

II-21 Web マイニング技術を用いた入札情報検索システムの開発

A Development of the Tenders Information Searching System based on Web mining technology

小俣尚泰¹・関根聡一²

Naoyasu Omata, Soichi Sekine

抄録：わが国では入札契約適正化法の施行以来，入札情報の Web 上での公開が進んでいる．しかし，公開方法の違いにより，その有効活用が限定的であると考えられるが，それを促進する一般の Web 検索エンジンでは，即時性や条件指定に問題があり，受注者側の業務には適用しにくい．そこで本研究では，Web を通じて公開された入札情報を横断検索できる入札情報特化型検索システムの構築をした．本システムは発注機関毎に異なる用語や文書形式の違いを吸収し，かつ収集しなければならぬ情報を判断する情報フィルタリング機構を特長とする．本システムを通じて，発注側・受注側双方にとって負担が少なく，効率の良い入札業務の改善を提言する．

キーワード：入札，検索エンジン，情報フィルタリング，Web クローラ，Web マイニング

Keywords : Tender, Web Search Engine, Information Filtering, Web Crawler, Web Mining

1. 開発の背景

(1) インターネット上の入札情報の増加

入札契約適正化法の施行以来，発注見通し・入札公告・落札結果等の入札情報のインターネット上での公開が進んでいる．国土交通省調査によると，2007 年時点では，入札情報のインターネット公開が国・都道府県機関では 100%，市区町村機関では 61.6%となっている．これは年々増大する傾向にあり，入札制度にも情報爆発の時代が到来したといえる（図-1）．しかしながら，公開のされ方によって情報の有効活用が限定的であると考えられ，溢れる入札情報をどのように処理していくかは喫緊の課題である．

(2) 発注機関の情報インフラ投資上の問題

インターネット上の情報公開手法について，標準化の流れとして国土交通省が運営する入札情報サービス（以下，PPI）¹⁾への統合が推進されている．このシステムを通じて全入札情報が PPI へ統合されるのが理想であるが，それに必要となる参加発注機関側でのシステム導入について，特に市区町村などの小規模な機関では費用対効果を期待できないという指摘がされており²⁾，情報インフラ投資面での問題が存在している．その結果，コンテンツ管理システムを通じ Web サーバに公開するのみに留まる機関が多く見受けられ，PPI への統合における問題となっていると考えられる．完全な入札制度の電子化と健全化を目指すには，発注機関におけるシステム導入コストの低減を図ると共に，その欠点を補う仕組みの導入が必要であると考えられる．

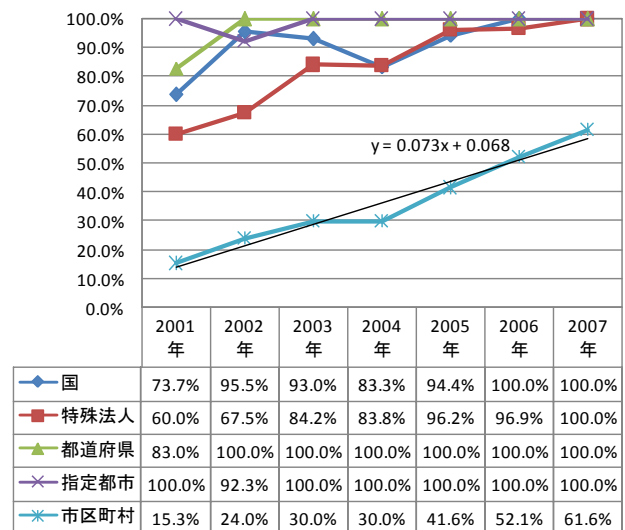


図-1 インターネット公開を行う発注機関数の推移

2. 入札情報検索システムの要件

本研究では，Web サーバを通じて公開される PDF 形式等の文書を収集・分析し，入札情報に特化した Web 検索エンジンの構築を目指す．そこで，入札情報の検索システムとしての要件は，①「情報の鮮度を保つこと」②「情報の正確に探し出せること」③「情報の網羅度を高くすること」④「利便性の高いシステムとすること」の大きく 4 つが挙げられると考える．以下では，これらのシステム要件について，一般的な Web 検索エンジンにて解決しようとする場合での問題点と比較を行いながら課題解決策を述べる．

1：正会員 修士(工学) 株式会社栗本鉄工所 技術開発本部 情報技術グループ
(〒559-0021 大阪府大阪市住之江区柴谷2丁目8番45号, Tel :06-6686-3149, E-mail : n_omata@kurimoto.co.jp)

2：非会員 修士(工学) 株式会社栗本鉄工所 技術開発本部 情報技術グループ
(〒559-0021 大阪府大阪市住之江区柴谷2丁目8番45号, Tel :06-6686-3149, E-mail : s_sekine@kurimoto.co.jp)

(1) 情報の鮮度を保つこと

情報の鮮度とは、情報の更新頻度を指す。例えば、入札公告は入札実施の1週間程度前に掲載され、入札が終了すると削除され入札結果の文書として更新をされる。そのため、Web 検索エンジンでチェックを行うと、いつのまにか入札は終わり入札結果が出ていたという事象が発生する。この問題の解決のため、より短い間隔で更新を行いながら、「情報はなかった」という情報を判断できる Web クローラを構築しなければならない。

(2) 情報の正確性を保証すること

入札情報という限定的な用途であるがゆえに、検索エンジンによる結果の正確性についても問題が生ずる。様々なニーズに対応すべく汎用的につくられているため、入札情報の調査業務に適用するには、検索時に設定できる項目が少ない。そのため得られる検索結果としては、再現率が高いが適合率が低いという状況になり、不必要な結果が多数混ざることとなる。例として、筆者らの所属機関でのアンケート結果（図-2、図-3）によると、工事名称と内容説明に対する全文検索や、予定価格の上下限範囲を指定しての検索などに強いニーズがある。また、入札結果に対しては、落札者や落札金額を指定して検索することにより競合他社の動向を分析したいというニーズも強い。

(3) 情報の網羅度を高くすること

情報の網羅度を考えると、検索するデータ母集団は、受注者側が営業範囲としている発注機関を網羅することが必要である。入手しなければならない情報は、発注見通し・入札公告・入札結果の3つ情報区分いずれに属するかと、どの発注機関のものが明らかな情報である。ここでは、対応する発注機関を増やせば増やすほど、システムが取り扱うデータの規模が大きくなる。クロウリングした Web 文書の網羅度のみを考えれば、一般の検索エンジンでも実現は可能であるが、先述の鮮度・正確性にて不都合が生じる。また、独自に Web クローラを設計した場合は、不要なデータが検索のデータ母集団となってしまうという、データベース構築上の問題が生じ、検索性能に影響する。

(4) 利便性の高いシステムとすること

その他の要件としては、利便性の高いシステムとすることが挙げられる。実現例として、筆者らが開発中のシステム画面を図-4に示し、利便性についての具体的な要件を以下に述べる。

a) 可用性が高いこと

可用性とは使いたいときに使えるかどうかという性質である。ここでは、特別なトレーニングを受けることなく誰でも簡単に利用できること、使用するにあたり特別な環境を要せずとも使用できることが求められる。したがって、配布が容易な簡易なユーザインター

必ず欲しいと思う	◎
あれば良いと思う	△

条件項目	発注見通	入札公告	入札結果
公表日・公開日	◎	◎	◎
発注時期	◎		
工事名称	◎	◎	◎
工事場所			
工期			
工事概要・内容説明	◎	◎	◎
発注方式(単体・JV等)			
参加資格			
申請書交付日			
申請書交付場所			
予定価格	◎	◎	◎
入札執行日	△	△	△
設計図書			
落札者			
落札金額			
入札参加者			
入札参加者の応札額			
入札日			
契約日			

図-2 入札情報検索システムに必要なと思う
検索項目（工事予算担当）

条件項目	発注見通	入札公告	入札結果
公表日・公開日			
発注時期			
工事名称	◎	◎	◎
工事場所			
工期		◎	△
工事概要・内容説明	◎	◎	◎
発注方式(単体・JV等)			
参加資格		◎	
申請書交付日			
申請書交付場所			
予定価格	△	◎	◎
入札執行日			
設計図書		△	
落札者			◎
落札金額			◎
入札参加者			
入札参加者の応札額			
入札日		◎	
契約日			

図-3 入札情報検索システムに必要なと思う
検索項目（設計担当）



図-4 システム例：検索結果画面

フェースを備える Web アプリケーションが望ましいと考える。

b) 可搬性があること

外出の多い営業職からは、モバイル機器からのアクセスについて強いニーズがある。自己の担当エリアの

営業方針に従った情報を、出先でも利用できる環境が整うことはビジネスのスピードを上げるためにも欠かせない項目となる。

c) 情報を自動的に知らせること

検索システムの実現を目指すものであるが、理想的にはユーザが欲する情報を自動配信することが望ましい。なぜなら、期中での営業方針の転換がない限りは、同一のユーザからは本システムへの検索条件指定は同じのものが多く指定されることになるからである。従って、ユーザの検索要求に対する適合率を高く保つことができれば、自動検索を通じてモバイル機器への配信機能の実現が可能となる。

3. 実現のための基礎技術の整理

(1) Web クローラ

a) 動作特性上の問題点

Web クローラとは、インターネット上の Web サイトからハイパーリンクを探索しながら Web ページを収集する機構をいう。図-5に示すように、探索の起点を定めて、次へ収集すべき Web ページを順次取得して処理が進んでいく。ここで問題となるのは、その動作の特性上、取得対象となった Web ページは全て集めてしまうことにより、不要な情報が検索データベースへ登録されてしまうことである。そのため、本稿で想定する検索システムを実現する上では、必要な情報のみを判断する機構を備えた独自の Web クローラが必要となる。

b) 情報フィルタリングの導入による解決

情報フィルタリングとは、大量の情報の中から、ユーザにとって必要な情報を取り出し、不要な情報を除外する処理を自動的に行う技術のことをいう。要・不要の2つに分けるほかに、情報に重要度や類似度などのメタ情報を加えて重み付けを行うものも含まれる。実用されている例としては、迷惑メールの除去フィルタ³⁾等がある。この情報フィルタリング機構を Web クローラが備えることで、情報の選別を自動的に行う入札情報フィルタを構成することができると考える。

(2) Web マイニングによる情報フィルタリング

前節の情報フィルタリングを搭載した Web クローラにて、どのような情報フィルタリング手法を確立するかが重要となる。近年では、Web システムに関わる多種多様なデータに潜むパターンやルールを発見を目標に Web マイニングと呼ばれる研究が行われている⁴⁾。Web マイニングは、その対象とするデータの特徴を踏まえて、以下の3領域に大きく分類されている。

a) Web コンテンツ・マイニング

Web コンテンツ・マイニングは、Web コンテンツ内の各種マルチメディアデータからルールを見つけ出す手

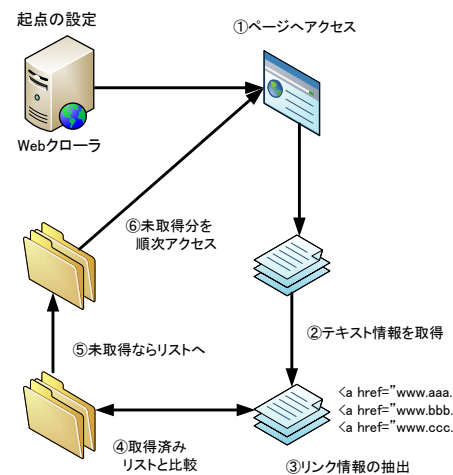


図-5 Web クローラの動作

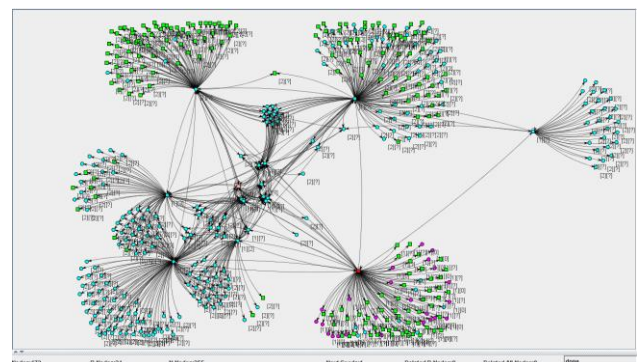


図-6 Web 構造マイニングの例

法であり、Web ページの特徴ベクトル生成を行う前処理を要する。そこで、非構造データとしてのテキストデータに対する自然言語処理技術を用いた解析や、HTMLやXMLによる半構造データのタグ記述を利用したキーワード抽出も行う技術である。内容をもって判断を行う技術として、入札情報フィルタの核となる。

b) Web 構造マイニング

Web 構造マイニングは、Web ページ群を結ぶハイパーリンクからなる Web グラフ解析に基づく特徴の発見の手法であり、興味を同じくする利用者群を見出すコミュニティ形成や、Web ページ群やコミュニティ内の代表ノードの選出に用いられる。図-6のように、ある発注機関の入札情報は、リンク構造上どこに良く現れるか等を解析することに用いることができる。

c) Web 利用マイニング

Web 利用マイニングは、不特定多数のユーザによる Web アクセスのログや、ブックマークなどに記録された行動履歴から、利用者のアクセスパターンやブラウジング目的などを見出すものである。コンテンツ連動型広告などレコメンデーション技術に実用されている⁵⁾。入札情報の調査業務の行動解析より、後述の事前情報の作成に応用できると考えられる。

4. 入札情報検索システムの構成

(1) 概要

本章では、上述の基礎技術を活用した入札情報検索システム（以下、本システム）の構成について述べる。

図－7に示すように、本システムでは処理内容により、①データ収集部 ②データ加工・解析部 ③データ認識部の3段構成を採る。以下、その具体的構成方法について述べる。

(2) 各部の構成

a) データ収集部

ここでは、Web クローラを通じて発注機関の Web サイトから入札情報を入手する処理を行う。クローラの処理結果から入札情報の Web ページとリンク構造のデータが得られる。

b) データ加工・解析部

ここでは、前項 a) で得られた Web ページとリンク構造のデータの解析を行う。具体的には発注機関と情報区分の組に対する特徴の分析を行い、入札情報フィルタを作成する。この特徴の分析を行う際には、得られている Web ページが、どの発注機関であり、どの情報区分に属するかを示す事前情報を作成する必要がある。

c) データ認識部

ここでは、前項 b) の手順により得られた情報フィルタを使用して、再度行われるクローリングの際に未知となる Web ページに対して、情報フィルタリングを行う。このような構成を採ることで、クローラが収集をしてくる不必要な情報を排除し、必要な情報のみを検索 DB に登録されるようにする。

(3) 本システム構成によるメリット

入札情報フィルタを搭載した Web クローラにより適切に情報仕分けを行うことで、一般的な Web 検索エンジンで問題となる情報の鮮度・正確性・網羅度の問題を解決することができる。また、Web クローラ方式により、対象となる入札情報は Web サーバへの Web 文書の公開を行うだけで良いため、発注者側での情報インフラへの投資効率の面で良い効果が得られる。

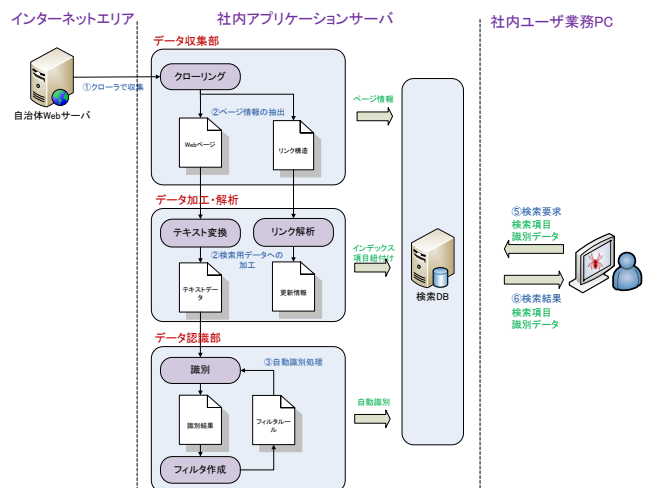
(4) 本システム構成が持つ課題

a) 情報フィルタの精度

情報フィルタの精度が全体の性能に大きく影響するため、その精度の維持は最重要の課題となる。精度維持のためには、情報フィルタで採用する Web マイニング技術の研究開発の発展が求められる。

b) 維持コスト

情報フィルタの精度を維持するためのもうひとつの要因として、4章で述べたデータ分析・加工部で必要とする事前情報の充実がある。この情報の作成は人手を要するため、本システムを維持するためのコスト要因となる。Web データの量は多いため、少量の作業で



図－7 システム構成

高い精度を作成できる情報フィルタの構成法または事前情報の作成方法が必要となる。この入札情報の選別作業は、受注者の各々が実施していることが現状であるため、この事前情報の作成作業を業界にて共同で行う枠組みができれば、前述した発注機関の情報インフラ投資メリットだけでなく、受注者のコスト軽減も実現できる可能性があると考えられる。このようにして、入札情報を得るための情報システムは、発注機関で整備するのではなく、受注者が協力体制を築いて構築していくというオープン志向への転換が望ましいのではないかと考える。また、Web 利用マイニング手法の活用により、事前情報をより簡易に生成することができる技術の開発を通じてのコスト低減も考えられる。

5. まとめと今後の課題

本稿では、入札情報の公開～取得の流れを全体で効率化することを主眼に置き、Web クローラ方式による入札情報に特化した Web 検索エンジンのシステム構成を提言し、プロトタイプシステムを開発した。また、そのメリット・デメリットに言及し、本システム構成は、発注機関・受注者双方にとって、情報インフラに対する投資効率を高める可能性があることを示した。

今後の課題としては、精度の高い情報フィルタの開発と維持コストの低減が挙げられる。

参考文献

- 1) 国土交通省：入札情報サービス、〈<http://www.i-ppi.jp>〉
- 2) (財)日本建設情報総合センター研究助成事業：中小都市における電子入札制度の効果報告書、〈<http://www.jaic.or.jp/kenkyu/5/5-2-2.pdf>〉、2003年11月
- 3) 安藤一憲：フィルタリング、情報処理学会誌, Vol. 46, No. 7, pp. 758-761, 2005年7月
- 4) 坂本比呂志, 有村博紀：ウェブ・マイニング, 人工知能学会誌, Vol. 16, No. 2, pp. 233-238, 2001年3月
- 5) 寺野隆雄：Web上の情報推薦システム, 情報処理学会誌, Vol. 44, No. 7, pp. 696-701, 2003年7月