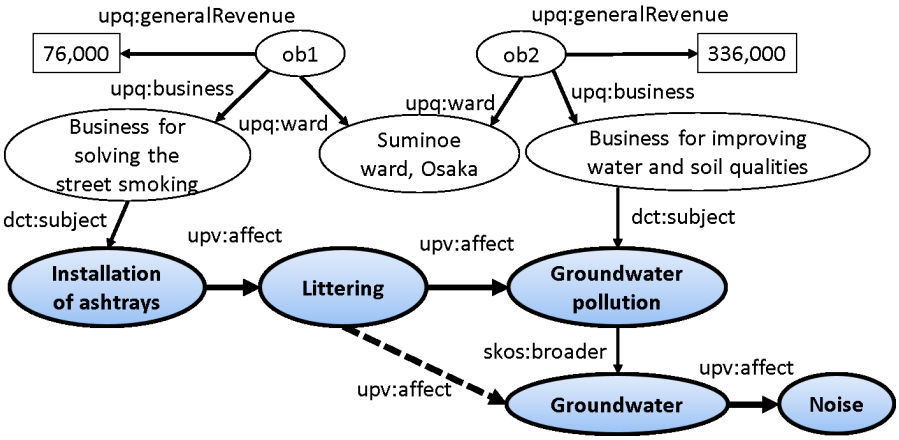

図 5　ポイ捨ての因果関係の部分グラフ

問題解決に繋がる可能性がある．路上喫煙対策事業の予算をより多く投じることで，路上喫煙の減少だけでなく，連鎖的に引き起こされるポイ捨て，地下水汚染，騒音問題の改善が期待される．一方で，地下水汚染が軽減されれば，土壌汚染・水質汚濁改善事業は予算を削減できるだろう．

5.3　郊外化の因果関係

図 6 は郊外化の因果関係とその予算情報について検索するクエリにより得られた部分グラフの例である．この図から，「郊外化→治安悪化→人口減少→郊外化」の悪循環が確認できる．郊外化（中心市街地の空洞化）により人通りの減少や空き家や増加することで治安が悪化する．治安悪化により犯罪が多発することでその地域からの人口流出，人口減少が加速し，さらなる郊外化（中心市街地の空洞化）に繋がる．また，郊外化に関するこの悪循環が繰り返される過程で，交通事故，シャッター通り，凶悪犯罪の発生などの他の問題に影響を与え，これらの二次的被害が拡大する恐れがあることがわかる．したがって，それ自体が深刻化するだけでなく，他の都市問題を次々と発生させる郊外化の悪循環は，対策が急がれる問題であると言える．

一方で郊外化の悪循環に突入する要因として不就

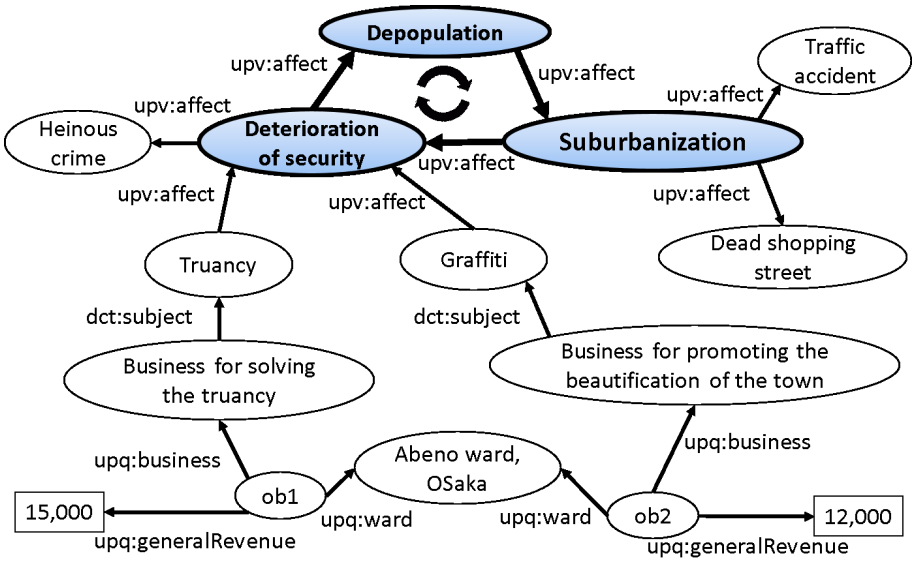

図 6　郊外化の因果関係の部分グラフ

学や落書きがある．不就学に関しては，素行不良の青少年が増加することが治安悪化につながると考えられる．落書きに関しては，落書きを放置しておくことで自治体が機能していないというイメージを与え，治安の悪化につながると考えられる．これは割れ窓理論 [15] に相当する現象である．しかしながら，大阪市阿倍野区において就学対策費は 15,000 円，まちの美化推進事業の予算は 12,000 円と極めて低く，大阪市全体で見ても就学対策費は平均 15,083 円，まちの美化推進事業の予算は平均 14,000 円とあまり重要視されていないことが伺える．これらの問題への対策予算を増加することで，郊外化の悪循環に入り込むリスクを減らすことが出来ると考えられる．

6　まとめと今後の課題

本研究では社会課題の因果関係とその予算を記述する Linked Data スキーマを設計し，自然言語処理とクラウドソーシングを用いて因果関係を抽出し，設計したスキーマに基いて LOD を構築した．さらに，LOD の検索例をもとに課題解決案および予算削減案を示し，LOD を使用することで自治体の抱える社会課題の解決策考案の一助と成る可能性を示した．本研究では独自に定義したルールを既存の OWL，RDF スキーマのルールとともに RDF ストア追加し，LOD に推論機能を付与して公開した．今後より高度な検索・推論機能を提供するため，独自に定義したルールを追加していく予定である．

本研究では社会課題の因果関係は行政による分析，社会学分野の研究成果，事実，人間の知識や連想などの情報に基づいて構築されている．従って，これらのソースデータを辿ることで全ての因果関係を説明することが可能である．今後，因果関係の理解が困難な場合に中間ノードを増やすことを検討している．

さらに，本研究で得られた成果を大阪市市民局に大阪市市民活動推進審議会を通じて説明し，社会課題の解決策を共同で検討する．実際に，筆者のうちの 1 人は市民局内の大阪市市民活動推進審議会の委員である．また，大阪市 ICT 戦略室とデータ提供に関して交渉を行う予定である．大阪市 ICT 戦略室はこれまでに国内のオープンデータコンテストに幾つかの行政データを提供しており，さらなるデータのオープンデータ化も見込まれるため，それらのデータを我々の LOD に追加することを検討している．

謝辞 本研究は JSPS 科研費 16K12411, 16K00419, 16K12533, 17H04705 の助成を受けたものです．

参 考 文 献

[1] Ahn, L. V.: Games with a purpose, *IEEE Computer*, Vol. 39, No. 6(2006), pp. 92–94.

[2] Augenstein, I., Padó, S., and Rudolph, S.: LODifier: Generating Linked Data from Unstructured Text, *Proc. of the 9th Extended Semantic Web Conference (ESWC)*, 2012, pp. 210–224.

[3] Celino, I., Contessa, S. annd Corubolo, M., Dell' Aglio, D., Valle, E. D., Fumeo, S., and Krüger, T.: Linking Smart Cities Datasets with Human Computation - The Case of UrbanMatch, *Proc. of the 11th International Semantic Web Conference (ISWC)*, 2011, pp. 94–108.

[4] Demartini, G., Difallah, D. E., and Cudré-Mauroux, P.: Large-scale linked data integration using probabilistic reasoning and crowdsourcing, *The International Journal on Very Large Data Bases*, Vol. 22, No. 5(2013), pp. 665–687.

[5] Egami, S., Kawamura, T., and Ohsuga, A.: Building Urban LOD for Solving Illegally Parked Bicycles in Tokyo, *Proc. of the 15th International Semantic Web Conference (ISWC)*, 2016, pp. 291–307.

[6] Etcheverry, L., Vaisman, A., and Zimányi, E.: Modeling and Querying Data Warehouses on the Semantic Web using QB4OLAP, *Proc. of the 16th International Conference on Data Warehousing and Knowledge Discovery (DaWaK)*, 2014, pp. 45–56.

[7] Fleiss, J. L. and Cohen, J.: The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability, *Educational and Psychological Measurement*, Vol. 33, No. 3(1973), pp. 613–619.

[8] Höffner, K., Martin, M., and Lehmann, J.: LinkedSpending: OpenSpending becomes Linked Open Data, *Semantic Web Journal*, Vol. 7, No. 1(2016), pp. 95–104.

[9] Milne, D. and Witten, I. H.: Learning to Link with Wikipedia, *Proc. of the 17th ACM conference on Information and knowledge management (CIKM)*, 2008, pp. 509–518.

[10] Nguyen, T. M., Kawamura, T., Tahara, Y., and Ohsuga, A.: Self-Supervised Capturing of Users' Activities from Weblogs, *International Journal of Intelligent Information and Database Systems*, Vol. 6, No. 1(2012), pp. 61–76.

[11] Pileggi, S. F. and Hunter, J.: An ontological approach to dynamic fine-grained Urban Indicators, *Procedia Computer Science*, Vol. 108(2017), pp. 2059–2068.

[12] Santos, H., Dantas, V., Furtado, V., Pinheiro, P., and McGuinness, D. L.: From Data to City Indicators: A Knowledge Graph for Supporting Automatic Generation of Dashboards, *Proc. of the 14th Extended Semantic Web Conference (ESWC)*, 2017, pp. 94–108.

[13] Shiramatsu, S., Tossavainen, T., Ozono, T., and Shintani, T.: Towards Continuous Collaboration on Civic Tech Projects: Use Cases of a Goal Sharing System Based on Linked Open Data, *Proc. of the 7th IFIP International Conference on Electronic Participation (ePart)*, 2015, pp. 81–92.

[14] Szekely, P., Knoblock, C. A., and Slepicka, J.: Building and Using a Knowledge Graph to Combat Human Trafficking, *Proc. of the 14th International Semantic Web Conference (ISWC)*, 2015, pp. 205–221.

[15] Wilson, J. Q. and K., G. L.: Broken windows, *Critical issues in policing*, (1982), pp. 395–407.

[16] Winkler, W.: The state record linkage and current research problems, *Technical report, Statistics of Income Division, Internal Revenue Service Publication*, (1999).