

博士論文

情報フィルタリングを用いた 地域ポータルサイトの開発

阿部 真也

博士（工学）

電気通信大学大学院情報システム学研究科

2017年9月

論文題目

情報フィルタリングを用いた
地域ポータルサイトの開発

博士論文審査委員会

主査 山本 佳世子 准教授

委員 栗原 聡 教授

委員 内海 彰 教授

委員 田原 康之 准教授

委員 小川 朋広 准教授

著作権所有者

阿部 真也

2017 年

Abstract

Limited local government management resources have prompted neighboring areas to collaborate. Consequently, portal sites are required that disseminate cross-boundary local information without using a local government framework. Thus, a portal site was developed that provides cross-boundary content and promotes cooperation between local governments.

Chapter 1 discusses current local portal sites, problems in local government cooperation, and the growing need for portal sites. Chapter 2 highlights public institutions that promote local industry and introduces a portal site designed for them. This chapter explains current sites' problems with maintainability, search accuracy, and quickness, and how improving these areas supports local government cooperation.

Chapter 3 discusses portal site development based on information filtering, the focus of this thesis. Strategies are introduced to improve maintainability via a web crawler, search accuracy via URL filtering, and quickness via an RSS parser. The web crawler monitors websites that publish local information, automatically updates the portal site, and has an internal information filtering function that selects the

desired web pages from a website URL. The RSS parser parses feeds from websites, publishes the newest information on the portal site, and triggers the web crawler.

Chapter 4 discusses implementation and assessment testing of the portal site for public institutions that promote local industry, and demonstrates the site's superior maintainability, search accuracy, and quickness. Chapter 5 considers expansion and search accuracy improvement for the portal site. Pattern matching from a term list incorporated into the web crawler is proposed for ensuring search accuracy, maintainability, and quickness after portal site expansion. Chapter 6 discusses implementation and assessment testing of a portal site for the tourism industry and shows that it also improves availability for foreign tourists.

Chapter 7 discusses how the portal site promotes cooperation in industry and tourism between local governments and identifies directions for future research.

概要

本論文は、保守性、検索精度、迅速性、可用性に優れたポータルサイトの開発に関する論文である。

近年、地方自治体の経営資源の制約から、公共サービスを含む様々な分野のサービスを、近隣の地域間で分担補完するための連携が進められている。ゆえに、地域の情報を発信するポータルサイトにおいても、自治体の枠にとらわれない横断的なコンテンツを提供すべきである。しかしながら、このようなポータルサイトは現状なく、その実現が急務である。そこで本論文では、横断的なコンテンツを提供するポータルサイトを開発し、地方自治体間の連携の推進に寄与することを目的としたポータルサイトの実現方法を提案する。

本論文は全7章からなる。第1章では、地域ポータルサイトの現状について、地方自治体間の連携の観点から言及する。地方自治体間の連携が社会的に一般化しつつあるため、連携に対応した横断的なポータルサイトの実現が急務であることを述べる。続いて第2章では、地域のものづくりを中心とした産業振興を担う公共機関に焦点を当て、そのポータルサイトを紹介し、地域ポータルサイトの現状と課題を具体的に掘り下げる。既存のポータルサイトは、保守性、検索精度、迅速性の面で劣ることを述べるとともに、それらを向上させることが、地方自治体間の連携の推進につなが

ることを示す。

第3章では、本論文の核である、情報フィルタリングを基盤としたポータルサイトの開発について述べる。ここでは、Web クローラによる保守性の向上策、URL フィルタによる検索精度の向上策、RSS パーサによる迅速性の向上策について述べる。開発したポータルサイトの特徴は、Web クローラの内部に情報フィルタリング機能を有している点である。Web クローラは、各地域情報が掲載された Web サイトを巡回し、ポータルサイトの情報を自動で更新するプログラムである。この機能によって、掲載情報の保守性を向上させる。また、この Web クローラ内部の情報フィルタリング機能は、Web ページの URL から必要な Web ページのみを選別する機能である。この機能によって、検索機能の精度を向上させる。RSS パーサは、Web サイトが配信する RSS を解析し、それをポータルサイトに反映させるプログラムである。この機能によって、新着情報のみを選別した上でポータルサイト上に掲載するとともに、Web クローラの実行のトリガーとすることで、情報発信の迅速性を向上させる。これらの機能を組み合わせることで、保守性、検索精度、迅速性に優れた、地域横断的なポータルサイトを実現する。

第4章から第6章では、開発したポータルサイトの社会実装と評価実験について述べる。まず第4章では、地域の産業を対象とした社会実装と評価実験について述べ、開発したポータルサイトが保守性、検索精度、迅速性の面で優れていることを定量的に示す。次に第5章では、第4章で社会実装を行ったポータルサイトの拡張と検索精度の向上策について述べる。ポータルサイトの拡張にともなう検索精度の低下に対する対策として、単語リストのパターンマッチングを提案する。この機能を Web クローラに組み込むことで、拡張前と同等の保守性、検索精度、迅速性を維持したま

ま拡張できることを示す。さらに第6章では、観光業を対象とした社会実装と評価実験について述べ、開発したポータルサイトが保守性、検索精度、迅速性の面で優れているとともに、外国人観光客にとっての可用性向上に資することを示す。

最後に第7章では、本論文で開発したポータルサイトによって、産業や観光業における地方自治体間の連携の推進に寄与したことを述べるとともに、今後の研究課題について言及する。

謝辞

本論文は、電気通信大学大学院情報システム学研究科社会知能情報学専攻に在籍中の研究成果をまとめたものです。

同専攻准教授 山本佳世子先生には、主任指導教員として本論文の執筆の機会を与えていただくとともに、研究の遂行にあたって終始ご指導いただきました。心より御礼申し上げます。

同専攻教授 栗原聡先生、同専攻准教授 田原康之先生には、指導教員として有益なご助言をいただくとともに、本論文の細部にわたりご指導いただきました。心より御礼申し上げます。

地方独立行政法人東京都立産業技術研究センター開発本部開発第三部情報技術グループ長 入月康晴氏、同センタープロジェクト事業推進部 IoT 開発セクター主任研究員 大原衛氏には、研究の遂行にあたって有益なご助言をいただきました。心より御礼申し上げます。

また、同センター経営企画部長 近藤幹也氏、同センター技術経営支援部技術経営支援室長 山田一徳氏には、社会実装の機会を与えていただくとともに、有益なご助言をいただきました。心より御礼申し上げます。

そして、電気通信大学大学院情報システム学研究科社会知能情報学専攻山本研究室の各位、地方独立行政法人東京都立産業技術研究センター開発本部開発第三部情報技術グループの各位には、日頃より有益なご討論、ご

助言をいただきました。心より御礼申し上げます。

最後になりますが，妻 幸恵，父 栄一，母 和子をはじめとした家族に支えられ，本論文の完成に至ることができました。この場をお借りして，御礼申し上げます。

2017年9月 阿部 真也

目次

第 1 章	序論	1
1.1	地域ポータルサイト	1
1.2	広域連携と横断型ポータルサイト	2
1.3	基礎となる研究事例	3
1.3.1	検索エンジン	4
1.3.2	情報フィルタリング	5
1.3.3	文書の自動分類	5
1.3.4	情報抽出	6
1.3.5	言語横断検索	7
1.3.6	マルチメディア検索	7
1.4	本論文の目的および手法	8
1.5	本論文の構成	9
第 2 章	地域ポータルサイトの現状と課題	11
2.1	公設試験研究機関とその変革	11
2.2	広域連携の必要性	14
2.3	ポータルサイトの現状と課題	16
2.3.1	東北 6 県公開設備・機器データベース	17

2.3.2	うおっち	18
2.3.3	関西ラボねっと	21
2.3.4	中国地域公設研究機関の開放機器	21
2.3.5	四国 4 県連携新製品開発支援協議会	23
2.3.6	九州・沖縄バーチャル公設試	24
2.4	第 2 章のまとめ	26
第 3 章	Web クローラと RSS パーサによるポータルサイトの開発	27
3.1	ポータルサイトの構成	27
3.2	検索エンジン	29
3.3	新着お知らせ	32
3.4	Google カスタム検索と比べた利点	33
3.5	第 3 章のまとめ	34
第 4 章	産業ポータルサイトへの応用	37
4.1	首都圏公設試験研究機関連携体	37
4.2	事業化に至る経緯	42
4.2.1	ポータルサイトの初期	42
4.2.2	編集用フォームの設置	43
4.2.3	XML による情報更新	44
4.2.4	提案サイトの事業化	46
4.3	評価	47
4.3.1	保守性の評価	47
4.3.2	検索精度の評価	49
4.4	第 4 章のまとめ	52
第 5 章	産業ポータルサイトの拡張	53

5.1	広域関東圏への拡張	53
5.2	拡張法の検討	54
5.2.1	検索エンジンの拡張	55
5.2.2	新着お知らせの拡張	57
5.3	拡張後の評価	60
5.3.1	拡張後の保守性	61
5.3.2	拡張後の検索精度	61
5.4	検索エンジンの改良	65
5.5	第 5 章のまとめ	67
第 6 章	観光ポータルサイトへの応用	69
6.1	観光業の広域連携	69
6.2	ポータルサイトの可用性	70
6.3	ポータルサイトの言語バリアフリー化	72
6.4	類似スポットの推薦	75
6.5	可用性の評価	75
6.6	第 6 章のまとめ	79
第 7 章	結論	83
7.1	本論文のまとめ	83
7.2	今後の研究課題	84
	参考文献	87
	関連研究の公表	99

目次

1.1	一般的な検索エンジンの構成	3
2.1	公設試の業務方法書の抜粋	12
2.2	今後の公設試のあるべき姿	14
2.3	公設試の広域連携のモデル	16
2.4	東北 6 県公開設備・機器データベース	17
2.5	うおっち	19
2.6	関西ラボねっと	20
2.7	中国地域公設研究機関の開放機器	22
2.8	四国 4 県連携新製品開発支援協議会	23
2.9	九州・沖縄バーチャル公設試	24
3.1	ポータルサイトの構成	28
3.2	Web クローラの構成	29
3.3	Web クローラの設定例	32
3.4	RSS フィード	33
3.5	新着リスト	34
4.1	首都圏テクノナレッジ・フリーウェイ	38

4.2	設備検索フォーム	39
4.3	技術検索フォーム	40
4.4	各機関からのお知らせ	41
4.5	技術相談フォーム	42
4.6	旧サイトの編集用フォーム	44
4.7	情報更新用 XML フォーマット	45
5.1	パターンマッチングを組み込んだ Web クローラの構成 . .	66
6.1	言語バリアフリー化した観光ポータルサイト	71
6.2	言語バリアフリー化した検索インターフェイス	72
6.3	類似スポットの推薦	74
6.4	観光スポットの類似度	76
6.5	類似度算出システムの構成	77
6.6	試作機	78

表目次

2.1	ポータルサイトの実現方法とその性質	26
3.1	地域の産業情報の配置箇所	31
3.2	地域の観光情報の配置箇所	31
4.1	設備情報とリンク構造の変更回数	47
4.2	評価環境・条件	49
4.3	クローリング開始点	50
4.4	クローリング対象	50
4.5	適合率	51
5.1	設備情報の文書形式	55
5.2	設備情報の配置箇所	56
5.3	RSS フィードの配信状況	59
5.4	設備情報とリンク構造の変更回数（拡張後）	60
5.5	評価環境・条件（拡張後）	62
5.6	クローリング開始点（拡張後）	63
5.7	クローリング対象（拡張後）	64
5.8	適合率（拡張前および拡張後）	65

5.9	適合率（パターンマッチング追加後）	67
6.1	回答結果	80
6.2	画面遷移数	81

第 1 章

序論

1.1 地域ポータルサイト

インターネットの普及により、様々な情報が、より速く、より多くの人々に届くようになった。事業者や住民などの地域の構成員が、情報の受発信者となり、オープンな情報交流を行うことが、地域の問題解決の一助となっている。さらには、住民の生活の質の向上や、地域経済の活性化をもたらす、活力ある地域社会の構築につながることを期待されている。その具体策の1つとして、地域の行政情報や農工商業の情報を集めたポータルサイトの運営が、各地で行われている。このようなポータルサイトを地域ポータルサイト [1] と呼ぶ。地域ポータルサイトという言葉は、狭義では公共情報を集約したものを指し、広義では公共情報に限らず地域情報を集約したものを指す。広義での地域ポータルサイトは、特に官民連携地域ポータルサイト [2] と呼ばれることもある。本論文では、地方自治体間の連携に焦点を当てることから、狭義の意味で地域ポータルサイトという言葉を用いる。

地域ポータルサイトの例として、つくば市 [3] や八王子市 [4, 5, 6] の

ポータルサイトが挙げられる。つくば市の例では、様々な Web サイト上で提供されているつくば市に関する情報を整理し、これらに容易にアクセスできる仕組みを提供している。主なコンテンツは、つくば市に関する情報を発信している Web サイトへのリンク集であり、公共機関、教育、福祉、交通などに関する情報を発信している Web サイトが主要なリンク先である。つくば市は研究学園都市であるがゆえに、教育研究機関に関する情報が特に充実している。

1.2 広域連携と横断型ポータルサイト

近年、地方自治体の経営資源の制約から、公共サービスを含む様々な分野のサービスを、近隣の地域間で分担補完するための連携が進められている。このような、地方自治体の枠を超えた連携を広域連携という。特に、産業や観光業において、広域連携に基づいたサービスの供給が行われ、管轄の地方自治体の枠にとらわれない利用が広まってきている。このような変化を背景に、地域ポータルサイトも、地方自治体の枠にとらわれない横断的なコンテンツを提供する必要がある。

しかしながら、広域連携に対応した地域ポータルサイトは、ほとんど実現されていないのが現状である。つくば市や八王子市のポータルサイトは、その地域のみを対象としており、広域連携に対応していないポータルサイトの例である。広域連携が社会一般になりつつある現在、横のつながりを重視した横断型のポータルサイトの実現は急務である。

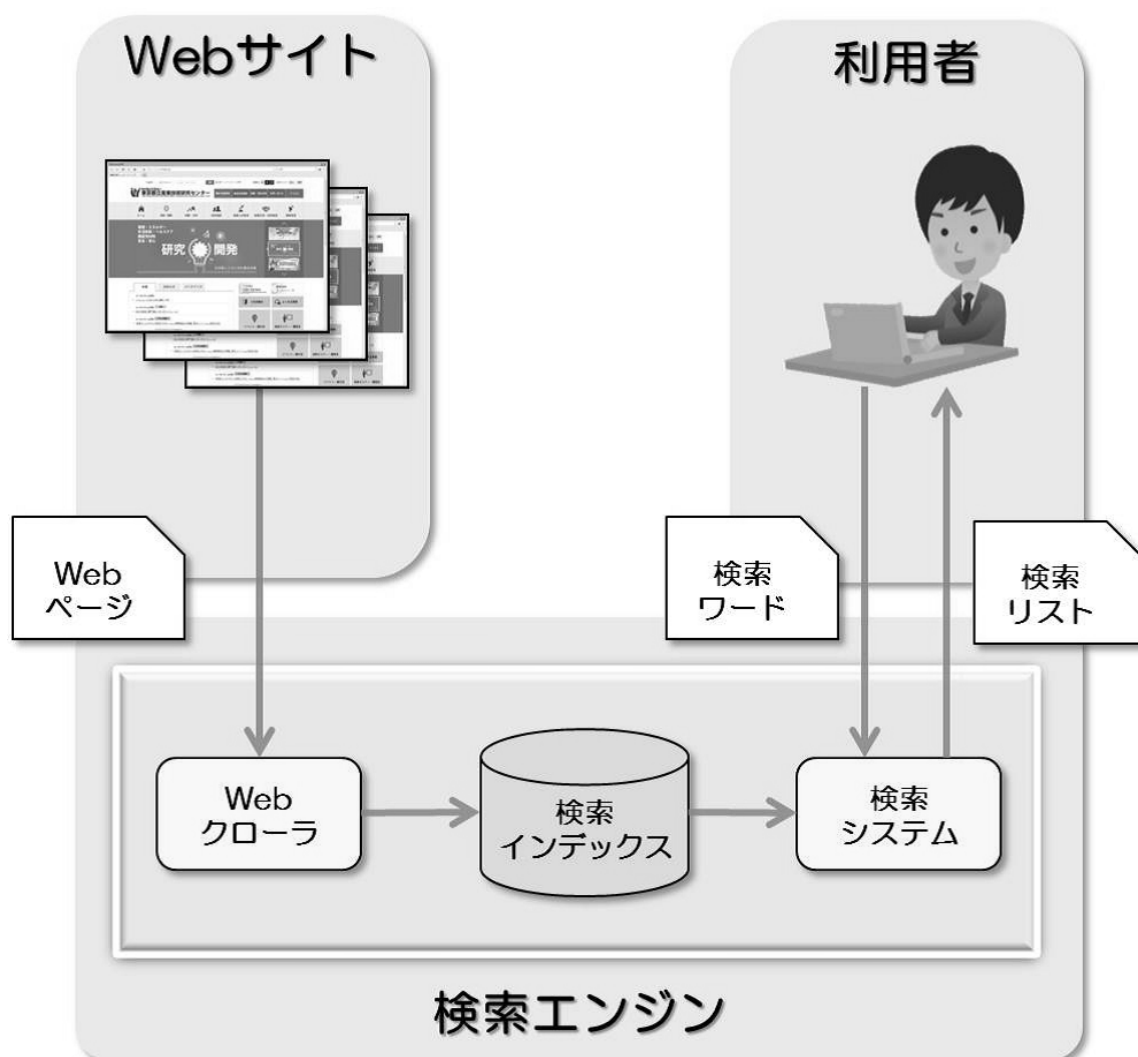


図 1.1. 一般的な検索エンジンの構成

1.3 基礎となる研究事例

ここでは、広域連携に対応した横断型のポータルサイトの実現に有効な研究事例を紹介する。

1.3.1 検索エンジン

1980年代に出現した World Wide Web, いわゆる WWW は, 1990年代に爆発的に普及し, 豊富な情報に容易にアクセスできるようになった. 同時に, 膨大な情報の中から必要な情報を入手する手段として, 検索エンジン [7, 8] が広く利用されるようになった.

検索エンジンは, 利用者が必要としている情報が, どの Web ページで提供されているかを特定する作業を支援するシステムといえる. 利用者が所望する情報に関連する検索ワードを入力すると, その検索ワードを含む Web ページ一覧を検索リストとして出力する. 利用者は, 検索リストを見て適当な Web ページを選択する.

一般的な検索エンジンの構成を図 1.1 に示す. 検索エンジンは, Web ページを高速に見つけ出すために, あらかじめ Web サイトからページを収集し, これを検索データベースに登録している. このページ収集を行うプログラムを, クローラ, ロボット, スパイダーなどと呼ぶ. また, 検索データベースに登録する際には, インデクシングを行い, 検索ワードがどのページのどの位置にあるかを高速に検索できるようなデータ形式に変換する. この検索データベースのことを検索インデックスと呼ぶ. 検索エンジンは, 検索ワードと検索インデックスを照合し, 検索リストを出力する.

現在, 多数の検索エンジンが, Web 上で検索サービスを提供している. 複数の検索エンジンに検索ワードを送り, これらから得られる検索リストを統合して出力するメタ検索エンジンも存在する.

1.3.2 情報フィルタリング

WWW の爆発的な普及により、膨大な情報にアクセスできるようになった。必要な情報はほぼ全てとあってよいほど、WWW のどこかには存在している。だがその一方で、その膨大さのあまり、必要としている情報を瞬時に見つけられないという状況を生み出している。このような状況を、情報爆発あるいは情報洪水と呼ぶ。

このような状況を回避するための技術が、情報フィルタリング [9] である。情報フィルタリングは、膨大な情報の中から、必要な情報のみを取り出し、不要な情報を排除する技術である。情報フィルタリングでは、あらかじめ選別したい情報の特徴をプロファイルに記述しておき、これを用いて必要な情報のみを選別する。

情報フィルタリングの具体例としては、多数の電子メールから重要性の高いものを選別する技術 [10] や、Web ニュースから利用者の興味に合う記事を推定する技術 [11]、インターネット上の有害サイトへのアクセス制御を行う技術 [12, 13] などがある。検索も、膨大な情報の中から、検索ワードに適合する情報のみを取り出す技術であり、一種の情報フィルタリングといえる。

1.3.3 文書の自動分類

情報アクセスを支援する 1 つの手段に、文書をその内容に応じて分類するというものがある。例えば、Web 上には Yahoo を代表とするポータルサイトが数多く存在する。これらのポータルサイトでは、Web ページを内容に応じて階層的に分類している。ただし、多くのポータルサイトは、人手で分類を行っているため保守性に問題がある。

ポータルサイトで扱うべき大規模な文書集合を分類するためには、コンピュータによる自動的な処理が不可欠である。そのための技術が、文書の自動分類である。文書の自動分類は、大きく2つに分けられる。1つは、文書の内容に応じて、あらかじめ用意したカテゴリのいずれに属するかを決定するものである。もう1つは、類似した文書をグループ化することにより、文書集合をいくつかのグループに分割するものである。

文書の自動分類は、2つの文書間の類似度、あるいは文書と文書カテゴリ間の類似度を計算することに帰着する。ゆえに、情報検索と類似の技術によって実現することができ、情報検索との親和性が高い。文書の自動分類の技術は、ベクトル空間モデルに基づく方法 [14, 15]、確率モデルに基づく方法 [16, 17]、規則に基づく方法 [18]、機械学習に基づく方法 [19, 20] などが提案されている。

1.3.4 情報抽出

膨大な情報の中から、あらかじめ指定した情報のみを効率的に取得する技術を情報抽出 [21] という。検索により取得した情報の中から、特に決まった情報のみを参照したい場合に情報抽出が役に立つ。

日本語文書を対象にした情報抽出としては、固有表現の前後に出現する単語情報をコーパスから学習して、固有表現を抽出する技術 [22]、電子メールや電子ニュース記事のスケジュールやイベント情報を抽出する技術 [23, 24, 25]、Web ページから新製品の情報を抽出する技術 [26] などがある。

情報抽出においては、抽出した表現の同一性や多義性を認識することが、今後の研究課題といわれている。同一性の認識は、電気通信大学と電通大のように、表記は異なるが同じ意味の言葉を同じものとして処理する

技術である。多義性の認識は、表記は同じだが異なる意味の言葉を異なるものとして処理する技術である。

1.3.5 言語横断検索

Web 上には様々な言語で記述された Web ページが存在する。知らない言語で記述されたページを読解するのが難しいのは当然である。しかしながら、そのようなページを能動的に探すのは、それ以上に難しい。知りたい情報を得るための検索ワードを全く思い付かず、自身の検索要求に有用なページを、読解以前にそもそも入手できないといった状況も考えられる。

このような状況では、言語の違いを意識せず、母国語で自由に検索することができれば便利である。これを実現する技術が、言語横断検索 [27] である。言語横断検索では、検索対象となる文書とは異なる言語の検索ワードを与えることができる。

言語横断検索を実現するための技術には、対訳辞書を用いる方法 [28]、機械翻訳を用いる方法 [29]、多言語の意味空間に基づく方法 [30]、クエリ拡張を用いる方法 [31] などが提案されている。

1.3.6 マルチメディア検索

計算機の高速化や記憶容量の大容量化にともない、情報のマルチメディア化が進んでいる。Web 上には、テキスト文書以外にも、画像や映像、音声などのマルチメディア情報が存在する。マルチメディア情報を対象とした検索をマルチメディア検索という。

マルチメディア検索の最も単純な実現方法は、人手でキーワードを付与する方法である。しかし、人手による方法では、作業量が多く、大規模な

システムを構築するのは困難である。このような理由から、マルチメディア情報そのものから得られる特徴量に基づいた検索手法が提案されている。特徴量に基づいた検索手法は、画像の検索 [32, 33] や映像の検索 [34], 音声の検索 [35] などに応用されている。

1.4 本論文の目的および手法

これまで、地域ポータルサイトの現状と課題について、地方自治体の広域連携の観点から述べた。現在、地方自治体では、近隣自治体との広域連携が進められており、ポータルサイトも地域の枠にとらわれない横断的な情報を提供すべきである。ゆえに、広域連携に対応した分野横断的なポータルサイトの実現が急務である。そこで本論文では、地方自治体の広域連携を推進することを主な目的としたポータルサイトの実現方法を提案する。

提案サイトは、Web クローラと情報フィルタリングを基にして構成される。Web クローラは、各地域情報が掲載された Web サイトを巡回し、ポータルサイトの情報を更新するプログラムである。また、この Web クローラは、内部にフィルタリング機能を有している。フィルタリング機能は、Web ページの URL と、出現する単語によって、必要な Web ページのみを選別する機能である。この機能によって、特定の分野に特化した情報のみを集約した地域横断的なポータルサイトを実現する。本手法は、Web クローラと情報フィルタリングが有効に働く場合であれば、地域ポータルサイトに限らず、広く応用が期待できる。

1.5 本論文の構成

第1章では、地域ポータルサイトの現状について、地方自治体の広域連携の観点から述べた。広域連携が社会的に一般化しつつあるため、広域連携に対応した横断的なポータルサイトの実現が急務である。

以降、第2章では、地域のものづくりを中心とした産業振興を担う公共機関に焦点を当て、そのポータルサイトを紹介し、それらが保守性、検索精度、迅速性の面で劣ることを課題として提起する。

第3章では、本論文の核である、情報フィルタリングを基盤としたポータルサイトの開発について述べる。ここでは、Web クローラによる保守性の向上策、URL フィルタによる検索精度の向上策、RSS パーサによる迅速性の向上策について述べる。

第4章から第6章では、開発したポータルサイトの社会実装と評価実験について述べる。まず第4章では、地域の産業を対象とした社会実装と評価実験について述べ、開発したポータルサイトが保守性、検索精度、迅速性の面で優れていることを定量的に示す。次に第5章では、第4章で社会実装を行ったポータルサイトの拡張と検索精度の向上策について述べる。ポータルサイトの拡張にともなう検索精度の低下に対する対策として、単語リストのパターンマッチングを提案する。この機能を Web クローラに組み込むことで、拡張前と同等の保守性、検索精度、迅速性を維持したまま拡張できることを示す。さらに第6章では、観光業を対象とした社会実装と評価実験について述べ、開発したポータルサイトが保守性、検索精度、迅速性の面で優れているとともに、外国人観光客にとっての可用性向上に資することを示す。

最後に第7章では、本論文で開発したポータルサイトによって、産業な

らびに観光業における地方自治体間の連携の推進に寄与したことを述べるとともに、今後の研究課題について言及する。

第 2 章

地域ポータルサイトの現状と課題

本章では，地域の産業振興を担う公設試験研究機関に焦点を当て，そのポータルサイトを紹介し，現状の課題を抽出する。

2.1 公設試験研究機関とその変革

公設試験研究機関，いわゆる公設試は，地方自治体が設置する試験研究機関であり，地域企業への技術指導を通して，その地域に資することを目的とする。公設試の業務の一例として，地方独立行政法人東京都立産業技術研究センターの業務方法書の抜粋を図 2.1 に示す。図 2.1 の第 3 条に示された業務を依頼試験と呼び，公設試の利用者から製品やその部品などの試験品を受け，その特性の測定や分析を請け負い，結果を証明書として発行する業務である。図 2.1 の第 7 条に示された業務を機器利用と呼び，公設試の利用者に設備や施設を貸し付ける業務である。依頼試験と異なり，顧客自身が設備を操作する。他に，産業技術に係る研究などの業務がある。

近年，地方自治体の経営資源の制約から，公設試の抜本的な組織改革の

第二章 産業技術に係る試験，研究及び調査

(試験に関する業務)

第3条 法人は，依頼に応じて，産業技術に係る試験（以下「依頼試験」という）を実施することができる。

2 法人は依頼試験を実施するときは，適正な対価を徴収するものとする。

(中略)

第四章 試験機器等の設備及び施設の提供

(試験機器等の設備及び施設の提供に関する業務)

第7条 法人は，依頼に応じて試験機器等の設備及び施設を貸し付けることができる。

2 法人は，試験機器等の設備及び施設を貸し付ける場合には，適正な対価を徴収することができる。

図 2.1. 公設試の業務方法書の抜粋

必要性が権田ら [36]，若生ら [37]，佐脇 [38] によって指摘されている。地方自治体から支出される科学技術関係諸費，すなわち理工系の高等教育機関や医療機関などの運営費を含む予算は，年々増加している。ただし，公設試に対する支出はほぼ横ばいであり，公設試運営の経費は相対的に減少している。

その一方で，公設試が産業構造の変化にともなうニーズの変化や，地域企業からの支援内容の多様化に対応していく必要性も，並河 [39] や福川 [40] によって指摘されている。1980年代までは，地域企業の多くは，近代

産業構造の中でも下流工程を担う受注型企業であった。製品の納入先は、その多くが系列会社や協力工場であったため、発注元から求められる基準をクリアすることが第一であった。この頃の公設試は、地域企業の製品の品質や性能を評価し、そしてそれを第三者として証明するという役割が主であった。加えて、製品の品質向上へ向けた技術指導も公設試の役割であった。Shapira[41]によれば、公設試による技術指導は、先端的ではない一般的な技術課題の支援であったが、先端的ではなかったことが、地域企業にとってむしろ好都合であったとされている。1980年代の公設試は、地域における特色などはあまりなく、ほぼ横並びで事業を展開していた。

1990年代になると、産業構造の変化にともない、地域企業のニーズも変化してきた。多くの地域企業の経営は、取引先企業に依存していたが、主要な取引先の海外シフトにともない、自立が求められる企業が増加した。この頃から地域企業は、自らの強みを認識し、技術獲得の機能をも内在する研究開発型企業へとシフトし、より先端的で分野横断的な技術指導を求めるようになった。

2000年代に入ると、分野横断的な技術指導の要求がさらに増してくる。以下の2つは、公設試に求められた分野横断的な技術指導の例である。

- X線CT装置を利用して、ある製品の部品を三次元画像化し、元の3D-CAD図面と比較しながら、加工の精度を評価したい。
- X線CT装置を利用して、ある製品の部品を三次元画像化し、それをさらに流体解析システムに入力して、放熱性を評価したい。

どちらも、X線CT装置による依頼試験業務であるが、X線CT装置に精通している職員のみでは対応できない。前者であれば、三次元モデリングの知識や、機械加工の知識も必要である。後者であれば、流体解析の知識

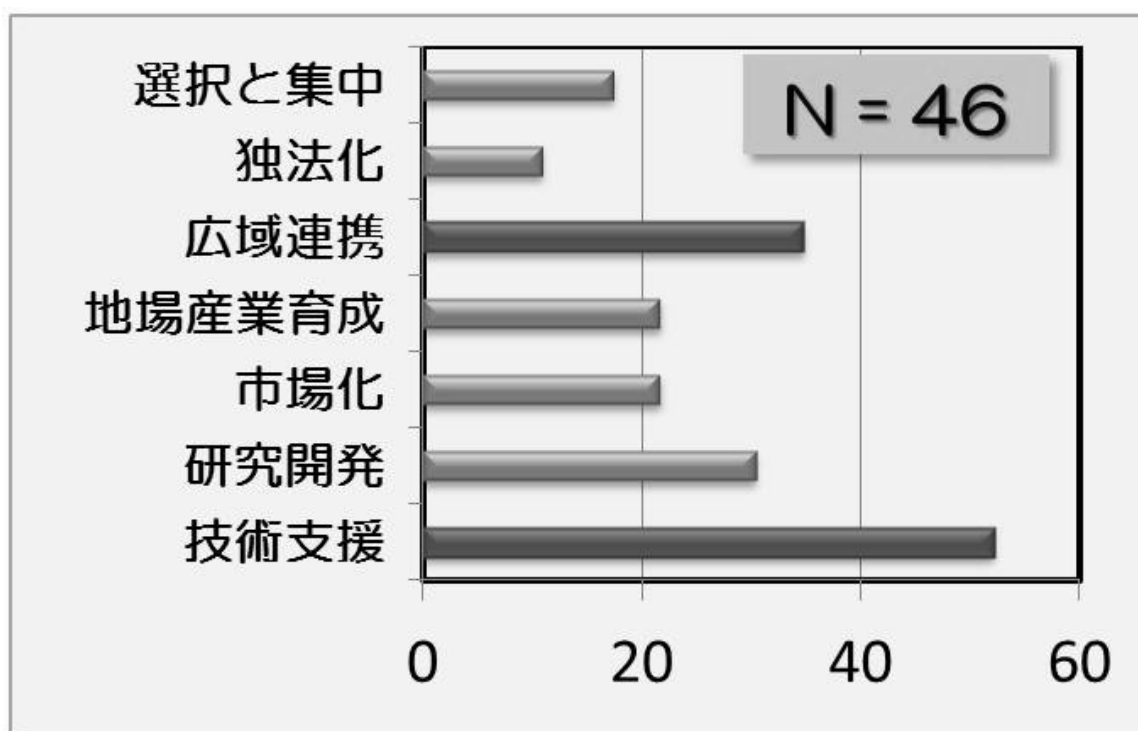


図 2.2. 今後の公設試のあるべき姿

が必要になる。また、この時期になると、異業種間のマッチングや、自社生産ラインの可視化など、より技術経営的な支援も求められるようになってきており、公設試単体では対応できない案件が増えてきた。

2.2 広域連携の必要性

これまで公設試では、地域社会と密接に結び付いてきたことと、地方自治体の予算により運営されてきたことから、自治体の枠を超えた連携はなされてこなかった。しかしながら、地方自治体の経営資源の制約がある中で、先端的で分野横断的な技術指導を、公設試が単体で行うことは困難であることが分かってきた。公設試の職員は高々50名程度であり、必要な専門性を有する職員が内部にいないことも多いためである。そのような経緯

から、公設試が自治体の枠を越えて連携する広域連携が求められるようになってきた。

中小企業庁 [42] は、今後の公設試のあるべき姿を発表している。この報告では、公設試および利用者へのアンケートを通して、公設試が今後強化すべき事業に関する集計が報告されている。これを筆者の解釈で図示したものを図 2.2 に示す。この図から、今後強化すべき事業の第一が地域企業に対する技術支援の拡充であり、第二が産学官連携や広域連携の推進であることが分かる。全日本地域研究交流協会 [43] は、公設試の定員は 30 名から多くとも 50 名強であり、今後も拡大が予想される産業全体を支援対象にすることは困難であるという理由から、広域連携による技術支援分野の分担補完の必要性を指摘している。また、大槻ら [44] や谷口 [45] による公設試の技術相談に関する調査では、6 割以上の公設試が技術分野の拡大に苦慮しており、他の公設試への問合せを要する案件が増えていることが明らかにされている。谷口は、これに対処する方法の 1 つとして、外部との連携による技術相談業務の拡充を挙げている。以上のように、公設試経営における直近の課題として、広域連携が求められている。

公設試の広域連携のあり方には、各機関の組織は独立で維持しつつ案件を相互に紹介するという入り口段階のものから、事実上 1 つの組織として人材や設備を共同運用するという踏み込んだものまで提案されている。図 2.3 に代表的な広域連携のモデルを示す。利用者は、製品開発や評価などにおいて、技術的課題が生じた場合、まずは最寄りの公設試に相談する。最寄りの公設試で対応できる案件であれば問題ないが、内容によっては最寄りの公設試では対応できず、他の公設試に問い合わせるべき場合がある。その場合、最寄りの公設試の職員は問い合わせるべき公設試を特定しなければならない。その際、各機関の Web サイトを 1 つ 1 つ順に参照す

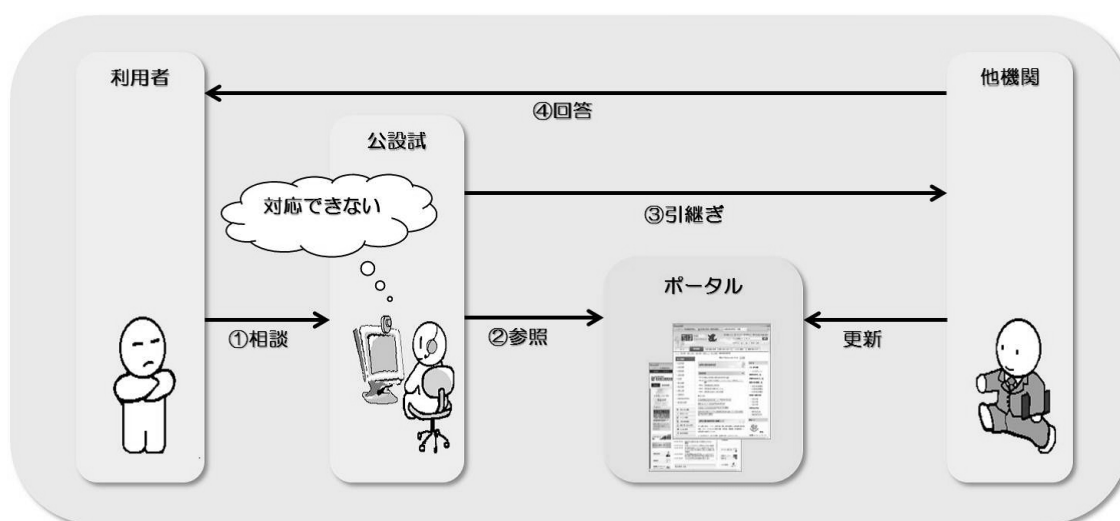


図 2.3. 公設試の広域連携のモデル

れば、問い合わせるべき公設試を特定できるが、このとき、もし、各機関の情報を1つに集約したポータルサイトがあれば、そのポータルサイトは職員にとって有用である。各地の公設試は、このような広域連携の推進を目的としたポータルサイトの開発を進めている。以降で、その具体例を紹介する。

2.3 ポータルサイトの現状と課題

公設試の広域連携を推進するにあたり、各機関の情報を1つに集約したポータルサイトの利用が有効である。各地の公設試は、このようなポータルサイトの開発を進めている。ここでは、その具体例を紹介し、現状のポータルサイトの課題を提示する。

なお、このようなポータルサイトは、公設試の職員だけでなく、公設試の利用者にとっても有用である。本論文では、公設試の職員と利用者を含めたものを、ポータルサイトの利用者と定義する。

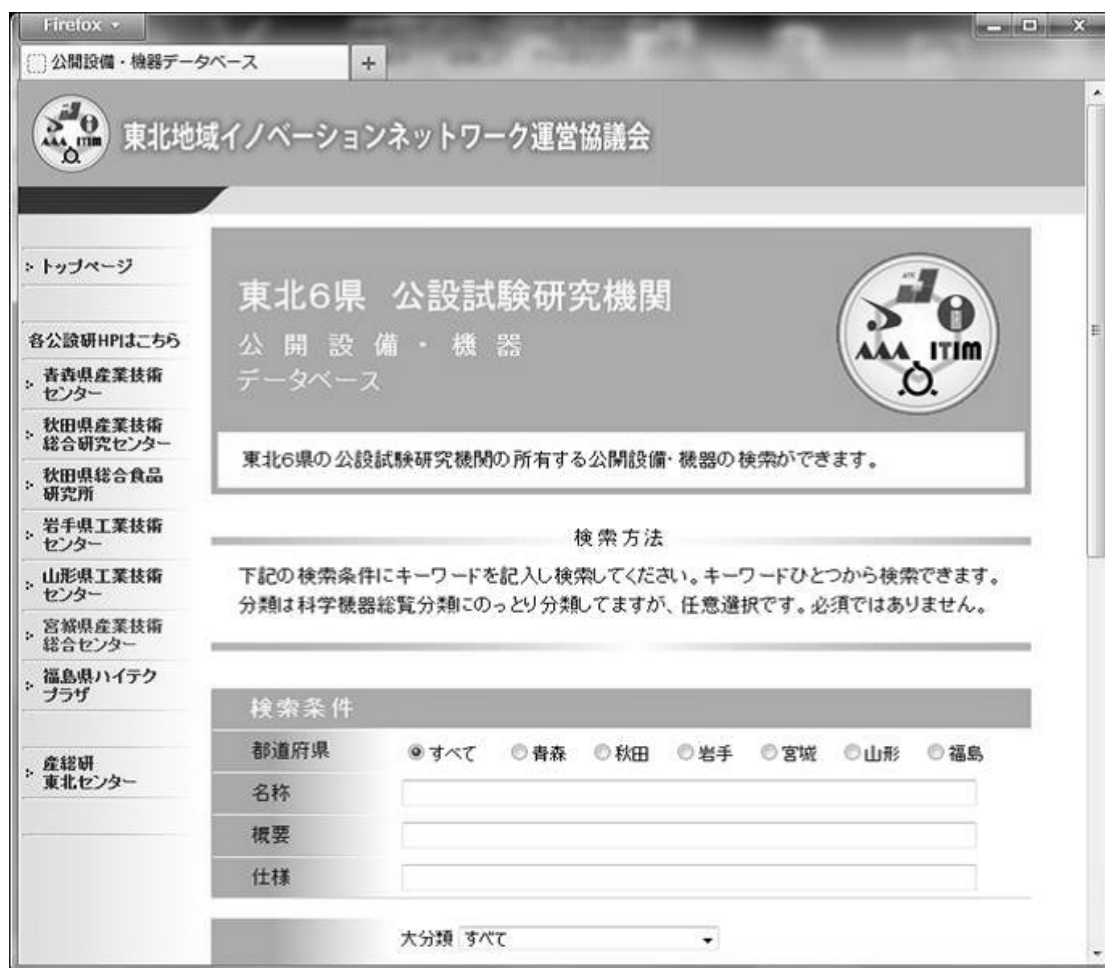


図 2.4. 東北 6 県公開設備・機器データベース

2.3.1 東北 6 県公開設備・機器データベース

東北 6 県公開設備・機器データベース [46] は、東北地域イノベーションネットワーク運営協議会が中心となり開発したポータルサイトであり、対象機関は以下の 7 機関である。

- 地方独立行政法人青森県産業技術センター
- 地方独立行政法人岩手県工業技術センター
- 宮城県産業技術総合センター

- 秋田県産業技術センター
- 秋田県総合食品研究センター
- 山形県工業技術センター
- 福島県ハイテクプラザ

図 2.4 に、東北 6 県公開設備・機器データベースのブラウザイメージを示す。東北 6 県公開設備・機器データベースは、対象機関の保有設備の情報を集約し、機器保有都道府県、機器名称、機器概要、機器仕様などの詳細な条件から目的の設備を検索できる。ポータルサイトの管理者によって情報が更新されているため、人的な誤りを除けば情報の正確性が担保されている。

ただし、設備の導入や廃止の度に管理者による更新を要するため、保守性や迅速性が低い。東北 6 県公開設備・機器データベースは、産業技術総合研究所の東北地域イノベーション創出共同体形成事業の一部として構築されたものであるが、2009 年 3 月に事業が終了したのを最後に情報の更新は行われていない。

2.3.2 うおっち

うおっち [47] は、東北経済産業局が中心となり開発したポータルサイトあり、東北 6 県公開設備・機器データベースに代わり運用されている。対象機関は以下の 8 機関である。東北 6 県公開設備・機器データベースの対象機関に加え、新潟県工業技術総合研究所も対象になっている。

- 地方独立行政法人青森県産業技術センター
- 地方独立行政法人岩手県工業技術センター
- 宮城県産業技術総合センター



図 2.5. うおっち

- 秋田県産業技術センター
- 秋田県総合食品研究センター
- 山形県工業技術センター
- 福島県ハイテクプラザ
- 新潟県工業技術総合研究所

図 2.5 に、うおっちのブラウザイメージを示す。うおちは、対象機関の保有設備とセミナー情報を一括検索できる。うおちは、Google カスタム検索を利用し、各機関のサイトから Googlebot が収集した情報を元に検索を行っている。東北 6 県公開設備・機器データベースと違い、Googlebot



図 2.6. 関西ラボねっと

を利用して更新を自動化することで、保守性を向上させている。

しかしながら、Googlebot が各機関の Web サイトを訪れるタイミングが不確定なため、迅速性に欠ける。Google Search Console のように、Googlebot の訪問を促すツールは存在するものの、更新の都度ツールを使用していたのでは保守性が高いとはいえない。すなわち、保守性と迅速性のトレードオフになっている。

なお、検索精度については、Google Search Console の設定によるので一概にいえない。

2.3.3 関西ラボねっと

関西ラボねっと [48] は、関西広域連合が中心となり開発したポータルサイトであり、対象機関は以下の 11 機関である。

- 大阪市立工業研究所
- 大阪府立産業技術総合研究所
- 京都市産業技術研究所
- 京都府織物・機械金属振興センター
- 京都府中小企業技術センター
- 滋賀県工業技術総合センター
- 滋賀県東北部工業技術センター
- 徳島県立工業技術センター
- 鳥取県産業技術センター
- 兵庫県立工業技術センター
- 和歌山県工業技術センター

関西ラボねっとのブラウザイメージを図 2.6 に示す。関西ラボねっとは、対象機関の保有設備を一括検索できるポータルサイトである。うおっちと同様に Google カスタム検索を利用しており、長所や短所も同様である。

2.3.4 中国地域公設研究機関の開放機器

中国地域公設研究機関の開放機器 [49] は、産業技術総合研究所が中心となり開発したポータルサイトであり、対象機関は以下の 6 機関である。

- 鳥取県産業技術センター



図 2.7. 中国地域公設研究機関の開放機器

- 島根県産業技術センター
- 岡山県工業技術センター
- 広島県立総合技術研究所
- 広島市工業技術センター
- 山口県産業技術センター

中国地域公設研究機関の開放機器のブラウザイメージを図 2.7 に示す。対象機関の保有設備を一括検索できるポータルサイトであり、Google カスタム検索を利用している。



図 2.8. 四国 4 県連携新製品開発支援協議会

2.3.5 四国 4 県連携新製品開発支援協議会

四国 4 県連携新製品開発支援協議会 [50] は、以下の 4 機関の連携体である。

- 徳島県立工業技術センター
- 香川県産業技術センター
- 愛媛県産業技術研究所
- 高知県工業技術センター

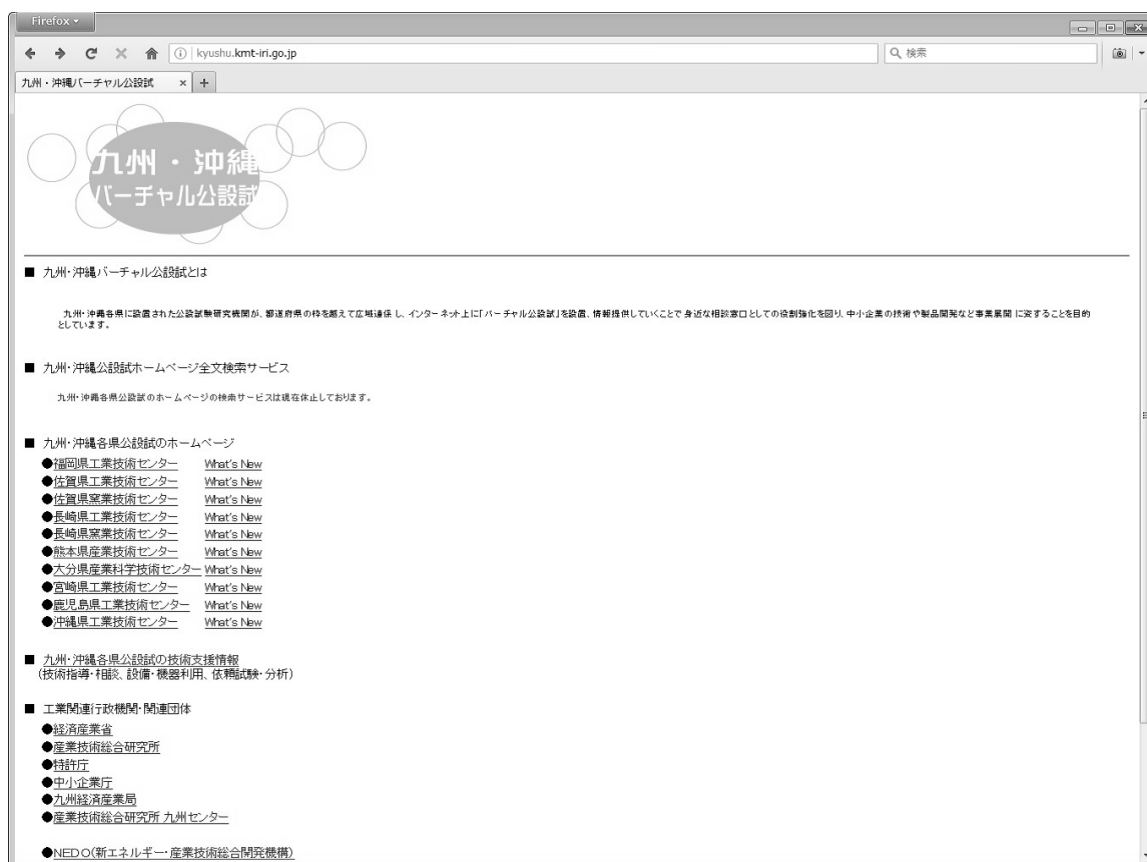


図 2.9. 九州・沖縄バーチャル公設試

四国 4 県連携新製品開発支援協議会の Web サイトでは、対象機関の保有設備を一括検索できる機能を提供しており、これがポータルサイトの役目を果たしている。そのブラウザイメージを図 2.8 に示す。検索機能の実現方法は Google カスタム検索である。

2.3.6 九州・沖縄バーチャル公設試

九州・沖縄バーチャル公設試 [51] は、熊本県産業技術センターが中心となり開発したポータルサイトであり、対象機関は以下の 10 機関である。

- 福岡県工業技術センター

- 佐賀県工業技術センター
- 佐賀県窯業技術センター
- 長崎県工業技術センター
- 長崎県窯業技術センター
- 熊本県産業技術センター
- 大分県産業科学技術センター
- 宮崎県工業技術センター
- 鹿児島県工業技術センター
- 沖縄県工業技術センター

図 2.9 にそのブラウザイメージを示す。九州・沖縄バーチャル公設試は、対象機関の Web サイトの情報を独自に収集し、検索できるようにしている。その仕組みは、対象機関の Web ページを定期的にダウンロードしてサーバ内に保存し、それを全文検索システム Namazu[52] を用いて検索するというものである。うおっちや関西ラボねっとと同様に更新が自動化されており、なおかつ更新のタイミングを任意に設定できるという利点がある。

ただし、あくまでも横断検索するだけであり、利用者にとって不要な情報が混在しやすい。たとえば、「光電子分光分析装置」を検索ワードとすると、装置の仕様や利用者講習会といった利用者にとって有益な情報以外に、関連機器の入札情報や導入年度の年報などが検索結果として得られるが、これらは職員や利用者にとっては不要な情報である。このことは、単に横断検索するだけでは、十分な検索精度が得られないことを意味している。

表 2.1. ポータルサイトの実現方法とその性質

実現方法	保守性	検索精度	迅速性
手動による情報更新	低い	高い	低い
全文検索システムの利用 (Namazu など)	高い	低い	高い
検索サービスの利用 (Google カスタム検索など)	トレード オフ	—	トレード オフ

2.4 第2章のまとめ

本章では、地域の産業振興を担う公設試のポータルサイトを紹介し、その実現方法と性質について述べた。それをまとめたものを表 2.1 に示す。ポータルサイトの運用において、保守性、検索精度、迅速性のいずれも高いことが理想であるが、そのようなポータルサイトは現状では実現されていない。

次章では、保守性、検索精度、迅速性に優れるポータルサイトの開発について述べる。

第3章

Web クローラと RSS パーサによる ポータルサイトの開発

前章では、地域の産業振興を担う公設試のポータルサイトを紹介し、その実現方法と性質について述べた。ポータルサイトの運用において、保守性、検索精度、迅速性のいずれも高いことが理想であるが、そのようなポータルサイトは現状では実現されていなかった。

そこで、保守性、検索精度、迅速性に優れるポータルサイトを実現するため、Web クローラと RSS パーサを基盤とするポータルサイト [53, 54, 55, 56] を開発する。本章ではその構成と実装について説明する。

3.1 ポータルサイトの構成

提案するポータルサイトの全体構成を図 3.1 に示す。提案サイトは、主に次の要素からなる。

- 検索エンジン
- 新着お知らせ

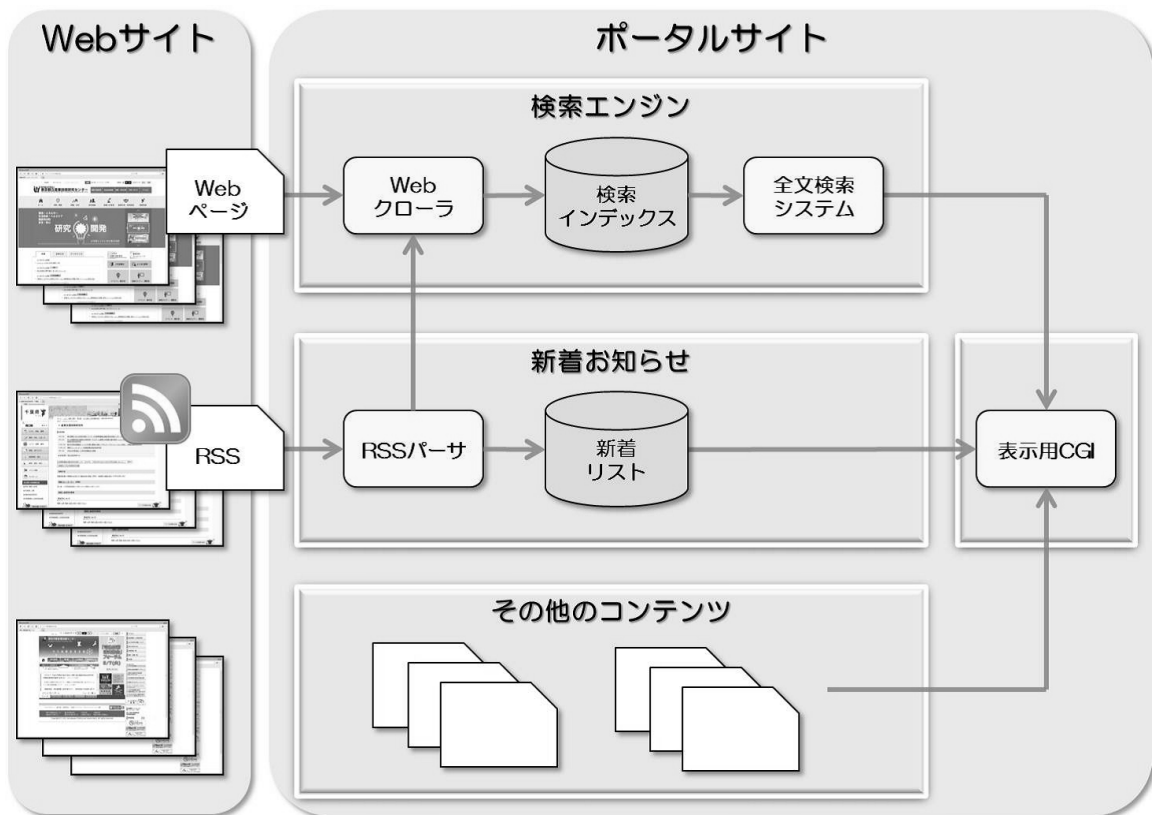


図 3.1. ポータルサイトの構成

- 表示用 CGI
- その他のコンテンツ

検索エンジンは、Web サイト内の情報を一括検索する機能である。Web クローラによって Web ページを収集し、検索インデックスを生成することで更新の自動化を図っている。利用者は、この検索エンジンを利用することで、複数の Web サイトを順に参照することなく、欲しい情報をワンストップで検索できる。

新着お知らせは、Web サイトの新着情報を一括表示する機能である。RSS パーサによって Web サイトが配信する RSS を解析し、それをポータルサイトに反映させることで、更新の自動化と迅速性の向上を図ってい

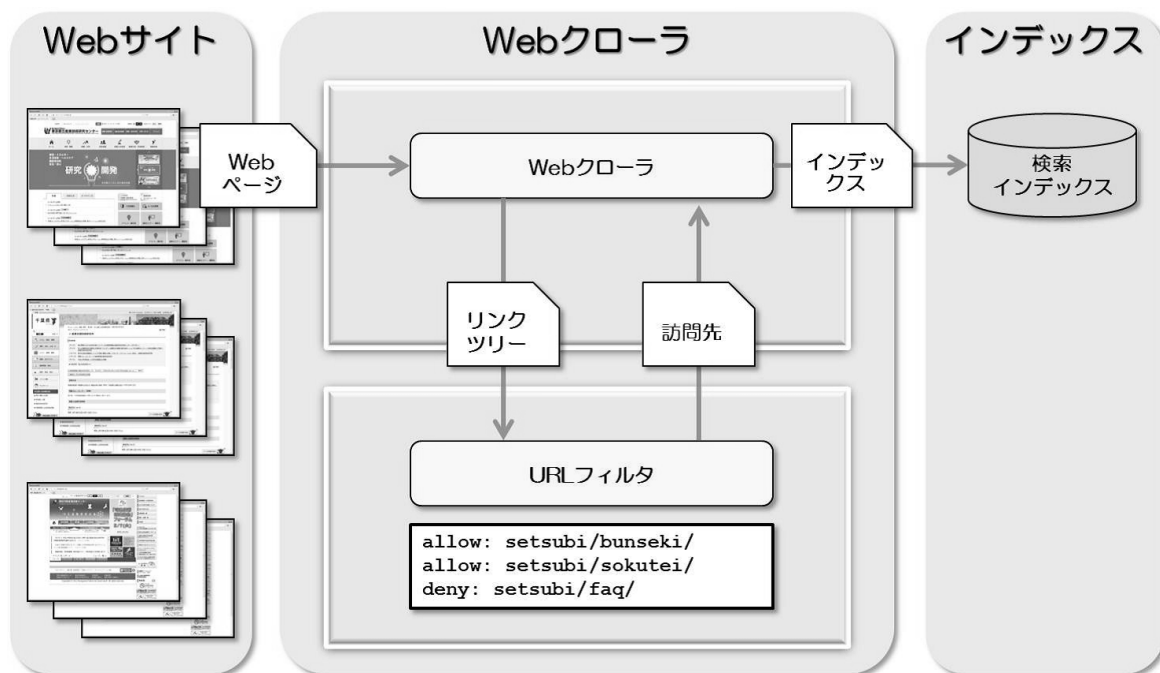


図 3.2. Web クローラの構成

る。利用者は、この新着お知らせを利用することで、複数の Web サイトを順に参照する必要がなくなり、素早く情報収集ができる。

以降では、提案サイトを特徴づける検索エンジンと新着お知らせの詳細について述べる。

3.2 検索エンジン

本検索エンジンは、Web サイト内の情報を一括検索する機能である。本検索エンジンは、全文検索システムと Web クローラからなる。全文検索システムの検索モデルは、ブーリアン検索モデルとベクトル空間モデルのいずれかを使用するか、もしくは両者を併用できるようにしている。インデックス方式は、分かち書きと文字バイグラムของいずれかを使用するか、もしくは両者を併用できるようにしている。各々の手法は、Manning

ら [57] および北ら [58] の定義に従う。

Web クローラは，Web サイトをクローリングし，検索インデックスを更新するプログラムである．Web クローラが対応する文書形式は，以下のとおりである．

- プレーンテキスト (.txt, .text, .asc など)
- ハイパーテキスト (.html, .xhtml など)
- 電子メール (.eml, .mime, .mht, .mhtml など)
- PDF (.pdf)
- MS-Office (.doc, .xls, .ppt など)
- DocuWorks (.xdw)

検索インデックスの更新にあたり，ポータルサイトの利用者にとって不要な情報が混在しないようにしなければならない．本提案では，Web クローラの内部に情報フィルタリング機能を持たせることで，これに対処する．Web クローラの構成を図 3.2 に示す．まず，Web クローラは，Web サイトから Web ページを収集し，リンクツリーを生成する．次に，リンクツリーに対して URL フィルタを適用し，クローリング対象とすべきかどうかを判断する．最後に，クローリング対象のページのみを，改めてインデクシングする．これにより必要な情報のみを選別した上でインデクシングできる．

このアイデアは，公共情報を発信している Web サイトでは，Web ページの意味によって構造化されているという事実に基づいている．表 3.1 に示すように，地域の産業情報は単一の Web ページもしくは特定の URL 配下にまとめて配置されている．表 3.2 に示す地域の観光情報も同様である．つまり，この配下のみをクローリング対象とすることで，産業情報あ

表 3.1. 地域の産業情報の配置箇所

地域	配置箇所
北海道	www.pref.hokkaido.lg.jp/105.htm
宮城県	www.pref.miyagi.jp/life/7/
東京都	www.metro.tokyo.jp/sangyo/
愛知県	www.pref.aichi.jp/life/5/
大阪府	www.pref.osaka.lg.jp/life/list3.php?ctg03_id=6
広島県	www.pref.hiroshima.lg.jp/life/16/
愛媛県	www.pref.ehime.jp/shigoto/
福岡県	www.pref.fukuoka.lg.jp/life/4/

表 3.2. 地域の観光情報の配置箇所

地域	配置箇所
北海道	www.pref.hokkaido.lg.jp/101.htm
宮城県	www.pref.miyagi.jp/life/3/11/
東京都	www.metro.tokyo.jp/kyoiku/kanko/kanko/
愛知県	www.pref.aichi.jp/life/2/11/
大阪府	www.pref.osaka.lg.jp/life/list3.php?ctg03_id=9
広島県	www.pref.hiroshima.lg.jp/life/12/
愛媛県	www.pref.ehime.jp/shigoto/kanko/
福岡県	www.pref.fukuoka.lg.jp/life/6/

るいは観光情報を選別してインデクシングできる。

検索エンジンの実装においては、オープンソースの全文検索システムである Hyper Estraier[59, 60] を利用する。Hyper Estraier には標準で

```
seed: 1.0 | http://www.pref.miyagi.jp/index.html
allowrx: ^http://www.pref.miyagi.jp/life/7/
allowrx: ^http://www.pref.miyagi.jp/life/3/11/
```

図 3.3. Web クローラの設定例

Web クローラが付属しており、その設定を適切に行うことで URL フィルタを実装できる。図 3.3 に Web クローラの設定例を示す。1 行目では、クローリングの種文書を与えており、ここで指定した URL がクローリングの開始点である。2 行目と 3 行目では、クローリング対象とする URL のルールを正規表現で与えている。この設定で Web クローラを実行すると、Web サイト `www.pref.miyagi.jp/` 内の `life/7/` 配下と `life/3/11/` 配下のみをクローリングする。

3.3 新着お知らせ

新着お知らせは、Web サイトの新着情報を一括表示する機能である。図 3.1 に示した RSS パーサは、Web サイトの RSS フィードを解析し、新着リストを生成するプログラムである。例えば、ある Web サイトの RSS フィードが、図 3.4 のようになっていたとする。RSS パーサは、RSS フィード内の各 item の情報を図 3.5 に示す新着リストに変換する。新着リストは、title と link の組で構成される。表示用 CGI は、このリストをもとにして HTML を生成する。

また、この新着リストは、検索エンジン側の Web クローラにも渡され、実行のトリガーにもなる。Web クローラは、新着リストの中にクローリ

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF
  (中略)
  <item rdf:about="http://www.xxx.com/seminar.html">
    <title>セミナーに参加される方へ</title>
    <link>http://www.xxx.com/seminar.html</link>
    (中略)
  </item>
  <item rdf:about="http://www.xxx.com/kenkyu.html">
    <title>研究発表会を開催します</title>
    <link>http://www.xxx.com/kenkyu.html</link>
    (中略)
  </item>
</rdf:RDF>
```

図 3.4. RSS フィード

ング対象として妥当な Web ページ，すなわち URL フィルタにマッチする Web ページがあれば，その時点でクロールを開始する。

新着お知らせの実装は，サーバサイドで動作するスクリプト言語と，そのライブラリを利用する。例えば Perl であれば XML::RSS モジュール，Ruby であれば RSS::Parser クラスが該当する。

3.4 Google カスタム検索と比べた利点

ここでは，提案サイトの利点を，Google カスタム検索を利用したシステムと比べながら述べる。検索対象としている Web ページには，セミ

”セミナーに参加される方へ”, ”http://www.xxx.com/seminar.html”
”研究発表会を開催します”, ”http://www.xxx.com/kenkyu.html”

図 3.5. 新着リスト

ナー開催情報など、迅速に発信すべき情報を含んでいる。このような情報を含む場合、Google カスタム検索の利用は不向きである。なぜならば、Web クローラが Web サイトを訪れるタイミングが不確定なため、迅速な情報発信ができない場合が想定されるからである。Google Search Console のように、Web クローラに更新を知らせるツールは存在するものの、更新の都度、Web クローラに知らせていたのでは、保守性が高いとはいえない。

一方で、提案サイトは、Web クローラを内部に持っており、任意のタイミングでクローリングできるため、Google カスタム検索のような問題は起きない。迅速に発信すべき情報を検索対象とする場合、Google カスタム検索を利用するよりも、提案サイトの方が適している。

3.5 第3章のまとめ

既存サイトは、保守性、検索精度、迅速性のいずれか 1 つ以上が欠けるものしかなかった。そこで本章では、保守性、検索精度、迅速性に優れたポータルサイトを実現するため、Web クローラと RSS パーサを用いたポータルサイトを開発した。Web クローラによる保守性の向上策、URL フィルタによる検索精度の向上策、RSS パーサによる迅速性の向上策に

ついて、一括して述べた。

次章では、提案サイトの社会実装と評価について述べ、保守性、検索精度、迅速性の面で優れていることを示す。

第4章

産業ポータルサイトへの応用

前章では，Web クローラと RSS パーサを用いたポータルサイトを開発し，保守性，検索精度，迅速性のいずれをも向上させるポータルサイトとして提案した．

本章では，提案サイトの社会実装である首都圏テクノナレッジ・フリーウェイ [61] を取り上げ，事業化に至った経緯と評価実験について述べる．

4.1 首都圏公設試験研究機関連携体

首都圏公設試験研究機関連携体 [62] は，首都圏の公設試の連携体である．対象機関は以下の5機関である．

- 埼玉県産業技術総合センター
- 千葉県産業支援技術研究所
- 地方独立行政法人東京都立産業技術研究センター
- 神奈川県産業技術センター
- 横浜市工業技術支援センター



図 4.1. 首都圏テクノナレッジ・フリーウェイ

首都圏公設試験研究機関連携体は、利用者サービスのワンストップ化、職員間の技術交流や技術協力を目的として発足した。連携体の活動内容の一部を以下に示す。

- 首都圏テクノナレッジ・フリーウェイの運営：
 - 各機関の情報を集約したポータルサイトである。
- パートナーグループの運営：
 - 研究分野ごとにグループを結成し、横断的な技術交流を図る。
- TKF フォーラム：
 - 各機関の職員が一同に集まり、連携体全体の交流を図る。



図 4.2. 設備検索フォーム

本連携体の主な事業の1つが首都圏テクノナレッジ・フリーウェイの運用である。首都圏テクノナレッジ・フリーウェイのブラウザイメージを図4.1に示す。以下は、首都圏テクノナレッジ・フリーウェイの主な機能である。

- 設備検索
- 技術検索
- 各機関からのお知らせ
- 技術相談フォーム

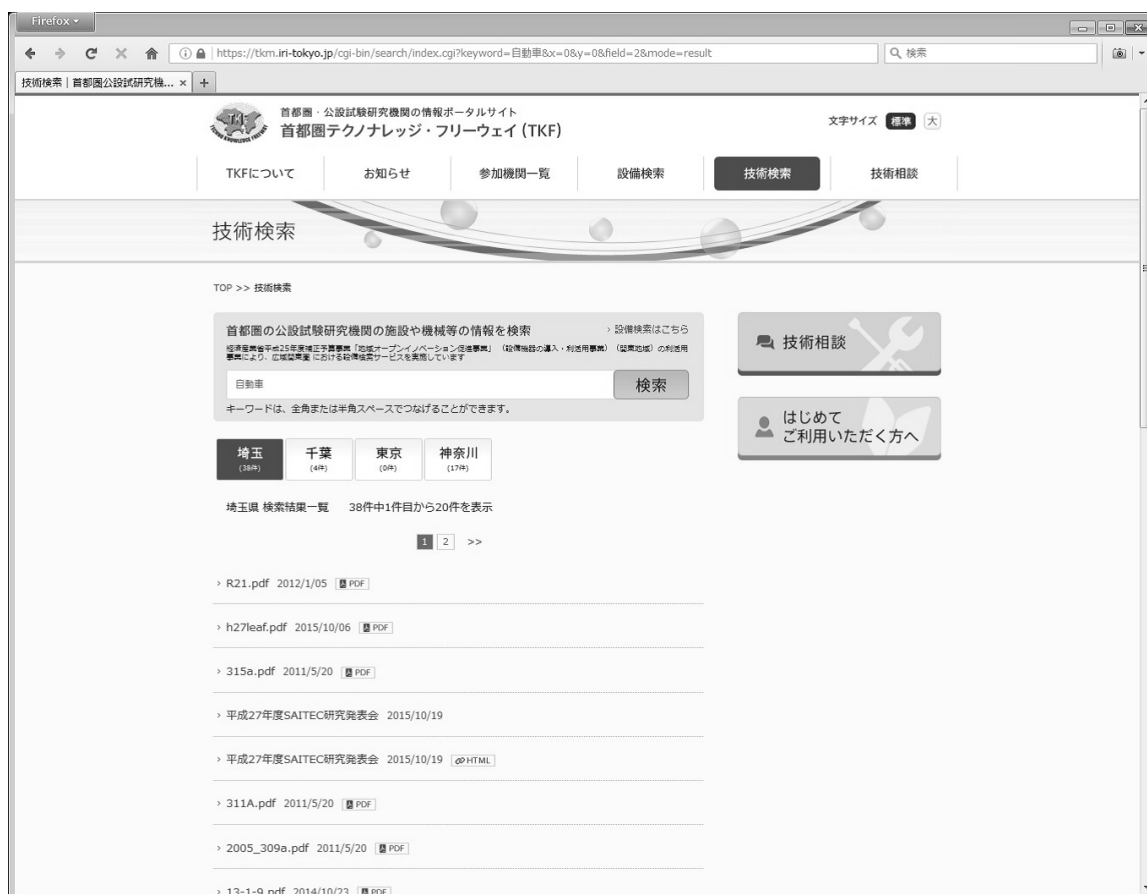


図 4.3. 技術検索フォーム

設備検索は、各機関が保有する設備情報を一括検索する機能である。図 4.2 に設備検索の検索フォームを示す。職員や利用者は、この検索機能を利用することで、目的の設備を保有する公設試をワンストップで特定できる。

技術検索は、各機関が保有する知的財産や研究報告などを一括検索する機能である。図 4.3 に技術検索の検索フォームを示す。ユーザインターフェイスを、設備検索のそれと全く同一にすることで、同様の手順で利用できるようにしている。

各機関からのお知らせは、各機関の新着情報を一括表示する機能である。図 4.4 にそのブラウザイメージを示す。利用者は、このページを参照



図 4.4. 各機関からのお知らせ

することで、各機関がいち早く伝えたい情報、例えばイベントの開催通知や料金表の改訂情報などを取得することができる。

技術相談フォームは、相談窓口を一本化した機能である。図 4.5 にそのブラウザイメージを示す。利用者がこのフォームから相談内容を送信すると、各機関の担当者へ相談内容がメールで一斉送信される。

これらの機能は、本論文の提案サイトによって実現されている。以降で、提案サイトが事業化に至った経緯について述べる。

図 4.5. 技術相談フォーム

4.2 事業化に至る経緯

本節では、本論文の提案サイトが事業化に至った経緯を時系列で述べる。ここで述べることは、あくまで広域関東圏の公設試における経験に基づくものであり、必ずしも全国の公設試に当てはまるものではない。

4.2.1 ポータルサイトの初期

首都圏公設試のポータルサイトの初期 [63] では、手動による情報更新が行われていた。具体的には、ポータルサイトの管理者が、各機関から情

報を保存した外部記録媒体の提供を受け、その情報をもとに更新するという運用方法である。

運用当初から挙げられていた問題は、保守性と迅速性が低いことであるが、特に迅速性の低さが問題視されていた。迅速性の低さは、各機関において、設備の廃止や料金の改定など、利用者に大きな影響を及ぼすような事業内容の変更が行われてから、その情報がポータルサイトに反映されるまでに遅延が生じることを意味している。さらに、この問題は、迅速性に欠けるというだけでなく、Web サイトとポータルサイトの間での情報の整合性が、長期に渡って担保されないという問題も含んでいた。

4.2.2 編集用フォームの設置

そこで、山田ら [64] は、迅速性を向上させるために、情報の更新をポータルサイトの管理者に一任するのではなく、図 4.6 に示す編集用フォームを設け、各機関がそれぞれ編集する運用に改めた。その結果、外部記録媒体が管理者の手元に届くのを待つことなく更新でき、遅延はある程度改善された。

しかしながら、新たな問題が指摘された。それは、将来的に外部ページの編集が禁止される公設試の存在である。ポータルサイトは、地方独立行政法人東京都立産業技術研究センターが保有する Web サーバ上にあつたので、それ以外の公設試から見れば外部ページであり、これも禁止対象となる。この制限は、地方自治体における情報漏洩防止策の 1 つとして計画されたもので、地方自治体の設置機関である公設試もそれに従わざるを得ない。したがって、各機関は、自身が保有するサーバ上にあるページのみ編集が許可されるという条件のもとで、運用方法を再検討せざるを得なくなった。

埼玉県産業技術総合センター 生産技術部

二次元座標測定顕微鏡

施設所在地

部署

機器名

仕様

用途

製造者

型番

導入年度

キーワード

画像ファイル

画像の削除 する しない



図 4.6. 旧サイトの編集用フォーム

4.2.3 XML による情報更新

北原ら [65] は、設備情報更新用の XML フォーマットを定め、これを利用して更新する方法を検討した。フォーマットの例を図 4.7 に示す。まず、各機関がフォーマットに則って更新情報をまとめた XML ファイルを

```
<?xml version="1.0" encoding="UTF-8"?>
<equipments pref="tokyo">
  <equipment>
    <name> 電子ビーム描画装置 </name>
    <spec> 加速電圧:100kV </spec>
    <spec> 最小線幅:8nm </spec>
    <spec> 継ぎ精度:40nm </spec>
    <model> ELS-7000Ac </model>
    <year> 2011 </year>
    <remarks 事前講習必須 </remarks>
  </equipment>

  <equipment>
    <name> Ge 半導体検出器 </name>
    <spec> 相対効率:kV36 </spec>
    <spec> 分解能:1.90keV </spec>
    <model> DSA1000 </model>
    <year> 1995 </year>
  </equipment>
</equipments>
```

図 4.7. 情報更新用 XML フォーマット

作成する。次に、以下のいずれかの方法で、ポータルサイトに更新情報を伝える。

- Web サーバ側に XML ファイルを追加する。ポータルサイトは、こ

のファイルを定期的にダウンロードする。

- Web サーバとポータルサイトの間で、XML/SOAP メッセージをやり取りする。

これらの方法であれば、外部ページを編集することなく、自身のサーバ上のファイルさえ編集すれば、ポータルサイトの情報を更新することができる。

ところが、この方法であっても対応できない公設試が存在した。それは、地方自治体が保有するコンテンツ管理システムを用いて Web サイトを運営している公設試である。地方自治体が運用するコンテンツ管理システムでは、公設試の裁量で編集可能なページは限られるので、図 4.7 のようなファイルを新規に追加できない。したがって、各機関は、情報更新用のファイルを新規に追加できないという条件も課せられることとなった。

4.2.4 提案サイトの事業化

ここまでで課せられた制約条件を以下に示す。

- 公設試の裁量で編集可能な Web ページは、自身の Web サイト上にあるページのみである。
- 公設試の裁量でファイルを新規に追加することはできない。

これらの打開策として、Web クローラを利用して、検索対象とする Web ページへのリンクを管理する方法を検討した。これが本論文の提案サイトの原型である。この方法であれば、外部ページの編集を要しないので、1つ目の条件を満たす。また、ファイルを新規に追加する必要もないので、2つ目の条件も満たす。さらに、Web ページのみ編集すれば、ポータルサイトにも自動で反映されるので、各機関から見ればメンテナンスフ

表 4.1. 設備情報とリンク構造の変更回数

	2012 年		2013 年	
	設備情報	リンク構造	設備情報	リンク構造
埼玉	14	0	5	0
千葉	3	0	5	0
東京	83	0	59	1
神奈川	3	0	8	0
合計	103	0	73	1

リーであるという利点もある。

以上を基に、先の 2 案と本案をまとめて首都圏公設試験研究機関連携体の最高意志決定機関である首都圏公設試連携推進会議に提案したところ、本案の採用が認められ、事業化に至った。

4.3 評価

本節では、提案サイトと既存サイトを比較検討することで、提案サイトが保守性と検索精度の面で優れていることを示す。

4.3.1 保守性の評価

ここでは、提案サイトの保守性を評価する。評価対象は、首都圏の 4 つの公設試とし、具体的には以下に挙げる機関の Web サイトである。また、既存サイトとの比較を容易にするため、既存サイトで既に実装されている保有設備の検索という観点から評価する。

- 埼玉県産業技術総合センター

- 千葉県産業支援技術研究所
- 地方独立行政法人東京都立産業技術研究センター
- 神奈川県産業技術センター

東北 6 県公開設備・機器データベースのような既存サイトは、設備情報が更新される度に、ポータルサイトの情報も更新が必要になる。すなわち、既存サイトの保守作業量は、設備情報の更新回数に依存する。一方、提案サイトは、Web クローラによって情報を収集するため、Web サイトの設備情報が変更されたとしても保守作業は必要ない。保守を要するのは、Web サイトのリンク構造が変更されたときであり、この場合は URL フィルタの設定変更が必要になる。すなわち、提案サイトの保守作業量は、リンク構造の変更回数に比例する。ゆえに、ある一定期間内の Web ページとリンク構造の変更回数を比較し、後者が少なければ、提案サイトの保守の優位性を示すことができる。

そこで、評価対象である 4 機関の Web サイトを調査し、設備情報とリンク構造の変更回数を集計したところ、表 4.1 の結果を得た。表 4.1 は、両者の変更回数を年度ごとに集計したものである。この表から、設備情報の変更回数は多い年で 103 回であり、既存サイトは、1 年間でこれだけの回数の保守を要する。一方、サイト構成の変更回数は多くとも 1 回であり、Web クローラの設定変更作業もやはり 1 回で済む。

以上より、保守作業量が設備情報の変更回数に比例する既存サイトよりも、リンク構造の変更回数に比例する提案サイトの方が、保守性が高いと評価できる。

表 4.2. 評価環境・条件

評価対象	埼玉県産業技術総合センター 千葉県産業支援技術研究所 地方独立行政法人東京都立産業技術研究センター 神奈川県産業技術センター
検索ワード	造形 万能試験機 耐候性 雷サージ Ge 半導体 三次元測定 3D スキャナ 分光 エミッション 赤外線
正解文書の判定	目視
検索モデル	ブーリアン検索モデル
インデックス方式	分かち書き

4.3.2 検索精度の評価

ここでは、検索精度の評価を行う。保守性の評価と同様、評価対象は4つの公設試の Web サイトとし、保有設備の検索という観点から評価する。

まず、検索精度の指標となる適合率と再現率について述べる。適合率は、正確性の指標であり、検索ヒット数（検索結果として得られた文書の総数）に対する正解ヒット数（検索結果として得られ、かつ検索要求を満

表 4.3. クローリング開始点

機関	クローリング開始点
埼玉	www.saitec.pref.saitama.lg.jp/kaihou/kikilist.html
千葉	www.pref.chiba.lg.jp/sanken/kikisetsubi/ichiran.html
東京	www.iri-tokyo.jp/setsubi/index.html
神奈川	www.kanagawa-iri.go.jp/equipment.html

表 4.4. クローリング対象

機関	クローリング対象
埼玉	www.saitec.pref.saitama.lg.jp/kaihou/
千葉	www.pref.chiba.lg.jp/sanken/kikisetsubi/
東京	www.iri-tokyo.jp/setsubi/
神奈川	www.kanagawa-iri.go.jp/equipment/

たす文書の総数)の比で表される。検索ヒット数を N 、正解ヒット数を R とすると、適合率は R/N となる。一方、再現率は、網羅性の指標であり、正解文書数（検索要求を満たす文書の総数）に対する正解ヒット数の比で表される。正解文書数を C 、正解ヒット数を R とすると、再現率は R/C となる。

評価環境や条件を表 4.2、評価に用いた Web クローラの設定を表 4.3 および表 4.4 に示す。評価対象は 4 機関の Web サイトである。評価に用いたキーワードは、地方独立行政法人東京都立産業技術研究センターの Web サイトにおける検索ワードランキングの上位 10 ワードとする。正解

表 4.5. 適合率

検索ワード	横断検索			URL フィルタ		
	N	R	R/N	N	R	R/N
造形	358	11	0.03	17	11	0.65
万能試験機	416	14	0.03	22	14	0.63
耐候性	162	11	0.07	13	11	0.85
雷サージ	83	9	0.11	12	9	0.75
Ge 半導体	44	1	0.02	2	1	0.50
三次元測定	477	19	0.04	30	19	0.63
3D スキャナ	11	1	0.09	2	1	0.50
分光	201	14	0.07	15	14	0.93
エミッション	130	10	0.08	11	10	0.91
赤外線	411	16	0.04	21	16	0.76

文書の判定は、具体的な設備情報、例えば機器分類、メーカー、型番、仕様などが文書に含まれているかどうかを目視によって行う。

この条件でクローリングと検索を実行したときの検索ヒット数 N 、正解ヒット数 R 、適合率 R/N を、横断検索と URL フィルタを設定した場合を比較して、表 4.5 に示す。横断検索の列は、全ての Web ページを検索対象としたときの結果であり、九州・沖縄バーチャル公設試と同様の手法である。URL フィルタの列はクローラに表 4.3 および表 4.4 の設定を行うことで、検索対象を限定した場合の結果である。九州・沖縄バーチャル公設試に合わせ、検索モデルはブーリアン検索モデル、インデックス方式は分かち書きとしている。いずれのキーワードにおいても、本検索エンジンの方が適合率が高いことが分かった。

次に再現率の比較を行う。まず、再現率を具体的な実測値として求める

ことは難しい。なぜならば、検索システム用テストコレクションなど特殊な場合を除き、正解文書数 C を把握することが困難であるためである。ただし、再現率の大小比較は可能である。検索対象とする文書集合が等しいならば C も等しいので、正解ヒット数 R の大小によって比較ができる。表 4.5 から、いずれのキーワードにおいても両者の R は等しく、したがってまた再現率も等しいことが分かった。

以上より、適合率は高く、再現率は両者で等しいので、既存サイトよりも提案サイトの方が、検索精度が高いと評価できる。

4.4 第4章のまとめ

本章では、提案サイトの社会実装である首都圏テクノナレッジ・フリーウェイについて述べた。首都圏テクノナレッジ・フリーウェイの機能である設備検索、技術検索、各機関からのお知らせ、技術相談フォームを紹介し、提案サイトを基礎として実現されていることを述べた。そして、提案サイトの評価実験について述べ、既存サイトと比べて保守性、検索精度、迅速性の面で優れていることを示した。

次章では、首都圏テクノナレッジ・フリーウェイの広域関東圏への拡張について述べる。

第 5 章

産業ポータルサイトの拡張

前章では、提案サイトの社会実装である首都圏テクノナレッジ・フリーウェイを紹介し、既存サイトと比べて保守性、検索精度、迅速性の面で優れていることを示した。

本章では、Web クローラと RSS パーサに加え、パターンマッチングを併用 [66, 67] することで、拡張前と同等の保守性、検索精度、迅速性を維持して拡張できることを示す。

5.1 広域関東圏への拡張

公設試は各地方自治体に存在するが、この潜在能力を広域関東圏で十二分に発揮するためには、地域を越えた相互利用の促進が重要である。具体的には、各都県の公設試が自身の得意分野を生かしながら役割分担し、広域関東圏内で整合性のとれた試験研究設備の導入や、相互利用の推進、実現性の高い開発テーマ選定、当該テーマに沿った技術支援の実施等を促進することが重要である。このような理由から、関東甲信越静の知事や政令指定都市市長などからなる関東地方産業競争力協議会 [68, 69] は、試験研

究機能の強化策として、各都県の公設試が保有する設備を一括検索できるポータルサイトの構築を提案した。

5.2 拡張法の検討

Web クローラと RSS パーサによるポータルサイトの構成法は、既存サイトよりも保守性、検索精度、迅速性に優れていた。しかしながら、評価対象が4つと少なく、十分な評価ができていなかった。全く同様の構成法で、より多くの Web サイトを対象にした場合であっても、同等の保守性、検索精度、迅速性が得られれば、提案サイトの有用性をより客観的に示すことができる。

そこでまずは、全く同様の構成法で広域関東圏の公設試、具体的には次に示す 13 機関の Web サイトへ拡張できるかどうかを検討する。

- 茨城県工業技術センター
- 栃木県産業技術センター
- 群馬県立産業技術センター
- 埼玉県産業技術総合センター
- 千葉県産業支援技術研究所
- 地方独立行政法人東京都立産業技術研究センター
- 神奈川県産業技術センター
- 横浜市工業技術支援センター
- 新潟県工業技術総合研究所
- 山梨県工業技術センター
- 山梨県富士工業技術センター
- 長野県工業技術総合センター

表 5.1. 設備情報の文書形式

機関	文書形式
茨城	ハイパーテキスト, PDF, MS-Office
栃木	ハイパーテキスト, PDF
群馬	ハイパーテキスト, PDF
埼玉	ハイパーテキスト, PDF, MS-Office
千葉	ハイパーテキスト, PDF
東京	ハイパーテキスト, PDF
神奈川	ハイパーテキスト, PDF
横浜	ハイパーテキスト, PDF, MS-Office
新潟	ハイパーテキスト, PDF
山梨	ハイパーテキスト, PDF, MS-Office
山梨富士	ハイパーテキスト, PDF, MS-Office
長野	ハイパーテキスト, PDF
静岡	ハイパーテキスト, PDF, MS-Office

- 静岡県工業技術研究所

以降では、検索エンジンと新着お知らせの拡張について述べる。

5.2.1 検索エンジンの拡張

ここでは、検索エンジンを拡張するための条件を考える。提案サイトに用いる検索エンジンは、全文検索システムと Web クローラからなり、Web サイトから情報を自動収集している。ゆえに、同様の構成で拡張するためには、Web サイト上に存在する文書の形式が、Web クローラに対応していることが条件となる。対応する文書形式を次に示す。

表 5.2. 設備情報の配置箇所

機関	配置箇所
茨城	www.kougise.pref.ibaraki.jp/equipment/
栃木	www.iri.pref.tochigi.lg.jp/index.php
群馬	mac.tec-lab.pref.gunma.jp/
埼玉	www.saitec.pref.saitama.lg.jp/kaihou/
千葉	www.pref.chiba.lg.jp/sanken/kikisetsubi/
東京	www.iri-tokyo.jp/setsubi/
神奈川	www.kanagawa-iri.jp/equipment/
横浜	www.city.yokohama.lg.jp/keizai/shien/sien-c/machines/
新潟	www.iri.pref.niigata.jp/equipment/
山梨	www.pref.yamanashi.jp/kougyo-gjt/instruments/
山梨富士	www.pref.yamanashi.jp/kougyo-fj/instruments/
長野	www.gitc.pref.nagano.lg.jp/hoyu_details.php? www.gitc.pref.nagano.lg.jp/hoyu_kubun.php?
静岡	iri.pref.shizuoka.jp/cgi/search/

- プレーンテキスト (.txt, .text, .asc など)
- ハイパーテキスト (.htm, .html, .xhtml など)
- 電子メール (.eml, .mime, .mht など)
- PDF (.pdf)
- MS-Office (.doc, .xls, .ppt など)
- DocuWorks (.xdw)

プレーンテキスト、ハイパーテキスト、電子メールは、そのままの形式でインデクシングできる。PDF やマイクロソフト社の Microsoft Office®、富士ゼロックス社の DocuWorks™ は、文書フィルタによって中間ファイルに変換することでインデクシングできる。つまり、Web サイト上に存在する文書の形式が、上記のいずれかであれば、同様の構成で実現可能である。そこで、13 機関の Web サイト上に存在する文書の形式を調査し、表 5.1 の結果を得た。この結果により、全ての Web サイトにおいて、Web クローラに対応した文書形式が存在することが分かった。

また、本検索エンジンは、URL フィルタによって、Web クローラが訪問するページを特定の URL 配下に限定することで、既存サイトよりも高い検索精度を得ている。ゆえに、設備情報の検索という観点で拡張前と同等の検索精度を求めるならば、設備情報が特定の URL 配下に配置されていることが条件となる。そこで、Web サイトの設備情報の配置箇所を調査し、表 5.2 の結果を得た。この結果により、全ての Web サイトにおいて、特定の URL 配下にまとめて配置されていることが分かった。

5.2.2 新着お知らせの拡張

ここでは、新着お知らせを拡張するための条件を考える。新着お知らせは、Web サイトが配信する RSS フィードを解析し、その内容を一括表示する機能である。この機能によって、提案サイトの情報発信の迅速性を向上させている。ゆえに、同様の構成で拡張し、なおかつ迅速性を担保するためには、Web サイトが RSS フィードで新着情報を配信している必要がある。

そこで、13 機関の RSS フィードの配信状況を調査したところ、表 5.3 の結果を得た。この結果により、全ての Web サイトにおいて、新着お知

らせを拡張できることが分かった。1つ注意を要するのは、新潟県工業技術総合研究所や山梨県工業技術センターのように、自らが保有する Web サイトではなく、県庁などの Web サイトから RSS を配信している場合である。この場合、RSS フィードの中に公設試と無関係の情報も含まれる。その対処として、<dc:creator>タグや<dc:contributor>タグを参照し、関連のある情報のみを抽出する工夫が必要となる。

表 5.3. RSS フィードの配信状況

機関	配信状況
茨城	www.pref.ibaraki.jp/chumoku.xml www.pref.ibaraki.jp/hotnews.xml
栃木	www.iri.pref.tochigi.lg.jp/index.php?id=253
群馬	www.tec-lab.pref.gunma.jp/rss.rdf
埼玉	www.pref.saitama.lg.jp/shinchaku/shinchaku.xml
千葉	www.pref.chiba.lg.jp/sanken/shinchaku.xml
東京	www.iri-tokyo.jp/shinchaku/shinchaku.xml www.iri-tokyo.jp/oshirase/oshirase.xml
神奈川	www.pref.kanagawa.jp/rss/10/list1.xml
横浜	www.city.yokohama.lg.jp/front/yokohama-latestnews-all.rss
新潟	www.pref.niigata.lg.jp/whatsnew.rdf www.pref.niigata.lg.jp/press_release.rdf www.pref.niigata.lg.jp/order.rdf www.pref.niigata.lg.jp/bosyu.rdf www.pref.niigata.lg.jp/shiken.rdf
山梨	www.pref.yamanashi.jp/shinchaku.xml
山梨富士	www.pref.yamanashi.jp/shinchaku.xml
長野	www.pref.nagano.lg.jp/chumoku/chumoku.xml
静岡	www.pref.shizuoka.jp/a_content/common/new.xml www.pref.shizuoka.jp/a_content/common/shinchaku.xml

表 5.4. 設備情報とリンク構造の変更回数（拡張後）

	2012 年		2013 年	
	設備情報	リンク構造	設備情報	リンク構造
茨城	5	1	8	0
栃木	4	0	5	0
群馬	5	0	4	0
埼玉	14	0	5	0
千葉	3	0	5	0
東京	83	0	59	1
神奈川	3	0	8	0
横浜	3	0	2	0
新潟	3	0	6	0
山梨	3	0	4	0
山梨富士	3	0	3	0
長野	6	0	20	0
静岡	6	0	4	0
合計	141	1	133	1

5.3 拡張後の評価

本節では、全く同様の構成で拡張した場合の保守性と検索精度を評価し、拡張時の問題点を抽出する。

5.3.1 拡張後の保守性

ここでは、拡張後のポータルサイトの保守性を評価する。拡張前と同様に、評価対象である 13 機関の Web サイトを調査し、設備情報とリンク構造の変更回数を集計した。集計結果を表 5.4 に示す。表 4.1 と比較すると、設備情報の更新回数は、拡張後に約 100 回増えており、拡張の影響を大きく受けることが分かった。一方で、リンク構造の変更回数は、2 年間で 1 回増えただけであり、拡張の影響は小さい。

以上より、保守作業量が設備情報の変更回数に比例する既存サイトよりも、リンク構造の変更回数に比例する提案サイトの方が、拡張時の保守性への影響は小さいと評価できる。

5.3.2 拡張後の検索精度

ここでは、拡張後のポータルサイトの検索精度を評価する。評価環境や条件を表 5.5、Web クローラの設定を図 5.6 および図 5.7 に示す。評価対象を 13 機関に拡張した以外は、拡張前と同様の条件である。

この条件でクローリングと検索を実行したときの検索ヒット数 N 、正解ヒット数 R 、適合率 R/N を、拡張前と拡張後を比較して、表 5.8 に示す。表 4.5 に示した拡張前 4 サイトでの横断検索の結果に比べると、拡張後の適合率は高いものの、提案ポータルサイトの条件で比較すると、5 個の検索ワードにおいて、拡張前よりも拡張後の適合率が低くなった。この評価では、公設試に補助金を提供した団体の紹介ページが検索ノイズとして現れ、それが多くを占めており、適合率を下げる主な原因となった。

表 5.5. 評価環境・条件（拡張後）

評価対象	茨城県工業技術センター 栃木県産業技術センター 群馬県立産業技術センター 埼玉県産業技術総合センター 千葉県産業支援技術研究所 地方独立行政法人東京都立産業技術研究センター 神奈川県産業技術センター 横浜市工業技術支援センター 新潟県工業技術総合研究所 山梨県工業技術センター 山梨県富士工業技術センター 長野県工業技術総合センター 静岡県工業技術研究所
検索ワード	造形 万能試験機 耐候性 雷サージ Ge 半導体 三次元測定 3D スキャナ 分光 エミッション 赤外線
正解文書の判定	目視
検索モデル	ブーリアン検索モデル
インデックス方式	分かち書き

表 5.6. クローリング開始点 (拡張後)

機関	クローリング開始点
茨城	www.kougise.pref.ibaraki.jp/equipment/01center.html www.kougise.pref.ibaraki.jp/equipment/02seni.html www.kougise.pref.ibaraki.jp/equipment/03yogyo.html
栃木	www.iri.pref.tochigi.lg.jp/index.php?id=489
群馬	mac.tec-lab.pref.gunma.jp/index.html
埼玉	www.saitec.pref.saitama.lg.jp/kaihou/kikilist.html
千葉	www.pref.chiba.lg.jp/sanken/kikisetsubi/ichiran.html
東京	www.iri-tokyo.jp/setsubi/index.html
神奈川	www.kanagawa-iri.go.jp/equipment.html
横浜	www.city.yokohama.lg.jp/keizai/shien/sien-c/machines/index.html
新潟	www.iri.pref.niigata.jp/equipment/equipment_list.html
山梨	www.pref.yamanashi.jp/kougyo-gjt/instruments/setsubi_all.html
山梨富士	www.pref.yamanashi.jp/kougyo-fj/instruments.html
長野	www.gitc.pref.nagano.lg.jp/hoyu_index.php
静岡	iri.pref.shizuoka.jp/cgi/search/eqp_all.html

表 5.7. クローリング対象 (拡張後)

機関	クローリング対象
茨城	www.kougise.pref.ibaraki.jp/equipment/
栃木	www.iri.pref.tochigi.lg.jp/index.php?
群馬	mac.tec-lab.pref.gunma.jp/
埼玉	www.saitec.pref.saitama.lg.jp/kaihou/
千葉	www.pref.chiba.lg.jp/sanken/kikisetsubi/
東京	www.iri-tokyo.jp/setsubi/
神奈川	www.kanagawa-iri.go.jp/equipment/
横浜	www.city.yokohama.lg.jp/keizai/shien/sien-c/machines/
新潟	www.iri.pref.niigata.jp/equipment/
山梨	www.pref.yamanashi.jp/kougyo-gjt/instruments/
山梨富士	www.pref.yamanashi.jp/kougyo-fj/instruments/
長野	www.gitc.pref.nagano.lg.jp/hoyu_kubun.php www.gitc.pref.nagano.lg.jp/hoyu_details.php
静岡	iri.pref.shizuoka.jp/cgi/search/

表 5.8. 適合率（拡張前および拡張後）

検索ワード	拡張前 4 サイト			拡張後 13 サイト		
	N	R	R/N	N	R	R/N
造形	17	11	0.65	24	17	0.71
万能試験機	22	14	0.63	42	30	0.71
耐候性	13	11	0.85	17	12	0.71
雷サージ	12	9	0.75	23	18	0.78
Ge 半導体	2	1	0.50	2	1	0.50
三次元測定	30	19	0.63	42	23	0.55
3D スキャナ	2	1	0.50	3	1	0.33
分光	15	14	0.93	20	19	0.95
エミッション	11	10	0.91	19	15	0.79
赤外線	21	16	0.76	25	18	0.72

5.4 検索エンジンの改良

本節では、URL フィルタのみでは除去できなかった検索ノイズへの対策について述べる。本論文では、Web クローラに対して Web ページのパターンマッチングを組み込むことで、これに対処した。

パターンマッチングを組み込んだ Web クローラの構成を図 5.1 に示す。まず、Web クローラは、Web サイトからページを収集する。次に、収集したページからリンクツリーと単語リストを生成する。そして、リンクツリーに対して URL フィルタを適用し、訪問候補を生成する。それと同時に、サポートベクターマシンによる単語リストのパターンマッチングを行い、前出とは別の訪問候補を生成する。この部分が、第 3 章で述べた Web

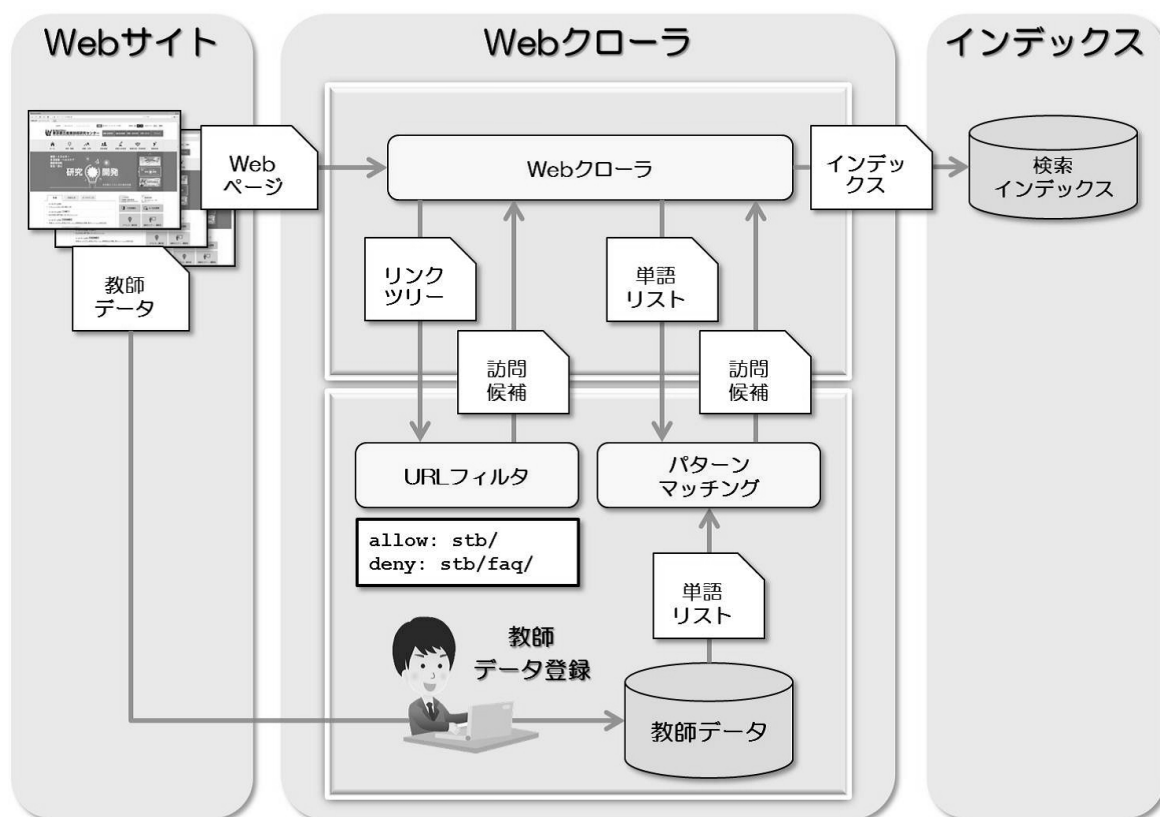


図 5.1. パターンマッチングを組み込んだ Web クローラの構成

クローラとの違いである。最後に、2つの訪問候補の積集合を求め、共通に含まれているページのみ、改めてインデクシングする。

次に、改良後の Web クローラおよび検索エンジンを、表 5.5、表 5.6 および表 5.7 の条件で改めて評価した。その時の適合率を、表 5.9 に示す。ここでは、表 5.8 と表 5.9 を比較しながら議論を進める。まず、両表の正解ヒット数 R は全て等しい結果となった。正解ヒットしたページを目視で確認したところ、全て同一のページであった。すなわち、パターンマッチングの誤判定により、設備情報のページであるにも関わらず、それが除外されるケースはなかった。また、7個の検索ワードにおいて、検索ヒット数 N が減少しており、検索ノイズを適切に除外できていることが

表 5.9. 適合率（パターンマッチング追加後）

検索ワード	N	R	R/N
造形	23	17	0.74
万能試験機	40	30	0.75
耐候性	14	12	0.86
雷サージ	20	18	0.90
Ge 半導体	1	1	1.00
三次元測定	33	23	0.70
3D スキャナ	3	1	0.33
分光	20	19	0.95
エミッション	19	15	0.79
赤外線	22	18	0.82

分かった。

以上のように、単語リストのパターンマッチングを併用することで、検索精度を大きく低下させることなく、広域関東圏の公設試へ拡張することができた。現在、首都圏テクノナレッジ・フリーウェイは、広域関東圏の公設試の総合カタログとして継続運営 [70, 71, 72, 73, 74] されている。

5.5 第5章のまとめ

本章では、首都圏テクノナレッジ・フリーウェイの広域関東圏への拡張について述べ、単語リストのパターンマッチングを併用することで、拡張前と同等の保守性、検索精度、迅速性を維持して拡張できることを示した。

次章では、提案サイトのもう1つの社会実装である観光ポータルサイトを紹介し、保守性、検索精度、迅速性に加えて、可用性が重要であること

を述べるとともに、その向上策と評価実験について説明する。

第 6 章

観光ポータルサイトへの応用

前章では，首都圏テクノナレッジ・フリーウェイの広域関東圏 13 機関への拡張について述べ，拡張前と同等の保守性，検索精度，迅速性を維持して拡張できることを示した。

本章では，提案サイトのもう 1 つの社会実装である観光ポータルサイト [75, 76, 77] を紹介する。観光ポータルサイトにおいては，保守性，検索精度，迅速性だけでなく，可用性が重要であることを述べるとともに，その向上策と評価実験について述べる。

6.1 観光業の広域連携

近年，情報化社会の進展や交通の利便性向上により，旅行者は多くの観光地や観光資源をより短い時間で巡ることができるようになってきている。複数の自治体区域にまたがって巡る広域的な周遊旅行を志向する傾向は，遠方から訪れる旅行者，特に外国人観光客に多い。ゆえに，観光庁 [78] は，観光業における広域連携によって，外国人観光客を主な消費者とした経済効果を，自治体の枠を超えて広域的に波及させることが重要であ

ると述べている。

産業の広域連携と同様に，観光業の広域連携においても，ポータルサイトの利用が有効である。そして，提案サイトを応用することで，観光ポータルサイトも同様に実現できる。つまり，クローリング開始点とクローリング対象を，集約したい観光情報に合わせて適切に設定すれば，保守性，検索精度，迅速性に優れた観光ポータルサイトの実現が可能となる。

6.2 ポータルサイトの可用性

東京オリンピック・パラリンピックを控え，今後10年間での外国人観光客の大幅増が見込まれている。このような背景から，外国人観光客にも利用しやすい観光ポータルサイトが望まれているが，彼らの言語格差が課題となっている。

言語格差を解消するためにはポータルサイトの多言語化が一般的である。例えば，外国人の訪日促進と受入環境整備に関する学術的動向 [79] を調査した結果，そのほとんど全てが多言語化に関するものであった。また，我が国の観光政策 [80] では，学術的な背景に基づいた外国人の受け入れ整備が進められているが，多言語化以外に具体的なアイデアは見られない。さらに，観光庁から配信されている外国人観光客向けのアプリケーション [81] は，2016年10月1日時点で18個存在するものの，やはり多言語化に終始しているのが現状である。しかしながら，多言語化によって必要十分な数の言語を網羅することは難しく，真の意味で言語格差を解消するには至らない。したがって，観光ポータルサイトでは，保守性，検索精度，迅速性の他に，言語格差を網羅的に解消する仕組み，すなわち外国人にとっての可用性が求められる。

そこで本論文では，言語による情報伝達には限界があるという立場に



図 6.1. 言語バリアフリー化した観光ポータルサイト

立った上で、言語によらない手段のみを利用した観光ポータルサイトを実現する。本ポータルサイトは、真に言語バリアフリーなポータルサイトである。

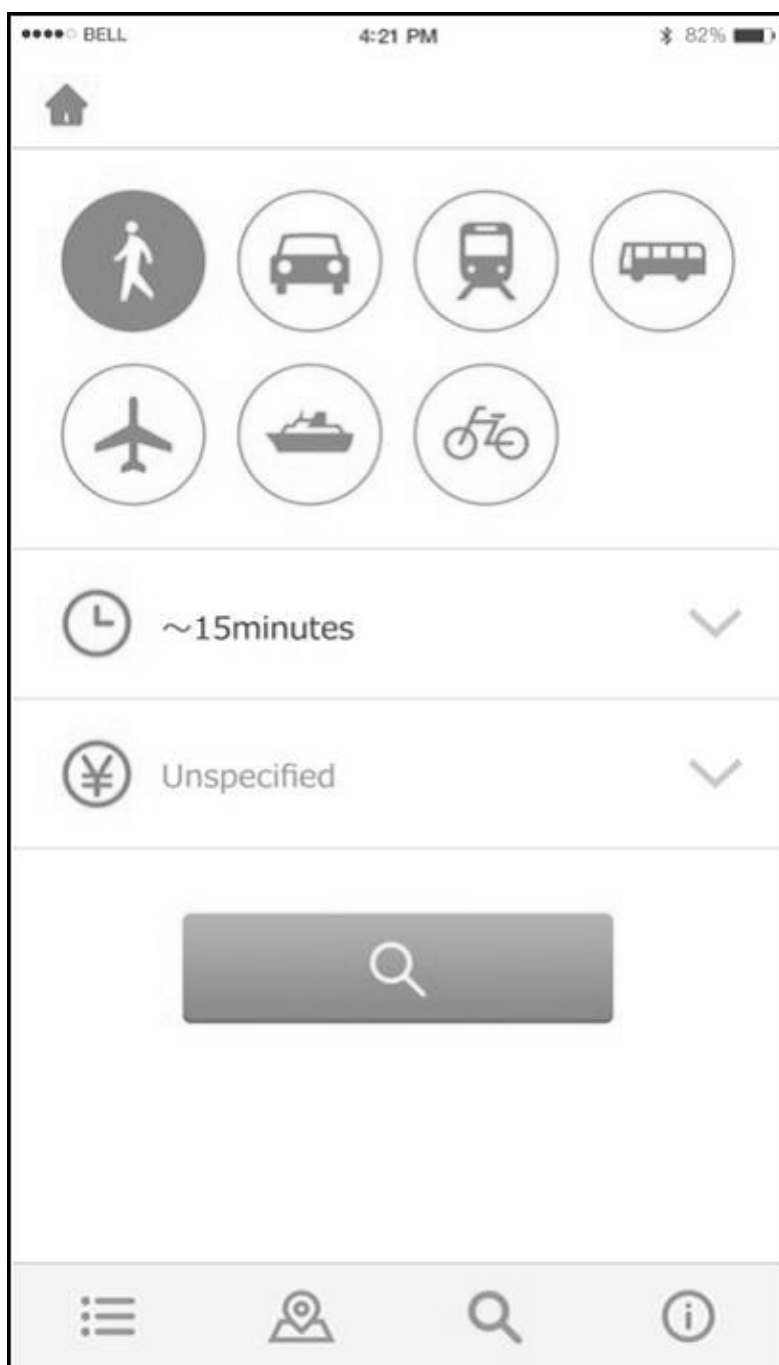


図 6.2. 言語バリアフリー化した検索インターフェイス

6.3 ポータルサイトの言語バリアフリー化

ここでは、ポータルサイトの言語バリアフリー化について述べる。本論文では、ピクトグラム、アラビア数字、空間イメージ、地理情報システム

を利用して，言語バリアフリー化する．図 6.1 に言語バリアフリー化したシステムを示す．ピクトグラムは JIS 規格 [82] で定められたものとし，観光スポットの種類や移動手段を示すために使用する．移動時間や料金などの表記は，世界共通の数字であることが望ましいため，アラビア数字のみを使用する．空間イメージは，観光スポットを容易に想像できるものとする．これらを地理情報システム上に重ねて表示することで，言語バリアフリー化を実現する．

観光スポットの検索インターフェイスについても，同様に言語バリアフリー化する．図 6.2 に言語バリアフリー化した検索インターフェイスを示す．観光スポットの種類や移動手段はピクトグラムから選択するボタン，移動時間や料金はアラビア数字から選択するドロップダウンリストで構成する．

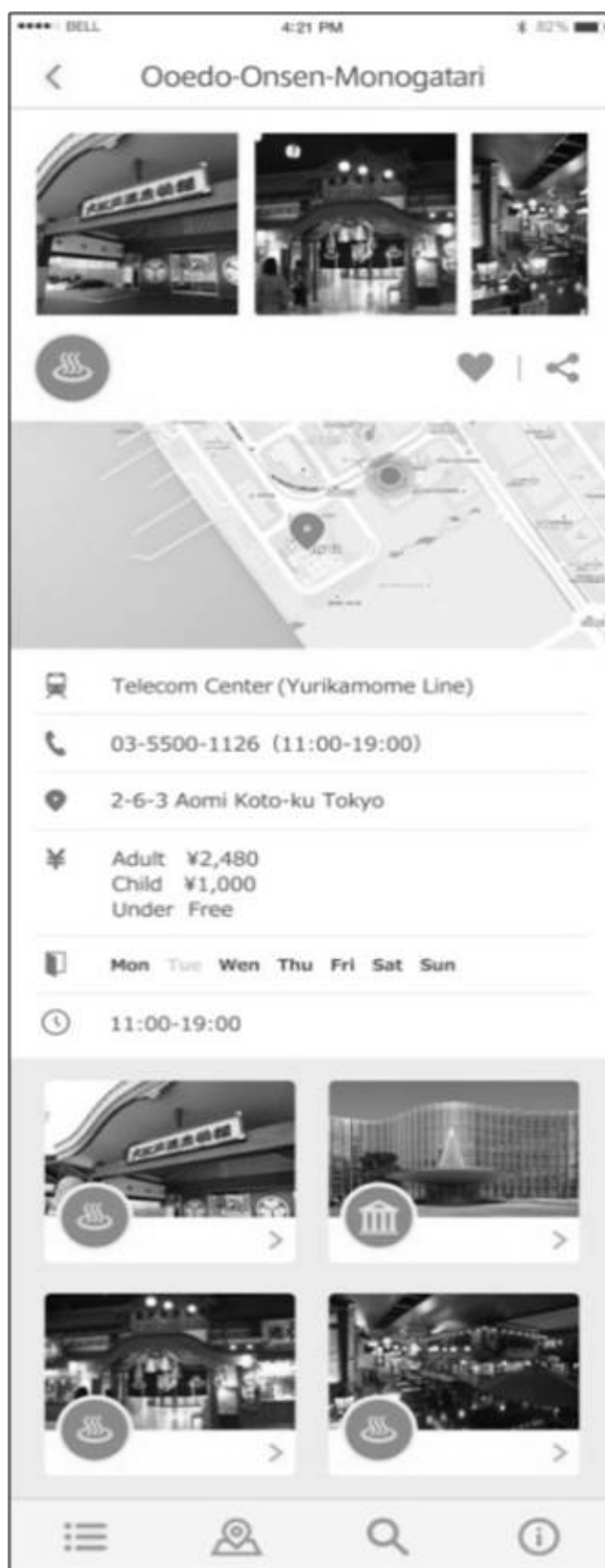


図 6.3. 類似スポットの推薦

6.4 類似スポットの推薦

言語バリアフリー化された検索インターフェイスは、キーワード検索に比べて検索語の入力の自由度が低いため、検索意図を反映しにくい。そこで、類似スポットの推薦機能を実現し、これに対処する。図 6.3 に推薦の様子を示す。図 6.3 下部の 4 つのスポットが、推薦された類似スポットである。もし、検索意図が反映されなかった場合でも、類似スポットから辿ることで、必要な情報を得ることができる。推薦される類似スポットも、図 6.2 の検索インターフェイスから絞り込むことができる。

次に、推薦の仕組みについて述べる。まず、図 6.4 のように、観光スポット間の類似度を前もって値付けしておく。データ構造は、観光スポットをノード、スポット間の類似度をリンクとするネットワークである。ユーザがある 1 つの観光スポットを選択すると、そのスポットと直接関連があるスポット、すなわち選択したスポットと直接リンクされているスポットを全て表示する。ユーザが絞り込みを実行したときは、リンクされているスポットの中から、条件に合うスポットのみを表示する。

登録された観光スポット数が少なければ、類似度の値付けは人手で行えばよい。しかしながら、登録された観光スポット数が多くなれば、保守性が低くなる。その場合は、図 6.5 のように、本論文で提案したパターンマッチングを応用することで解決できる。

6.5 可用性の評価

ここでは、観光ポータルサイトの可用性を定量的に示すことで、言語によらない手段であっても、観光スポットの検索や経路案内が可能であるこ

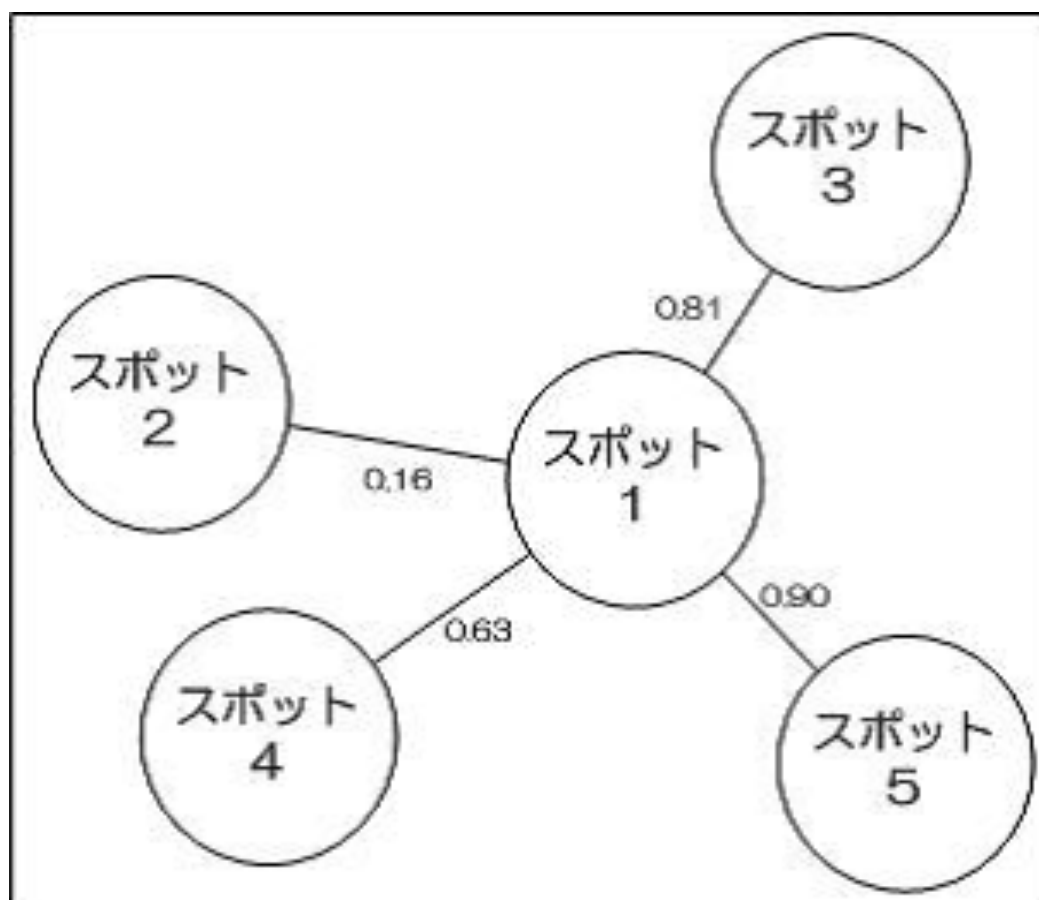


図 6.4. 観光スポットの類似度

とを実証する。定量的指標には画面遷移数を用い、検索や表示などの操作において、手戻りが何回発生したかを評価する。手戻りは、一般に少ないほどよいとされる。

実験の方法は次のとおりである。監督者は、被験者に対して、特定の観光スポットを探すよう指示する。被験者は、監督者の指示に従って、図 6.6 に示した試作機を操作して観光スポットを探し、見つけるまでの画面遷移数をカウントする。監督者から被験者への指示は、次の順に行った。

指示 1 検索機能を使って「すみだ水族館」を探してください。

指示 2 推薦機能を使って「レジャー施設」を 2 つ探してください。

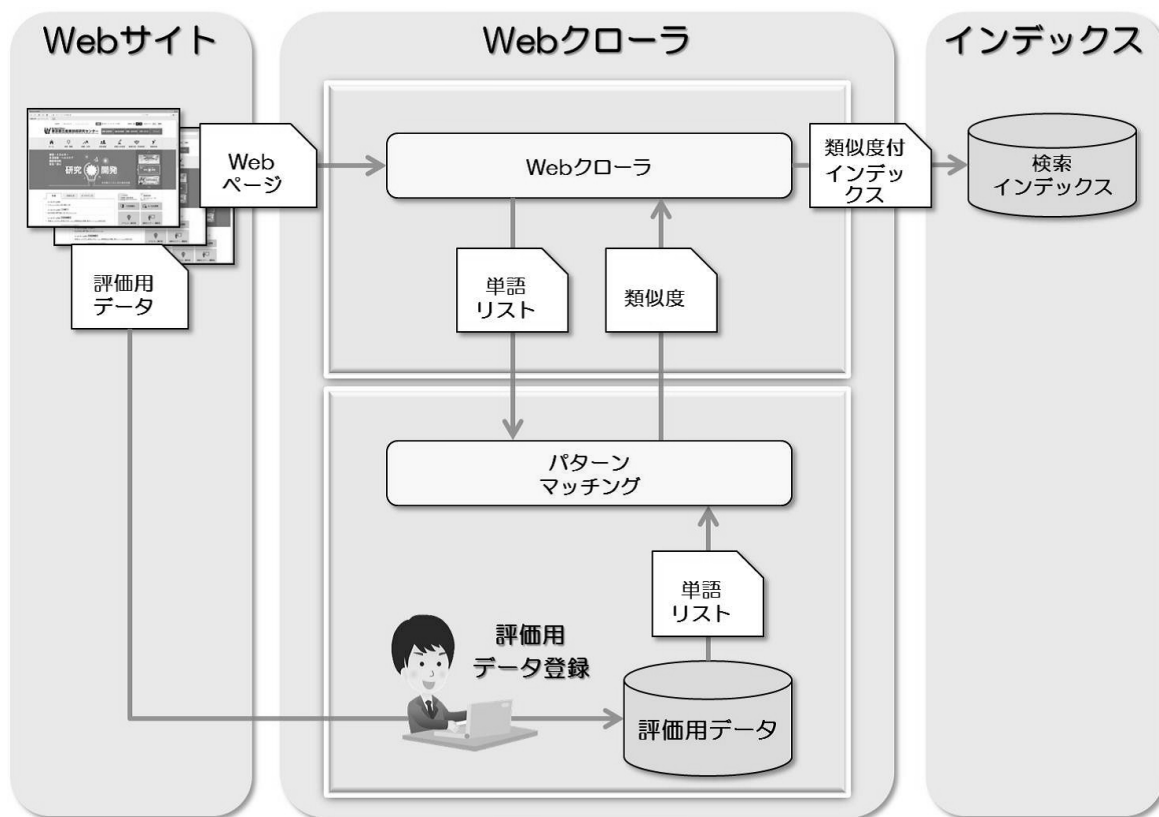


図 6.5. 類似度算出システムの構成

指示 3 検索や推薦を自由に使って「近隣の宿泊施設」を探してください。

表 6.1 に回答結果，表 6.2 に回答に至るまでの画面遷移数を示す．表 6.1 は，見つけた観光スポットが，監督者の指示に合致するものであったかを示す表である．表 6.2 は，正解したか否かに関わらず，被験者が観光スポットを見つけるまでの画面遷移数を示している．

指示 1 は，目的地が確定している状況を想定した実験である．この実験の画面遷移数は最大 7 回，平均 2 回程度であった．観光客が目的地の名前をキーワードに検索できるなら，1 回の検索実行でヒットするので，キーワード検索の方が優れている．一方で，名前が分からない，あるいは名前が読めないといった場合では，平均 2 回程度の操作で発見できることが分

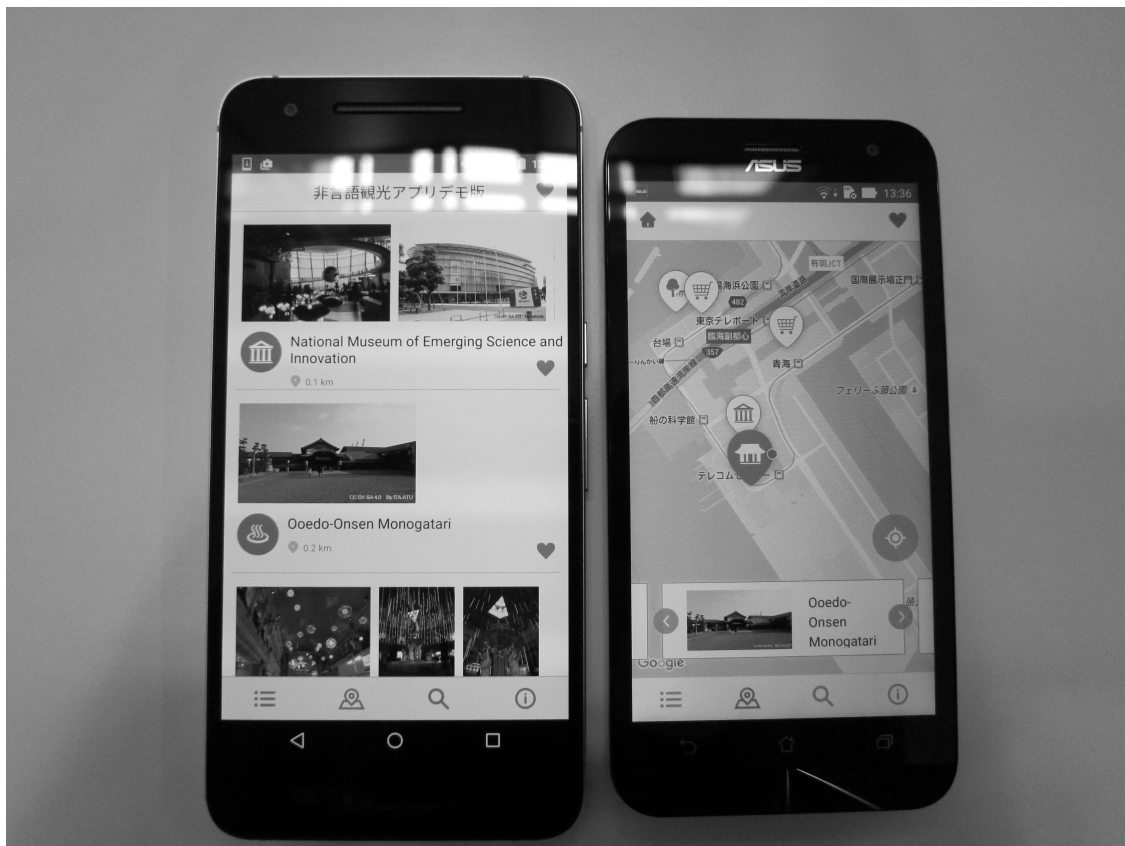


図 6.6. 試作機

かった。

指示 2 は、目的地が確定していない状況を想定した実験である。この実験では、画面遷移数は 1 回程度であり、キーワード検索に比べても同等以上の操作性であった。これは、レジャー施設というカテゴリから、複数の観光スポットを発見する際に、推薦機能が有効に働いたことを示している。

指示 3 は、より実際の状況に近い実験であり、新たな課題を抽出するためのものである。ある観光スポットに赴いた後、そこから近隣の施設を探す状況を想定した実験である。この実験では、一部の被験者の画面遷移数が多い傾向があった。これは、JIS のピクトグラムが紛らわしく、喫茶・

軽食施設を宿泊施設と勘違いした，あるいは宿泊施設を病院と勘違いしたことが原因であった．

今後は，JIS 改正委員会の動向 [83] に注意し，必要なピクトグラムはドラフト段階であっても採用していくことが望ましい．また，必要なピクトグラムが現状なく，JIS の改正案にも含まれないことが分かった場合は，ピクトグラムを新たに作成の上，JIS 改正委員会に提案していくことも重要である．

6.6 第6章のまとめ

本章では，言語による情報伝達には限界があるという立場に立った上で，言語によらない手段のみを利用した観光ポータルサイトを実現した．本ポータルサイトは，真に言語バリアフリーな観光ポータルサイトである．ポータルサイトの可用性を被験者実験によって評価し，言語によらない手段であっても，観光スポットの検索や経路案内が可能なことを実証した．また，観光スポット間の類似度の算出については，本論文で提案したパターンマッチングの手法を応用することで，自動化できることを示した．しかしながら，一部のピクトグラムが紛らわしく，可用性に欠けることがあった．今後は，JIS 改正委員会の動向に注意しつつ，必要不可欠なピクトグラムを委員会に提案していくことが重要である．

次章では，本論文のまとめと今後の研究課題について述べる．

表 6.1. 回答結果

被験者番号	国籍	指示 1	指示 2	指示 3
被験者 1	中国	○	○	○
被験者 2	日本	○	○	○
被験者 3	日本	○	○	○
被験者 4	中国	○	○	○
被験者 5	中国	○	○	○
被験者 6	イラン	○	○	○
被験者 7	日本	○	○	○
被験者 8	日本	○	○	○
被験者 9	日本	○	○	○
被験者 10	日本	○	○	○
被験者 11	日本	○	○	○
被験者 12	日本	○	○	○
被験者 13	日本	○	○	○
被験者 14	日本	○	○	○
被験者 15	日本	○	○	○
被験者 16	日本	○	○	○
被験者 17	日本	○	○	○
被験者 18	中国	○	○	○
被験者 19	中国	○	○	○
被験者 20	ネパール	○	○	○
被験者 21	ドミニカ	○	○	○
被験者 22	中国	○	○	○
被験者 23	中国	○	○	○

表 6.2. 画面遷移数

被験者番号	国籍	指示 1	指示 2	指示 3
被験者 1	中国	2	1	7
被験者 2	日本	2	1	8
被験者 3	日本	1	1	2
被験者 4	中国	2	2	3
被験者 5	中国	1	1	3
被験者 6	イラン	7	3	10
被験者 7	日本	1	1	4
被験者 8	日本	2	2	3
被験者 9	日本	2	1	5
被験者 10	日本	3	1	4
被験者 11	日本	1	1	7
被験者 12	日本	2	2	3
被験者 13	日本	2	1	6
被験者 14	日本	1	1	2
被験者 15	日本	3	2	2
被験者 16	日本	2	2	8
被験者 17	日本	2	1	7
被験者 18	中国	1	1	3
被験者 19	中国	1	1	4
被験者 20	ネパール	1	1	2
被験者 21	ドミニカ	1	1	2
被験者 22	中国	1	1	2
被験者 23	中国	2	1	1

第 7 章

結論

7.1 本論文のまとめ

本論文は，保守性，検索精度，迅速性，可用性に優れたポータルサイトの開発に関するものであった。

近年，地方自治体の経営資源の制約から，公共サービスを含む様々な分野のサービスを，近隣の地域間で分担補完するための連携が進められている。ゆえに，地域の情報を発信するポータルサイトにおいても，自治体の枠にとらわれない横断的なコンテンツを提供すべきである。しかしながら，このようなポータルサイトは現状なく，その実現が急務であった。そこで，地方自治体間の連携を推進することを主な目的とした横断的なコンテンツを提供するポータルサイトを開発した。

提案サイトは，Web クローラと情報フィルタリングを基にして構成される。Web クローラは，各地域情報が掲載された Web サイトを巡回し，ポータルサイトの情報を更新するプログラムである。また，Web クローラは，内部にフィルタリング機能を有している。フィルタリング機能は，Web ページの URL と，出現する単語によって，必要な Web ページのみ

を選別する機能である。これらの機能によって、特定の分野に関する情報のみを集約した、地域横断的なポータルサイトを実現した。

さらに、地域の産業と観光業を対象とした社会実装と評価実験を行い、提案サイトの保守性、検索精度、迅速性、可用性が、既存サイトと比べて優れていることを定量的に示した。

以上の研究成果により、地方自治体間の連携の推進に寄与した。

7.2 今後の研究課題

今後、地方自治体間の連携をさらに推進していくためには、下記の研究課題に取り組む必要がある。

1つ目は、検索結果の順位付けである。本論文では、適合率と再現率を用いて検索精度を評価したが、検索結果の順位付けや評価は行っていなかった。より実用的なポータルサイトにするためには、検索結果の順位も考慮していく必要がある。この課題に対して、考えられる工夫は次のとおりである。産業や観光業に関する文書では、その他の文書に比べて固有名詞が多く出現し、かつ固有名詞が文書の特徴をよりよく表していることが分かっている。そこで、固有名詞の索引語に重みを付け、スコアの高い文書から順に表示すれば、より実用的なポータルサイトになる。

2つ目は、検索精度の評価方法の再検討である。本論文では、検索にヒットした全ての文書を対象に、検索精度を評価した。ところが、実際の検索システムでは、上位数十件程度の文書まではユーザがアクセスするが、それ以降の文書はほとんどアクセスしないことが分かっている。そこで、上位 k 個の適合率を用いて評価することで、より実用的な側面から評価できる。

3つ目は、社会実装と評価実験の拡大である。本論文では、広域首都圏

の公設試験研究機関向けポータルサイトと、都内の観光スポットを対象とした観光ポータルサイトの2つについて、社会実装および評価実験を行った。これを、日本全国、さらには海外へと展開することで、よりグローバルな観点から評価を行うことができる。特に、観光ポータルサイトについては、言語バリアフリーであることが大きな強みであり、海外へ展開することで一層強みを生かすことができる。

参考文献

- [1] 田村信之, 地域ポータルサイトについて, 情報の科学と技術, Vol.55, No.2, pp76-80, 2005.
- [2] 藤田昌弘, 官民連携地域ポータルサイトに求められる要件と信用の源泉, 大手前大学論集, Vol.10, pp.197-211, 2009.
- [3] 西沢明, 高橋秀知, つくばポータルサイト, 情報の科学と技術, Vol.55, No.2, pp.81-85, 2005.
- [4] 上野亮, 平本一雄, 自治体ポータルサイトの全国動向と構築事例 都市運営のための情報プラットフォームの動向と事例研究 その1, 日本建築学会学術講演梗概集, pp.511-512, 2008.
- [5] 森下知行, 平本一雄, 上野亮, 自治体ポータルサイトの全国動向と構築事例 都市運営のための情報プラットフォームの動向と事例研究 その2, 日本建築学会学術講演梗概集, pp.513-514, 2008.
- [6] 上野亮, 平本一雄, 地域において総合情報サービスを行うポータルサイトのあり方に関する研究—八王子市を事例とした考察—, 日本社会情報学会学会誌, Vol.22, No.1, pp.17-30, 2010.
- [7] Brin, S., Page, L., The anatomy of a large-scale hypertextual Web search engine, *Proceedings of the 7th International WWW Conference (WWW-98)*, pp.107-117, 1998.

- [8] 原田昌紀, WWW サーチエンジンの作り方, 情報処理, Vol.41, No.11, pp.1280–1283, 2000.
- [9] 森田昌宏, 速水治夫, 情報フィルタリングシステム—情報洪水への処方箋, 情報処理, Vol.37, No.8, pp.751–758, 1996.
- [10] 獅々堀正幹, 藤井誠, 安藤一秋, 青江順一, 多属性項目の履歴情報に基づく電子メール文書のフィルタリング手法, 情報処理学会論文誌, Vol.41, No.8, pp.2299–2308, 2000.
- [11] Foltz, P. W., Dumais, S. T., Personalized information delivery: An analysis of information filtering methods, *Communications of the ACM*, Vol.35, No.12, pp.51–60, 1992.
- [12] 井ノ上直己, 帆足啓一郎, 橋本和夫, 文書自動分類手法を用いた有害情報フィルタリングソフトの開発, 電子情報通信学会論文誌, Vol.J84-D-II, No.6, pp.1158–1166, 2001.
- [13] 中村健二, 田中成典, 山本雄平, 安彦智史, 共起関係の抽出範囲を考慮した有害情報フィルタリング手法, 情報処理学会論文誌, Vol.54, No.2, pp.571–584, 2013.
- [14] Huffman, S., Acquaintance: Language-independent document categorization by N-grams, *Proceedings of the Fourth Text REtrieval Conference (TREC-4)*, pp.359–371, 1995.
- [15] Lewis, D. D., Schapire, R. E., Callan, J. P., Papka, R., Training algorithms for linear text classifiers, *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-96)*, pp.298–306, 1996.
- [16] Li, H., Yamanishi, K., Document classification using a finite mix-

-
- ture model, *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL-EACL '97)*, pp.39-47, 1997.
- [17] McCallum, A., Nigam, K., A comparison of event models for naive Bayes text classification, *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization*, pp.41–48, 1998.
- [18] Cohen, W. W., Text categorization and relational learning, *Proceedings of the International Machine Learning Conference (ICML-95)*, pp.124–132, 1995.
- [19] Joachims, T., Text categorization with support vector machines: Learning with many relevant features, *Proceedings of the 10th European Conference on Machine Learning (ECML-98)*, pp.137–142, 1998.
- [20] D’Orazio, V., Landis, S. T., Palmer, G., Schrod, P., Separating the Wheat from the Chaff: Applications of Automated Document Classification Using Support Vector Machines, *Political Analysis*, Vol.22, No.2, pp.224–242, 2014.
- [21] 関根聡, テキストからの情報抽出, 情報処理, Vol.40, No.4, pp.370–373, 1999.
- [22] 内元清貴, 馬青, 村田真樹, 小作浩美, 井佐原均, 最大エントロピーモデルと書き換え規則に基づく固有表現抽出とその評価, 言語処理学会第6回年次大会ワークショップ論文集, pp.1–8, 2000.
- [23] 佐藤円, 佐藤理史, 篠田陽一, 電子ニュースのダイジェスト自動生成, 情報処理学会論文誌, Vol.36, No.10, pp.2371–2379, 1995.

- [24] 長谷川隆明, 高木伸一郎, 文書構造の認識と言語の特徴の利用に基づく電子メールからのスケジュールと ToDo の抽出, 情報処理学会論文誌, Vol.40, No.10, pp.3694–3705, 1999.
- [25] 山田渉, 菊地悠, 落合桂一, 鳥居大祐, 稲村浩, 太田賢, マイクロプログラムを用いたイベント情報抽出技術, 情報処理学会論文誌, Vol.57, No.1, pp.123–132, 2016.
- [26] 梶井文人, 福本淳一, 製品情報一覧の自動提示のための情報抽出, 言語処理学会第 6 回年次大会ワークショップ論文集, pp.56–63, 2000.
- [27] 藤井敦, 言語横断検索への入門 —翻訳と検索の統合が生み出す新たな可能性—, 情報処理, Vol.42, No.3, pp.327–329, 2001.
- [28] 奥村明俊, 石川開, 佐藤研治, コンパラブルコーパスと対訳辞書による日英クロス言語検索, 自然言語処理, Vol.5, No.4, pp.77–93, 1998.
- [29] 藤井敦, 石川徹也, 質問翻訳と文書翻訳を統合した日英言語横断情報検索, 電子情報通信学会論文誌, Vol.J84-D-II, No.2, pp.362–369, 2001.
- [30] Rehder, B., Littman, M. L., Dumais, S., Landauer, T. K., Automatic 3-language cross-language information retrieval with latent semantic indexing, *Proceedings of the Sixth Text REtrieval Conference (TREC-6)*, pp.233–239, 1997.
- [31] Agrawal, A., Agrawal, A. J., Designing Cross-Language Information Retrieval System using various Techniques of Query Expansion and Indexing for Improved Performance, *International Research Journal of Engineering and Technology*, Vol.3, No.2, pp.1505–1508, 2016.

-
- [32] 串間和彦, 赤間浩樹, 紺谷精一, 山室雅司, 色や形状等の表層的特徴量にもとづく画像内容検索技術, 情報処理学会論文誌: データベース, Vol.40, No.SIG3(TOD1), pp.171–184, 1999.
- [33] 栗山繁, イラスト画像のスタイル識別子生成, 情報処理学会論文誌, Vol.56, No.8, pp.1657–1666, 2015.
- [34] Ardizzone, E., Cascia, M. L., Automatic video database indexing and retrieval, *Multimedia Tools and Applications*, Vol.4, No.1, pp.242–285, 1997.
- [35] Foote, J., An overview of audio information retrieval, *Multimedia systems*, Vol.7, No.1, pp.2–10, 1999.
- [36] 権田金治, 森川晴成, 公設試験研究機関の現状と課題: 過去7年間の推移の観点から, 研究・技術計画学会第14回年次学術大会講演要旨集, pp.460–465, 1999.
- [37] 若生彦治, 奥村皓一, 工業系公設試験研究機関の経営課題についての考察, 研究・技術計画学会第18回年次学術大会講演要旨集, pp.686–689, 2003.
- [38] 佐脇政孝, 公設試験研究機関の運営の現状と課題, 研究・技術計画学会第19回年次学術大会講演要旨集, pp.517–520, 2004.
- [39] 並河良一, 産業構造の変化に伴う地域公設試験研究機関の変遷: 農村工業系公設試を事例として, 研究技術計画, Vol.20, No.3, pp.226–238, 2005.
- [40] 福川信也, 地域イノベーションシステムにおける公設試験研究機関の位置づけと戦略, 研究・技術計画学会第22回年次学術大会講演要旨集, pp.645–648, 2007.
- [41] Shapira, P., Modernizing small manufacturers in Japan: The role

- of local public technology centers, *Journal of Technology Transfer*, Vol.17, No.1, pp.40–57, 1992.
- [42] 中小企業庁，公設試経営の基本戦略～中小企業の技術的支援における公設試のあり方に関する研究会中間報告～，http://www.chusho.meti.go.jp/keiei/gijut/2005/download/051220kousetushi_senryaku_houkokusho.pdf, accessed 2010-06-03.
- [43] 全日本地域研究交流協会，地域の産学官連携への公設試の効果的な取組みに関する調査研究―地域イノベーションの加速を目指して―，<http://www.jarec.or.jp/pdf/cyosa/18-chiikino.pdf>, accessed 2010-06-15.
- [44] 大槻眞一，谷口邦彦，小川裕策，下田正憲，柁村哲男，外部機関との連携による公設試験研究機関の技術相談・技術支援業務の拡充，研究・技術計画学会第15回年次学術大会講演要旨集，pp.247–250，2000.
- [45] 谷口邦彦，公設試験研究機関の役割：技術集積機関と中堅・中小企業との橋渡し，研究技術計画，Vol.15，No.3，pp.162–167，2003.
- [46] 地方独立行政法人青森県産業技術センター，地方独立行政法人岩手県工業技術センター，宮城県産業技術総合センター，秋田県産業技術センター，秋田県総合食品研究センター，山形県工業技術センター，福島県ハイテクプラザ，東北6県公開設備・機器データベース，<http://www.nita-tohoku-6ptc.jp/r2db/portal.php>, accessed 2011-12-15.
- [47] 経済産業省東北経済産業局，地方独立行政法人青森県産業技術センター，地方独立行政法人岩手県工業技術センター，宮城県産業技術

- 総合センター，秋田県産業技術センター，秋田県総合食品研究センター，山形県工業技術センター，福島県ハイテクプラザ，新潟県工業技術総合研究所，うおっち，http://www.tohoku.meti.go.jp/s_sangi/setsuken/setsuken.html，accessed 2015-07-30.
- [48] 関西広域連合，大阪市立工業研究所，大阪府立産業技術総合研究所，京都市産業技術研究所，京都府織物・機械金属振興センター，京都府中小企業技術センター，滋賀県工業技術総合センター，滋賀県東北部工業技術センター，徳島県立工業技術センター，鳥取県産業技術センター，兵庫県立工業技術センター，和歌山県工業技術センター，関西ラボねっと，<http://kansai-kousetsushi.jp/summary-eachengine.html>，accessed 2016-08-02.
- [49] 産業技術総合研究所，鳥取県産業技術センター，島根県産業技術センター，岡山県工業技術センター，広島県立総合技術研究所，広島市工業技術センター，山口県産業技術センター，中国地域公設研究機関の開放機器，<http://unit.aist.go.jp/chugoku/kaihou-kiki/index.html>，accessed 2016-08-02.
- [50] 徳島県立工業技術センター，香川県産業技術センター，愛媛県産業技術研究所，高知県工業技術センター，四国4県連携新製品開発支援協議会，<http://www.shikoku-itc.jp/index.php>，accessed 2016-08-02.
- [51] 福岡県工業技術センター，佐賀県工業技術センター，佐賀県窯業技術センター，長崎県工業技術センター，長崎県窯業技術センター，熊本県産業技術センター，大分県産業科学技術センター，宮崎県工業技術センター，鹿児島県工業技術センター，沖縄県工業技術センター，九州・沖縄バーチャル公設試，<http://kyushu.kmt-iri.go>.

- jp/index.html, accessed 2012-01-16.
- [52] 高林哲, Namazu: 全文検索で文書の山に立ち向かう, 情報処理, Vol.41, No.11, pp.1227–1232, 2000.
- [53] 阿部真也, 北原枢, 五十嵐美穂子, 山田一徳, 近藤幹也, 吉野学, 片岡正俊, 設備データベースと新着情報機能を有する公設試広域連携 Web サイトの開発, 情報処理学会論文誌: データベース, Vol.6, No.4, pp.58–69, 2013.
- [54] 阿部真也, 北原枢, 五十嵐美穂子, 山田一徳, 近藤幹也, 吉野学, 片岡正俊, 設備データベースと新着情報機能を有する公設試広域連携 Web サイトの開発, 第 5 回 Web とデータベースに関するフォーラム, 2012.
- [55] 阿部真也, ポータルサイト向け情報検索エンジン, 第 5 回 Web インテリジェンスとインタラクション研究会, 2014.
- [56] 阿部真也, 公設試験研究機関向けポータルサイトの開発, 平成 26 年度電気・電子・情報関係学会東海支部連合大会, 2014.
- [57] Manning, C. D., Raghavan, P., Schütze, H. (著), 岩野和生, 黒川利明, 濱田誠司, 村上明子 (訳), 情報検索の基礎, 共立出版, 2012.
- [58] 北研二, 津田和彦, 獅々堀正幹, 情報検索アルゴリズム, 共立出版, 2002.
- [59] 平林幹雄, 江渡浩一郎, N.M-gram: ハッシュ値付き N-gram 索引による全文検索の一手法, 情報処理学会論文誌: データベース, Vol.48, No.7, pp.29–37, 2007.
- [60] 平林幹雄, Hyper Estraier, <http://fallabs.com/hyperestraier/index.html>, accessed 2017-05-25.
- [61] 埼玉県産業技術総合センター, 千葉県産業支援技術研究所, 地方

- 独立行政法人東京都立産業技術研究センター，神奈川県産業技術センター，横浜市工業技術支援センター，首都圏テクノナレッジ・フリーウェイ，<http://tkm.iri-tokyo.jp/index.html>，accessed 2014-07-16.
- [62] 片岡正俊，首都圏の連携体「テクノナレッジ・フリーウェイ」，産学官連携ジャーナル，Vol.7，No.10，pp.6-7，2011.
- [63] 山田一徳，土屋敏夫，宮田勝雄，中小企業向けイントラネット環境の構築，東京都立産業技術研究所研究報告，No.2，pp.103-108，1999.
- [64] 山田一徳，土屋敏夫，Web による簡易データベースの構築，東京都立産業技術研究所研究報告，No.3，pp.147-148，2000.
- [65] 北原枢，土屋敏夫，バーチャル公設試の XML 化，東京都立産業技術研究所研究報告，No.7，pp.91-92，2004.
- [66] 阿部真也，富山真一，大平倫宏，公設試験研究機関向け技術検索エンジン，電子情報通信学会論文誌，Vol.J99-D，No.5，pp.549-558，2016.
- [67] 阿部真也，富山真一，大平倫宏，公設試験研究機関向け技術検索エンジン，情報アクセスシンポジウム 2014，2014.
- [68] 経済産業省関東経済産業局，関東地方産業競争力協議会，<http://www.kanto.meti.go.jp/seisaku/kyosoryoku/index.html>，accessed 2014-09-05.
- [69] 関東地方産業競争力協議会，関東地方産業競争力強化戦略－関東地方全体の「連携」で産業競争力を強化する－，<http://www.kanto.meti.go.jp/seisaku/kyosoryoku/data/20140331senryaku02.pdf>，accessed 2014-09-05.
- [70] 経済産業省関東経済産業局，公設試験研究機関が保有する試験研究機

- 器の検索システムを利用開始します, <http://www.kanto.meti.go.jp/annai/hodo/data/20141114openinnov.pdf>, accessed 2014-11-20.
- [71] 地方独立行政法人東京都立産業技術研究センター, 広域首都圏の公設試験研究機関が保有する試験研究機器の検索システムの利用を開始, <https://www.iri-tokyo.jp/uploaded/attachment/2869.pdf>, accessed 2014-11-20.
- [72] 経済産業省関東経済産業局, 平成 25 年度補正予算事業「地域オープンイノベーション促進事業」(設備機器の導入・利活用事業)(関東地域)の募集について, http://www.kanto.meti.go.jp/seisaku/gizyutsu/25fy_hosei_openinnov_setsubi_koubo.html, accessed 2014-06-10.
- [73] 地方独立行政法人東京都立産業技術研究センター, 平成 25 年度補正予算事業に採択「地域オープンイノベーション促進事業」(設備機器の導入・利活用事業)(関東地域), <https://www.iri-tokyo.jp/uploaded/attachment/2883.pdf>, accessed 2014-07-16.
- [74] 地方独立行政法人東京都立産業技術研究センター, 平成 25 年度経済産業省補正予算事業「地域オープンイノベーション促進事業」のご紹介, TIRI News, No.12, p.5, 2014.
- [75] Abe, S., Miki, D., Yamamoto, K., A Tourism Information System with Language-Barrier-Free Interfaces for Foreign Visitors, *Proceedings of International Conference on Marketing and Tourism (MAT2017)*, pp.58–66, 2017.
- [76] 阿部真也, 山本佳世子, 三木大輔, アリアバディ・ファラハニ・マルジエ, 情報検索方法, 情報検索プログラム, 情報検索用端末および情

-
- 報検索装置, 特願 2016-155123, 2016.
- [77] Abe, S., Miki, D., Aliabadi, F. M., Yamamoto, K., A Tourism Information System with Language-Barrier-Free Interfaces, Japan Academic Society of Tourism, pp.56-57, 2016.
- [78] 観光庁, 観光圏整備事業のノウハウに関する基礎資料, <http://www.mlit.go.jp/common/000161119.pdf>, accessed 2015-07-02.
- [79] 藤田礼子, 観光情報学 : 2. 観光政策における ICT の活用について, 情報処理, Vol.53, No.11, pp.1140-1145, 2012.
- [80] 観光庁, 観光立国実現に向けた多言語対応の改善・強化のためのガイドライン, <http://www.mlit.go.jp/common/001029742.pdf>, accessed 2015-08-04.
- [81] 観光庁, スマートフォンアプリ等の提供, <http://www.mlit.go.jp/kankocho/shisaku/kokusai/appli.html>, accessed 2016-10-01.
- [82] 日本工業規格, 案内用図記号, JIS Z 8210:2016, 2016.
- [83] 経済産業省, 2020 年東京オリ・パラに向けて案内用図記号の JIS 改正委員会を開催します, <http://www.meti.go.jp/press/2016/07/20160704001/20160704001.pdf>, accessed 2016-11-07.

関連研究の公表

学術雑誌審査付論文

- [1] 阿部真也，北原枢，五十嵐美穂子，山田一徳，近藤幹也，吉野学，片岡正俊，設備データベースと新着情報機能を有する公設試広域連携 Web サイトの開発，情報処理学会論文誌：データベース，Vol.6，No.4，pp.58–69，2013.
- [2] 阿部真也，富山真一，大平倫宏，公設試験研究機関向け技術検索エンジン，電子情報通信学会論文誌，Vol.J99-D，No.5，pp.549-558，2016.

国際会議審査付論文

- [3] Abe, S., Miki, D., Yamamoto, K., A Tourism Information System with Language-Barrier-Free Interfaces for Foreign Visitors, *Proceedings of International Conference on Marketing and Tourism (MAT2017)*, pp.58–66, 2017.

国内会議審査付論文

- [4] 阿部真也, 北原枢, 五十嵐美穂子, 山田一徳, 近藤幹也, 吉野学, 片岡正俊, 設備データベースと新着情報機能を有する公設試広域連携 Web サイトの開発, 第 5 回 Web とデータベースに関するフォーラム, 2012.
- [5] 阿部真也, 富山真一, 大平倫宏, 公設試験研究機関向け技術検索エンジン, 情報アクセスシンポジウム 2014, 2014.

口頭発表

- [6] 阿部真也, ポータルサイト向け情報検索エンジン, 第 5 回 Web インテリジェンスとインタラクション研究会, 2014.
- [7] 阿部真也, 公設試験研究機関向けポータルサイトの開発, 平成 26 年度電気・電子・情報関係学会東海支部連合大会, 2014.
- [8] Abe, S., Miki, D., Aliabadi, F. M., Yamamoto, K., A Tourism Information System with Language-Barrier-Free Interfaces, Japan Academic Society of Tourism, pp.56-57, 2016.

特許

- [9] 阿部真也, 山本佳世子, 三木大輔, アリアバディ・ファラハニ・マルジエ, 情報検索方法, 情報検索プログラム, 情報検索用端末および情報検索装置, 特願 2016-155123, 2016.

著者略歴

阿部 真也

岩手大学大学院工学研究科博士前期課程修了。電気通信大学大学院情報システム学研究科博士後期課程在学。2009年 地方独立行政法人東京都立産業技術研究センター研究員。2016年 同センター主任研究員。情報検索システム，観光情報システムに関する研究開発に従事。情報処理学会第70回全国大会学生奨励賞，情報処理学会創立50周年記念第72回全国大会大会奨励賞受賞。情報処理学会，電子情報通信学会，日本データベース学会，日本観光学会，観光情報学会各会員。