



道路政策の質の向上に資する技術研究開発
成果報告レポート
No. 20-1

研究テーマ

サービスイノベーション型空間情報社会基盤
に関する研究開発

研究代表者： 東京大学特任准教授 関本 義秀
共同研究者： 東京大学特任教授 山田 晴利
東京大学協力研究員 中條 覚
東京大学特任研究員 南 佳孝
東京大学特任助教 薄井 智貴

平成 23 年 7 月

新道路技術会議

目次

概要（様式3の公表用資料）	1
第1章 はじめに	3
第2章 道路更新情報に関するニーズと情報提供の実態について	5
2.1 道路更新情報の定義	5
2.2 道路更新情報に対するニーズ	6
2.3 情報提供の実態について	7
第3章 道路更新情報の収集・提供システム構築	11
3.1 システムのスコープ	11
3.2 システムの全体構成と各機能の要件	11
3.3 路線ジオコーディングの手法	13
第4章 実証実験（1）：Webマイニング手法を用いた工事発注見通し情報の全国収集と提供	18
4.1 実験概要	18
4.2 工事発注見通し情報	18
4.3 データ処理の考え方	19
4.4 実験	23
4.5 まとめ	25
第5章 実証実験（2）：SVM手法を用いた工事入札公告情報による道路更新状況推定	27
5.1 道路更新の事前推定方法の検討	27
5.2 検討の手順について	28
5.3 基本となるモデルの構築	29
5.4 他自治体への適用可能性	30
5.5 道路管理者の協力による精度向上可能性	30
5.6 案件名称のみを用いたモデルの構築	32
5.7 データの蓄積による精度向上の可能性	32
5.8 大規模工事のみの抽出可能性	32
5.9 まとめ	32
第6章 実証実験（3）：道路工事図面の収集実験	34
6.1 実験概要	34
6.2 Webによる管理者側からのアップロード	34
6.3 情報公開請求を用いた方式	35
6.4 図面からのメタデータの抽出試行	35

6.5 国の枠組みを用いた全国化への展望.....	36
第7章 道路管理者の提供形態のあるべき姿.....	37
7.1. 道路管理者との協力体制構築可能性の検討.....	37
7.2. 道路更新情報流通実現へ向けた取り組みの基本方針.....	37
7.3. 管理者（国・地方公共団体等）への要望事項.....	37
7.4. 産学官連携による実現方策の提案.....	38
第8章 まとめ.....	39
参考文献.....	40
付録 本研究開発の体制.....	42

研究成果概要

平成20年度採択分
平成23年7月20日作成

研究課題名 サービスイノベーション型空間情報社会基盤に関する研究開発

研究代表者及び共同研究者

- ・研究代表者氏名 関本義秀（せきもとよしひで）
- ・共同研究者氏名 山田晴利（やまだはるとし）、中條覚（なかじょうさとる）、南佳孝（みなみよしたか）、薄井智貴（うすいともたか）、史云（しゆん）

所属研究機関・役職 東京大学空間情報科学研究センター・特任准教授

【研究の概要】

地方自治体を含めたデジタル道路地図の迅速かつ持続安定的な更新を実現するため、道路更新情報の自動収集・提供システムを開発し、様々なユーザに利用してもらった。

【キーワード】

デジタル道路地図、道路更新情報、自動収集技術、道路管理者、自律分散システム

（研究開始当初の背景・動機）

デジタル道路地図は1980年代にカーナビゲーションシステムの基盤データとして作成されたが、近年はGoogle Map等のWeb技術の進展、GPS機能を備えた携帯端末の普及によりデジタル地図をもとにしたビジネス戦略が大きく発展している。また低廉なPND（Personal Navigation Device）の普及も急速に進んでいる。一方、カーナビも現在2700万台普及しているものの、従来のカーナビの基本的な機能は殆ど変化しておらず、ユーザーからの苦情の第一は「利用されている地図が古い」ことである。

（研究の目的）

こうした情勢を踏まえて、地方自治体とも協調しつつ、デジタル道路地図の迅速な更新を契機とするサービスイノベーションにつながるフラッグシップ的な実験プロジェクトを立ち上げることを目的として、(1)道路地図コンテンツの流通活性化のためのニーズと課題の整理、(2)地図更新情報の自動収集技術の開発、のふたつを実施する。

（研究の方法）

上記(1)については、別途、産官学による

「道路更新情報流通推進研究会」を平成20～22年度まで開催し、道路更新情報に関する民間企業等のニーズをまとめるとともに、国・都道府県・市町村等の整備・提供の実態を調査し、課題等を整理した。

(2)は、道路更新情報を構成する工事発注見通し情報や入札公告情報に対して、自動収集・整理するシステムを構築し、実際に全国的に収集を行うとともに、125法人の参加があった「地理空間情報流通実験コンソーシアム」の中でこれらの収集データを実際に使ってもらった。

（研究の主な成果）

1. 道路更新情報の体系化
道路更新情報の体系化を行い、道路事業の過程で提供される様々な情報（工事発注見通し、入札公告、工事竣工図、開通情報等）の位置づけを明確にした。

2. 道路更新情報のニーズと提供実態整理
各情報のニーズや提供実態を整理した（図-1は提供タイミングをまとめたもの）。

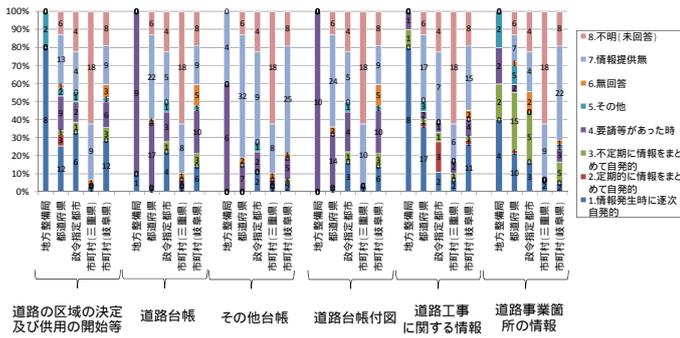


図 - 1 各道路情報の提供タイミング

3. 道路更新情報収集・提供システム構築

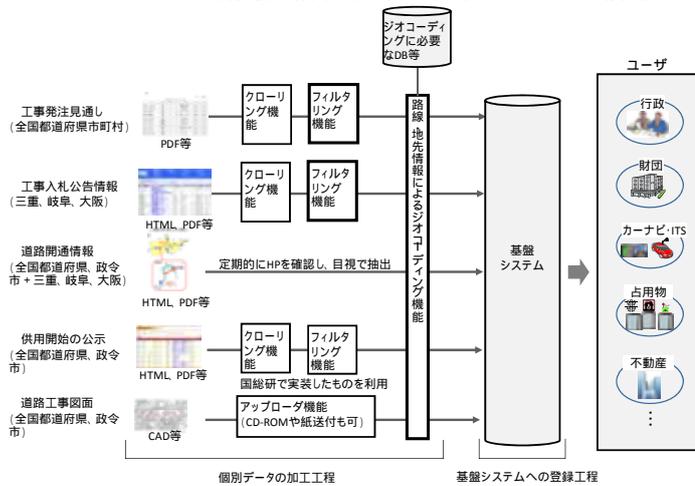


図 - 2 道路更新情報収集・提供の仕組み

4. 実際の収集

全国ベースで工事発注見通し情報は約 2.7 万件(図 - 3), 供用開始情報は約 1600 件, 道路開通情報は約 300 件, 工事入札公告情報は 3 県を対象に約 4700 件, 工事図面情報は 25 道府県で約 2000 件収集することができ, 上記コンソーシアム参加者からも利用価値のあるデータとの評価を得た。(主な発表論文)

- 1) 薄井, 関本, 金杉, 南他: 地理空間情報活用推進に向けた流通実験システムの開発と適用, 土木学会土木情報利用技術論文集, 第 19 巻, pp. 125-132, 2010.
- 2) 南, 関本, 中條, 柴崎: 路線情報を加味した道路関連情報の位置特定に関する研究~工事入札公告を事例に, 第 30 回交通工学研究発表会論文集, pp. 265-268, 2010.

- 3) 中條, 関本, 南, 柴崎: 道路更新情報に関するニーズと情報提供の実態について, 第 29 回交通工学研究発表会論文集, pp. 305-308, 2009.
- 4) Nakajo, Sekimoto, Minami, Yamada, et al.: Getting broad overview of road update from procurement notices of road constructions, Proceedings of the 15th world Congress on ITS, Stockholm, CD-ROM (6 pages), 2009.

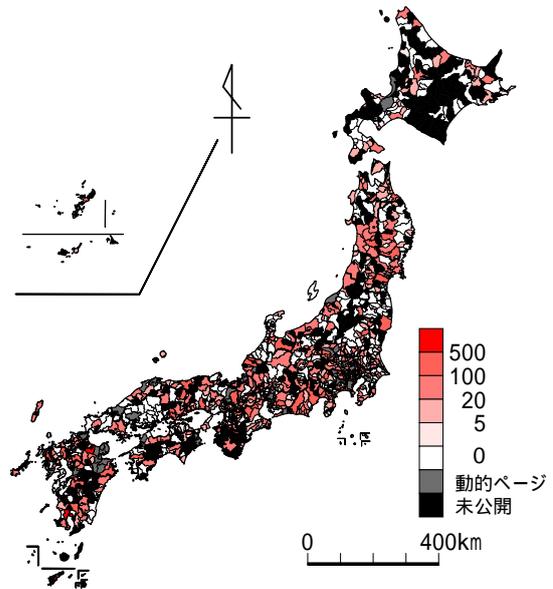


図 - 3 各自治体 HP から抽出できた道路更新に関わる工事発注見通し件数 (今後の展望)

今後, 継続的に収集・提供が実施できるよう, 自動収集のさらなる精度向上のための技術改良を進めるとともに, 情報提供の際の管理者サイドの記載の揺れ等を極力少なくするよう, 国と連携して標準化の推奨等を地方自治体に対して行う必要がある。(道路政策の質の向上への寄与)

本研究成果により, 国と地方自治体それぞれの情報提供の自律性を保ちつつも, ユーザーからは一元的に利用できるようなシステム連携が可能になり, 道路更新情報の迅速な提供が実現できる。

ホームページ等

<http://i.csis.u-tokyo.ac.jp/old/research/road.htm>

第1章 はじめに

人の快適な移動を支援するものとして、1990年代～2000年代前半のカーナビゲーション普及を第一世代とすると、近年、安全運転支援、電気自動車によるエコドライブ、電動自転車による登坂支援、PNDによるマンナビゲーションなど、モビリティの形も急速に多様になってきており、まさに第二世代に突入したと言えよう。

一方、それらの制御・判断を支える現地の状況を記録した地理空間情報、とくにデジタル道路地図の社会基盤としての重要性は、ますます高まっており、官民それぞれの主体で道路に関連したデータベースは充実し、高精度化・高鮮度化の流れは続くと思われ、ITSの切り口からは政策推進や提言^{1),2)}なども行われている。

しかし、これらが多様化するサービスに対する競争領域とすると、データ更新のための負担は各主体共通の悩みの種であり、道路構造の変化のような基本的な情報（詳細は2章で述べるがこのような変化情報を本研究では「道路更新情報」と呼ぶ）は、道路が公共事業を主体にした人口構造物である限り、事業主体である道路管理者に問い合わせが行くケースが多い³⁾。最終的には現地調査を行うことが多いのだが、現地調査の負荷も極力減らしたく、道路管理者に電話・メール・情報公開請求などを通じ、様々な問い合わせがみられるようになってきた。

道路管理者側も「道路開通情報ホームページ」⁴⁾や「道路図面情報提供サービス」⁵⁾のように徐々に情報をオープンにするケースが増えてきたが、まだまだ形式がばらばらであったり、あくまで予算を伴う事業説明を趣旨としているために、詳細に変化を伝えていくとはいえない。しかしその反面、行政で対応できる範囲も限界があり、地方自治体を含め、全国ベースで強制的に情報提供を行うことは現実的にも難しいため、なるべくITを活用し、官以外でも行えることは効率的に行い、サービスレベルが段階的に向上できるような仕組みが必要である。具体的には、行政側で予算や制度の制約があって提供ができない場合は、公開情報をベースに自動抽出技術を

組み合わせてある程度の品質でも道路更新情報を提供しつつ、道路管理者から未公開情報が公開された場合はすみやかに切り替えて高精度な情報を提供できるような、自律的な枠組みとそれを支えるシステムが必要である。

従って、本研究では、道路管理者側の情報提供の現状を明らかにするとともに、現状をふまえた上で、ある程度の品質以上の道路更新情報が得られるような収集システムを機能要件とともに提案し、いくつかの重要な機能についてはプロトタイプで実証を行う（図-1）。具体的には2章で道路更新情報の定義、体系化を行うとともに道路管理者から提供される情報の実態をまとめる。3章では、それらをもとに収集システムの機能要件等をまとめ、説明を行った。4章以降は実証実験としていくつかの種類のデータの収集等を行った。まず4章では、工事発注見通し情報を全国ベースでクローリングから道路更新情報に関する情報の抽出まで一連の自動収集を行った。次に、5章では工事入札公告情報を3県を対象に収集を行い、とくに一部で入手することができた道路更新に関連したかどうかの実績情報をベースに機械学習（SVM：Support Vector Machine）の手法を用いて、道路更新に該当するかどうかの推定を行った。また、6章では道路工事図面情報の収集作業そのものの効率化や図面のメタ情報作成作業の効率化など試みた。こうした実証実験をもとに、7章で道路管理者のあるべき情報提供の姿をまとめ、8章でまとめを行った（図-2）。

なお、関連研究として、筆者らを中心に道路更新情報のニーズや提供実態をまとめたもの⁶⁾、道路工事の電子納品により道路更新情報の蓄積を試みたもの⁷⁾、工事入札情報から道路更新情報の抽出を試みたもの⁸⁾、道路の公示情報をもとに道路更新情報の抽出を試みたもの⁹⁾などがあるが、本研究は、これらの様々なデータソースによる更新情報収集の試みを経て得られた知見をもとに、地域の実情に適したデータ提供により実現する自律発展的なシステムのあり方をまとめたものである。

これらについては、とくに東京大学空間情報科学研究センターで設置し、議論を行ってきた道路更新情報流通推進研究会（詳細は付録を参照のこと）の

成果の一端でもある。

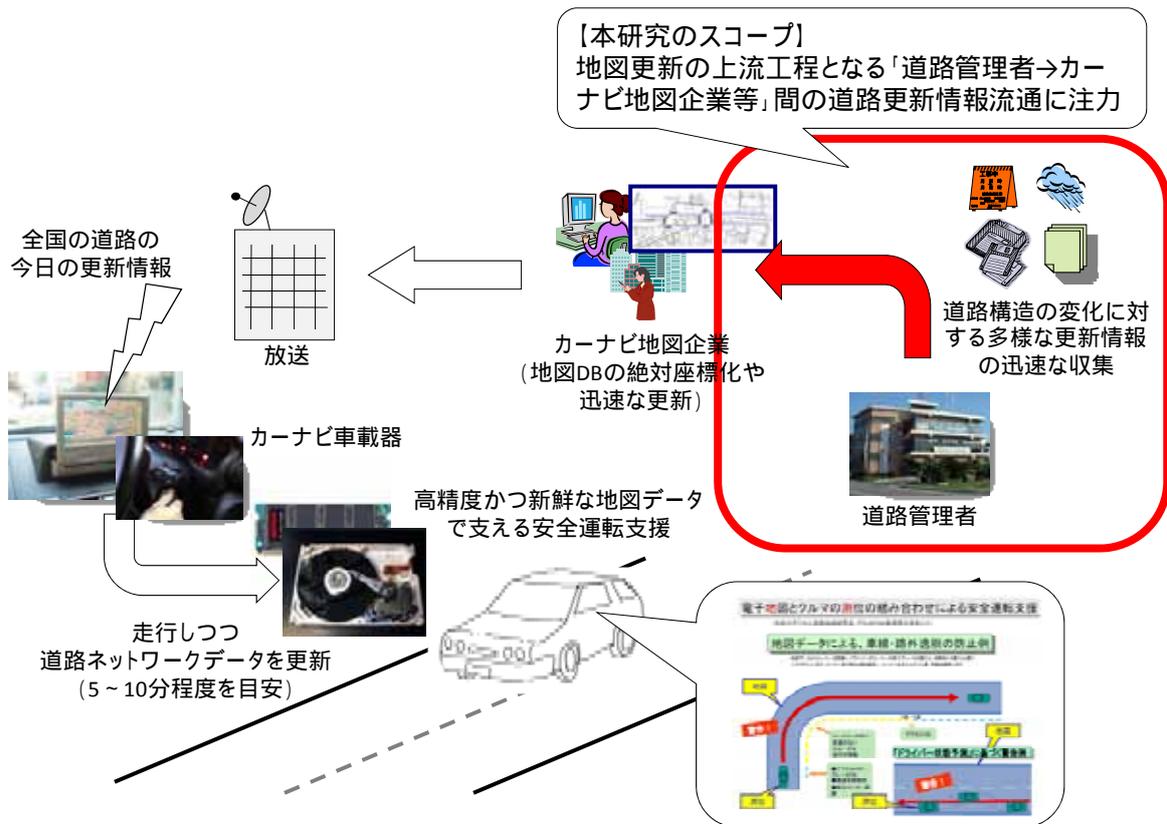


図-1 本研究の位置づけ

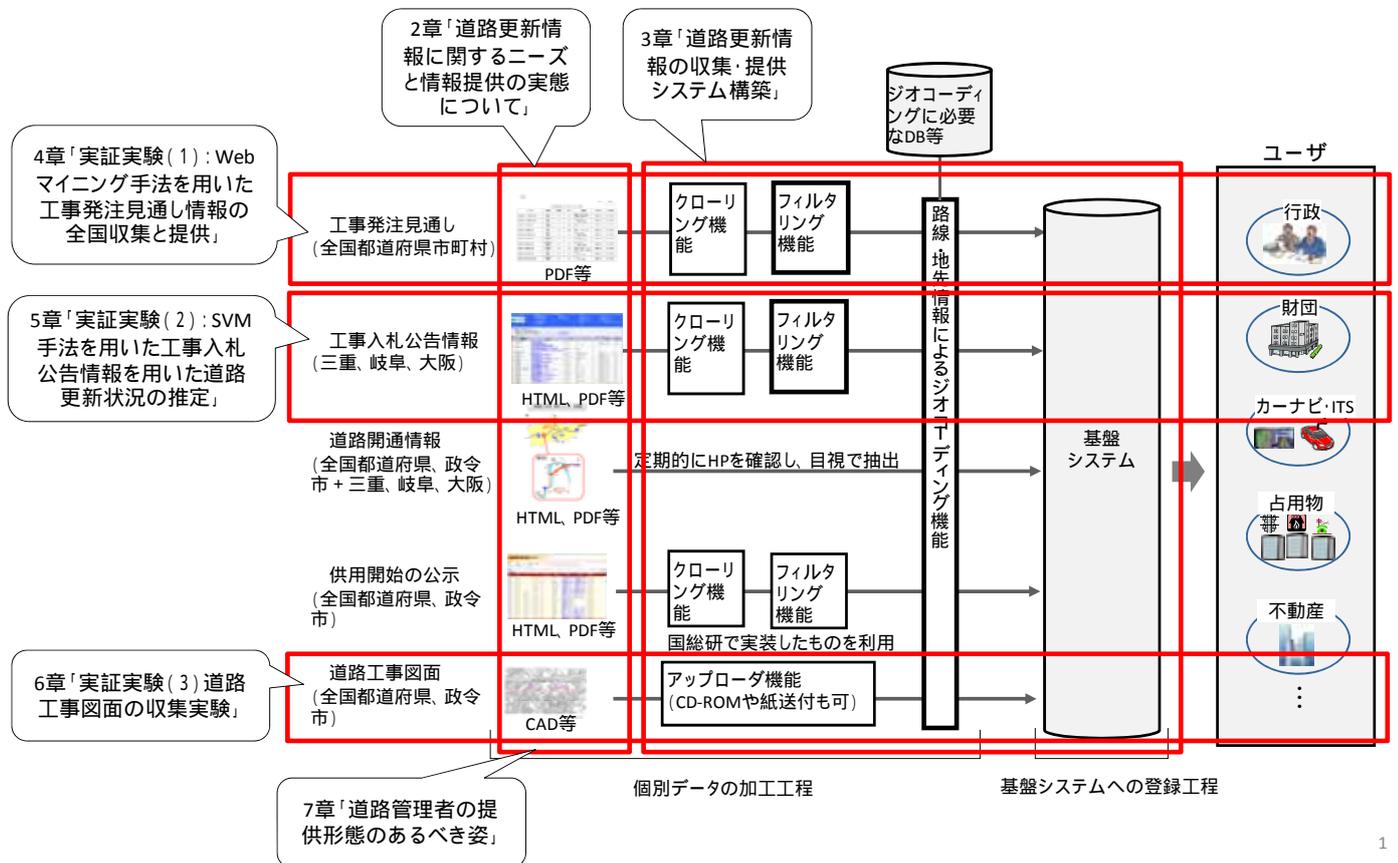


図-2 本研究の章構成

第2章 道路更新情報に関するニーズと情報提供の実態について

2.1 道路更新情報の定義

まずここでは、道路に関する情報が色々ある中で道路の変化・更新に関する情報がどのような位置づけに当たるか整理を行う。多くの道路情報は道路に関する位置・形状と何らかの属性情報から構成されるが、表-1は道路構造そのものの情報かそれに付随する情報かという属性情報の内容の軸と、簡単な位置情報（キロポストや地先名など）か形状を伴うものかという位置・形状の表現レベルの2つの軸で整理したものである。

それらには当然、道路台帳やデジタル道路地図データなどのように、まとまった単位で網羅性を持つような全体情報としてのデータベース（以降、データベースと呼ぶ）もあるし、その一方で、道路開通情報や入札情報のような断片的な個別の変化を伝える情報もある。我々は、1章で述べてきたように様々なデータベースの更新を活性化する意味で、これらのうち、「断片的な個別の変化を伝える情報」を「道路更新情報（road update information）」と呼ぶことと

する。また、本研究の中では、とくに基盤となるような道路構造に関する更新情報を主体に論ずることとする。

道路構造の更新情報の提供を考察する際、中心となるのは道路工事であり、工事の種類によって提供内容やタイミングは変わってくる。これらを概観したのが図-3であるが、新規路線・バイパス建設のような大規模な工事は、数が少ないものの、事業規模の面からも説明責任度合いが高いため、計画公表から調査・用地取得・設計などについても情報提供が行われ、供用開始の公示や台帳の新規作成についても確実に行われると言えよう。

一方で、災害復旧、維持修繕あるいは付帯設備単独の工事については、現状に対する軽微な変更であるためあまり大々的な情報提供はなく、工事入札・供用に至り、台帳としては軽微な修正となるため、実際の情報更新の対応は遅くなることも多いようである。

表-1 道路更新情報の概観

		簡単な位置情報を伴うもの	形状を伴うもの	
			粗いレベル（精度レベル25000以下）	細かいレベル（精度レベル500～2500程度）
道路構造情報	まとまった単位で網羅性を持つ全体情報（データベース）	道路台帳（道路法 ¹⁰ 第28条に基づく）	デジタル道路地図（DRM）データ ¹¹	道路台帳付図（道路法・施行規則第4条の2に基づく）、道路管理データ ¹²
	断片的な個別の変化情報（道路更新情報）	工事発注見直し情報（公共工事入札契約適正化法 ¹³ 第4～9条に基づく） 工事入札公告情報（公共工事入札契約適正化法 ¹³ 第4～9条に基づく） 供用開始の公示情報（道路法第18条に基づく）	工事発注図 工事完成図（土木工事共通仕様書 ¹⁴ 等に基づく） 道路開通情報 ¹⁵	現地計測
付随する情報	まとまった単位で網羅性を持つ全体情報（データベース）	道路交通センサス ¹⁶ 、デジタル交通規制データ ¹⁷ 、交通事故統計データ ¹⁸ 、道路の走りやすさマップ ¹⁹	-	-
	断片的な個別の変化情報（道路更新情報）	VICS ²⁰ 、プローブ情報	-	-

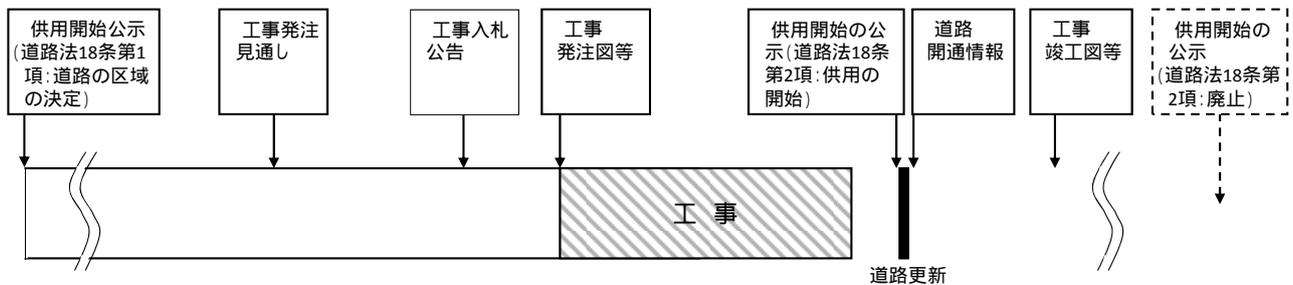


図-3 道路工事に関連して道路更新情報が発生するタイミング

また、電気・ガス・通信設備のような地下埋設物を扱う道路管理者以外の占用工事については、関係者間の利用調整会議を通じて、情報共有が行われるもののそれらの情報が外部に出ることはあまり多くなく、最終的に道路管理データに反映される程度である（図-4）。

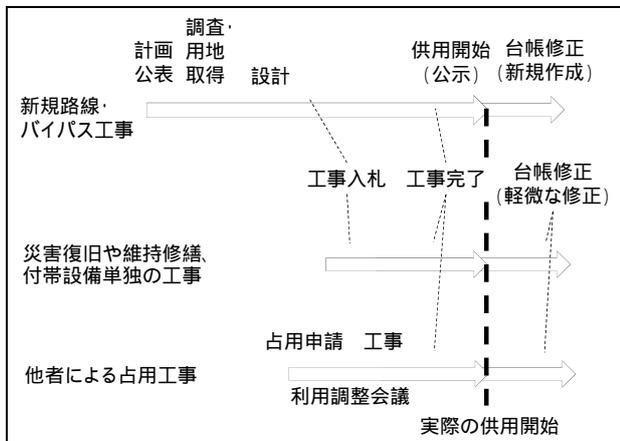


図-4 工種に応じた道路更新情報提供のタイミング

2.2 道路更新情報に対するニーズ

本節では、ニーズの広がりについて、地図調整業者へのヒアリングおよび「道路更新情報流通推進研究会（概要は付録を参照のこと）」における議論を通じ、利用者側および管理者側それぞれの潜在的なニーズの広がりに関する規模の推計を行った。

利用者側については、現状では、カーナビへの活用へ向けた地図調整業者等による国（直轄国道）および都道府県（計48主体）への問い合わせが顕在化している。問い合わせを行っている地図調整業者等は一業者あたり年間約1,000件の図面等を管理者から入手している。収集にかかるコストの関係から、この件数は業者が本来、国および都道府県から入手したい件数の約1/3となっている。

また、利用者側から見ると、1,776の市町村に対する問い合わせは現状では行えておらず、約37倍の主体それぞれから約3倍の図面等の入手を行うことによる、現状の約100倍のニーズが潜在的に存在すると考えられる。

一方の管理者側では、地図調整業者等約5社から年間それぞれ約20件の問い合わせに対応を行っているのが現状である。しかし、道路更新情報に興味を持つ民間はカーナビ（地図調整業者）に限らず、沿道沿いの小売業（駐車場、GS、コンビニ、ファミレス

等）や流通業など様々に渡ると考えられる。5業種の大手10社それぞれにニーズが存在すると考えれば、主体数は約10倍となり、それぞれが現状の約3倍の問い合わせを行うとすると、一主体で必要となる対応は約30倍となる。さらに、これらの対応を市町村も含めて行うとすると、主体数は約37倍となることから、管理者側での延べでの対応負担は現状の約1,100倍になると考えられる。

表-2 道路更新情報の潜在的なニーズ

	管理者	利用者
現状	・約20件/年×5社程度に対応【都道府県47団体】	・都道府県から1000件/年の図面情報等を入手【カーナビ(地図調整業等)5社程度】
広がり(潜在的ニーズ)	・約60件/年×50社程度に対応【都道府県+市町村 約1800団体】 1団体あたりの対応規模約30倍, 対応団体数約37倍	・都道府県+市町村から3000件/年の図面情報等を入手【駐車場、GS、コンビニ、スーパー、ファミレス、流通業等、50社程度】 1団体あたりの対応規模約100倍、団体数約10倍

こうした試算に加え、「道路更新情報流通推進研究会」で議論した内容をもとに、これらを利用する「地理空間情報流通実験コンソーシアム（概要は付録を参照のこと）」の会員に対し、任意回答で道路更新情報の価値に関するアンケートを行ったところ、有効回答60のうち、58（97%）が「道路更新情報が流通することは良いことだと思う」と回答している。さらに、有効回答37のうち、28（76%）が支払う価値があるとも答えている（図-5）。

また、個別の情報についても利用価値を尋ねたところ、どれも6～8割程度で利用価値があると回答している。とくに利用価値が高い電子納品の情報は図面等豊富なデータを持っていることを意味し、道路開通情報は開通に関するピンポイントな情報を表していることからユーザー層にとっては利便性が高いという認識があることがわかる。ただし、実態としては電子納品は一般的には未公開であることが多いことや道路開通は大規模な道路を対象としていることなどが一般的であり、必ずしも入手のやすさや網羅性とは一致していない。むしろ入札公告の方が公表状況も高く、件数も多い点から、こうした情報も有効活用し、出所等の品質を明記した上で道路更新情報として提供することでニーズに合わせていくことができる必要性がある。

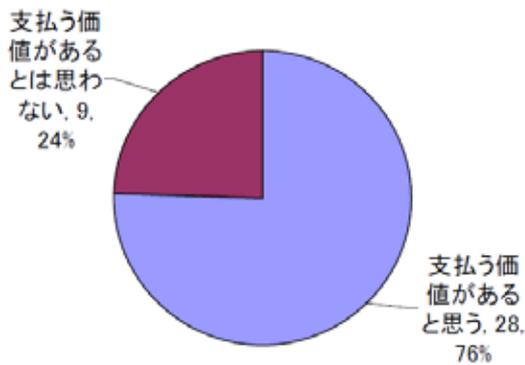


図-5 道路更新情報に関する支払い意思

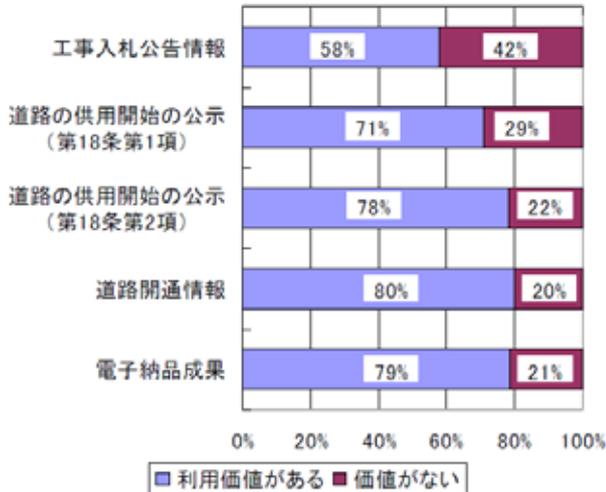


図-6 様々な道路更新情報の利用価値の認識

2.3 情報提供の実態について

(1) 道路管理者の提供認識

まずは、道路管理者自身はどの程度情報提供の認識があるかについて、アンケートにより調査を行った。具体的には、平成21年3～4月の間に道路更新情

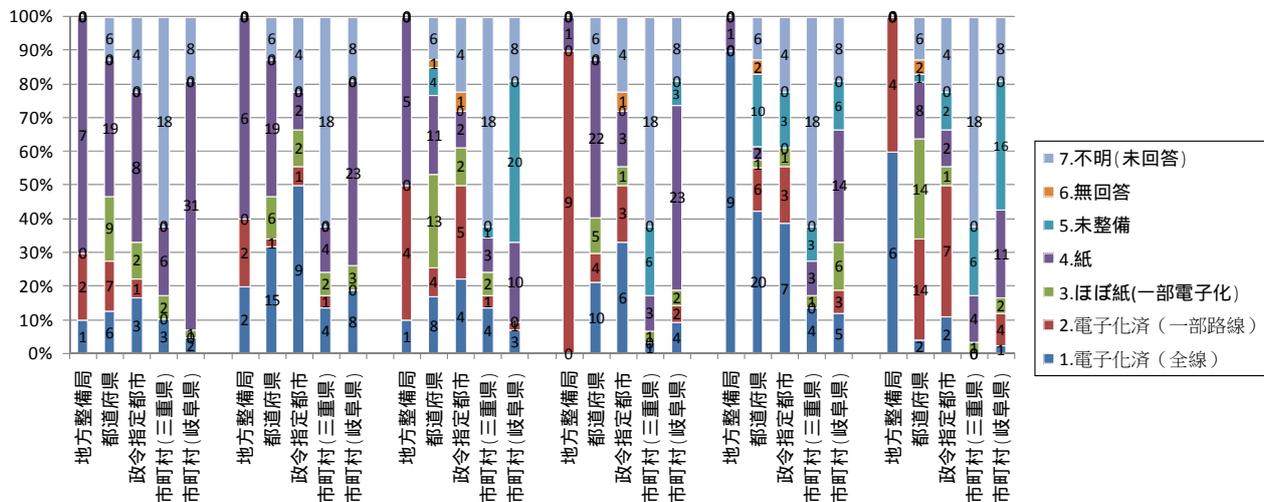


図-7 各道路情報の整備状況(道路管理者による回答)

報流通推進研究会の委員である国土交通省道路局を通じ地方整備局、都道府県、政令指定都市に対して、岐阜県・三重県を通じ市町村に対してアンケートの送付・回収を行い、データベースそのものである「道路台帳」「その他台帳」「道路台帳付図」と、更新情報である「道路の区域の決定及び供用の開始等」「道路工事(入札)に関する情報」「道路事業箇所の情報」に関して、それぞれの整備状況(図-7)、提供状況(図-8)、提供タイミング(図-9)をまとめた。

これらからわかることは、やはりデータベースそのものよりは更新情報の方が、整備・提供の傾向は一般的に高いことが読み取れる。また、その中で整備と提供の傾向の違いとしては、データベースはとくに、紙・電子含め整備はしているものの、提供している割合はかなり低い状況であった。さらに道路管理者ごとと比較を行うと、大まかに言えば、国>都道府県 政令指定都市>市町村(都道府県によるばらつきあり)と言えそうである。

次に、道路更新情報である「工事入札公告情報」「道路開通情報」「公示情報」について、Webサイトの確認を行い、実際の提供状況の調査を行った(図-10～図-13)。具体的には、平成21年3～4月に都道府県、政令指定都市、三重県・岐阜県の市町村のWebサイトを調査して確認するとともに、公開されているコンテンツのレベルをまとめた。

(2) 工事入札公告情報

図-10の「工事入札公告情報」では、都道府県・政令指定都市までは100%でPPIデータベースから公開を行っており、市町村では半数程度である。また、図面や工事内訳書の添付の有無は大雑把に言えば公開しているうちの半数程度と言える。また、工事案件名における住所記載レベルは、かなりばらつきがあるが、大字・町名までが一番多いようである。

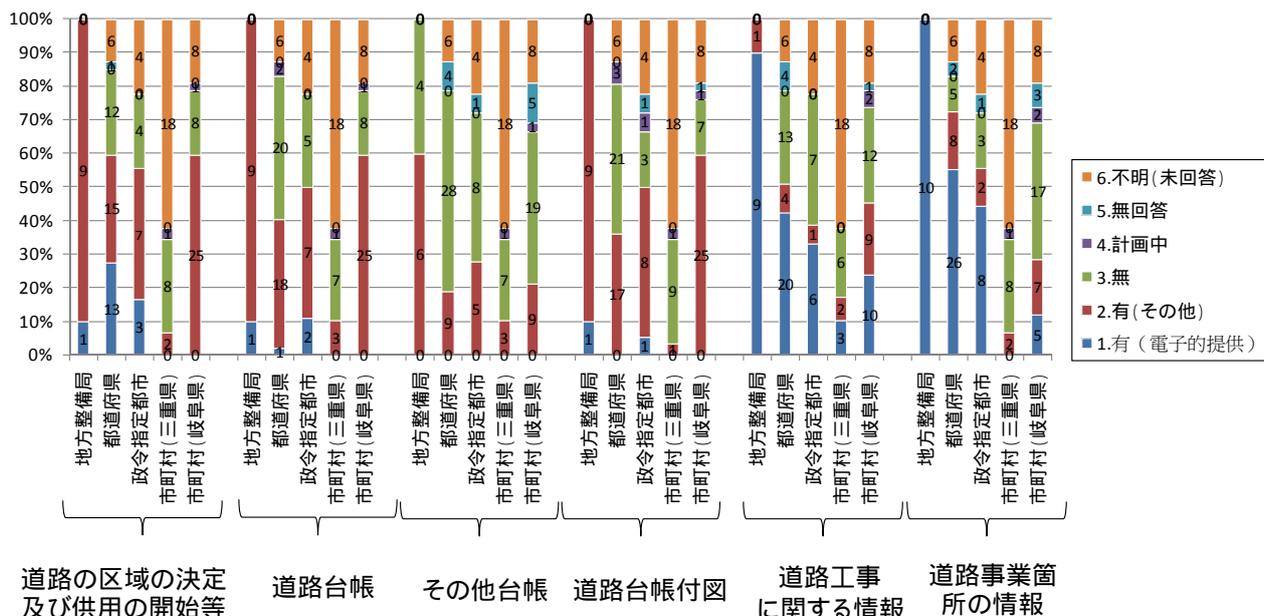


図-8 各道路情報の提供状況(道路管理者による回答)

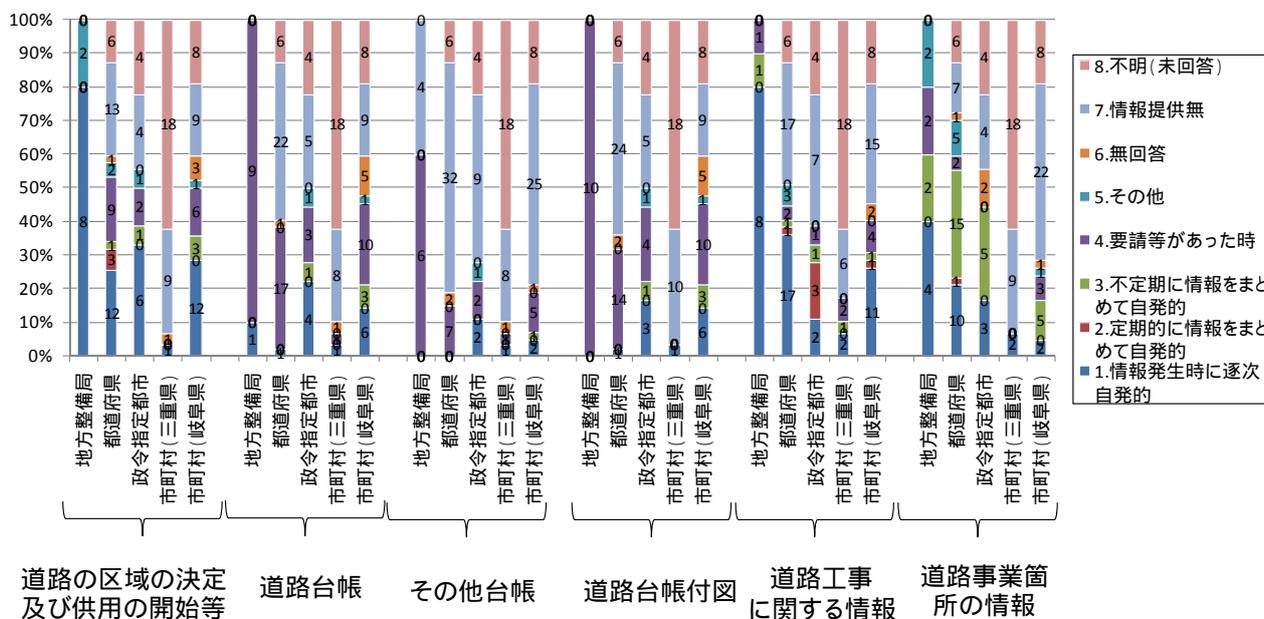


図-9 各道路情報の提供タイミング(道路管理者による回答)

(3) 道路開通情報

の「道路開通情報」では、都道府県は8割程度が提供を行っているが、政令指定都市となると4割弱、市町村となると5%弱と一気に少なくなってしまう。これはそもそも新規路線・バイパス工事の主体が、市ベースではあまり存在しないことが原因だと思われる。また、路線名については100%記載をしており、町名・地先等や開通日などについても、提供している中では8割程度の確率で掲載している。

(4) 公示情報

図-12の「公示情報」は、布施らの論文によるデータであるが、都道府県・政令指定都市では概ね公開

しているが、都道府県では、週一度、期日当日に情報を掲載している状況で、揃っているようである。一方で、政令指定都市では、公開頻度・掲載時期ともかなり幅があるようである。

(5) 工事発注見通し情報

図-13は工事発注見通しの状況である。内容は表形式で表現されておりシンプルであるため、7割程度が公開している。また、公開している中ではPDFが8割程度を占め、残りはEXCEL, HTMLが多いが、一方で、件数が多いことから動的ページを構成しているケースも1割強存在している。

(6) まとめ

これらを情報提供状況と記載状況，道路更新に近い情報かどうかなどの観点から総合評価をすると，表-3のようになる．こうした状況を踏まえ，4～6章

の実証実験では工事発注見通し，工事入札公告，公示図面等を用いて行うこととする．

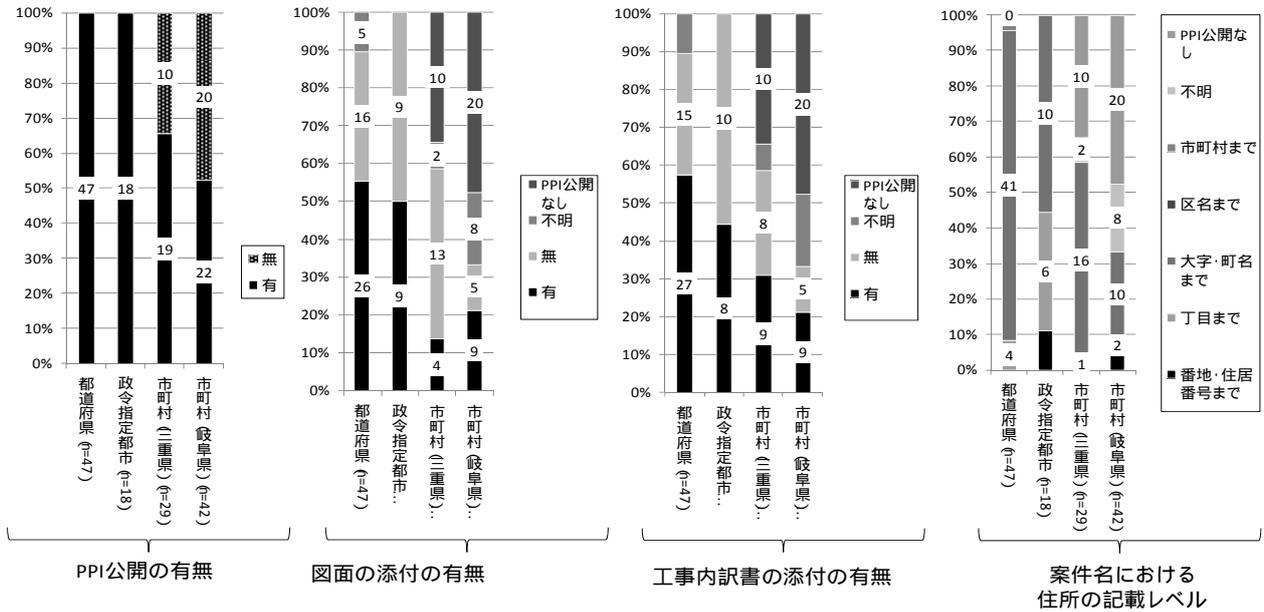


図-10 道路工事入札情報の内容 (HP を調査)

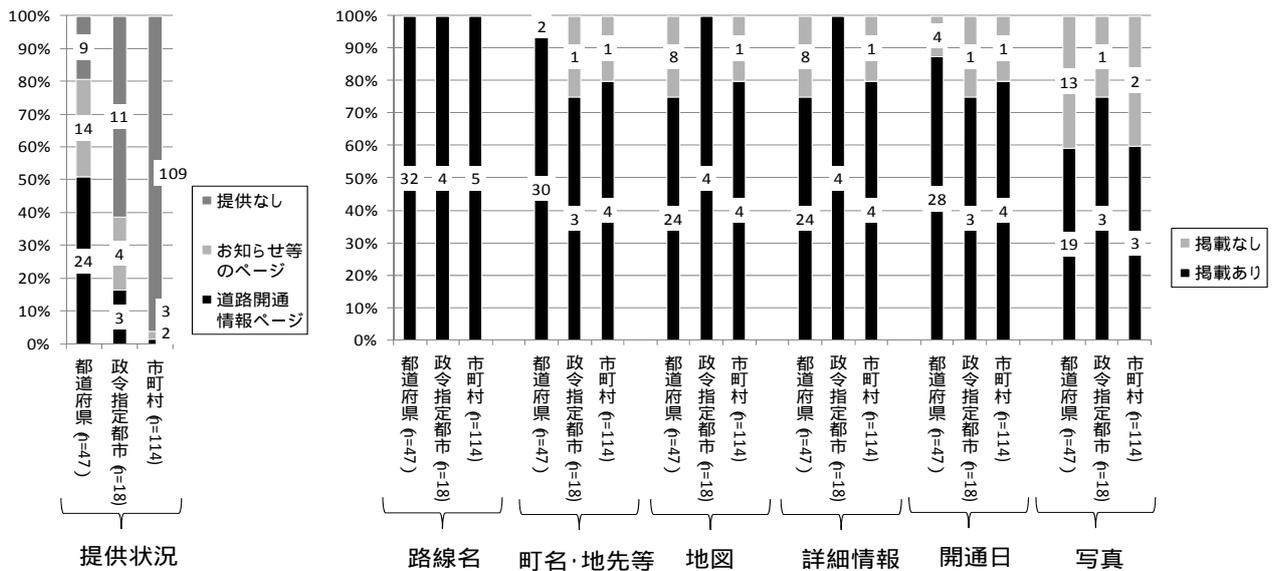


図-11 道路開通情報の内容 (HP を調査)

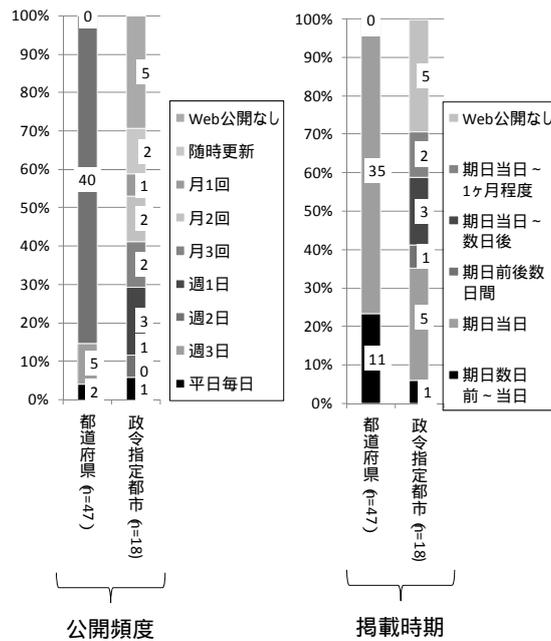
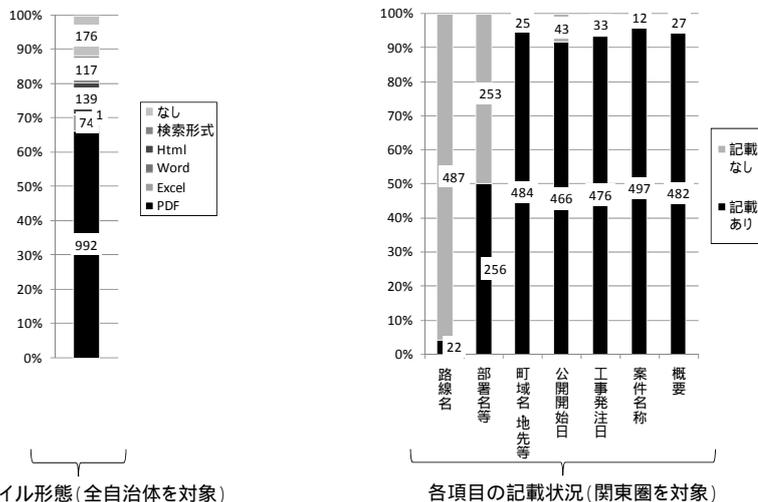


図-12 公示情報の内容 (HPを調査)



提供ファイル形態(全自治体を対象)
一部の自治体では複数の部課室から公開

各項目の記載状況(関東圏を対象)

図-13 発注見通し情報の内容 (HPを調査：関東)

表-3 道路更新情報のいくつかの視点での評価

	網羅性・公開状況	更新情報の抽出しやすさ	道路更新情報としての価値
供用開始の公示 (図-12より)	(都道府県は確実に公開, 政令市は2/3程度公開, 市町村はあまりない)	(公開の形式が複雑な表形式のケースもある)	(道路計画のかなり初期段階なので, その後の予定がなかなか見えにくい)
工事発注見通し情報 (図-13より)	(都道府県, 政令市までは確実に公開, 一般市町村でも2/3程度は公開)	(表形式であることがほとんどのため抽出しやすい. ただし用語のブレがあるため辞書化必要)	(発注見通しであるため, 入札が行われない等, 変更になるケースがある)
工事入札公告情報 (図-10より)	(都道府県, 政令市までは確実に公開, 一般市町村ではそれほど多くない. 一部, 民間のクロージングサービスもあり)	(公開の形式が複雑な表形式のケースもある)	(一つの道路更新箇所が複数の工事に分かれていることもある. 入札段階なので, 開通時期がはっきりはわからない)
道路開通情報 (図-11より)	(ほとんどが国や都道府県管理の道路以上のみ. その範囲は国が「道路開通情報」サイトを設置している)	(記者発表のPDFレベルであることが多く, 様式は多様. 先サイトを使えば難しくない)	(道路の新規開通なので, 道路更新に関わる案件の一部であり件数が多いが, わかりやすいためニーズは大きい. ただし公開情報が主なので, すでに流通している)
道路工事図面情報 (図-8より)	(都道府県・政令市は1/3程度は公開, 市町村はほとんどない)	(図面からの更新情報の抽出は一定の画像認識等の処理が必要)	(完成図面等の入手後となるので, タイミングは遅いが, 詳細な形状を記載している価値は高い)

第3章 道路更新情報の収集・提供システム構築

本章では、2章での道路管理者側からの情報提供の実態に基づき、自律的に精度の向上が望まれるような収集システムについて、システムのスコープ、各機能の要件を明らかにしつつ、全体の設計を示す。

3.1 システムのスコープ

(1) 蓄積するデータ構成

まず蓄積する道路更新情報の共通的なメタデータは、収集対象となるデータの「カテゴリー」「タイトル」「登録年月日」「著作者」「登録者」「位置情報（緯度経度）」などの基本的な情報と、関連する情報をファイルとし、また、位置情報と表裏一体的に「行政区域（市町村まで）」を蓄積することとする。

(2) 収集対象とするデータ

(1)のデータを蓄積するためにまずは、2章の図-3で整理してきたような、断片的な個別の変化の直接の原因である道路工事があったことを押さえられる情報を対象とする。また、位置情報についても、簡単な位置情報を伴うものとある程度詳細な形状を伴うものに大別したが、今まで見てきたように、簡単な位置情報を確実に取得できることを最優先とし、位置情報を推定できるような情報を対象とする。実際に本研究で収集したものを整理したのは表-4の通りである。

(3) 利用イメージ

また、道路更新情報の利用イメージとしては、欲しい箇所の欲しい種類のデータが円滑に検索・ダウンロードできるような基本的な機能に注力し、それ以外の機能は本論文では検討対象外とする

(4) 中心となる機能

本システムでは、今まで述べてきたように道路管理者が持つ高精度な道路更新情報を提供できるようになった段階で簡易に登録・アップロードできるような機能と、それに至るまでの間にある程度の誤差を許容しつつも関連する公開情報から道路更新情報

を抽出する機能がまず必要である。次にそれらの道路更新情報のある程度の精度で道路上の位置に特定する必要がある。

3.2 システムの全体構成と各機能の要件

前節の議論に基づきシステムの全体構成を図-14のようにまとめた。例えば、工事入札情報や供用開始の公示情報のような、道路管理者がHP上で構造化されずに公開しているような情報（PDFやhtmlなど）は、システム側で収集・構造化の必要がある。場合によっては、道路開通情報のようなかなり多様な提供形態の場合は、定期的にHPを確認し、目視で抽出するような手作業の可能性もある。

一方で、最近HP上のニュース・お知らせの中で、RSSを用いてある程度構造化した上で、情報提供する場合は、そうしたものは不要であり、路線ジオコーディングしさえすればよい。

また、図面情報のような未公開情報の場合は、公開というよりは、アップロードの機能となる。それらを用いて、最後は道路上の位置に落とし、データベースに落とす。

また、これまでの議論をもとに各機能の要件を表-5にまとめた。クローリング機能は道路更新情報を含むであろう対象ファイルを自動的に取得する機能であり、フィルタリング機能は取得した各ファイルから道路更新に関するメタ情報を文字形式で取得し、とくに画像を経由する場合はOCR機能等を用いて、最終的にはデータ辞書などと照合し類似度の高いものを抽出する。アップロード機能はクローリングできない未公開情報等を同意の上で提供してもらう形で最近では、CMS（Contents Management System）のようなシステムも出ており、近い趣旨と言える。さらに路線ジオコーディングの機能としては、道路に関連する事象の場所を地先名等、多様な形で受け入れつつも、路線上の位置（緯度経度）で特定できるように近似的に変換する機能を有することでかなり場所を具体的に提示することができる。とくに、クローリング、アップロード、蓄積、提供機能につい

ては、既存の様々なソフトウェアの利用で問題ないと思われるが、フィルタリング、路線ジオコーディ

ング機能については、道路固有の知識を踏まえたアルゴリズムを構築する必要がある。

表-4 本研究で収集した道路更新情報

ID	カテゴリー	データ作成者(収集範囲のみ記載)	データ収集者/管理者	データ対象時期	提供範囲	提供形式	データ数	その他共通で持つデータ項目
1	工事入札公告情報	都道府県,市町村	東京大学空間情報科学センター(公開情報を収集)	2009~2010年度分	三重県,岐阜県,大阪府(HPで公開している府県下の市町村含む)	TXT,PDF等	4,713	路線名,地先名,所属自治体名等(DB登録時には)緯度・経度
2	道路開通情報	都道府県,市町村	同上	同上	同上	同上	312	
3	供用開始公示情報	都道府県,政令指定都市	国土技術政策総合研究所(公開情報を収集)	2009年2~3月の発行ペース	全範囲	CSV	1,622	
4	工事発注見通し情報	都道府県,市区町村	東京大学空間情報科学センター(公開情報を収集)	2010年度分	公開している全範囲	CSV	約24,000	
5	道路工事図面情報	国土交通省,都道府県,市町村	東京大学空間情報科学センター(未公開情報を収集)	2008年,2009年度分	全地方整備局と25都道府県の県と政令指定都市分,岐阜県,三重県の一部市町村	CSV,CAD,PDF,XML	約2,000	

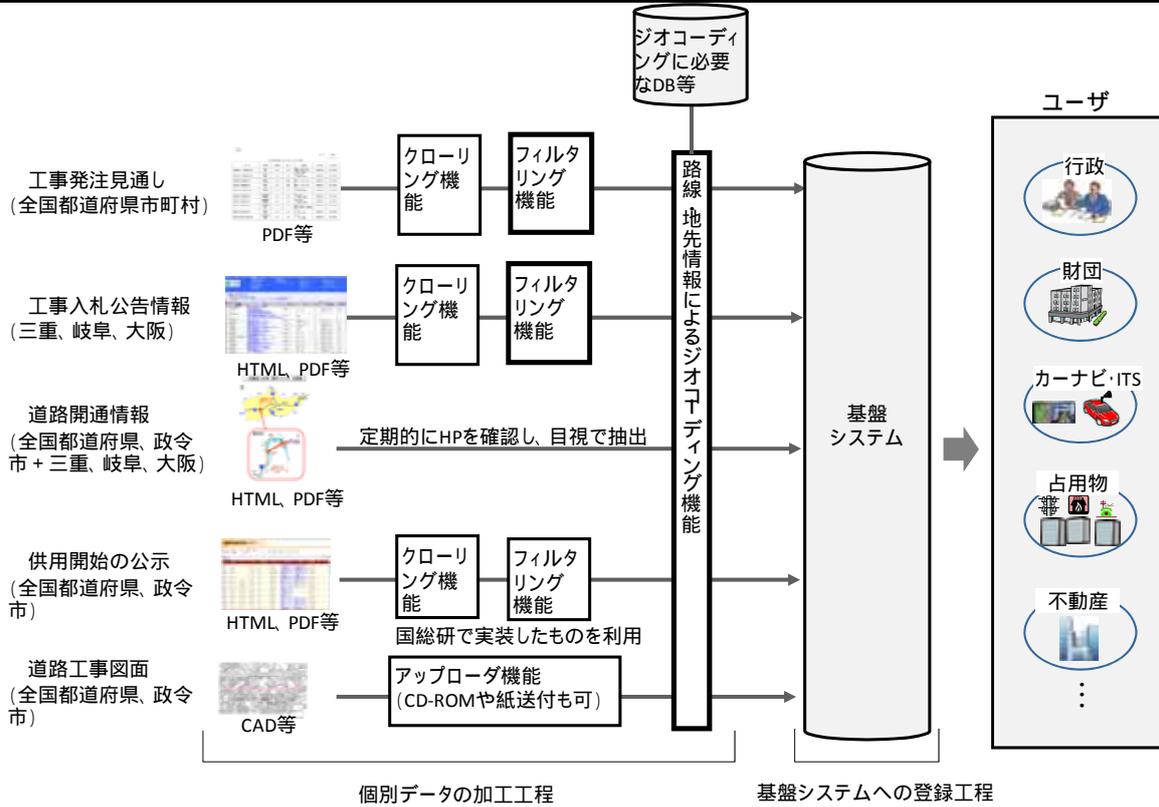


図-14 全体システムの構成

表-5 各機能の要件

	内容
クローリング	工事入札情報・道路開通情報・供用開始の公示等,道路事業に関する公開サイトを対象としクローリングを行いつつ,道路更新情報を含む対象ファイルを取得する機能を持つこと.
フィルタリング機能	取得した各ファイルから道路更新に関連するメタ情報を文字形式で取得し,データ辞書等に併せて類似度判定等を行い抽出すること.

アップロード機能	未公開の図面情報については,50MByte程度までのCADデータやPDFデータなどが現実的な時間(1~2分以内)でアップロードでき,その際に位置情報もアップロードできること.
路線ジオコーディング機能	道路に関連する事象の場所を地先名等,多様な形で受け入れつつも,路線上の位置(緯度経度)で特定できるように近似的に変換する機能を有すること.

3.3 路線ジオコーディングの手法

(1) 考え方

路線ジオコーディングでは、道路関連情報が示す道路上の位置を特定するために、図-15に示すように、次の3つのステップを順に処理する。まず、地先名を緯度経度に変換し、その点を「基点」とする。次に、路線名に該当する路線を特定する。そして、基点から最も近い路線上の点を算出し、その点を「補正点」とする。

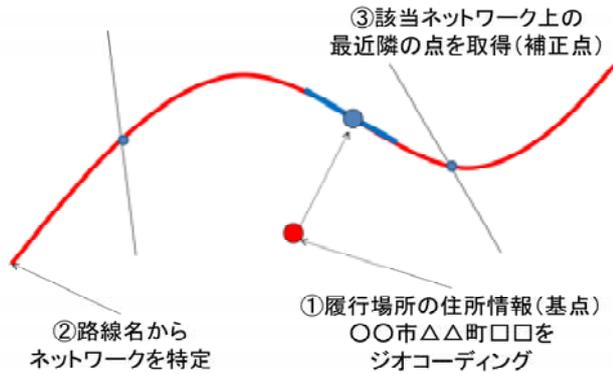


図-15 位置特定の手法

路線ジオコーディングでは、汎用的に多様な道路関連情報を対象とするために、特定の形式を対象とせず、道路関連情報に含まれている地先名を緯度経度に変換する。地先名を緯度経度に変換するジオコーディングについては、東京大学空間情報科学研究センターのCSVアドレスマッチングサービス²¹⁾を利用する。このサービスでは、iConfとiLvlの2つのコードが出力される。iConfは、変換の信頼度を示すコードで、変換誤りが無いかどうかの指標となる。iLvlは、変換された住所階層レベルを表すコードである。

路線の特定には、財団法人日本デジタル道路地図協会のDigital Road Map (DRM) 2003を用いる。DRMにおける道路網は、交差点その他道路網表現上の結節点などを示すノードと、ノード間の道路区間を示すリンクで構成されている。このリンクに路線番号のデータが含まれており、このデータと道路関連情報に含まれる路線名とを照合することによって、路線を特定する。路線名と路線番号の対応については、DRMのrmt形式のファイルに1対1で記録されている。また、路線名が記録されているデータは、市町村道以下はほとんど無く、都道府県道や主要地方道以上

であった。

路線への補正には、基点と路線を用いる。具体的には、次のような処理を行う。まず、該当する路線上の任意の隣り合う2点を端点とする線分に、基点から垂線を下ろす。それら垂線のうち、最短の物を選択し、その垂線の足を補正点とする。路線上の点は、DRMのノードと、ノード間の補間点を用いる。補間点とは、DRMでノード間の道路形状を示すために補われる点のことである。

(2) 路線ジオコーディングの評価方法

路線ジオコーディングの距離精度の算出には、実際に工事があった場所と、その工事に対応するデータを用いる。このデータに路線ジオコーディングを適用し、実際に工事があった場所との距離を算出し、評価する。この距離精度については、住所階層レベルによって、その精度が異なると考えられる。市区町村レベルの場合、全国の市区町村の平均面積は約200km²であることから、その平方根を取り、14km程度の誤差が発生すると想定する。字レベルの場合、字レベルの平均面積は約1.5km²であることから、同様に平方根を取ると、1.2km程度の誤差が発生すると想定する。この住所階層レベルは、対象となる情報に含まれている場所を示す表現が指す領域と対応する。したがって、誤差が、場所を示す表現が市区町村レベルの場合は14km、字レベルの場合は1.2km程度であれば、路線ジオコーディングの処理として期待される正確度を満足すると考える。

現状でどの程度の割合のデータが位置を特定できるかを把握するために、表-6に示す道路関連情報に対して、路線ジオコーディングを適用する。路線ジオコーディングの適用結果には、図-16に示す位置精度指標を用いて評価の基準とする。位置精度指標を定義する基準として、都道府県レベル、市区町村レベル、字レベルの座標が得られたかどうかの基準と、路線を特定して道路上に補正できたかどうかの基準を用いた。位置精度指標を評価基準とすることによって、道路関連情報の位置を特定できた割合を把握し、路線ジオコーディングと各道路関連情報を評価する。位置を特定できなかったデータについては、その原因を検証し、路線ジオコーディングと道路関連情報の改善点を検討する。

表-6 道路関連情報の位置情報と路線情報

情報の種類	位置情報	路線情報
発注見直し情報	石井地内	市道111号線道路改良事業
工事入札公告情報	和泉市父鬼町	一般国道 480号 道路改良工事(第2トンネル)
道路工事図面	本巢市根尾平野～本巢市根尾日当	871203線
供用開始の公示情報	「区間始点」海部郡美波町西の地字山神78番3地先「区間終点」海部郡美波町西の地字山神84番1	日和佐小野線
道路開通情報	「都道府県名」茨城県「市区町村名」小美玉市「町域名地先等」与沢～野田地内	百里飛行場線

「」は情報項目名

位置精度指標の定義

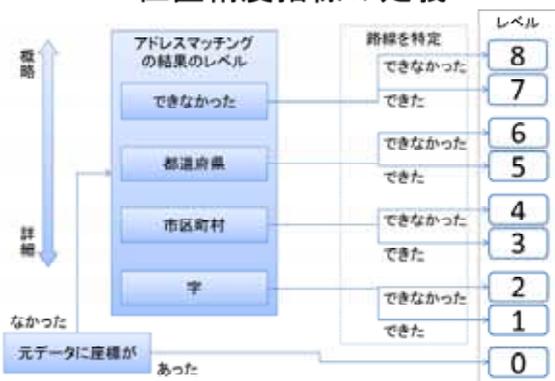


図-16 位置精度指標

(3) 実験に用いる道路関連情報

本研究で利用する道路関連情報は、2章で説明したように、発注見通し情報、工事入札公告、道路工事図面、道路の供用開始の公示、道路開通情報、台帳修正箇所など、道路工事に関連する情報とした。

発注見通し情報は、工事発注見通し情報を用いた全国における道路工事関連箇所の自動抽出手法構築²⁵⁾において入手した情報のうち、道路更新関係の工事として抽出できた23,836件のデータを対象とした。発注見通し情報の内容を分析してみると、位置情報については、「石井地内」など、住所で記載されていることがわかった。路線情報については、「市道111号線道路改良事業」など、案件名に含まれている場合があることがわかった。

工事入札公告情報は、本研究では、2009年12月1日から2010年3月4日までの間にWeb上で公開されている入札情報を大阪府と岐阜県、三重県から収集できた2,946件のデータを対象とする。工事入札公告情報の内容を分析してみると、位置情報については、「関市稲口地内」など、住所で記載されていることがわかった。住所の記載については、区間を示す2つの住所が含まれている場合もあった。路線情報については、「一般国道480号道路改良工事」など、案件名に含まれている場合があることがわかった。

道路工事図面は、岐阜県と岐阜県の7市町村、三重県、三重県の8市町村から、平成20年度から平成21年12月までに開始している工事について、道路工事図面を貸与いただいたデータを対象とする。件数は、岐阜県が19件、岐阜県の市町村が、85件、三重県が99件、三重県の市町村が23件であった。道路工事図面の内容を分析してみると、位置情報については、住所で記載されていることがわかった。住所の記載については、区間を示す2つの住所が含まれている場合があることや図面に記載されている場合があることがわかった。路線情報については、案件名や図面に含まれている場合があることがわかった。

道路の供用開始の公示情報は、国土技術政策総合研究所が収集した⁹⁾データを貸与いただいた。このデータは、2009年3月1日から2009年3月17日までの間に収集されたデータで、1623件あった。道路の供用開

始の公示情報の内容を分析してみると、位置情報については、区間始点と区間終点という2つの情報項目に、住所で記載されていることがわかった。路線情報については、路線という情報項目に路線名が記載されていることがわかった。

道路開通情報は、全国の自治体がWeb上で提供する情報である。本研究では、2009年12月1日から2010年3月4日までの間にWeb上で公開されたデータを収集できた224件のデータを対象とする。道路開通情報の内容を分析してみると、位置情報については、住所で記載されていることがわかった。住所の記載については、区間を示す2つの住所が含まれている場合や複数の情報項目にわたって記載されている場合があることがわかった。路線情報については、案件名に含まれている場合があることがわかった。

台帳修正箇所は、実際に道路構造が変化した後修正される道路台帳の情報である。この情報については、三重県から、2007年度の工事入札公告が2030件とそれに対応する台帳修正箇所のデータが179件あった。本研究では、既研究⁸⁾で工事入札公告と対応が把握できた台帳修正箇所のデータで、位置図が付属していた41件を対象とする。

(4) 路線ジオコーディングの精度の評価結果

路線ジオコーディングの距離精度評価には、三重県からいただいた2007年度の工事入札公告のデータに路線ジオコーディングを行い、その結果を台帳修正箇所のデータと比較することによって行った。既研究²²⁾では、この工事入札公告と台帳修正箇所を比較・分析し、手作業でマッチングすることによって、道路構造に変化のあった工事を特定した。これらの工事において、台帳修正箇所の位置図から、その工事箇所の中心付近の緯度経度を取得し、それを「真値」とした。真値の取得には、地名や店舗名など特徴のある記述と工事箇所の道路の形状を把握し、それらの情報を元に目視で地図を検索した。今回対象とした工事のデータのうち、位置図が付属していたのは41件であった。また、その位置図から緯度経度を取得できたものは、37件であった。

精度検証の結果、真値を取得できた37件のデータのうち、基点を取得できた29件では981m、さらにそのうちの信頼度の高い124件では、722mの精度が得られた。想定した誤差は、字レベルで1.2kmであり、実際の工事区間の長さは想定した距離よりも短い場合がほとんどである。また、さらにその工事区間の中心を取得しているため、真値を決定する場合の手作業による誤差は数m程度であり、想定した誤差に対して無視できるレベルだと考える。これらから、路線ジオコーディングの処理手法で算出した結果は、想定した誤差の内側であるため、妥当手法だと言える。

また、精度検証の過程で、地先名や路線名など、表記揺れに対応することで、精度が向上することがわかった。具体的には、県名などの補完、「一般国道」などの文字列の削除、ハイフンと長音記号などの統一などである。

さらに、路線ジオコーディングの精度向上を考え、対象としたデータを確認した。すると、住所表記が全て字レベルより大きい区分であった。これが、番地表記レベルであれば、さらに精度が向上すると考えられる。

(5) 各道路関連情報に適用した場合の評価結果

本研究では、道路関連情報のうち、既研究²³⁾で対

象とした工事入札公告情報、道路工事図面、道路の供用開始の公示情報、道路開通情報に加え、発注見直し情報にも路線ジオコーディングを適用した。位置精度指標を用いて、各道路関連情報から算出した結果を比較し、その結果を図-17と図-18に示す。



図-17 各情報の位置精度指標のレベル

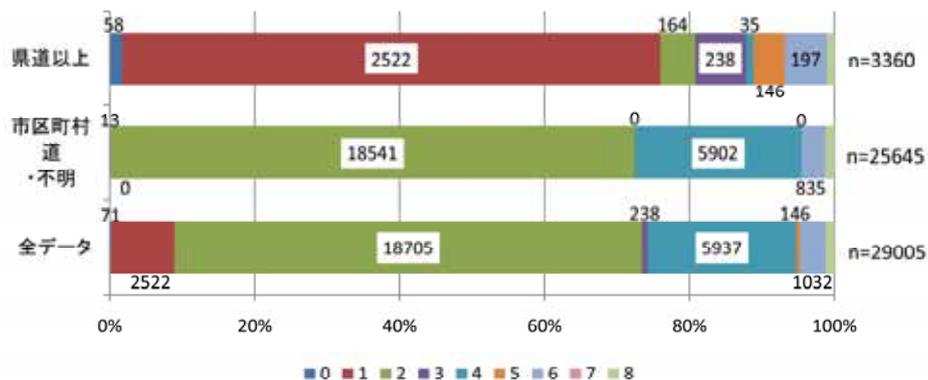


図-18 県道以上とそれ以外に分けた場合

図-17に示すように、道路関連情報毎に結果を比較すると、それぞれのレベル2(字レベル)以上の割合は、発注見直し情報が74.5%、工事入札公告が52.3%、道路工事図面が83.6%、供用開始の公示情報が94.5%、道路開通情報が73.2%であった。また、レベル4(市区町村レベル)以上の割合は、いずれのデータについても80%以上であった。さらに、図-18に示すように、県道以上のデータとそれ以外のデータを比較すると、レベル2(字レベル)以上の割合は、県道以上が81.7%、県道以外でも72.4%であった。

全体として、73.4%のデータがレベル2(字レベル)以上という結果が得られたが、よりピンポイントに位置を特定するには、さらに高い割合で結果が得られることが重要である。そこで、レベル2未満のデータについて、原因を検証した。

その結果、次の4種類の原因がわかった。1点目は、住所表記に郡が抜けている場合や市町村合併により

住所表記が変更している場合などのCSVアドレスマッチングサービスで未対応の住所表記である。2点目は、そもそも元の住所表記が市区町村レベルの場合である。3点目は、インターチェンジ名や橋梁名、警察署など、住所以外の位置情報で記載されている場合である。4点目は、住所に無関係な語句や記号が含まれている場合である。

そこで、本研究では、道路関連情報毎に、上述した4種類全ての原因に対応して住所を修正し、再度路線ジオコーディングを適用した。修正作業は、群抜けや住所以外の施設名についてはWebやCSVアドレスマッチングサービスを利用して完全な住所表記に修正し、無関係な語句については位置精度指標のレベルが低いものを対象に目視で検索した。その結果を図-19と図-20に示す。また、道路関連情報毎の修正内容を表-7に示し、次に述べる。

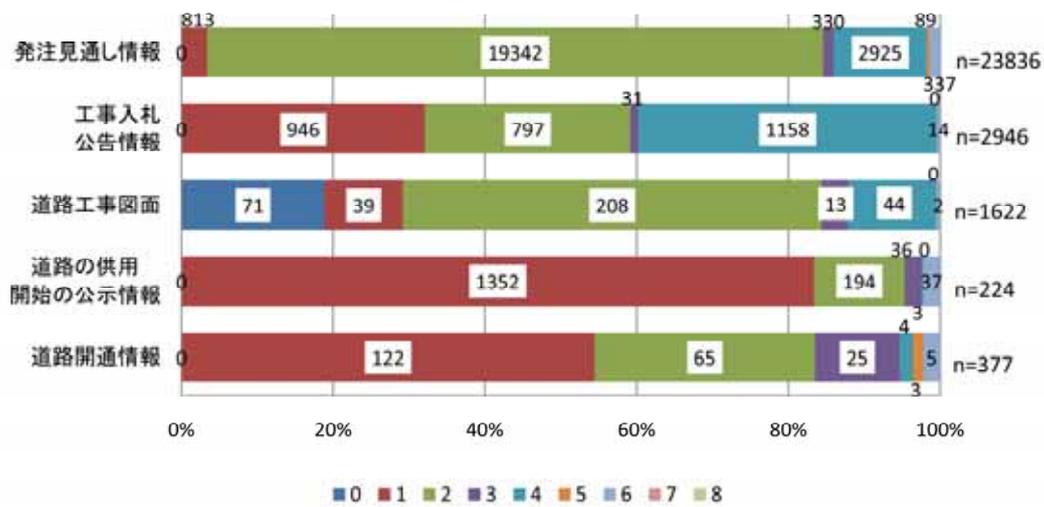


図-19 各情報の位置精度指標のレベル（修正後）

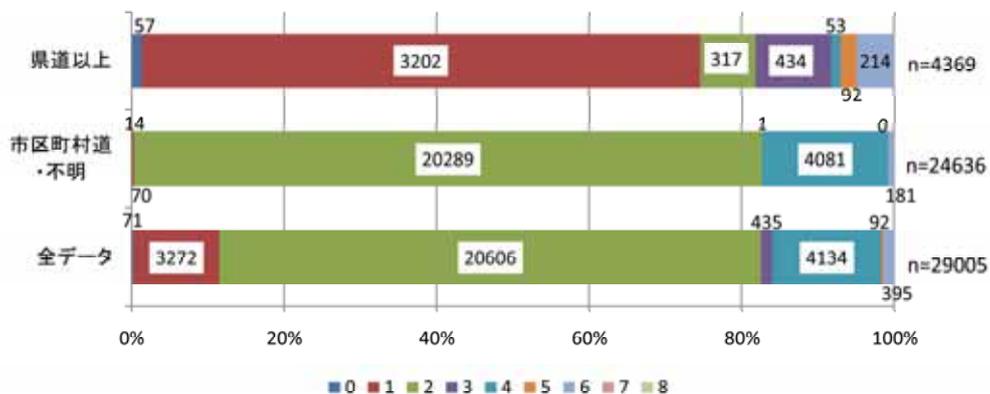


図-20 県道以上とそれ以外に分けた場合（修正後）

発注見通し情報については、既研究で自治体名を付加する処理を行っていたが、都道府県の工事などの場合、郡が付加されないため、不完全な住所となっている場合が416件あった。そこで、このような住所に郡を付加し、完全な住所になるように修正した。また、自治体によって、住所に路線名も記載している場合が692件あったため、そのような住所については、路線名を抽出し、住所から路線名を削除した。次に、沢や川、橋の名称が住所に記載されている場合が11件あったため、住所からそれらの名称を削除した。そして、工事名に括弧書きで住所が記載されている場合が21件あったため、括弧内の住所を採用した。その結果、レベル2(字レベル)以上の割合は、74.5%から84.6%に増加した。

工事入札公告については、まず、郡が抜けている住所となっている場合が109件あったため、郡を付加し、完全な住所になるように修正した。また、括弧

内に詳細な住所が記載されている場合が22件あったため、括弧を削除した。そして、警察署や水道工事センターなど住所以外の位置情報が含まれている場合が42件あったため、当該施設の住所に置換した。その結果、レベル2(字レベル)以上の割合は、52.2%から59.2%に増加した。

道路工事図面については、まず、郡が抜けている住所となっている場合が4件あったため、郡を付加し、完全な住所になるように修正した。次に、住所の前に数字が記載されている場合や、「(自)」といった文字列が付加されている場合が21件あったため、それらの文字列を削除した。そして、住所に誤字がある場合が27件あったため、それらを修正した。その結果、レベル2(字レベル)以上の割合は、83.6%から84.4%に増加した。

表-7 道路関連情報毎の修正内容

	修正内容	例
発注見直し 情報	郡抜きの補完	埼玉県加須市栄 → 埼玉県北埼玉郡北川辺町栄
	住所から路線名を抽出・削除	福島県国道289号下郷町大字南倉沢地内 → 住所:福島県下郷町大字南倉沢地内 路線名:国道289号下郷町大字南倉沢地内
	沢や川、橋の名称を削除	福島県塩竈川只見町大字塩ノ岐地内 → 福島県只見町大字塩ノ岐地内
	括弧内の住所への置換	奈良県大野向瀨地区(宇陀市室生区向瀨) → 奈良県宇陀市室生区向瀨
工事入札 公告情報	郡抜きの補完	三重県南伊勢町 → 三重県度会郡南伊勢町
	括弧の削除	岐阜県(土岐市泉町地内) → 岐阜県土岐市泉町地内
	警察署などの施設を住所へ変換	三重県津南警察署 管内 → 三重県津市久居明神町2501番地1
道路工事 図面	郡抜きの補完	岐阜県垂井町 → 岐阜県不破郡垂井町
	不要な文字の削除	02-1 三重県鳥羽市安楽島町字南山1312番48地先～ 1312番15地先02-2 → 三重県鳥羽市安楽島町字南山 1312番48地先～1312番15地先02-2
	誤字修正	岐阜県可茂郡白川町白山地内 → 岐阜県加茂郡白川町白山地内
供用開始の 公示情報	不要な文字の削除	後A:佐倉市寺崎字仲反町1709番4地先 → 佐倉市寺崎字仲反町1709番4地先
	括弧内の住所への置換	官公有無番地先(南松浦郡新上五島町宿ノ浦郷字浜迫 1252番21地先) → 南松浦郡新上五島町宿ノ浦郷字浜 迫1252番21地先
	路線名の漢数字の変換	一九三号 → 193号
	路線名に「線」を追加	石鏡公園 → 石鏡公園線
道路開通 情報	郡抜きの補完	岩手県川井村 → 岩手県下閉伊郡川井村

供用開始の公示情報については、まず、住所とは無関係の文字やスペースがある場合が16件あったため、それらを削除した。また、工事名に括弧書きで住所が記載されている場合が2件あったため、括弧内の住所を採用した。次に、路線名について、漢数字で記載されている場合が46件あったため、数字に変換し、路線を特定できるよう修正した。また、路線名に「線」の文字が無い場合が47件あったため、付加して路線を特定できるよう修正した。その結果、レベル2(字レベル)以上の割合は、94.5%から95.3%に増加した。

道路開通情報については、郡が抜けている住所となっている場合が25件あったため、郡を付加し、完全な住所になるように修正した。その結果、レベル2(字レベル)以上の割合は、73.2%から83.4%に増加した。

また、図-19に示すように、県道以上のデータとそれ以外のデータを比較すると、レベル2(字レベル)以上の割合は、県道以上が81.7%から81.9%に、県道以外が72.4%から82.7%に増加し、全体として82.6%がレベル2(字レベル)以上となった。

表-6に示した修正内容について、この結果を活用するには、道路管理者などの道路関連情報の提供元で作成する時点でデータの質を向上させるという方法と、本研究で分析した内容をシソーラスとして整備し、道路関連情報の雑多な文字列情報から位置に関する語句と路線名に関する語句を抽出するシステムを開発する方法が考えられる。前者の場合、質の高いデータの作成には、道路関連情報の提供者の負担の増加や作成指針などの規定の取り決め、規定の周知徹底など、障害は多い。後者の場合、今回対象とした5種類の道路関連情報においてさえ、それぞれに異なる対応が必要となったことを考えると、新たに道路関連情報の種類や対象範囲を増加すると、それに伴って新たに対応が必要になることが容易に想定でき、そのような対応を自動的に行うことは困難である。そこで、今後の展開として、前者と後者をバランスよく実施することを考える。具体的には、前者の対応としては、表記揺れを軽減するような入力支援ツールの開発、後者の対応としては、対象とする道路関連情報を増加し、分析した結果をシソーラスに追加が今後の課題となる。

第4章 実証実験(1): Webマイニング手法を用いた工事発注見通し情報の全国収集と提供

4.1 実験概要

本章では、カーナビ等、地図更新の効率化につながるよう、道路の変化情報(以下では道路更新情報と呼ぶ)を、全国レベルで無理なく把握できることを目的として、「工事発注見通し情報」のPDF、HTML、EXCELファイルを対象に、Webクローリング、OCR技術、テキスト解析技術、シソーラス作成技術、ジオコーディング等を用いて、道路更新に関する工事を抽出し、位置や工事概要の情報をデータベース化し、数万件レベルの情報を利用できるようにするものである。具体的には、4.2で発注見通し情報の概要を述べ、4.3で全体のデータ処理フローの考え方を説明する。4.4では処理システムを実装し、全国の自治体サイトを対象に実験を行い、その妥当性を確認し、4.5でまとめを行う。

なお、本章の関連研究として、情報分野では、組織のWebサイトから発信される公式情報を収集したもの²⁵⁾がある。とくに、Webからの表形式の情報抽出の研究として板井²⁶⁾、増田²⁷⁾あるいは田仲²⁸⁾などがある。

一方、建設分野における課題解決という意味では、筆者らを中心に道路更新情報のニーズや提供実態をまとめたもの⁶⁾、道路工事の電子納品により道路更新情報の蓄積を試みたもの⁷⁾、工事入札情報から道路更新情報の抽出を試みたもの⁸⁾がある。また、OCR技術を用いて道路の供用開始・廃止の公示情報をもとに道路更新情報の抽出を試みたもの⁹⁾などがある。さらに、マルチエージェント技術を用いて建設情報のDB統合を試みた研究²⁹⁾などもある。しかし、全国に約1800近くある地方自治体から広く収集できる可能性のあるデータをもとに実用を意識した大規模な実験を行った研究はなく、そのような意味で本研究は新しい試みと言える。

4.2 工事発注見通し情報

2章でも発注見通し情報について説明したが少し

詳細に述べる。まず入札適正化法施行令第五条では、表-8で示すように、公表時期、対象とする工事、公表事項、公表方法について定めている。

また、図-21のa)は自治体のWebサイトで掲載される典型的な発注見通し情報である。HTMLファイル上で項目を立て、実際に発注見通し情報が書かれたページにリンクする構造が普通である。通常リンク先は表形式で、b)で見られるようにPDFファイルで公開することが多い。表形式の並びはとくに決められておらず、岐阜県神戸町の例では、「工事名」、「工事場所」、「工期」、「工事概要」、「入札・契約方法」、「入札・契約時期」を掲載している。また、c)のようにHTMLファイルでセル構造を表現している場合も見られる。

こうした記載は個別の自治体で見るとばらつきがあるため、日本全国を対象に収集を行う前に、記載内容の傾向をつかむ必要がある。2章の図-13で示したように、「路線名」、「部署名等」を除く5項目については、9割以上該当する項目が掲載されていた。一方で部署名等は約半分、路線名は道路工事以外の工事も多数含まれていたため、5%程度であった。また全国的なデータ形式の内訳としてもPDFは半数以上を超え、PDFを対象とする重要性が確認できるとともに、HTMLやEXCELを併せて対応すれば公開しているものの9割程度はカバーできるであろうことは確認できた。

さらに、図-22は各主体の工事件数を示したものである。合計は41,967件に及ぶが、最も頻度が高い工事件数は0~10件、10~20件、100~150件が17%ずつとなっている。また、最も件数が多いのは東京都建設局の1,787件であった。

これらのことから、本研究では、発注見通し情報の中から、「町域名・地先名等」、「案件名称」、「概要」をもとに該当工事の位置や種別を特定しつつ、それ以外の「工期」、「部署名等」、「路線名」などについても、情報を取得することとした。

表-8 入札適正化法施行令第五条で定める主な事項

項目	内容
公表時期	毎年度、四月一日（当該日において当該年度の予算が成立していない場合にあつては、予算の成立の日）以降遅滞なく。
対象とする工事	当該年度の発注することが見込まれる公共工事（予算価格が二百五十万円を超えないと見込まれるもの及び公共の安全と秩序の維持に密接に関連する公共工事であつて当該地方公共団体の行為を秘密にする必要があるものを除く。）

公表事項	<ul style="list-style-type: none"> 公共工事の名称、場所、期間、種別及び概要 入札及び契約の方法 入札を行う時期（随意契約を行う場合にあつては、契約を締結する時期）
公表方法	<ul style="list-style-type: none"> 公報または時事に関する事項を掲載する日刊新聞紙 公衆の見やすい場所に掲示し、又は公衆の閲覧に供する方法（閲覧所を設けること又はインターネットを利用して閲覧に供する方法）



a) 岐阜県神戸町入札情報HP:
<http://www.town.godo.gifu.jp/content/areaguide/admin/nyusatsu.htm>



b) 岐阜県神戸町発注見通し情報HP (PDFファイル):
<http://www.town.godo.gifu.jp/content/areaguide/admin/nyusatsu/minaoshi.pdf>

c) 三重県尾鷲市発注見通し情報HP (Htmlファイル):
http://www.city.owase.lg.jp/contents_detail.php?co=cat&frmId=5077&frmCd=2-15-2-0-0

図-21 工事発注見通し情報例

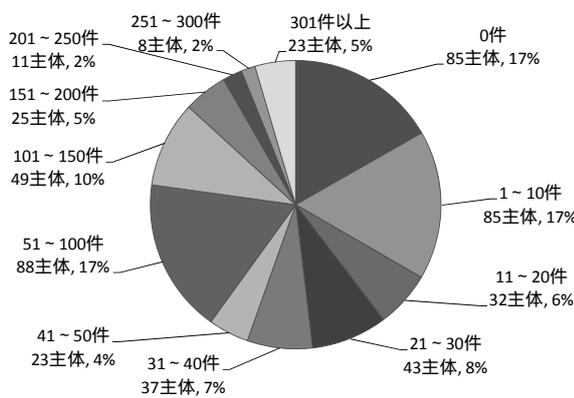


図-22 主体別の公表件数分布状況（関東圏を対象）

4.3 データ処理の考え方

(1) 処理フローの概要

本節では、まず全体の処理フローについて述べる。最初に各自治体の発注見通しに関するサイトの

URL 情報収集を行う。これは手作業で記録することとしたが、ほとんどの自治体では公共事業のサイトから簡単に見つけることができる。これを入力データとして、Step0 ではファイル取得を行う。ただし、四半期発注見通しのように年間でも複数のファイルが存在する場合もあるため、本研究では経過なども追跡できるように該当ディレクトリにある、全てのファイルを取得している。そして Step1 では取得したファイルの中で主要な形式である PDF, HTML, EXCEL について、一度 PDF 形式にした後、Step2 で OCR 等の機能を用いた画像データを経て表形式を読み取り CSV ファイルに変換する。また Step3 では、各自治体によって、記載している項目の表現やその並びの順序が異なるため、表形式の各項目の正規化を行う。さらに Step4 では、正規化を行った工事リストの中から関連キーワードによって道路更新に係る工事の抽出を行う。また、Step5 ではデータベースに登録するためのデータを作成する。最後にそ

れらによって抽出された工事の位置を明確にするため、Step6で路線に沿ったジオコーディングを行う。これらをまとめたものが図-23であるが、次節以降ではStep2からStep6の詳細を説明する。



図-23 工事発注見通し情報の処理フロー

(2) PDF ファイルの取得と表構造の抽出

Step2では、Step0で取得したPDF、HTML、EXCELファイルからPDF形式にした後、表構造を認識してその内容のCSVファイルを抽出する。HTMLやEXCELはもともとある程度構造化されているものの、それ自体にはタイトルや説明、フレーム用の表構造等、該当する表構造以外の要素が多数含まれ、必ずしも抽出が容易ではないため、本研究では一度PDFに一本化した後CSVにしている。HTMLからPDFについてはGoogle社提供の変換ツールwkhtmltopdfを用い、EXCELからPDFへの変換は市販ソフトで一部バッチ処理を行った。表-9はHTMLとEXCELファイルの全国分の結果を表しており、動的サイトの構成である等の理由からコンテンツが取得できないケースもごく一部見られるが概ね取得でき、PDFへも変換できている。

表-9 HTMLとEXCELファイル取得、PDFへの変換状況

	HTML	EXCEL
処理対象サイト数	139	74
取得可能サイト数	131 (1)	69 (1)
ファイル取得状況	131	450 (2)
PDF変換状況	129	504 (3)

- 1 動的サイトに行っているためコンテンツが取得できない、あるいはファイルの提供形式が調査後に変更しているなどによる
- 2 指定のサイトに複数のEXCELファイルが存在
- 3 EXCELの印刷設定等でPDFファイルが複数に分かれる

また、PDFにした後の手順は大きく分けると、OCR処理により罫線を判定する、罫線に囲まれている個々のセルに含まれる文字列を抽出する、表の記載様式のまま構造化する、という三段階の処理が必要である。こうした機能は、1章で述べたように情報分野の研究などもあり、まだ多くはないものの、最近では商用ソフトなども出つつあるため、本研究ではそれらを利用することとした。具体的には、日本語ファイルに対して前述の処理をサーバーサイドでバッチ処理として行えるプログラムライブラリとして、メディアドライブ(株)の「活字文書OCRライブラリ」やパナソニックソリューションテクノロジー(株)の「活字認識ライブラリー」があるが、これらを比較し、本研究ではPDFファイルの画像変換機能の有無や開発言語(VB.NET)のサポート状況、費用面から「活字文書OCRライブラリ ver.6.0」の中の罫線抽出ライブラリを採用した。

大規模な実験の前に、性能確認のため、岐阜県下市町村のうちPDFで情報を公開している16自治体の工事発注見通し情報全1,348件に対して抽出処理を実施したところ、表-10で示すように、表形式のセル自体は全て抽出できた。ただし、抽出したセル内の文字については、一部文字化けがあり、その割合は1.55%であった。文字化けのエラー要因はソフトの特徴に依存するが、表-11に挙げたようにいくつかの文字については、認識性能が劣る特徴があった。

表-10 PDFから表構造の抽出状況(岐阜県の例)

抽出対象自治体	レコード数	セル数	文字数	誤認文字数	誤認率(%)
神戸町	42	294	2135	21	0.98
郡上市	116	928	7016	277	3.95
多治見市	132	1188	8947	81	0.91
羽島市	55	495	3257	29	0.89
可児市	53	477	2652	26	0.98
海津市	70	560	3402	20	0.59
関市	146	1314	6385	40	0.63
恵那市	106	848	5971	113	1.89
御嵩町	31	279	1457	19	1.30
高山市	230	1840	14513	347	2.39
瑞浪市	50	450	2950	66	2.24
中津川市	176	1408	10784	62	0.57
飛騨市	17	170	887	8	0.90
美濃市	27	243	1269	13	1.02
北方町	17	136	662	21	3.17
本巣市	80	640	2992	22	0.74
合計	1348	11270	75279	1165	1.55%

飛騨市は都市整備課分、本巣市は産業建設部分

表-11 文字抽出時のエラー頻度(岐阜県の例)

正	誤	頻度(回数)
=	二, 二, -	85
0	0	37
-	-	31
m ²	m, r m ² , 2	23
o.		20
m	1 n, r n	9
ボ	ボ	5
工	工	3

複数の市町村で合計3回以上エラーが発生したものを抽出

(3) 各データ列の正規化

CSV で出力した発注見通し情報は、自治体ごとに各データ列のタイトルの表現や並びが統一されていないため、正規化する必要がある。本研究では予め集めた、関東（1都6県）、三重県、岐阜県、大阪府の発注見通し情報をもとに、手動で対象項目に関する辞書（シソーラス）を作成した。表-12 はそれらの一部を抜粋したものであるが、「案件名称」は合計44種類、「町域名・地先等」は34種類、「工期」は57種類、「概要」は25種類、「入札予定時期」は98種類、「部署名等」は54種類、「路線名」が4種類存在した。日本全国に対して適用する際には、このシ

ソーラスに対して、いずれかの文字列に完全一致したものを採用し、当該項目とした。

(4) 関係工事の抽出

前節で正規化したデータは公共工事全般にわたり、学校建設や河川工事あるいは道路関連工事の中でも除草工事等、道路更新情報に直接関係しない工事も多数含まれるため、該当する工事のみを抽出する必要がある。そこで、「部署名等」、「案件名称」、「概要」の記載をもとにキーワードでフィルタリングを行っている。この処理の具体的なフローを図-24 に示す。

表-12 発注見通し情報の関連シソーラス（関東圏、三重県、岐阜県、大阪府のデータをもとに作成）

各項目	関連すると思われる発生項目
案件名称	「工事件名」「件名」「件名・施設名」「工事概要」「工事件名（予定）」「工事名称」「工事の名称」「公共工事の名称」「名称（件名）」「工事名」「予定工事名」「調達案件名称」「名称」「工事（業務）名称」「業務名（履行・業務場所）」「工事等の名称」「工事名（委託名）」「案件名」「事業名（工事名）」「事業名」等 合計44種類
町域名・地先等	「工事場所」「履行場所」「施工場所」「場所」「工事の場所」「公共工事の場所」「工事箇所」「施工箇所」「予定施工場所」「案件場所」「場所（地区）」「工事（業務）場所」「業務名（履行・業務場所）」「工事箇所（概略位置）」「工事等の場所」「工事場所（履行場所）」「施行場所」「工事名称 工事場所」「施工予定箇所」等 合計34種類
工期	「工期」「工期・履行期限」「予定工期（月）」「予定工期」「履行期間」「履行日数」「工事期間」「工期始（予定）工期終（予定）」「工期日数」「工期（月）」「期間」「工事の期間」「予定期間」「予定工事期間」「工事期間（開始月） 工事期間（終了月）」「入札・契約」「案件期間」「工期 開始完了」等 合計57種類
概要	「工事概要」「概要」「施工内容」「概要（内容）」「工事の概要」「公共工事の概要」「予定工事概要」「案件概要」「事の概要」「工事内容」「種別及び概要」「工事（委託）概要」「事業内容」「事業概要」「工事等の概要」「工事等概要」「調達案件の概要」「工事概要工事概要」「概要、規格、数量等」「工事概要・仕様等」等 合計25種類
入札予定時期	「入札予定時期」「公表予定時期」「入札時期」「入札予定」「契約時期」「入札（契約）を行なう時期」「時期」「発注時期」「発注予定時期」「入札時期（予定）」「入札予定時期又は状況」「発注予定」「入札の時期」「入札・随意契約時期」「入札を行なう時期又は契約を締結する時期」「開札日」「予定入札時期」等 合計98種類
部署名等	「発行部署名」「発注部署名」「担当課」「主管課」「課・部」「工事担当課」「課・係名」「課名・係名」「課所名」「発注課」「発注部署」「課名」「所管課」「課（所・局）名」「担当部課（所）名」「所属名称」「所属名」「部課名」「担当課名」「担当」「発注課名」「発注担当課」「担当課担当係」「担当部署」「見積課名」等 合計54種類
路線名	「工事場所」「路線河海名」「路河川等」「道路・河川名」等 合計4種類

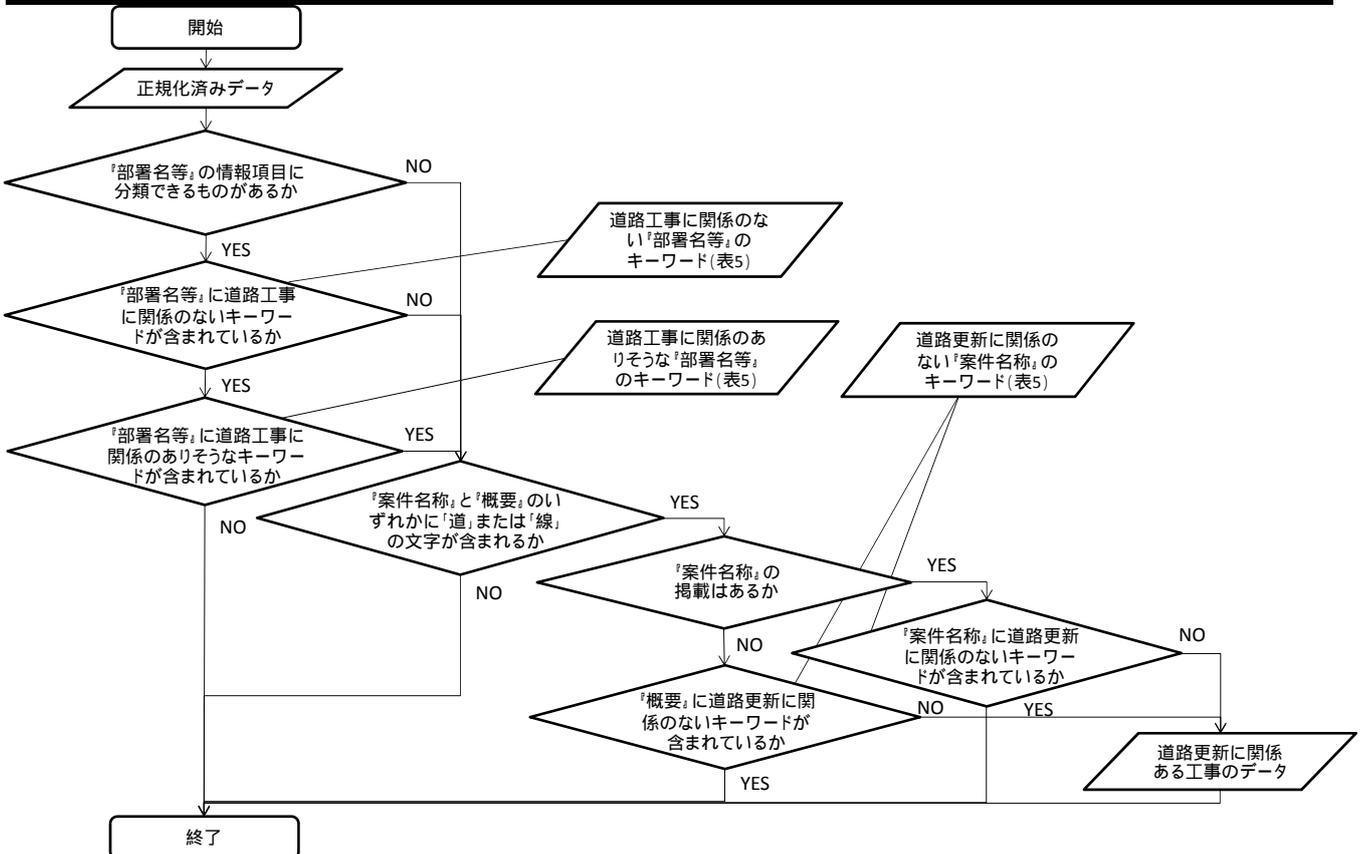


図-24 道路更新に関する工事抽出時の処理フロー

まず道路工事に関係のない「部署名等」(例えば水道・学校等)を含まないものや直接道路工事に関係する「部署名等」(例えば道路・土木等)を含むものを抽出し、その後、「案件名称」と「概要」のいずれかに「道」または「線」が含まれるものを抽出する。最後に、それらから「案件名称」または「概要」に道路更新情報に関係のないキーワードが含まれていないものを抽出する。ここで関係のないキーワードとしては、契約形態では道路更新情報に関係のない「委託」、「単価契約」等のキーワードが該当し、道路の維持管理に関する工事を示す「除草」、「街路樹」等のキーワードなども同じである。本研究では、上記抽出過程で用いたシソーラスは、三重県・岐阜県・大阪府のデータをもとに手動で作成したものである(表-13)。

三重県、岐阜県、大阪府のデータを対象として、図-24の提案手法を検証した結果が表-14である。表側が提案手法により判別したもので、表頭が目視で確認した真値としている件数である。実際には3,899件中3,736件(96%)が正しく判別された。判別が正しく行えなかったものとして、目視で「道路更新に関する工事」であると判定されたにも関わらず提案手法では「その他」になったものは96件あったが、そのうちの95件は「案件名称」と「概要」どちらにも「道」、「線」を含まなかったものである。一方、目視では「その他」と判定されたが提案手法では「道路更新に関する工事」と判定された67件には、「部署名の記載がなく、概要に「道」が含まれるもの」などがあったが提案手法で判別された968件の7%弱であったためこのまま関係工事として含めている。

(5) 登録データの作成

前節までのプロセスで取得・整理された発注見通しのソースデータを最終的にDBに登録する。これらのデータは別途進めている「地理空間情報流通実験コンソーシアム」における「流通実験プラットフォーム」

フォーム」に蓄積し、登録ユーザーはデータを検索・ダウンロードできるようにしている
(<http://parma.csis.u-tokyo.ac.jp>)。

表-13 道路工事抽出における関連キーワードの設定
(三重県、岐阜県、大阪府のデータをもとに作成)

道路工事に 関係のない 「部署名等」 の キーワード	道路工事に 関係ありそ うな「部署名 等」のキー ワード	道路更新に関係のない 「案件名称」のキーワード	
水道	道路	契約形態として関係 しない工事	委託
学習	土木		単価契約
学校	建設		水道
教育	交通		汚水
福祉	整備		排水
管繕			送水
総務			水路
危機			用水路
消防			布設替
図書館			ポンプ場
博物館		道路構造物に関する 工事	照明灯
研究所			電気 設置
センター			設備 施設
文化 住民			施設
スポーツ		維持管理に関する 工事	体育館
管理室			耐震
こども		工事準備段階のもの	除草
子ども			街路樹
		「道」や「線」の文 字列が含まれるが道 路ではない工事	測量
			軌道
			無線
			武道
			道の駅

表-14 抽出フローによる判別状況
(三重県、岐阜県、大阪府の例)

フロー による 判別	道路更新に関 係ある工事	目視による判別		
		道路更新に関 係ある工事	その他	合計
道路更新に関 係ある工事	901	67	968	
その他	96	2,835	2,931	
合計	997	2,902	3,899	

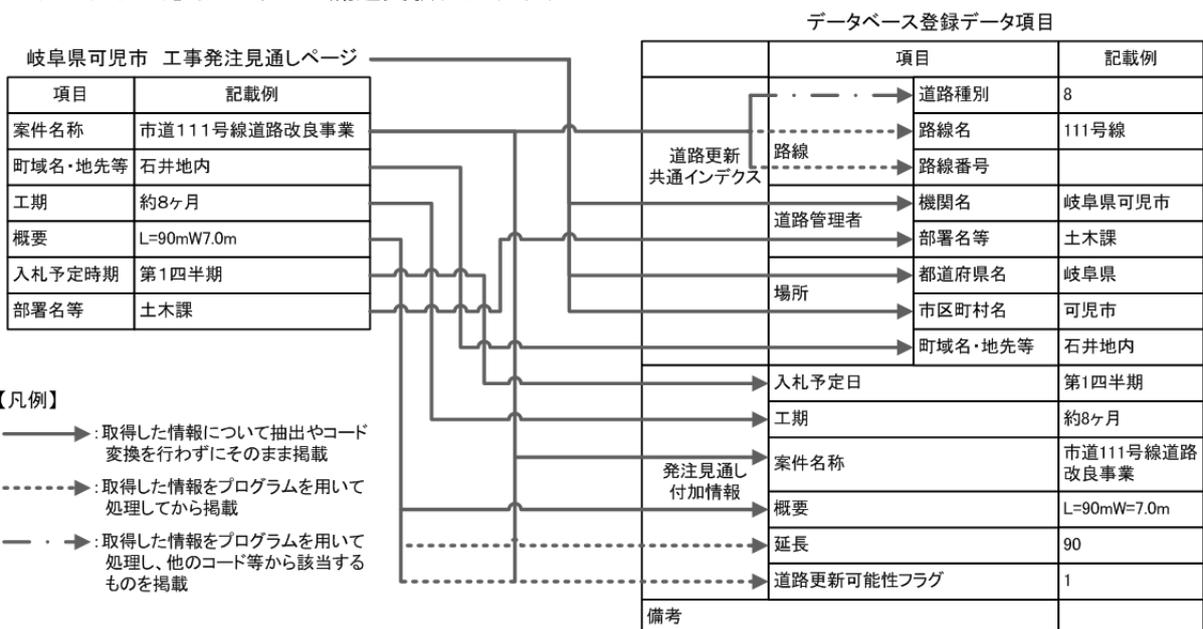


図-25 工事発注見通しデータとデータベース項目の対応関係

とくに道路関係のデータは数種類の蓄積を行っているため、共通のメタデータなども定めている。図-25の左にソースデータ、右にDBの項目を挙げ、その対応関係をまとめた。そのまま取得した情報をDBに格納できるものが多いが、一部、該当文字列を抽出したり、判定処理を行いコード化するものなどもある。なお、上記プラットフォームに関する研究の詳細は薄井³⁰⁾を参照されたい。

(6) 路線ジオコーディング

ここでは、抽出した該当工事について、町域・地先名や路線名から具体的な場所を特定する。本研究では、南¹⁵⁾が行っている路線ジオコーディングを用いた。これは通常の住所をベースにしたジオコーディング（ここでは、東京大学空間情報科学研究センターのCSVアドレスマッチングサービス：<http://newspat.csis.u-tokyo.ac.jp/geocode/>を利用）に対して、さらに路線情報を用いてマップマッチングを行うものである。ただし、路線名は、とくに市町村道などの細かい道路の場合、ベースとなるシソーラス等もなく、該当路線を特定できないケースも多い。そのため、図-26で示すような段階的な位置情報の精度レベルを提案しており、実際の道路更新箇所に近い場所がある程度具体的に把握できることを重視すると実用上は、字レベルで特定できるレベル2以上が望ましいと考える。

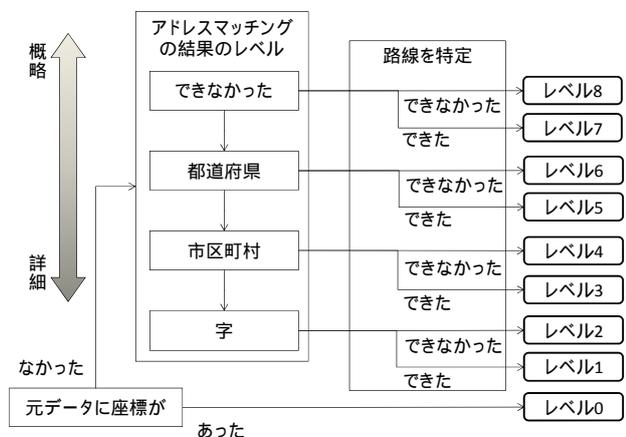


図-26 路線ジオコーディングの考え方(南¹⁵⁾を基に作成)

また、路線ジオコーディングそのものはシンプルな方法なため、入力する地先名や路線名等の表記ゆれ等、雑音の傾向によっても精度が変わる。南¹⁵⁾によると、いくつかの都道府県のデータをもとに「工事入札公告情報」、「供用開始の公示情報」、「道路開通情報」、「道路工事図面情報」についてそれぞれ数百件から三千件程度のデータを対象に路線ジオコーディングを行ったところ、レベル2以上の達成状況がそれぞれ、50%強、90%強、70%強、80%強であり、ある程度の品質は保たれている。

4.4 実験

(1) 適用実験

本節では、前章までの処理フローにもとづき、日本全国の工事発注見通し情報への適用を行った。表-15が実験環境の諸元である。また計算にかかった時間は、表-16のとおり、Step1とStep2で各1日程度、Step3からStep5で約10分、Step6で約4時間かかっている。

さらに、全体の実験結果をフローとともにまとめたものが図-27である。まず全国の自治体の発注見通しサイト情報を手動で収集したところ、公表状況はPDFが992サイト、HTMLが139サイト、EXCELが74サイトであり、公表していない自治体を含めると6割程度のカバー率であるが、何らかの公表している自治体の中でのカバー率は9割以上であり概ねカバーしていると言える。ただし、同一の組織でも異なる部課室ごとに違うサイトで掲載している場合はそれぞれカウントした。この割合は2章の図-17で概観した関東圏での割合とほとんど同じである。

また、Step0ではおおむねPDFを取得できたが、11%ほどのサイトからは、PDFが取得できなかった。この理由として、サイト情報が調査時点から変更されたか、サイトが取得プログラムによって自動的にファイル取得をできない構造になっていることが原因であると考えられる。また、Step1は3(2)で述べたとおりなのでここでは割愛する。

次に、Step2では、約12,700ファイル中、約10,400ファイル、すなわち約82%がCSVとして表構造が抽出できている。一方でCSVに変換できない約18%はもともと工事発注見通し情報と同一サイト（ディレクトリ）に全く別のファイルを置いているケースもあり、Step0ではそうしたものも取得してしまい、表構造を持たない無関係なファイルはCSVに変換できず除去される。

さらに、Step3の各ファイルの正規化では、約半数の約5,100ファイルが正規化ができたが、これは想定よりやや低いように思える。これは3章の(3)で説明した「案件名称」、「町域名・地先等」、「工期」、「概要」、「入札予定時期」、「部署名等」、「路線名」の全てが、表-12の10都道府県をベースにしたシソーラスがカバーしているデータ列に完全一致する必要があり少し厳しい条件だったからと考えられる。この課題に対しては、全都道府県を対象とした自動収集結果からシソーラスを更新し、そのカバー範囲を広げるとともに、近い表現についても何とか採用できる枠組みを用意する必要がある。いずれにしてもこの段階で約5,100ファイルから総工事件数は13.8万件となっている。Step4の道路更新の関係工事抽出では、関係工事として約20%が抽出されている。これは、表-14の3府県をサンプルとして適用した結果（フローによる判別の場合道路更新の関係工事が約25%）とあまり変わらない比率であり妥当な結果と思われる。

最後のStep6の路線ジオコーディングでは、レベル2以上で位置特定されたものが約74%となった。これらのデータは、最終的には約2.7万件の道路更新関係の工事としてDBに登録された。

表-15 実験の諸元

	Web サーバ	DB サーバ
ミドルウェア	Apache 2.2.13 Tomcat 6.0	PostgreSQL 8.3 PostGIS 1.3.6
OS・ハードウェア	OS: Windows 2008 Standard CPU: インテル Xeon 3GHz メモリ: 2GB	

表-16 処理時間の概要

各処理項目	およその時間
Step1. PDF ファイルの取得	約 19 時間
Step2. 表構造の抽出	約 32 時間
Step3. 各データ列の正規化	約 10 分
Step4. 道路更新関係工事の抽出	
Step5. 登録データの作成	
Step6. 路線ジオコーディング	約 4 時間

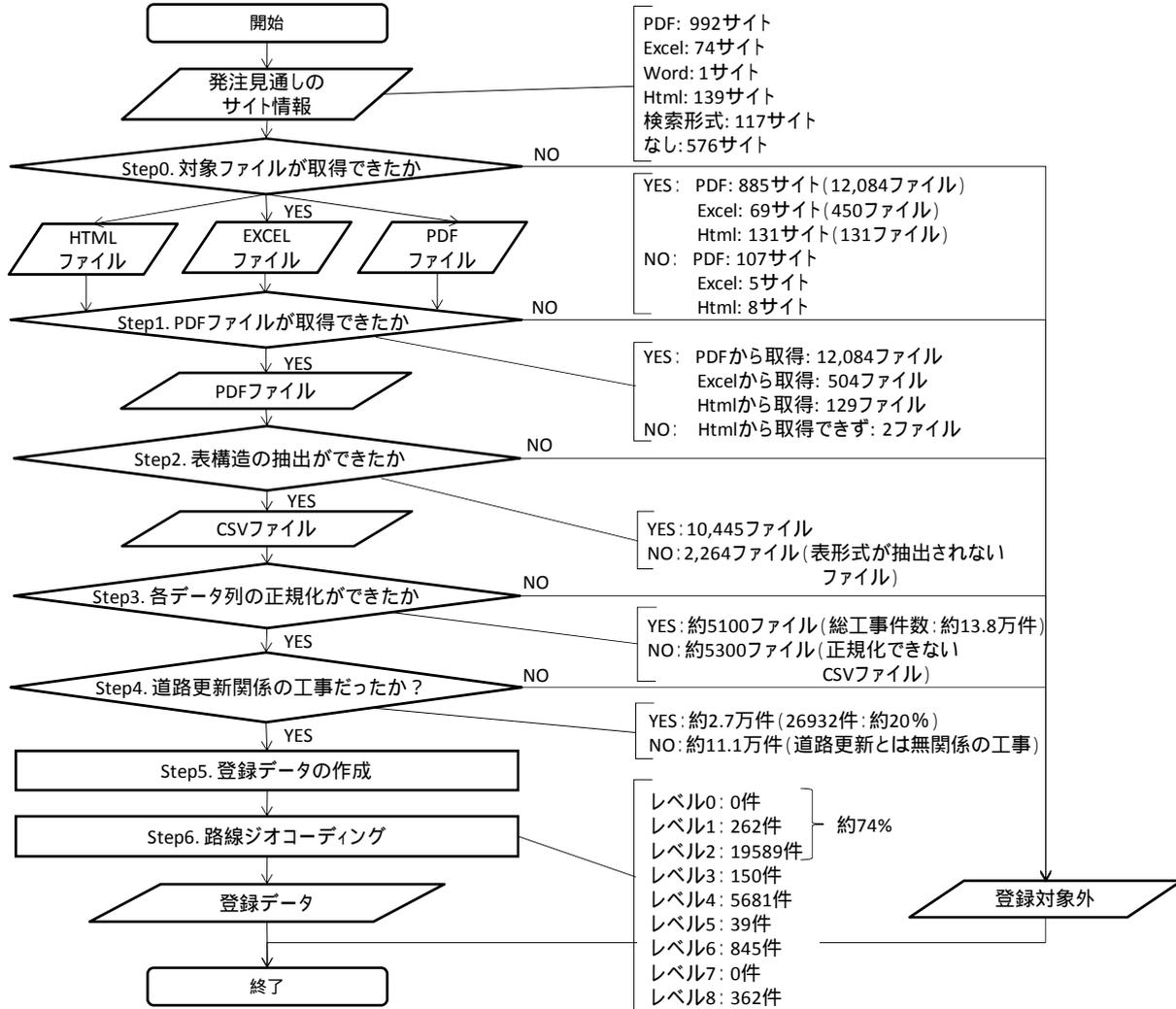


図-27 適用結果

(2) 抽出結果の検証

ここでは、前節で抽出された結果が道路更新情報として妥当なものであったか、ファイル取得ができた約 1 割に該当する 100 の自治体を実験として抽出を行い、各サイトの公表内容を目視で確認することにより確認した。その結果が表-17 であるが、100 自治体のうち、Step1 で PDF が取得できなかったのが 12 自治体、Step2 の表構造が抽出できなかったのが、1 自治体、Step3 の正規化がうまくできなかったのが 15 自治体であった。これは先ほど図-27 で述べた全体の結果と概ね一致している。残った自治体については Step4 以降の道路更新に関する工事件数も約 2,500 件となっており、全体の割合と概ね一致している。

さらに Step3 の正規化でエラーが発生する原因をエラーメッセージ等から探ってみたところ表-18 のようにいくつかの要因に分類された。「シソーラス定義に合う列がなかった」が最も多いが、これは前節で述べたように、全ての列のヘッダがシソーラス定義に合うことを条件としていたことが正規化のエラーの主要因だったことを示している。

(3) 抽出結果の視覚化

また、前節の抽出処理以降は、3(5)で述べたようにデータがプラットフォーム上に登録され、検索・閲覧・ダウンロードできるようにしている。例えば三重県の範囲で検索・表示した例が図-28 であり、1,335 件が登録されていることを確認できる。また、

さらに全国での登録状況について、自治体ごとの数を濃淡で表現したものが図-29である。黒色はもともと工事発注見通し情報を提供していない自治体で、灰色が動的なページで提供しているため、本研究では抽出が行えなかった自治体である。一方で数の多少は赤色の濃淡で表現しているが、白色は実際には何らかの情報提供があるものの、うまく抽出できなかったものということになる。

4.5. まとめ

本研究では、日本全国の年間の道路の変化箇所を抽出するために、国や地方自治体を対象とした統合的なシステムを何年もかけて作るのではなく、国や各地方自治体が Web サイトで公開している PDF, HTML, EXCEL ベースの工事発注見通し情報を Web マイニングに関わる様々な技術を組み合わせて抽出する手法を提案した。また、提案手法を、全国の地方自治体を対象に適用し、約 2.7 万件の道路更新情報のデータベース化を行った。このように道路管理者が Web サイトで公表している情報について、一部の地域ではなく全国を対象に、また、各ステップを自動化して全体の自動収集を実現した例はなく、新規性が高い。これにより、道路管理者以外に道路更新情報を必要とする主体が必要な情報を迅速に入手することが可能になるとともに、道路管理者サイドの情報提供に関する必要以上の負担も減らすことに道筋をつける事ができた。

一方で今後の課題として、いくつかのステップでは改良すべき事項もある。例えば、発注見通しのサ

イト情報そのものを今後定期的に自動更新していく方法は各主体ごとに発注見通し情報の提供サイトに変更があった場合に必要な対応である。また、正規化については、構築したシソーラスで正規化できずエラー扱いになったケースもかなりあったため、今後継続してデータ収集を進めることによるシソーラスの自動更新や近い用語でも確率的に当たっている可能性が高いと判定するなどの方法も必要である。さらに路線ジオコーディングについても、入力する地先名や路線名そのものの記載が不十分な場合で特定できないケースも多々あり、他の情報項目から関連する場所の候補を絞っていくような試みも必要であろう。

表-17 100 自治体における検証結果

処理結果		自治体数
PDF が取得できなかった		12
PDF が取得できた (Step1)	CSV が取得できなかった	1
	CSV が取得できた	15
	正規化が正しくできなかった (Step2)	72
	正規化が正しくできた (Step3)	

表-18 正規化等によるエラー分類 (一つのファイルに複数種類のエラーが存在するケースもあり)

エラー内容	ファイル数
一行目と列数が違う	976
シソーラス定義に合う列がなかった	4835
案件名称と概要がありません	78
案件名称に文字がありません	3
案件名称の文字数が 100 文字以上あります	10

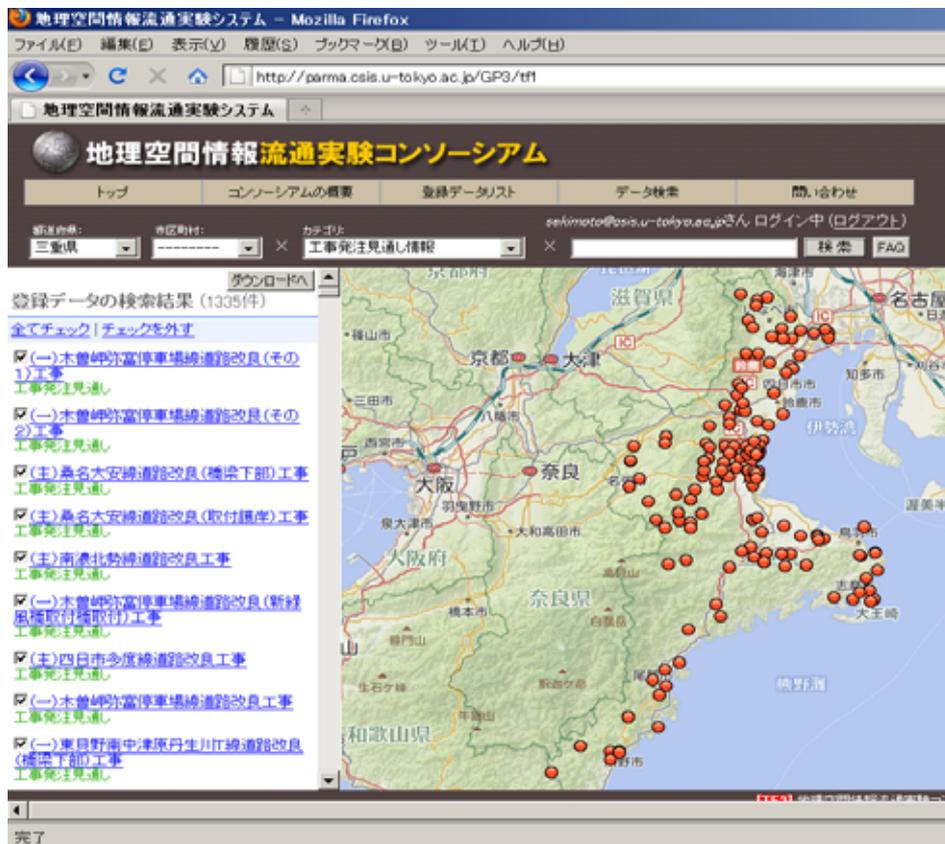


図-28 Web 画面での表示状況 (三重県の 1,335 件が検索された例: ただし表示は一部のみ)

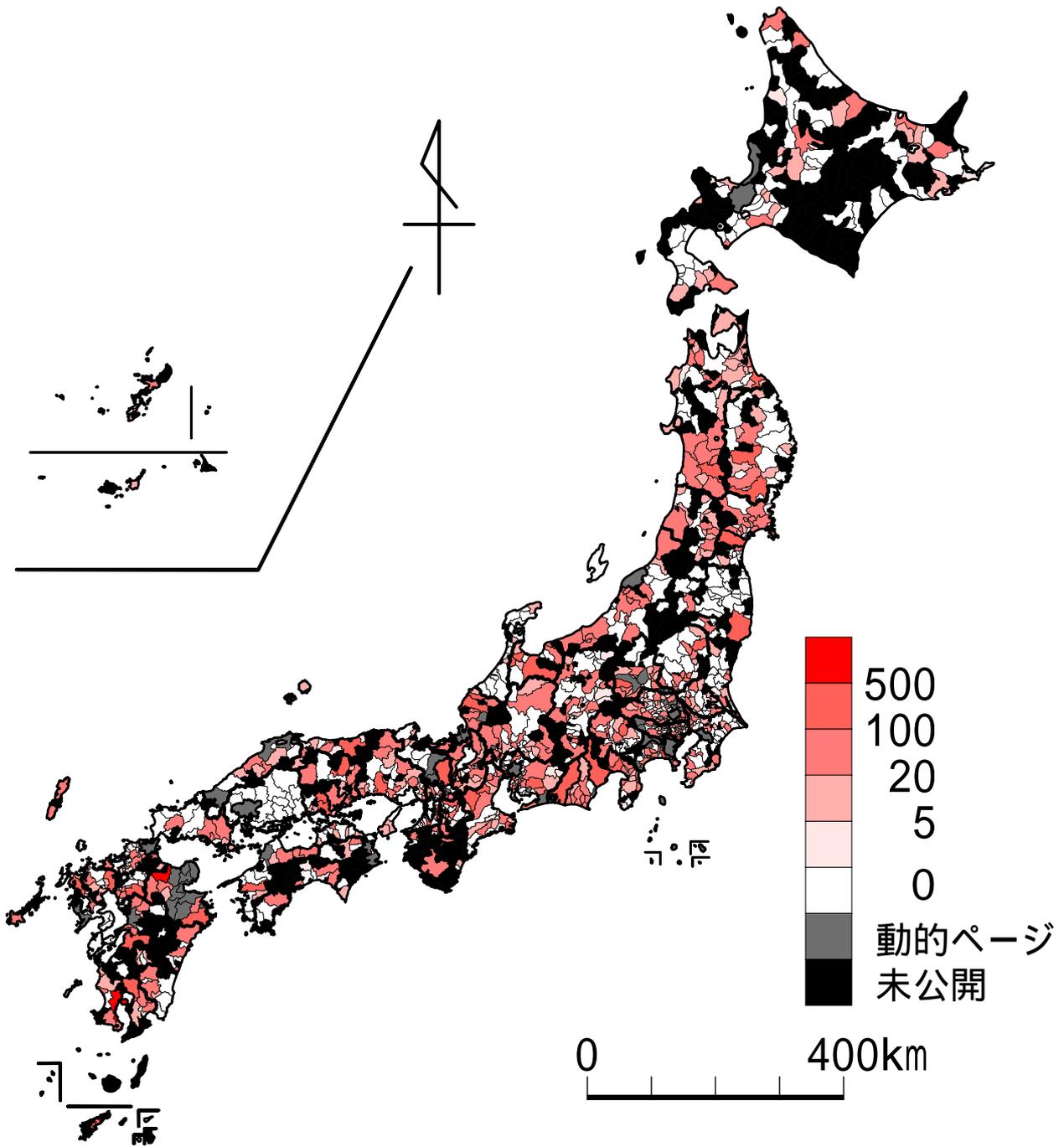


図-29 各自治体ごとの抽出できた工事発注見通し情報件数

第5章 実証実験(2): SVM手法を用いた工事入札公告情報による道路更新状況推定

5.1. 道路更新の事前推定方法の検討

道路更新を事前に把握することが可能と考えられる情報は2章の図-3に示したとおり複数存在するが、情報提供のタイミング、情報提供の実施状況等はそれぞれで異なる。その中で、工事入札公告は実際の道路更新に先立つ情報である点、比較的全国規模で電子化が進んでいる情報である点、他の道路更新情報と比較し含まれる情報量が多い傾向にある点等から、工事の内容により道路更新に資するかどうかを判断することが可能と考えた。

上記を踏まえ、本節では「工事入札公告」をもとに、道路更新の有無を事前に推定する方法の検討を行った。工事入札公告をもとに道路更新の有無を推定する際には、機械学習の手法を用いることとした。機械学習の手法として、ニューラルネットワーク、ブースティング(ADAブースト)等が挙げられるが、今回は主に計算量の観点からSVM (Support Vector Machine)^{31), 32)}を採用した。機械学習の手法を取り入れた道路更新自動判別モデルの構築イメージは、図-30に示すとおり。初期の段階では推定結果の確認を必要とするが、将来的には工事入札公告から自動的に道路更新フラグを付与するモデルの構築を目指した。

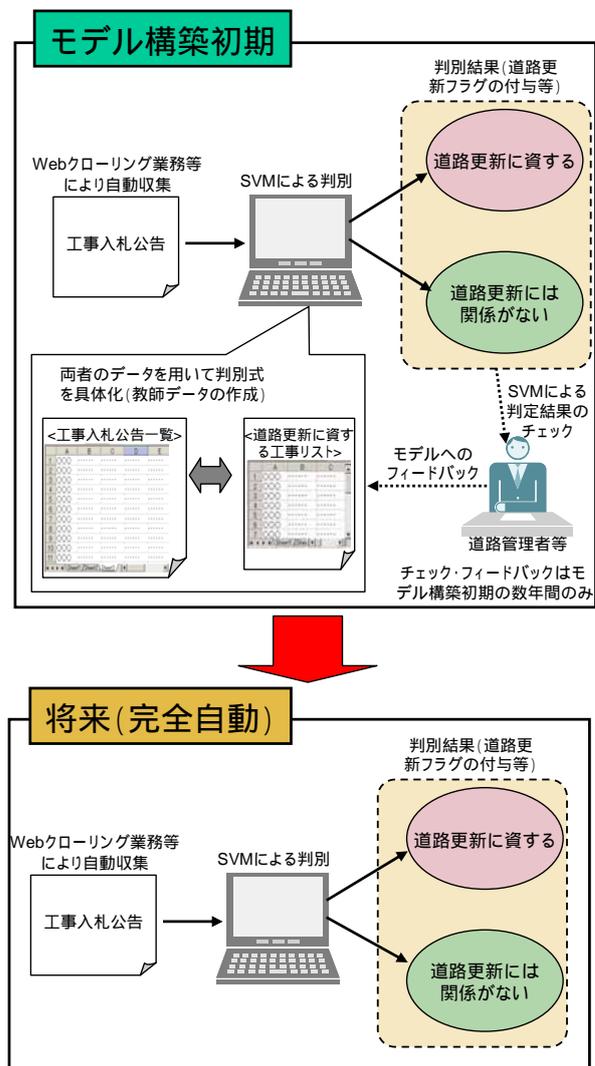


図-30 道路更新自動判別モデル構築イメージ

5.2. 検討の手順について

検討は、工事入札公告からキーワードの抽出を行ったうえで、機械学習によるモデルの構築・検証を行った。機械学習では図-31に示すとおり、モデルの構築・検証を行うために、予めそれぞれの工事入札公告が道路更新を伴うものであったかを明らかとしたデータが必要となる。今回は、工事入札公告と道路更新の有無の特定が可能なデータを準備できた三重県のデータを対象として分析を実施した。なお、工事入札公告は、三重県の入札情報サービス（工事関係）で公開されている工事入札公告のうち、件名に「道」もしくは「線」が含まれている工事を対象とした。

また、各工事入札公告が道路更新を伴うものは、

三重県から貸与いただいた道路台帳のデータと財団法人日本デジタル道路地図協会から貸与いただいた道路台帳修正箇所明細データをもとに、住所、案件名、延長等から予め判定をした。

機械学習を行うため、各工事入札公告に含まれる語句（キーワード）を抽出する必要があることから、キーワードの抽出を行った。工事入札公告は、図-32に示すとおり、HTMLで記述されWeb上に公開されている概要を示す情報とPDFファイル等で公開されている詳細情報の2つで構成されることから、本検討では両者をテキストデータ化した上で、形態素解析を行った。

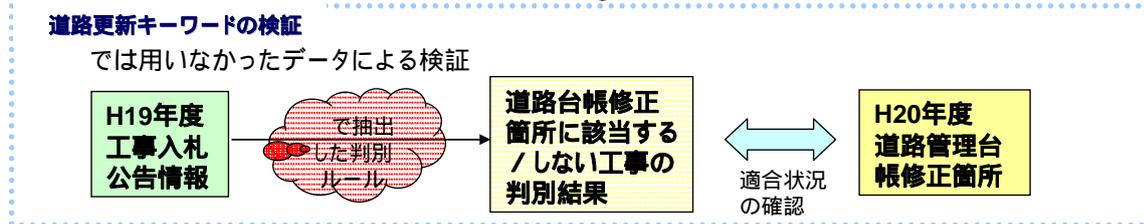
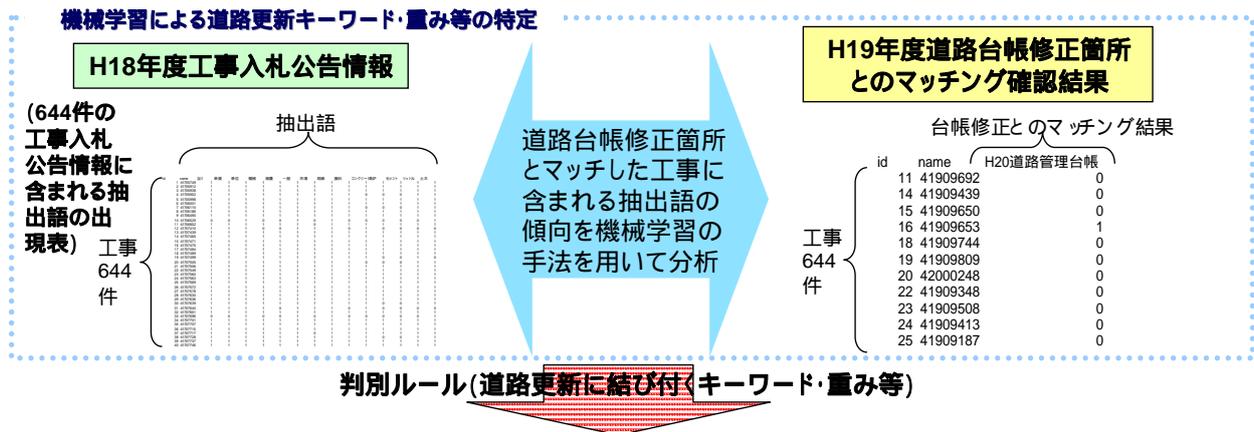


図-31 機械学習の手順

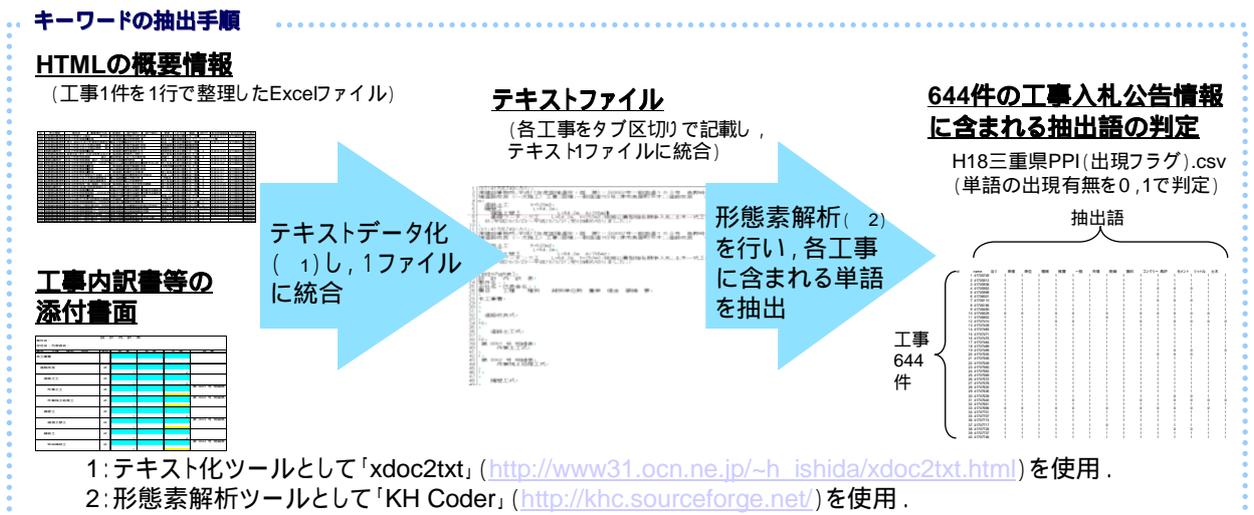


図-32 キーワードの抽出方法

形態素解析では、表-19に示すように、工事入札公告に含まれる語句を抽出することが可能であるが、機械学習に全ての用語を用いると、道路更新に関連しない用語の影響により、機械学習の精度が低下する可能性がある。そこで本検討では、「名詞」および「サ変名詞（名詞+するの形で構成され、動詞化する名詞）」の語句のみを抽出した。

ただし、形態素解析のツールに予め設定されている辞書を用いた場合、抽出される語句が細分化されすぎてしまう傾向があったことから、表-20に示すとおり道路工事に関係する語句やひとつながりとなる語句を目視で確認し辞書登録した。また、形態素解析ツールで抽出された語句に、路線名、地名が含まれている場合には、その語句は機械学習のキーワードには用いないこととした。

機械学習は、基本となるモデルを構築したうえで、他自治体への適用可能性を検討した。そのうえで、精度向上方策の検討を実施した。機械学習に関する検討は次項以降に示す。

表-19 分析の対象とする品詞と抽出語の例

品詞名	抽出語の例
名詞	土木、側溝、道路、車道、構造
サ変名詞	施工、供用、設置、舗装、掘削
形容動詞	特殊、普通、自由、安全、小規模
固有名詞	平成、JIS、久居、白山
組織名	ゼブラ、アスカ
人名	厚、整
地名	三重、松阪、伊勢
副詞可能	前、直接、今、場合
未知語	、W/C、
感動詞	オー、ノー、ま、ウン
動詞	埋める、均す、戻る、固める、掘る
形容詞	厚い、均しい、深い、細い、高い
副詞	別途、普通、又、一層、別に
名詞B	さく、かご、さび、なべ、まち
動詞B	とりこわす、わく、する、とぶ、いる
形容詞B	ない、こわい
副詞B	ふと、また
名詞C	工、価、額、要
否定助動詞	ない、ぬ、ん

品詞名の区分は形態素解析ツール「KH Coder」による。

表-20 辞書登録した語句

カテゴリ	語句の例	登録語句の数
新土木工事積算体系(H19改訂版)事業分野「道路」レベル2・工種	・道路土工 ・地盤改良工	84
H21年度、道路更新判別キーワード	・側溝工 ・道路改良	3
ランキング上位200キーワードを目視で確認して抽出	・ホーリング ・路側部	7
H18、19年度のキーワードを目視で確認して抽出	・災害防除 ・切削オーバレイ工	21

5.3. 基本となるモデルの構築

SVMモデルを構築する際の最適なカーネル関数（linear, polynomial, RBF, sigmoid）やパラメータ（cost, gamma）を設定するために、三重県の工事入札公告をもとにモデルを構築・検証し、比較を行った。具体的には、平成18年度から平成20年度の

工事入札公告をもとにモデルを構築したうえで、平成21年度の工事入札公告を用いて検証を行った。モデル及び検証結果より最適なカーネル関数、パラメータをモデル及び検証結果の的中率と網羅率の観点から確認し設定した。なお、モデル構築（学習）の的中率、網羅率はクロスバリデーション（交叉妥当化）の結果とした。的中率とは、真値に対してモデルが判別した結果がどれだけ正しかったかを表した比率であり、モデルの基本性能を表現するものと考えた。また、網羅率とは、真の道路更新のうちモデルで道路更新と判別したものの比率であり、本モデルの算出結果によりどれだけの道路更新を網羅できるかを表現した指標である。なお、モデル構築は、学習時にはクロスバリデーション（交叉妥当化の検証）を行った上で最適なパラメータを設定し、これを学習には用いなかったデータに適用することよりの的中率、網羅率の評価を行っている。

各カーネル関数でモデル構築・検証を行った結果は表-21に示すとおり。なお、パラメータは初期のパラメータとしている。表-21に示すとおり、学習時の的中率、検証時の的中率とも高い「linear」をカーネル関数として設定した。次いで、パラメータ（cost, gamma）を設定するため、複数のパラメータの組合せで繰り返し計算を行い、的中率を比較することにより、最適なパラメータの探索を実施した（グリッドサーチ）。グリッドサーチでチューニングしたパラメータでモデル構築・検証を行った結果を表-22に示す。

表-21 各カーネル関数の比較

予測値	真値	学習(H18-20)			検証(H21)		
		道路更新	以外	的中率	道路更新	以外	的中率
linear	道路更新	513	35	0.951	71	133	0.691
	以外	84	1792		110	472	
polynomial	道路更新	0	0	0.754	0	0	0.770
	以外	597	1827		181	605	
RBF (ガウシアン)	道路更新	0	0	0.754	0	0	0.770
	以外	597	1827		181	605	
sigmoid	道路更新	0	0	0.754	0	0	0.770
	以外	597	1827		181	605	

表-22 パラメータの設定による結果の比較

予測値	真値	学習(H18-20)			検証(H21)		
		道路更新	以外	的中率 / 網羅率	道路更新	以外	的中率 / 網羅率
Linear (チューニング前:再掲)	道路更新	513	35	0.951/ 0.859	71	133	0.691/ 0.392
	以外	84	1792		110	472	
Linear (グリッドサーチの設定)	道路更新	401	62	0.894/ 0.671	63	87	0.739/ 0.348
	以外	196	1765		118	518	
Linear (網羅率重視の設定)	道路更新	562	797	0.657/ 0.941	143	243	0.642/ 0.790
	以外	35	1,030		38	362	

表-22に示すとおり、的中率は向上したものの、道路更新を伴う工事入札公告を網羅できた率である網

率率が低下した。そのため、道路更新の重み(Weight)を増やして機械学習を行った。道路更新とそれ以外のWeightとcostの様々な組合せのうち最適な設定値を検討した。最適な設定値でモデル構築・検証を行った結果は表-22に示すとおり。グリッドサーチでチューニングしたパラメータで構築したモデルにくらべ、的中率を大幅に落とすことなく、学習、検証の結果何れも網羅率が向上した。

5.4. 他自治体への適用可能性

前項で構築した三重県モデルを他自治体へ適用することの可能性を検討した。

モデルの検証を行うために必要となる各工事入札公告が道路更新を伴ったものであったかの判定データは、表-23に示すとおり平成21年度と平成22年度の三重県下の11市町村と岐阜県下の8市町村の工事入札公告を対象に、各自治体の道路管理者に作成頂いたものを用いている。

三重県下市町村と岐阜県下市町村のいずれも三重県で構築した学習結果と同等的中率を示したものの網羅率は三重県に比べ劣る結果となった。このため、市町村独自のモデル構築を試みることにした。具体的には、三重県下と岐阜県下の市町村の工事入札公告を用いた場合のモデルの構築を試みた。

市町村データは数が少ないため、年次によって学習用と検証用に区別してモデルの構築・検証を行うのではなく、全体の中から3/4のデータを用いて学習、残りの1/4のデータを用いて検証を実施する方法を行った。

表-23 対象市町村

都道府県	対象市町村(順不同)	工事入札公告情報の件数
三重県	四日市市, いなべ市, 伊賀市, 桑名市, 菟野町, 志摩市, 川越町, 津市, 名張市, 鈴鹿市, 松阪市*	930 件
岐阜県	羽島市, 可児市, 岐阜市, 郡上市, 恵那市, 御高町, 高山市, 本巣市	203 件

*松阪市(59件)は添付書面無し。

結果として、三重県下市町村で構築したモデル、岐阜県下市町村で構築したモデル、三重県と岐阜県の市町村を統合して構築したモデルのいずれもキーワード群が安定しているためか、パラメータのチューニングを行う必要のない結果となった。三重県市町村単独モデルの検証結果を表-24に、岐阜県市町村単独モデルの検証結果を表-25に、三重県・岐阜県市町村統合モデルの検証結果を表-26に示す。

表-24 三重県市町村単独モデルの検証結果

予測値	真値	学習			検証		
		更新	以外	的中率 網羅率	更新	以外	的中率 網羅率
linear	道路更新	123	4	0.987	26	20	0.842
	以外	5	564	0.961	17	171	0.605

表-25 岐阜県市町村単独モデルの検証結果

予測値	真値	学習			検証		
		更新	以外	的中率 網羅率	更新	以外	的中率 網羅率
linear	道路更新	45	0	1.000	15	1	0.885
	以外	0	106	1.000	4	32	0.789

表-26 三重県・岐阜県市町村統合モデルの検証結果

予測値	真値	学習			検証		
		更新	以外	的中率 網羅率	更新	以外	的中率 網羅率
linear	道路更新	146	15	0.950	44	18	0.874
	以外	27	659	0.844	18	206	0.710

三重県市町村、岐阜県市町村それぞれ単独で構築したモデルの、市町村レベルでのモデルの汎用性を検討するため、例えば三重県市町村で構築したモデルを岐阜県市町村のデータで検証するといった、相互の検証を実施した。三重県市町村単独モデルを岐阜県市町村のデータで検証した結果を表-27に、岐阜県市町村単独モデルを三重県市町村のデータで検証した結果を表-27に示す。

検討の結果より、都道府県が異なることによりモデルの網羅率が低下しており、モデルのパラメータ設定は、県道・市町村道別かつ都道府県別に設定することにより、それぞれの的中率、網羅率が高いモデルが構築可能であることがわかった。

表-27 三重県市町村モデルを岐阜県市町村のデータで検証した結果

予測値	真値	学習(再掲)			検証(岐阜県下)		
		更新	以外	的中率 網羅率	更新	以外	的中率 網羅率
linear	道路更新	123	4	0.987	18	3	0.759
	以外	5	564	0.961	46	136	0.281

表-28 岐阜県市町村モデルを三重県市町村のデータで検証した結果

予測値	真値	学習(再掲)			検証(三重県下)		
		更新	以外	的中率 網羅率	更新	以外	的中率 網羅率
linear	道路更新	45	0	1.000	54	119	0.746
	以外	0	106	1.000	117	640	0.316

5.5. 道路管理者の協力による精度向上可能性

5.3で構築した三重県モデルを実運用する場合を想定し、道路管理者に何らかの道路更新を示唆するキーワードやフラグを工事入札公告に記載してもらった場合のモデルの精度向上可能性を検討した。今回は道路更新を伴う工事入札公告の20%、50%、80%、100%にフラグを付与する場合で精度の向上が見られるかどうか検討を行った。

モデルの構築は平成18年度の三重県の工事入札公告、検証には平成19年度の三重県の工事入札公告を用いた。

なお、実際に道路管理者にキーワード(フラグ)の記載を依頼した場合に記載ミスが生じる可能性もあることから、付与したフラグの5%を道路更新の無い工事につけ間違っただけと想定して検証を行

った。検証結果は表-29 に示すとおり。

いずれの場合でも、フラグの付与割合に応じて確率的に中率は向上する結果となり、工事入札公告に何らかの信頼性の高いヒントが記載されていればモデルの精度は向上するといえる。

表-29 道路更新フラグを付与した場合の検証 (5%付与間違い)

道路更新フラグ追加 Case4 (辞書 + 1ヵ年出現)ベース =0.03125(全ケースで同じ結果) CV: Cross Validation		学習 (H18)		検証 (H19) フラグ付与なし		検証 (H19) 20%付与		検証 (H19) 50%付与		検証 (H19) 80%付与		検証 (H19) 100%付与	
		真値		真値		真値		真値		真値		真値	
		更新	以外	更新	以外	更新	以外	更新	以外	更新	以外	更新	以外
フラグ 付与 なし	道路更新	201	14	55	138	55	138	55	138	55	138	55	138
	以外	13	416	161	613	161	613	161	613	161	613	161	613
	cost=0.25, CV=0.623	0.958		0.691		0.691		0.691		0.691		0.691	
フラグ 20% 付与	道路更新	203	11	50	119	55	119	71	119	86	119	100	120
	以外	11	419	166	632	161	632	145	632	130	632	116	631
	cost=0.25, CV=0.735	0.966		0.705		0.710		0.727		0.743		0.756	
フラグ 50% 付与	道路更新	201	4	38	94	52	95	79	95	104	95	119	99
	以外	13	426	178	657	164	656	137	656	112	656	97	652
	cost=0.25, CV=0.660	0.974		0.719		0.732		0.760		0.786		0.797	
フラグ 80% 付与	道路更新	211	2	27	86	46	86	87	87	125	90	145	94
	以外	3	428	189	665	170	665	129	664	91	661	71	657
	cost=0.25, CV=0.833	0.992		0.716		0.735		0.777		0.813		0.829	
フラグ 100% 付与	道路更新	214	0	18	48	48	49	93	52	138	55	167	57
	以外	0	430	198	703	168	702	123	699	78	696	49	694
	cost=1.0, CV=0.888	1.000		0.746		0.776		0.819		0.862		0.890	

5.6. 案件名称のみを用いたモデルの構築

工事入札公告の中には公開されている情報が、案件名称のみといった情報量の少ない自治体も存在することから、5.3で構築した三重県モデル構築の際に対象とした工事入札公告の案件名称のみを用いてモデルを構築し、検証を行った。なお、モデルの構築に当たっては、道路更新を重視するパラメータのチューニングを行っている。

案件名称以外の情報を含めて構築を行った場合に比べ、学習時点で精度が上がらない結果となった。これは判定材料が不足しているためと推察される。検証結果も的中率57%・網羅率70%と共に下がる結果となった。三重県データを用いて検証した結果を表-30に示す。

また、市町村についても案件名称のみを用いてモデルを構築した場合も検証を行った。三重県市町村単独モデル（案件名称のみ）の検証結果を表-31に、岐阜県市町村単独モデルの検証結果（案件名称のみ）を表-32に、三重県・岐阜県市町村統合モデル（案件名称のみ）の検証結果を表-33に示す。

表-30 案件名称のみを用いたモデルの検証結果(三重県データ)

真値 予測値		学習(H18-20)				検証(H21)			
		道路更新	以外	的中率	網羅率	道路更新	以外	的中率	網羅率
Linear	道路更新	576	861	0.636	0.965	126	284	0.569	0.696
	以外	21	966			55	321		

表-31 三重県市町村単独モデルの検証結果(案件名称のみ)

真値 予測値		学習			検証		
		更新	以外	的中率 網羅率	更新	以外	的中率 網羅率
linear	道路更新	115	42	0.921	38	20	0.893
	以外	13	526	0.898	5	171	0.884

表-32 岐阜県市町村単独モデルの検証結果(案件名称のみ)

真値 予測値		学習			検証		
		更新	以外	的中率 網羅率	更新	以外	的中率 網羅率
linear	道路更新	43	3	0.967	18	4	0.903
	以外	2	103	0.956	1	29	0.947

表-33 三重県・岐阜県市町村統合モデルの検証結果(案件名称のみ)

真値 予測値		学習			検証		
		更新	以外	的中率 網羅率	更新	以外	的中率 網羅率
linear	道路更新	131	39	0.904	43	24	0.850
	以外	42	635	0.757	19	200	0.694

5.7. データの蓄積による精度向上の可能性

構築したモデルを実際に運用する場合を想定し、運用年次が増えるにつれデータが蓄積した場合の精度向上の可能性を検討した。

三重県の平成18年～20年度までのデータを蓄積

した場合の的中率を比較した。3ヵ年分のデータ蓄積の学習、検証結果を表-34に示す。なお、パラメータの設定はグリッドサーチで設定した値とした。学習時はモデル構築の対象を増やすと的中率は低下するが、検証ではモデル構築の対象を増やすことでの中率が向上することが明らかとなった。したがって、モデル構築の対象を増やしていくことにより、学習時の的中率に検証時の的中率が近づいていくと推察される。

表-34 3ヵ年のデータ蓄積による検証結果

			学習		的中率	検証		的中率
			真値			真値		
			道路更新	以外	道路更新	以外	道路更新	以外
添付 画面 あり	1年のみ(H18)学習 (H19検証)	道路更新	208	9	0.977	51	124	0.701
		以外	6	421		165	627	
	2年分(H18,19)学習 (H20検証)	道路更新	343	44	0.919	67	97	0.758
		以外	87	1137		100	549	
	3年分(H18～20)学習 (H21検証)	道路更新	436	60	0.909	67	89	0.742
		以外	161	1767		114	516	

5.8. 大規模工事のみの抽出可能性

道路更新を推定する際に、推定の対象を限定することにより、推定の精度が向上するかを検討するため、道路更新に結びつく工事のうち大規模な工事に絞った場合の精度向上可能性を検討した。工事入札公告に対応する道路管理台帳の修正データを特定のうち、修正区分から判断して大規模な工事と考えられるもの(三重県の道路管理台帳修正区分がA、B、E)で、モデルを構築し検証した。

なお、三重県における道路台帳修正の区分Aは現道に沿い拡幅施工するもの(用地買収のあるもの)、区分Bは道路改良工事のうちバイパスおよび都計街路等、新しく道路図を作成するもの。またはルート変更によるもの、区分Eはルート変更のうち道路現況図(1/1,000、図巾100m)があるものと定義されている。

大規模工事のみを抽出する場合の精度向上可能性の検証結果を表-35に示す。なお、パラメータの設定はグリッドサーチで設定した値とした。道路更新に結びつく工事が減った結果、道路更新以外に判定されてしまう工事が増加した。学習では高い的中率となるが、検証でやや下がる傾向は、三重県モデルと同様の結果といえる。

表-35 大規模工事のみを抽出する場合の検証結果

	学習(H18,19,20データ)			検証(H21データ再作成)		
	真値		的中率	真値		的中率
	道路更新	以外		道路更新	以外	
道路台帳 修正区分 がABEの 工事	153	4	0.978	14	27	0.894
	49	2,218		56	689	

5.9. まとめ

三重県の平成18～20年度の工事入札公告でモデル構築を行い平成21年度の工事入札公告で検証した結果、ある工事入札公告が道路更新に資するかどうかを判別する率(的中率)約65%、モデルにより

道路更新に資すると判別された情報が道路更新の真値を網羅する率（網羅率）約80%のモデルを構築することができた。同じデータで単純なキーワード検索により判定を試みた場合の的中率，網羅率はそれぞれ約50%，約75%であり，手作業での判別と比較し精度の高いモデルが構築できた。また，データの蓄積によりの中率や網羅率がさらに向上することも明らかとなっており，さらなる精度向上のためにはより多くのデータ蓄積が望まれる。また，三重県の工事入札公告で構築したモデルにより，サンプルとして抽出した他の自治体の工事入札公告で検証しても，網羅率は低いものの，同等の的中率を示すことを明らかとしており，構築したモデルを他地域へ展開できる可能性があることがわかった。また，添付書面が公開されていない自治体を対象にモデルを構築した結果，的中率は概ね1割程度低下する傾向にあることも分かった。

なお，運用面で工事入札公告を公開する際に一部の道路管理者に道路更新の有無をフラグやキーワードとしてあわせて提供して頂くことが仮に実現された場合の検討を行った結果，モデルを構築する工事入札公告のうち20%の工事であっても，道路管理者の協力が得られれば，全体の的中率は向上し，また，協力が得られる割合が向上するほどの中率の精度も向上することがわかった。

第6章 実証実験(3): 道路工事図面の収集実験

6.1 実験概要

6章では、道路工事図面の収集実験について述べる。これは2章でも述べたように、大きく分けると2つの側面があり、ある場所の詳細形状そのものを伝える場合と道路更新情報そのものをメタデータとして伝える際の参考資料として添える場合である。とくに道路工事図面は未公開なケースが多いため、道路管理者と協力関係を築きながら進めることが重要である。これらには既存の情報収集の枠組みにのるか、新たな枠組みで実施するか、いくつかタイプがあるが、本章はこうした試行錯誤の過程をまとめたものであるが、6.2では道路管理者に直接アップロードしてもらう方式について試行した。また、6.3ではアップロードしてもらった図面からいかに自動的に道路更新情報を抽出するかの実験を行った。さらに、6.4では情報公開請求を用いた場合の調査を行った。最後に、6.5では国の「デジタル道路地図更新業務」に関わる枠組みを用いた収集実験とこれらを踏まえた今後の全国展開への考察を述べた。

6.2 Webによる管理者側からのアップロード

ここでは、新たな枠組みで実施することとして、岐阜県と県下の市町村に参加頂き説明会を実施し、道路更新情報の重要性を理解頂くとともに、任意参加で趣旨に賛同してもらった場合は、道路更新に該当する工事の図面情報をWebサイトからアップロードしてもらった。

図-33がその画面であるが、入力を最小限に留めるように、担当者の情報以外に、タイトルと位置に関する情報を自由記述で必須とした。ただしプロードバンド回線が使えないあるいは、一部あるいは全部の電子データがない等、何らかの原因による場合は郵送等も認める方式で行ったところ、65%の市町村は参加をしてもらい、そのうちの2/3にあたる42%はアップローダーを利用して工事図面情報を送って頂く事ができた(図-34)。それ以外は図面の電子データを入れたCDROM等の電子媒体あるいは紙の送付であったり(合計13%)、実際には対象工事が存在し

ない、あるいは何も返信がないなども10%存在した。

また、送付したサイズの容量についてまとめたものは図-35であるが、8割のファイルが5Mバイト以内、9割のファイルが10Mバイト以内に収まっており、当初CADファイル等は容量が大きいのでは、という声が説明会でも見られたが、概ね問題がなかった。

これらのことからWebサイトを通じて管理者からアップロードをしてもらう事は技術的にはそれほど問題がないことは確認できたものの、説明会をして趣旨を直接伝えることや県サイドの強い理解があつてこそ進められた面もあり、サステナビリティの観点からも、新たな枠組みで実施するというよりは、既存の何かの枠組みと関連しながら、こうした方法を実施することが重要に思われた。

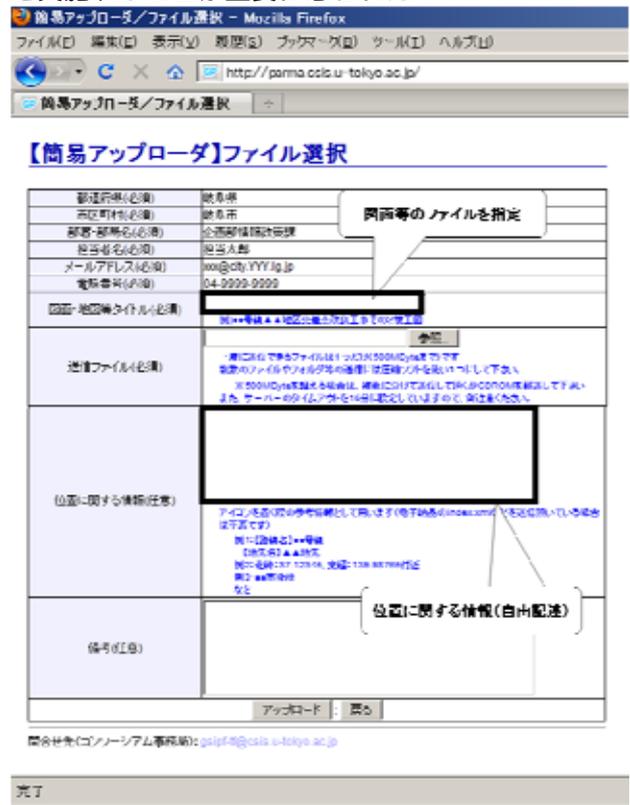


図-33 アップローダー画面

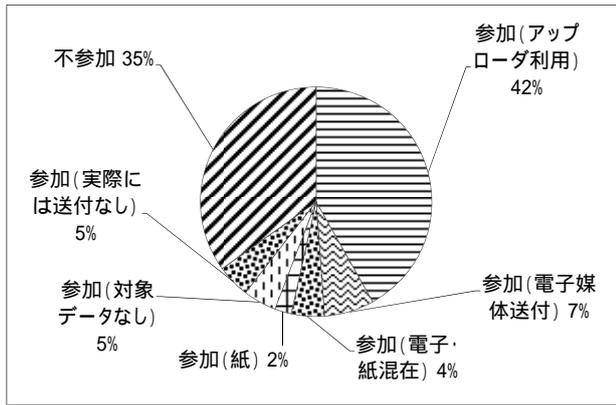


図-34 アップローダの利用状況

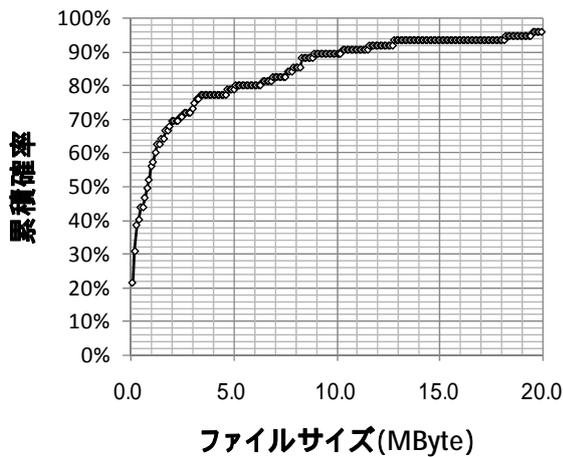


図-35 ファイルの容量の関係

6.3 情報公開請求を用いた方式

ここでは、既存の枠組みとして情報公開請求を用いた方式を行った。すなわち情報公開法にもとづいて、総務関係の窓口へ請求を行うものである。一見これは一番よさそうに思えるものの、意外と簡単ではない。図-36は関東1都7県と三重県、岐阜県の市町村の情報公開請求への対応状況を各HPから調べてまとめたものである。図からわかるように外部からの請求を認めないことや窓口以外の郵送やメールでの申請を認めない団体も少なからずあるという点は効率化という点ではボトルネックになり得る。

ただし注意点としてはこれらはHPからの調査であるため、HPには記載していないものの、実際に確認すると上記のようなケースでも請求可能というケースはある。実際に岐阜県では県サイドから各市町村に確認してもらったところ、全市町村で情報公開制度は存在し、8割程度の市町村では外部からの請求を認め、7割程度では郵送またはメールでの請求を認めているとの回答があった。

また、もう1点留意点がある。情報公開請求の対象となり得るのは公文書のみという点である。工事図面そのものは2章で述べてきたように道路法や工事契約関連で直接対象となっているものなので公文書であるため、理屈上個別には請求可能であるが、対象工事等を指定する必要があり実質的には対象工事

を網羅的にわからないため、指定することが困難である。一方で6.1で述べたような道路更新情報の一覧のようなものは、公文書でないため、入手が基本的には不可能である。

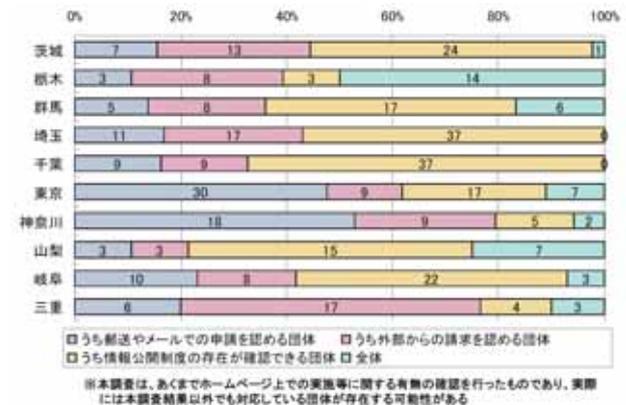


図-36 情報公開請求への対応状況

6.4 図面からのメタデータの抽出試行

ここでは、道路工事図面が仮に入手できたと想定して、そこから自動的に様々な道路更新に関するメタデータが生成できるかどうかを試行した。具体的には、入手できた岐阜県と三重県とその市町村の図面合わせて約850枚から図-37のような4章の工事発注見通し情報の抽出と類似したプロセスを経て、メタデータの抽出を行った。その結果のキーワードの発生状況を表したのが表-36の通りであり、ある程度最適なキーワードに発生が集まっていたようである。また、表-37が抽出状況であるが、CADファイルの国の標準フォーマットであるP21をPDFに変換するのは大変困難であったため、今回は対象外としたが、PDFに変換したものから何らかのキーワードが抽出できたものは5~6割存在し、今後メタデータの自動生成について何らかの可能性があると見える。

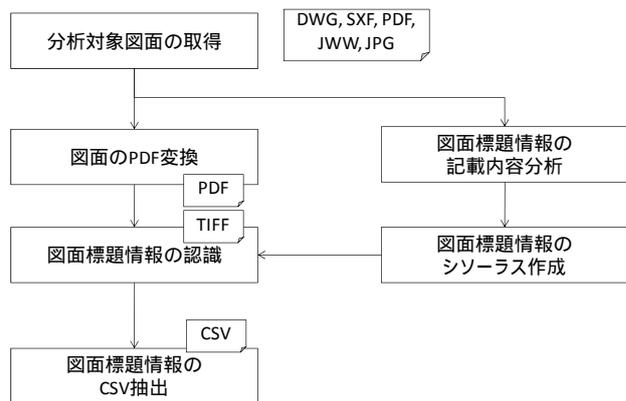


図-37 道路工事図面の処理フロー

表-36 図面からのメタデータ抽出結果

項目	該当キーワードの発生状況
工事名	業務名(2), 業務名/工事名(16), 工事名(72), 事業名(1)
図面名	図面の種類(64), 図面の名称(2), 図面種類(1), 図面名(13), 図面名称(10), 名称(2)
年月日	作成年月(1), 作成年月日(3), 設計年月(2), 設計年月日(10), 測量年月日(1), 年月日(8)
尺度	尺度(9), 縮尺(81),
図面番号	図番(2), 図面番号(91)
会社名	会社名(31), 測量業者名・設計業者名(2), 調査機関名(1)
事務所名	事業者名(3), 事業所名(1), 事務所名(69), 発注者(1),

	法務局名(1)
施工箇所名	箇所名(1), 工事箇所(5), 工事場所(7), 施工箇所(8), 施工箇所名(52), 施工箇所名(3), 施工場所(3), 場所(2)
路線・河川名等	河川・路線名等(1), 河川名 路線名(2), 路線・河川名等(14), 路線名(14), 路線名・河川名等(5)
事業名	事業名(1)
工事番号	工事番号(4)
設計者	設計者(3), 設計製図(2), 調査者(1)
担当者	課長(1), 課長補佐(1), 係長(1), 測量担当者(2), 管理担当者名(2), 照査技術者名(2), 担当者(1)

表-37 図面からのメタデータ抽出結果

	三重県			岐阜県		
	抽出対象数	PDF変換成功数	抽出数	抽出対象数	PDF変換成功数	抽出数
DWG	31	44	34	29	127	51
SFC	13			42		
P21	201	対象外	---	---		
JWW	---	---	---	55		
XDW	---	---	---	1		
PDF	321	321(変換なし)	153	87	87(変換なし)	81
JPG	72	対象外	---	---	---	---
合計	638	365	187(51%)	214	214	132(62%)

6.5 国の枠組みを用いた全国化への展望

さらに、他の既存の枠組みとして、国がデジタル道路地図更新業務の一環で地方整備局単位で、都道府県と政令指定都市から図面を収集する通達が存在している。それはあくまで国-地方自治体間のやり取りであり、他者の利用は想定していないが、既存の収集ルートがあるとすると、その枠組みを使えば、地方自治体側が改めて図面等を他者に提供するような手間をかなり減らせる可能性があるため、それを目指すこととした。

具体的には全都道府県に対し、道路更新情報の趣旨を説明した上で、『現在、都道府県が地方整備局に提出している図面を他者(この場合、東京大学空間情報科学研究センターが事務局を行う「地理空間情報流通実験コンソーシアム」)が公益の目的で利用することとし、国あるいは管理している公益法人「デジタル道路地図協会」から提供すること』に対して、

承諾してもらえるかどうかを聞いたところ、半数以上の25道府県が承諾してもらうことができた。この結果、表-38で示すように、平成21年度収集分について全国で1,460件分(約3,000強の図面ファイル)を登録することができた。

これらから言えることは、本節で述べた道路管理者の既存の枠組みに乗ることは有効な点である。ただし、この枠組みは現在は市町村はまだ任意参加であり、加わっている状況は低いため、とくに市町村に対しては何らかインセンティブを持って加わってもらう状況が必要である。そうした時に6.1や6.2で試行してきた技術と組み合わせることによるメリットが出てくるように思われる。従って、今回の試行により全国展開に可能性が開ける組み合わせを見出せたことは大変大きいと言えよう。

表-38 図面情報の収集状況(平成21年度国土交通省デジタル道路地図更新業務関連)

	更新データリスト	位置図	平面図	ファイル数/登録件数	登録件数内訳
北海道	Excel(24),PDF(3)	PDF	PDF,TIFF,CAD	359 / 164	開発局(116), 北海道(48)
東北	PDF(5)	JPEG	TIFF	471 / 130	整備局(19), 青森県(44), 岩手県(17), 山形県(17), 福島県(33)
関東	Excel(19),PDF(14)	PDF,TIFF,Excel	PDF,TIFF,CAD,BMP	1950 / 579	整備局(55), 群馬県(6), 山梨県(39), 千葉県(29)
北陸	Excel(8),PDF(2)	PDF,TIFF	PDF,TIFF,CAD		整備局(45), 富山(24), 新潟(106)
中部	Excel(30)	PDF,TIFF	PDF,TIFF,JPEG,BMP		整備局(66), 岐阜県(52), 三重県(103), 静岡県(54)
近畿	PDF(2)	PDF	BMP,PDF	90 / 87	整備局(87)
中国	PDF(6)	Excel	BMP,TIFF	87 / 115	整備局(27), 広島(27), 山口(14), 鳥取(25), 島根(22)
四国	PDF(2)	PDF	BMP,PDF,CAD	98 / 105	整備局(73), 愛媛県(52)
九州	PDF(7)	PDF	BMP,PDF,CAD, Word	215 / 236	整備局(47), 宮崎(49), 熊本(40), 佐賀(29), 鹿児島(33), 大分(26), 長崎(12),
沖縄	PDF(4)	PDF	BMP,PDF	37 / 24	沖縄県(24)

第7章 道路管理者の提供形態のあるべき姿

7.1. 道路管理者との協力体制構築可能性の検討

道路管理者との協力体制構築の可能性として、情報公開制度を活用する手法とデータ登録システムを活用する手法を検討した。その結果、情報公開制度は、当該自治体在住者のみを申請対象としているケースや、申請や情報の受け取りに現地まで出向く必要があるケースが存在し、本制度の活用のみでは網羅的に情報を収集することは困難であることがわかった。

また、データ登録システムを活用した情報の登録は、岐阜県の協力を得て試験的に市町村の管理者にデータ登録の協力を依頼際に、継続的に運用するためには登録者へのインセンティブ付与が必須であること等がわかった。

7.2. 道路更新情報流通実現へ向けた取り組みの基本方針

道路更新情報の流通を実現するために、第一には、電子的な情報の流通量を拡大することが重要である。特に市区町村では電子的な情報流通量が圧倒的に少ないため、重点的に情報の流通を促進する必要がある。第二には、様々な方策・情報を組み合わせることにより、トータルとして道路更新情報の流通拡大を実現することである。実際に運用することを考えると、一つの方策・情報だけでは完全な道路更新情報の流通を実現することは難しい。様々な方策・情報の流通を拡大しておき、これらのうちのどれかを用いて道路更新が把握できるような仕組みを構築することが望ましい。

7.3. 管理者（国・地方公共団体等）への要望事項

(1) 要望：電子的な情報流通の推進

管理者における道路更新に関する様々な情報は、可能な限り電子的な形式での公開を求める。情報の提供形式は、中のテキスト情報を読み取る

ことも可能であることから、単純なPDF形式で構わない。加えて、例えば工事入札公告情報であれば、工事内訳書や図面等の詳細情報も関係者限りの電子入札システムではなく、一般に広く情報が収集可能な環境への公開をお願いしたい。

電子的な情報の公開により、Webクローリングによる自動収集・情報の集約化が可能となることで、人手による問い合わせの削減が期待できる。また、電子的に公開されている情報量の増加に伴い、情報が蓄積されることで高い精度の道路更新判定が可能となる。

(2) 要望：必要な情報を備えた形式での情報流通の推進

管理者における道路更新に関する様々な情報は、可能な限り道路更新か否かを判断する情報や位置情報を含めての情報提供を求める。

供用開始の公示情報は道路更新情報であるのか、隅切りを直した程度であるのかを示してほしい。このことにより、法的根拠に基づき網羅的な道路更新情報提供手段である供用開始の公示情報をより利用価値の高い形式に運用上で改善することができる。

また、道路更新箇所的位置を地図上でデータとして特定できるようにしてほしい。具体的には様々な図面等に位置情報を追加すること、住所の情報は省略せず、かつ住居表示法に則った住所を記載すること、例えば道路の共通位置参照方式を活用する等して道路にユニークなIDを付与すること等を想定している。提供された情報の「位置」がデータとして特定できるようになることで、様々な情報の組合せ可能性が高まる。併せて、これらの必要な情報を備えることにより、問い合わせ件数の削減が期待できる。

(3) 要望：標準フォーマットに則った形式での情報流通の推進

電子的なデータの再利用可能性向上を図るため、

予め標準となるメタ情報を定めた上で、これらの情報をRSS形式で表現した情報流通を求める。

具体的には、メタ情報の構成案を図-38に、RSSのChannel定義及びitem定義の素案を表-39に示す。

標準となるメタ情報のフォーマットを定めることで、情報の活用可能性の向上とRSS形式での情報流通推進により、問い合わせに頼らない道路更新情報流通の実現を可能にする。

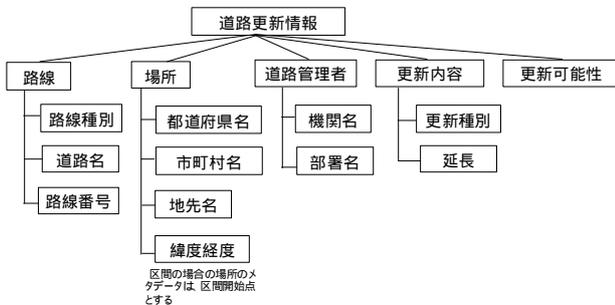


図-38 道路更新に関するメタ情報の構成案

表-39 Channel 及び item 定義素案

要素	記載内容	記載例
title	「道路更新情報」と記載。	<title>道路更新情報</title>
link	当該情報についてのリンク先。システム（配信サイト）に対応するウェブサイトのURL	<link>http://www.xx.com</link>
description	システム（配信サイト）の概要	<description>県の道路開通情報サイト</description>
要素	記載内容	記載例
Title	当該道路更新の『路線』と『更新内容』を続けて記載。	<title>三重県道 号線新規開通</title>
description	当該道路更新の『場所』『管理者』『道路更新可能性』をカンマ区切りで記載	<description>三重県 町 番地先,三重県,1</description>
georss:point	道路更新の位置。緯度経度をGeoRSSでホワイトスペースで区切られたペアとして半角英数で記載。	<georss:point>43.082931 141.520386</georss:point>

7.4. 産学官連携による実現方策の提案

(1) 情報流通基盤の構築・運用

情報を一元的に集約する基盤の構築・運用を提案する。

情報流通基盤として、本研究会および地理空間情報流通実験コンソーシアムで構築したシステムの更なる改良および継続的な運用を求める。ただし、継続的な運用にあたっては、道路更新情報収集の更なる自動化を図るための技術面の課題と、システム運用モデルの具体化を図るため、費用負担を含めた運用体制構築の課題解決を図る必要がある。

一元的な集約を実現することにより、市区町村等

多数に及び情報提供主体からの情報を利用者が一括して入手できることや、道路更新情報以外の情報と組み合わせることができる等、ワンストップでの情報提供が期待できる。

(2) 継続的な道路更新情報流通に関する実証実験の実施

一時的な「実験」ではなく、継続的な運用を前提とした取り組みの実施を提案する。

一部地域を対象とした、管理者と利用者（および提供地域等）にとってメリットのある先進的な具体化事例の構築を求める。

これら事例の構築により、管理者や利用者をはじめとする、様々な管理主体のメリットにつながる情報流通のあり方を具体化することができる。例えば、管理者や利用者の双方が一つのデータを修正し合うような運用もイメージする。

また、実証実験のほか、他分野や海外における成功事例・検討例の整理等も有効であると考えている。

(3) 市民や関連主体への周知・広報活動の実施

様々な関連主体を幅広く巻き込むオープンな活動、ムーブメントの構築を提案する。

道路更新流通に関する主要施策や主要イベントとの連携や道路以外の更新情報流通への展開を求める。

公の団体以外にも、全国に散らばる個人を含めた有志をつなぐコミュニティの形成等を通じて本活動の推進者や支援者を増強し、活動を面で支える体制の構築を目指す。

第8章 まとめ

本研究では、新しいステージの ITS を支える基盤としてのデジタル道路地図を地方自治体の道路を含めた迅速な更新の仕組みを目指すものとして、現状の道路管理者の情報提供の実態や民間企業等利用者のニーズを明らかにしつつ、道路更新情報を自動的に収集・提供するシステムを構築し、いくつかの実証実験を行った。

具体的には、1章で研究全体の構成を概観し、2章で道路更新情報の定義、体系化を行うとともに道路管理者から提供される情報の実態として、国、都道府県、市町村の間のかなりの差を明らかにした。また、民間企業等利用者のニーズを明らかにし、地図調整業者等は一業者あたり年間約千件の図面等を道路管理者から入手していることがわかった。

3章では、それらをもとに、区域の供用開始告示、発注見通し、入札公告、道路竣工図、道路開通情報等、道路更新情報の収集・提供システムの機能要件等をまとめ、クローリング、フィルタリング、路線ジオコーディングなどの機能を構築した。例えば、路線名と地先名を組み合わせ、路線上の位置を特定することを目指した路線ジオコーディングは、県道以上ではレベル1以上（路線が特定され字レベルに特定）が75%程度で、市町村道ではレベル2以上（もともとデータベースが存在しないため、路線は特定できないが、字レベルで特定できる）が75%でかなりの詳細なレベルで位置が特定できることが明らかになった。

また、4章以降は実証実験としていくつかの種類の方法の収集等を行った。まず4章では、工事発注見通し情報を全国ベースでクローリングから道路更新情報に関わる情報の抽出まで一連の自動収集を行い、公開している約1200の自治体を対象に2.7万件の道路更新に関連すると思われる発注見通し情報を登録した。次に、5章では工事入札公告情報を3県を対象に収集を行い、とくに一部で入手することができた道路更新に関連したかどうかの実績情報をベースに機械学習（SVM：Support Vector Machine）の手法を用いて、道路更新に該当するか

どうかの推定を行い、74%の的中率を得ることができた。また、6章では道路工事図面情報の収集作業そのものの効率化について、道路管理者の自主的な提供、情報公開請求による提供、国のデジタル道路地図更新業務に関連した参考提供等、いくつかの制度に基づき収集実験を行い、国のデジタル道路更新業務に関連した枠組みと連携することで1,460件の図面を収集することができ、この方法が効率的という結論を得た。また、図面から標題欄にある道路更新情報に関連する工事名、場所、時期、発注者名等、図面のメタ情報作成作業の効率化など試み、今後の全国展開への可能性を検討した。また、こうした実証実験をもとに、7章で道路管理者のあるべき情報提供の姿をまとめた。

謝辞：本研究は国土交通省新道路技術会議から「サービスイノベーション型空間情報社会基盤に関する研究開発」というテーマで支援を頂き、評価委員・事務局の皆様には様々な有益なアドバイスを頂いた。また、担当の国土技術政策総合研究所高度情報化研究センター情報基盤研究室の皆様には色々と調査等に関わる多大な御協力を頂いた。さらに、検討の過程では、東京大学空間情報科学研究センターで開催した「道路更新情報流通推進研究会」や「地理空間情報流通実験コンソーシアム」において、柴崎亮介委員長を始めとして、多数の委員の皆様から様々な有益な御意見を頂いた。これらの皆様に心より感謝を致したい。

参考文献

- 1) ITS Japan : 安全・環境に資する走行支援サービス実現のための道路情報整備と流通へ向けた提言 ,
<http://www.its-jp.org/topics/topics069.htm>
- 2) ERTICO : Feedmap プロジェクトホームページ ,
<http://www.ertico.com/en/subprojects/feedmap/home/>
- 3) 関本義秀, 金澤文彦, 松下博俊, 次世代デジタル道路地図のあり方に関する研究, 国土技術政策総合研究所資料, ISSN1346-7328, No.372, 2007.
- 4) 国土交通省 : 道路開通情報ホームページ ,
<http://www.mlit.go.jp/road/kait-su/index.html>
- 5) 沖縄総合事務局 : 道路図面情報提供サービス ,
<http://www.dourozumenjoho.dc.ogb.go.jp/road/counserv/counservice.aspx>
- 6) 中條覚, 関本義秀, 南佳孝, 柴崎亮介, 道路更新情報に関するニーズと情報提供の実態について, 第 29 回交通工学研究発表会論文集, pp.305-308, 2009.
- 7) Yoshihide Sekimoto, Katsumi Uesaka, Prompt development and updating of Road GIS data integrated into the public works, Proceedings of 11th World Congress on Intelligent Transport Systems, San Francisco, CD-ROM, 2005.
- 8) Satoru Nakajo, Yoshihide Sekimoto, Yoshitaka Minami, Harutoshi Yamada, Ryosuke Shibasaki, Getting broad overview of road update from procurement notices of road constructions, Proceedings of 15th World Congress on Intelligent Transport Systems, Stockholm, CD-ROM, 2009.
- 9) 布施孝志, 松林豊, 中條覚, 高橋香織, 脇嶋秀行, 山口章平: 公示情報に基づく道路更新情報のクロージングシステムの検討, 土木情報利用技術論文集, Vol.18, pp.281-290, 2009.
- 10) 道路法 : <http://law.e-gov.go.jp/htmldata/S27/S27HO180.html>
- 11) 財団法人日本デジタル道路地図協会 : デジタル道路地図データベース , <http://www.drm.jp/>
- 12) 財団法人道路管理センター : 道路管理システム ,
<http://www.roadic.or.jp/>
- 13) 公共工事の入札及び契約の適正化の促進に関する法律 , 2000 .
- 14) 例えば国土交通省関東地方整備局 : 土木工事共通仕様書 ,
<http://www.ktr.mlit.go.jp/kyoku/tech/shiyousho/shiyousho.htm>
- 15) 国土交通省 : 道路開通情報ホームページ ,
<http://www.mlit.go.jp/road/kait-su/index.html>
- 16) 国土交通省 : 道路交通センサからみた道路交通の現状, 推移 , http://www.mlit.go.jp/road/ir/ir-data/data_shu.html
- 17) 財団法人日本交通管理技術協会 : デジタル交通規制データ ,
<http://www.ktr.mlit.go.jp/kyoku/tech/shiyousho/shiyousho.htm>
- 18) 財団法人交通事故統計分析センター : 交通事故統計データ , <http://www.itarda.or.jp/data/syukei.html>
- 19) 例えば国土交通省九州地方整備局 : 道路の走りやすさマップ , <http://www.qsr.mlit.go.jp/kyukan/map05/model.html>
- 20) 財団法人道路交通情報通信システムセンター : Vehicle Information and Communication System (VICS) ,
<http://www.vics.or.jp/>
- 21) 相良毅, 有川正俊 : 「日本の住所体系に適した分散アドレスマッチングサービス」, 地理情報システム学会講演論文集, Vol.9, pp.183-186, 2000.
- 22) 南佳孝, 関本義秀, 中條覚, 柴崎亮介 : 路線情報を加味した道路関連情報の位置特定に関する研究 ~ 工事入札公告を事例に , 第 30 回交通工学研究発表会論文集, 交通工学研究会, pp.265-268, 2010.9.
- 23) 南佳孝, 関本義秀, 中條覚, 柴崎亮介 : 路線情報を加味した位置特定の手法に関する研究-道路関連情報の比較-, 地理情報システム学会講演論文集, CD-ROM, 2010.10.
- 24) 関本義秀, 中條覚, 南佳孝, 山口章平, 山田晴利, 布施孝志 : 工事発注見通し情報を用いた全国における道路更新情報の自動抽出に向けた試み, 土木学会論文集 (投稿中) .
- 25) 例えば, 鶴田雅信, 関根聡, 増山繁 : 企業の公式 Web サイトからの基本情報抽出, 人工知能学会全国大会論文集, Vol. 23, 2009.
- 26) 板井久美, 高須淳宏, 安達淳 : HTML からの情報抽出と統合, NII journal, Vol. 6, pp.9-19, 2002.
- 27) 増田英孝, 塚本修一, 安富大輔, 中川裕志 : HTML の表形式データの構造認識と携帯端末表示への応用, 情報処理学会論文誌. データベース, Vol. 44, No. 12, pp.23-32, 2003.
- 28) 田仲正弘, 石田亨 : 表構造の一般化に基づくオントロジの獲得, 情報処理学会論文誌, Vol. 47 No.5 pp. 1530-1537, 2006.
- 29) 佐藤郁, 渡邊英一, 古田均, 宮口智樹 : マルチエージェントによる建設情報データベース統合化に関する研究, 土木学会論文集 F, Vol. 62 No.1 pp. 13-24, 2006.
- 30) 薄井智貴, 関本義秀, 金杉洋, 南佳孝, 柴崎亮介 : 地理

空間情報活用推進に向けた流通実験システムの開発と
適用, 土木学会土木情報利用技術論文集, Vol.19, pp.
125-132, 2010.

- 31) V.N.Vapnik, Statistical Learning Theory, John Wiley & Sons (1998).
- 32) N.Cristianini, J.S-Taylor, An Introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 2000.

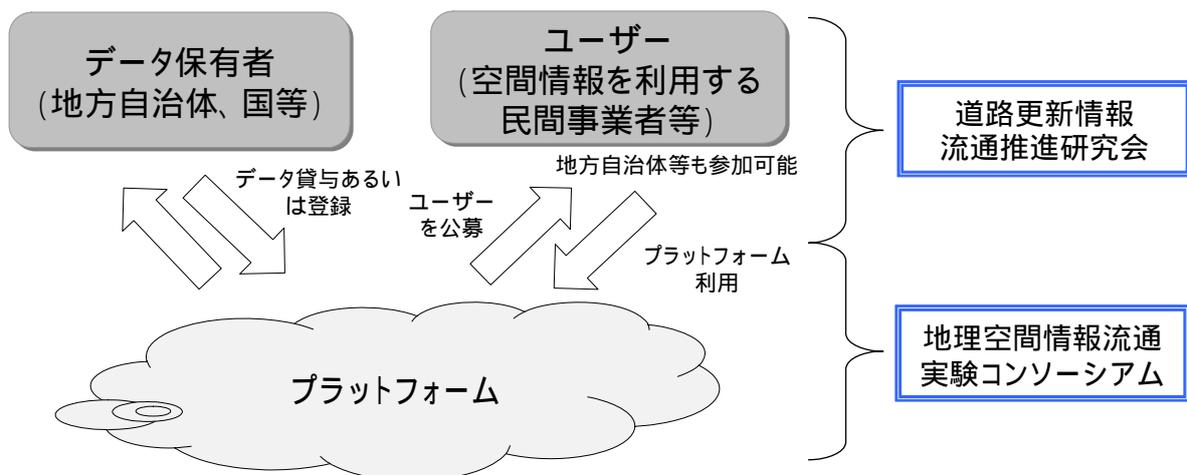
付録 本研究開発の体制

本研究開発では、デジタル道路地図の迅速な更新を契機としてサービスイノベーションにつながるフラッグシップ的な実験プロジェクトを立ち上げることを目的に据え、道路地図コンテンツの流通活性化のためのニーズと課題の整理や地図更新情報の自動収集技術の開発等の研究を行った。研究に際しては、東京大学空間情報科学研究センターに設置された「道路更新情報流通推進研究会」での検討を踏まえるとともに、「地理空間情報流通実験コンソーシアム」と連携し実証を行った。それぞれの活動の概要は以下に示すとおり。

道路更新情報流通推進研究会は2008年から3ヶ年設置され、付録図に示すとおりデータ保有者の現状、ユーザの道路更新情報に対する実態、ニーズの整理を行った。なお、道路更新情報流通推進委員会は産学官の委員25名（第13回研究会時点）で構成し、本研究開発に関してご議論頂いた。

一方、地理空間情報流通実験コンソーシアムは、

2009年から2ヶ年設置され、付録図に示すとおり、地理空間情報を流通するための仕組みの検討、構築を行った。構築したシステムに本研究で開発した技術を組込むとともに、システムを活用して道路更新情報を収集・提供する実証を行った。なお、オープンな場でシステムの検討、構築を行うため、コンソーシアム参加者を広く公募し、125団体182名（1次、2次募集分計）の参加を頂いた。システムへログインしたユーザからは、期間中に平均6.1回のアクセスがあった。



付録図 本研究開発の体制