

博士論文

メタデータを活用したユーザ指向型行動支援情報 取得・収集手法に関する研究

新井 イスマイル

2008年3月31日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学) 授与の要件として提出した博士論文である.

新井 イスマイル

審査委員：

砂原 秀樹 教授	(主指導教員)
山口 英 教授	(副指導教員)
藤川 和利 准教授	(副指導教員)
下條 真司 教授	(大阪大学)

メタデータを活用したユーザ指向型行動支援情報 取得・収集手法に関する研究*

新井 イスマイル

内容梗概

無線 LAN や携帯電話などの無線通信技術の発展によりユーザはいつでもどこでも Web 情報検索ができるようになり、移動先での飲食店情報や天気予報など、自らの位置やスケジュールといった様々な状況を反映した行動支援情報を求めるようになった。本研究ではユーザが膨大な Web 文書から行動支援情報を省作業で的確に取得できることを目的とする。

現状の携帯情報端末は小さい画面に少数のボタンなどといったユーザインターフェイスの制約により、迅速な情報検索が困難である。また、現在地やスケジュールなどの状況情報はキーワードによる的確な記述が困難となる。さらに、今後の爆発的なコンテンツの増加に対して検索処理時間が安定しなければならない。

従来の検索手法では検索結果に的確さを要求するほど、検索作業が複雑となる問題がある。またユーザの要求の多様化およびコンテンツの増加は検索時の演算量を増加させることに繋がり、従来のサーバ集中型のアーキテクチャでは対応できない。そして、的確な検索のためのコンテンツへの意味付け作業は従来は主にコンテンツ作成者の手作業に委ねられていたが、これについてもユーザの要求の多様化、コンテンツの爆発的な増加によって限界が生じる。

本論文ではユーザの行動支援情報の要件を「実空間との対応がとられた情報、個々のユーザに適した情報を提供すること」、「様々な検索要求に対して瞬時に検索結果が得られること」、「ユーザのこれからとる行動を示唆するもの」に整理

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 博士論文, NAIST-IS-DD0461003, 2008 年 3 月 31 日.

した。そして、個々に該当する従来の研究を踏まえて取り組むべき課題を「ユーザのサービス利用における許容性」、「システムのリアルタイム性と柔軟性」、「正確なコンテンツメタデータの維持性」の3点に整理し、ショッピングモールでの商品推薦を行う実証実験および Web サイトからクローリングの実験を通してそれぞれの課題に取り組んだ。

ユーザの位置・現在時間といった状況をリアルタイムに読み取り、年齢や性別などのユーザプロフィールをあらかじめ登録しておき、情報取得端末にカードをかざすのみで要求されるクエリをすべて自動的にクエリ生成する手法、およびそのクエリを理解しメタデータマッチングを行う推薦アルゴリズムを提案し、検証した結果、ユーザにとって許容性のあるサービスを実現できたことを確認した。

システムのリアルタイム性と柔軟性の課題については、位置にもとづくオーバーレイネットワーク上に、コンテンツとメタデータマッチング機能をもつエージェントを配置することによって、複数属性の演算を負荷分散し、推薦アルゴリズムの変更が必要な場合はエージェントを組み替えるだけで良いアーキテクチャを提案した。リアルタイム性については実証実験の規模から証明するに至らなかったが貴重なコンテンツの実統計情報を得ることができ、今後の検証の指針を得ることができた。

正確なコンテンツメタデータの維持性の課題については、個人サイトからメタデータを抽出するモデルを提案した。特に行動支援情報を見出すために重要となる時間情報についてはブログ中に含まれる写真の Exif 情報に着目しそれを抽出する手法、位置情報については文書の形態素解析結果のパターンマッチングにより住所文字列を抽出する手法を提案し、評価した結果、有利な適合率によって目的とする情報の抽出が実現できた。

キーワード

コンテキストウェア、メタデータ、情報検索、エージェント、オーバーレイネットワーク

A study for managing behavioral support information for users utilizing metadata*

Ismail Arai

Abstract

Users can search information through the Internet anytime anywhere, by using highly developing wireless communication technologies. In this situation, users expect the system to provide behavioral support information which reflects users' preferences and circumstances. In this paper, we aim to realize that users can search their behavioral support information easy and accurately from the huge amount of Web pages.

Due to poor user interface, less buttons, and smaller monitor of mobile terminals, it's hard to search such information rapidly. And, it's difficult to input temporal and positional information correctly by manually inputting some keywords. Moreover, the calculation time of matching contents must be scalable against increasing contents.

The operation of information retrieval becomes complicated when the system has to supply accurate results to the users. Also, conventional architectures such as centralized servers cannot adapt for diversification of user's requirements and contents explosion, because the calculation cost would be higher. Furthermore, contents distributors become hard to add semantics to the contents manually because of the same reason as above.

In this thesis, I classify the requirements of users' behavioral support information as "supplying information that links real-space and adapts each users",

*Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0461003, March 31, 2008.

“supplying results of search immediately for various queries”, and “supplying information that suggest user’s future behavior”. Then, I arrange requirements “admissibility of users”, “scalability and flexibility of systems”, and “managing contents metadata automatically and continuously” as the results of the surveys on the related work. And I solved these requirements through some experiments as follows.

To be able to read the user’s position and time , and store the user’s profile (such as his/her age, gender and occupation), the proposed method enables users to obtain their behavioral support information by only putting RFID card on the information terminals. So, the proposed recommendation algorithm satisfies the users’ demand.

To solve problems of scalability and flexibility, I propose a agent which filters contents by comparing a user profile and contents metadata, on a location based overlay network. The overlay network enables users to search location information quickly. And the agent enables users to filter contents by not only a location but also other preferences. It also allows the system to replace recommendation algorithms without rebooting. Even though I couldn’t proof the scalability, I could obtain the real statistics of contents in the mall. The statistic would help a lot of developers to design a future overlay network.

In order to keep correct contents metadata automatically and continuously, I propose Geocrawler and Time-crawler. Geocrawler collects location information by pattern matching of morphological analysis outputs. Time-crawler collects temporal information by extracting shooting time of photos from Exif data of the photos on blogs. As the results of evaluation, I confirm that the proposed method is effective to provide correct metadata for providing behavioral support information for users.

Keywords:

Context-aware, Metadata, Information Retrieval, Agent, Overlay Network

目次

第1章 はじめに	1
第2章 関連研究の考察と本研究の課題	7
2.1. 実空間情報取得手法の分類・考察	7
2.1.1 全文検索	7
2.1.2 Web ディレクトリ	8
2.1.3 嗜好に基づく推薦	9
2.1.4 メタデータによる情報検索	10
2.1.5 コンテキストウェアサービス	11
2.1.6 情報検索手法まとめ	12
2.2. リアルタイム性を考慮した情報流通基盤の分類・考察	12
2.2.1 サーバ集中型	13
2.2.2 エージェント型	13
2.2.3 P2P 型	13
2.2.4 情報流通基盤まとめ	15
2.3. 行動支援情報の情報源の分類・考察	15
2.3.1 商用サイトと個人サイトの比較	15
2.3.2 投稿型ポータルサイトとブログサイトの比較	18
2.4. 本研究の目的・課題	18
第3章 行動支援情報提供を考慮したメタデータマッチング手法	22
3.1. ユーザの許容性を考慮したメタデータマッチング手法	22
3.2. 行動支援情報提供メタデータマッチングの提案	23
3.2.1 RFID カードを活用したゼロクリック推薦	23

3.2.2	ユーザプロファイルの設計	24
3.2.3	コンテンツメタデータの設計	25
3.2.4	コンテンツ推薦アルゴリズムの設計	27
3.3.	ショッピングモールでの実証実験による評価と考察	29
3.3.1	実験仕様	29
3.3.2	許容性の評価・考察	32
3.3.3	推薦の影響	34
3.3.4	コンテンツ提供者のメタデータ作成負荷に対する考察	37
3.4.	まとめと今後の課題	40
第 4 章	行動支援情報をリアルタイムに提供するシステムアーキテクチャ	41
4.1.	リアルタイム性と柔軟性を考慮したアーキテクチャ	41
4.2.	分散エージェントによるコンテンツ推薦システムの提案	42
4.2.1	位置に基づくオーバレイネットワークの有効性	42
4.2.2	柔軟性を考慮した, 分散エージェントによる機能構成	43
4.3.	分散エージェントによるコンテンツ推薦システムの設計	44
4.4.	ショッピングモールでの実証実験による評価と考察	46
4.4.1	柔軟性の考察	47
4.4.2	処理時間の考察	48
4.4.3	コンテンツ分布の考察	49
4.5.	まとめと今後の課題	51
第 5 章	個人サイトからのメタデータ抽出モデルの提案	52
5.1.	個人サイトに含まれる評価情報	52
5.2.	Web ページからのメタデータ抽出モデル	54
5.3.	Geocrawler の設計と評価	55
5.3.1	Web からの位置座標抽出の現状	56
5.3.2	形態素解析結果を活用した位置座標抽出	56
5.3.3	抽出精度の評価・考察	58
5.4.	Time-crawler の設計と評価	60

5.4.1 イベント生起時間の重要性	61
5.4.2 Exif に着目したイベント生起時間抽出	62
5.4.3 抽出精度の評価・考察	64
5.5. まとめと今後の課題	67
第 6 章 おわりに	69
謝辞	71
参考文献	73
研究業績	82

目 次

1.1	サービスイメージ	4
2.1	“What’s Open” での検索例	16
2.2	“SHOOTI” での検索例	17
2.3	行動支援情報提供モデル	19
3.1	ユーザプロファイルの記述例	24
3.2	cp:keyword から対象ユーザ特性を連想するリスト	25
3.3	コンテンツメタデータの記述例	26
3.4	コンテンツ推薦アルゴリズム	28
3.5	コンテンツ推薦システムの画面イメージ	30
3.6	ユーザの性別・年齢別分布	32
3.7	「このサービスが普及したら使いたいですか？」に対する回答 (年 齢別)	33
3.8	「端末は何が使いやすいですか？」に対する回答 (年齢別)	33
3.9	「おすすめ情報は嗜好にあったか」に対する回答 (男性・女性)	35
3.10	コンテンツ閲覧時間のヒストグラム	36
3.11	コンテンツ数の推移と更新頻度	38
3.12	コンテンツ提供者に対するアンケート結果	39
4.1	PIAX のアーキテクチャ	43
4.2	コンテンツ推薦サービスの概要	45
4.3	コンテンツ推薦アルゴリズム	46
4.4	平均処理時間およびコンテンツアクセス数	47
4.5	処理時間についてのアンケート結果	48

4.6	各ピアが保持するコンテンツ割合の推移	49
4.7	各ピアが保持するコンテンツ量の分散の推移	50
5.1	メタデータ抽出モデル	54
5.2	ブログの例	62

表 目 次

3.1	ショッピングモールでの実証実験の詳細	31
3.2	コンテンツの閲覧状況	35
5.1	収集された Web サイトの記述内容	53
5.2	茶笥を使った解析例	57
5.3	Exif 情報の例	63
5.4	ブログ中に掲載される写真の内訳	64
5.5	写真と本文の内容の関連	64
5.6	グルメキーワード毎の適合率	66

第1章 はじめに

Webが登場して以来、ユーザは多くの情報取得をインターネットを通じて行うようになった。初期は既知のURLに直接アクセスした後にHTML文書内のハイパーリンクを辿る閲覧方法が主であったが、世界中に膨大な量のWebコンテンツが作成されるようになってからは、これらコンテンツをインデックス化し検索する手法が主流となった。Web文書の形態素解析により単語を分割・抽出し、自動的にインデックスを作成することが可能となった。数10億ものWeb文書群から瞬時にキーワードによるパターンマッチングが行えることにより、ユーザの知識獲得ツールの1つとして多くの広いユーザ層に利用されている。

一方、モバイル通信技術などの発達による携帯端末の普及に伴い、ユーザがいつでもどこでもインターネットに接続できる環境が整いつつある。このような環境を本論文ではユビキタス環境と呼ぶ。ユビキタス環境下では、従来の全文検索による文書の検索の場合とは異なり、計算機がユーザの状況を読み取り、それにあったサービスを自動的に提供するサービスに期待がかかる。膨大なWeb情報からユーザの求める情報（状況・嗜好等にあった情報）を的確・迅速・省作業に提供できることが望ましい。ユーザの移動中もしくは外出中の状態を考慮すると提示すべき情報は特に実空間上のものと一致するものが多くなる。

この要求について、特にユーザの状況を反映するコンテキストウェアサービスが積極的に議論されている。コンテキストとは人や物の状況を特徴づけるのに使用できるすべての情報である。そして、このコンテキストを入力として、その変化に応じた結果を出力するサービスがコンテキストウェアサービスである。初期にこのサービスを提案したWeiser[1]によると、“door open only to the right badge wearer”といったように、近づく人が誰かということに依存してドアの開閉サービスが異なって提供される。その後、化学物質コンテナ[2]や飛行機の整備

用の道具 [3] に RFID (RadioFrequency IDentification) タグを取り付け, 常に安全なあるいは正しい位置に物がない状態になった場合にアラートを発するサービス, 友人が近づいたらアラートを発するサービス [4], 子供が離れたらアラートを発するサービス [5] など RFID による存在検出手法を応用したサービスの提案が目立った. 活用するセンサの多様化につれて, 看護師が患者の前に薬を持ってきたときにそれが医師の指示通りのものであれば患者の前のディスプレイに薬の詳細な情報が提示されるサービス [6] や, GPS (Globla Positioning System) センサを搭載した端末を所持したユーザが重要な位置に訪れたときに周辺の店舗情報等をプッシュ配信するサービス [7] が提案されるなど, コンテキストの種類の増加, 状態推定アルゴリズムの高度化によってコンテキストの次元が高められていった.

このような従来のコンテキストウェアサービスで取り扱うサービスのほとんどは, ユーザの行動を補助するものとなっている. あらかじめユーザが得たいサービスを選択・登録しておき, しかるべき状態に計算機環境が自動的にユーザに知らせるといったように“気付き”の情報を提供するものがほとんどである. これはユーザにとっては“既知”の情報である. しかし, 背景にも挙げたように, ユーザが Web から情報取得する場合には, 多くのものがユーザにとって“未知”の情報である. ユーザにはレストラン, 買い物情報や天気予報といったユーザの行動を支援する情報を外出中に知りたいという要望がある. しかし, 既存のセンサ等によって得られるコンテキストはユーザの要望を的確に捉えるには未だに不十分である. ユーザの現在地周辺の店舗情報が得られたとしても, その検索時刻に閉店している店舗情報であっては, そのユーザにとっての価値は低い. また, 位置, 時間といった状況に合った情報が提示されたとしても, ユーザの嫌いな食べ物が推薦された場合には, 同じく価値が低い. 本論文ではユビキタス環境において, このようなコンテキストウェアサービスにおいて提供し得るユーザにとって未知な, ユーザの行動を示唆する, ユーザの“状況”や“嗜好”にあった情報を行動支援情報と定義し, その取り扱いについて議論する.

行動支援情報提供サービスとして, 多数の店舗や施設があるショッピングモールでの商品情報取得を例に挙げると, ユーザが「現在」:(時間), 「飲食店売り場周辺」:(位置) で, 「買い物」:(目的) をしていた場合, 「和食の総菜」:(ユーザ

の嗜好)の、「特売」(限られた期間で販売したい:コンテンツ提供者の要望)の情報を取得するといったように、ユーザの状況や嗜好を考慮した情報取得の実現が望まれる。従来の全文検索では検索結果の正確さは、検索キーワード記述の的確さに依存していたが、上記の要求については時間や位置など、もはやキーワードによって表現できない情報が含まれている。

上記の括弧内に表されるような、ユーザの情報検索におけるクエリの元となる要素が個別に機械処理可能となればユーザの行動支援情報検索は実現性が増す。これにはユーザの現在地や現在時間といったユーザの状況、そのユーザはどのような特徴をもつ人物か、そのユーザはどのような情報を求めているかといったことを、ユーザが情報取得の度に計算機に手入力で伝えるのではなく、計算機が自動的に読み取ることが望ましい。

また、コンテンツにもそのユーザの要求に合致しているか否かを判断するための構造化された情報が必要になる。文書の意味を示すには、“データのデータ”であるメタデータを記述する手法がある。キーワードによるパターンマッチングで実現できなかった、文書の意味を解釈した的確な情報検索を行うにはこのメタデータを十分に用意する事が重要となる。

さらに、ユーザは行動支援情報を外出中にその場で求めることが予想され、リアルタイム性が要求される。今後コンテンツが爆発的に増加すると、上記のメタデータによる高度な演算処理時間はリアルタイム性の要求に対して障害となる可能性がある。したがって、行動支援情報提供基盤はコンテンツ増加に対して演算負荷の分散を考慮する必要がある。

本論文は以上の問題に取り組み、関連研究の考察から得られた課題を解決する手法を提案し、実験を通して得られた結果・知見について述べる。

ユーザが行動する時に、多くのユーザは計画をもって行動をとる。出張や旅行時にはこれから訪れる情報を収集する(図1.1)。このような場合には、事前に観光本や観光情報サイトを調査することによってスケジュール調整ができ、ユーザはそのスケジュールを記憶もしくはメモに残し計画された行動をとる。しかし、調査が十分でなかったときに現地で情報を収集することは現地の人に尋ねるもしくは知人に教えてもらう程度の手段しかなく困難となる。本研究ではこのような、

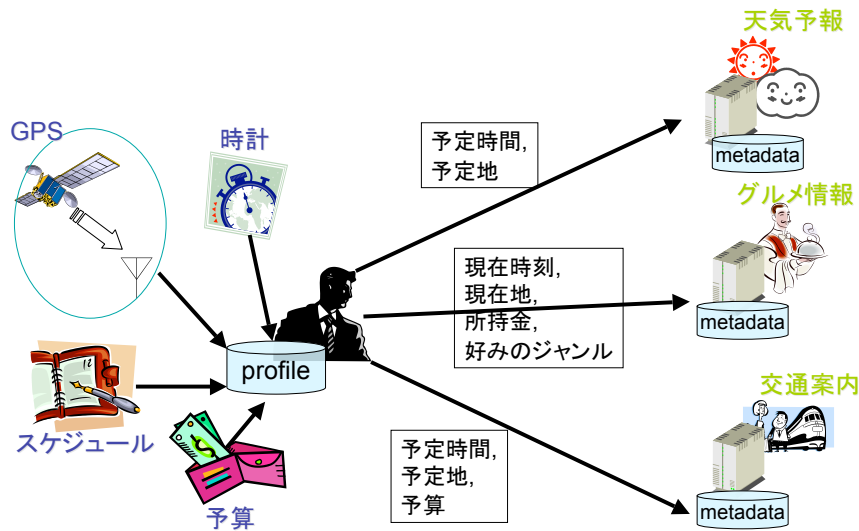


図 1.1 サービスイメージ

ユーザがおかれている時刻, 地点といったその場にあった情報を行動支援情報として取り扱う. 行動支援情報をユーザに提供するにあたっての要件を問題点と課題とともに述べる.

- 実空間との対応がとられた情報, 個々のユーザに適した情報を提供すること
全文検索等の従来の情報検索手法では携帯情報端末にキーワードを手入力するなどの手間がかかり, 検索労力が問題となる. 移動中のユーザにとって情報の検索労力は最小限であることが望ましい. 個々のユーザの嗜好や属性にあった情報を自動的に絞り込むことができれば, ユーザの検索労力は十分に少なくなると期待できる. また, 提供する情報はユーザのその場の状況にあったもののみに絞られることが望ましい. Web 上にある膨大な情報のほとんどは URL やリンク情報しか意味を持たないため, これらを位置・時間といった実空間との対応でインデックス化し, 検索範囲をそのインデックスに基づいて定める必要がある. 本論文では従来の検索手法として, 全文検索, 嗜好に基づく検索, メタデータによる検索, コンテキストウェアサービスを取り上げ, それぞれの特徴を整理した上で, コンテキストウェアの

状況対応性の長所およびメタデータ検索の属性・嗜好対応性の長所を兼ね備えたメタデータマッチング手法を提案し、その有効性を検証する。

- 様々な検索要求に対して瞬時に検索結果が得られること

ユーザは徒歩中など何らかの行動中に行動支援情報を求める。そのため、検索結果が要求を出してからいつまでたっても返事が来ないことは利便性を大きく損なう。ユーザの行動に対する割り込みは極力短いことが望まれる。近年の爆発的なコンテンツ増加によってサーバへのアクセスは増加し、サーバにおける検索処理の応答速度が著しく低下する恐れがあるため、これらに対処した情報流通基盤について検討する必要がある。本論文では、従来の情報流基盤としてサーバ集中型、エージェント型、P2P 型を取り上げ、それぞれの特徴について整理する。規模性を考慮して P2P 型を採用した上で、様々な検索要求に柔軟に対応するエージェントを P2P ノード上に配置する手法を新たに採用する。そしてショッピングモールでの実証実験を通して得られた知見について述べる。

- ユーザのこれからとる行動を示唆するもの

ユーザがこれから訪れる場所、購入するもの、飲食するものを決定するために、それらの説明記述をユーザに提供する必要がある。商業者が発信する情報は物やサービスの売り込みであり、個人が発信する情報には共感を促す体験談などがあり、ユーザはその情報を元に行動を決定する。本論文ではこの個人サイトと商用サイトの特徴について整理し、ユーザにとって親近感のある口コミ要素が強い個人サイトを情報源の対象とする。個人サイトは的確な検索を行うためのコンテンツの構造化が不十分なため、無秩序に記述された Web ページからコンテンツの意味を記述するデータであるメタデータを加工・抽出する方法を提案する。特に行動支援情報検索で重要となる位置・時間情報についてそれぞれ形態素解析結果のパターンマッチングおよび Exif データを活用したメタデータ抽出手法を提案し、適合率を評価した結果について述べる。

本章に続き、第 2 章では本研究に関連する既存の情報検索手法について考察し、

本研究で取り組むべき課題を、「ユーザのサービス利用における許容性」、「システムのリアルタイム性と柔軟性」、「正確なコンテンツメタデータの維持性」の 3 点に設定する。第 3 章では、主に「ユーザのサービス利用における許容性」の課題について取り組み、ユーザが外出中に情報検索する際に許容できる限られた操作量からの的確な行動支援情報をいかに推薦するかについて議論し、ショッピングモールでの実証実験の結果を踏まえて考察する。第 4 章では、主に「システムのリアルタイム性と柔軟性」の課題について取り組み、位置に基づいたオーバレイネットワークによる位置検索のリアルタイム性を確保するアーキテクチャ、そのオーバレイネットワーク上にエージェントを配置することによって柔軟性を確保するコンテンツ推薦システムを提案し、同上の実証実験を通して得られた結果および考察について述べる。第 5 章では「正確なコンテンツメタデータの維持性」の課題に取り組み、個人サイトからコンテキストウェアサービスの実現に必要なメタデータを自動抽出するモデルを提案する。そのモデルの性能の要となるメタデータ生成クローラにおいて、主に時間と位置について取り組む。ブログ上の写真に埋め込まれた Exif 情報に着目した時間情報の抽出性能、および文書の形態素解析結果のパターンマッチングにより住所文字列を見出すことによる位置情報の抽出性能について評価実験を行い、得られた結果と考察を述べる。最後に第 6 章にてまとめる。

第2章 関連研究の考察と本研究の課題

本研究の目指す行動支援を実現する上で、情報検索・情報推薦という観点において関連する情報取得手法、情報流通基盤の現状を考察し、本研究の課題をまとめる。

2.1. 実空間情報取得手法の分類・考察

第1章で述べた行動支援の第1要件であった実空間との対応がとられた電子情報を提供するにはどのような情報取得手法が望ましいかを見出すために、既存の情報取得手法を分類し考察する。

2.1.1 全文検索

Web上の文書を検索する手法として、現在最もユーザに利用されているのが全文検索である。ユーザに対するサービス利用の事前登録作業が不要なく、検索結果が即時に返ってくるため、机上のコンピュータで利用することに適している。Web文書をあらかじめ茶筌[8]やMeCab[9]などの形態素解析によって分かち書き、もしくはn-gramによって分割し、単語に基づくインデックスを作成し、ユーザが検索時に入力したキーワードと、このインデックスに存在するキーが一致した場合に検索結果がユーザに返される。したがって、ユーザが要求する情報を取得できるか否かは入力するキーワードがインデックスに存在することと、文書中のキーワードがユーザの意図する内容と一致する必要がある。キーワードのみが一致し内容がユーザの要求に合わない問題がしばしば起こる。n-gram抽出結果の学習

によって単語の途中の記述から完全な単語を連想する手法を用いることによってキーワードの記述ミスを修正する方法はあるが、同音異義語のような意図のミスマッチについては防ぎきれない。

また、より価値のある Web ページを優先的に検索結果として提示できることが検索サービスに求められる。何らかの基準によって検索結果を並び替えることを“**スコアリング**”と呼ぶ。現在最も多用されている全文検索エンジン Google[10]では、このスコアリング手法は PageRank[11] が基本となっている。PageRank では、ランク値によってスコアリングする。ある Web ページが他の Web ページに参照されればされるほど高スコアとなる。さらに参照元のページのランク値を重み付けてランク値を計算することによって、信頼度の高い Web ページから多く参照されるほど、その Web ページのランクが上がる。荒谷ら [12] による研究では、この PageRank によるスコアリング結果の高ランク値のみに着目し、その類似ページを中心に高いスコアを動的に付与する手法が提案され、よりユーザの意図にあった全文検索を実現している。また、遺伝的プログラミングを用いて、ユーザが入力するキーワードによってスコアリングの挙動を動的に変更することによってユーザの状況を考慮した検索結果を提示する手法 [13] がある。

いずれにおいても全文検索における大きな問題はキーワードによるクエリのみしか対応していないことにある。時間や場所などユーザがキーワードでどのように表現すればよいか分からない場合がある。また、基本的にランク値によってソートがなされ、1つの Web ページに対して持つ属性値が他に更新時間くらいしかないことから、任意の視点によるデータの並び替えが難しい。

2.1.2 Web ディレクトリ

実際に人間の編集者が Web ページを手動で確認して登録された情報の集まり (リンク集) で、リンクを登録するにはあらかじめ分類されたカテゴリ項目の中から最もふさわしいカテゴリを選択する。代表的なものに、Yahoo![14]、Excite[15]、Netscape[16] がある。このリンク集に対して全文検索することによって検索者は目的の Web ページのリンクを得る。リンク集が編集者によって登録、管理されているため無駄な情報が少ない。また編集者が最適と判断したカテゴリに登録する

ため、興味がある分野に関するウェブサイトのカテゴリを通して多数を簡単に発見できる。しかし Web コンテンツ数が爆発的に増加している中で、人を介して情報を登録することは検索対象の絶対数が全文検索に比べ少なくなり非現実的である。また検索対象が登録されたタイトルや概要文程度の限られた情報であるため、Web 文書内に目的の情報が存在していた場合にも、検索結果として得られない場合がある。

2.1.3 嗜好に基づく推薦

嗜好に基づく推薦の代表的なものは協調フィルタリング [17] である。協調フィルタリングはユーザが興味のあるコンテンツに対して評価付けを行い、他のユーザとその評価付けデータを照らし合わせることによって、似た傾向の評価をつけるユーザを同じ興味をもつユーザとしてグルーピングし、同じグループ内のユーザで未評価のコンテンツがあった場合にそれを推薦する。評価付けの手法は Web の閲覧履歴を利用したもの [18] や、購買履歴を利用した Amazon の商品推薦 [19]、ユーザが明示的にコンテンツに対して点数付けをする GroupLens[20] などがある。特に Amazon は、通信販売を主たるサービスとしてサービス利用開始時期からサービスを満足に受けられる状態があった上で、購入および閲覧した商品が蓄積され、商品が推薦されるといった形態をとることによって、継続的な運用が実現できている。

協調フィルタリング以外にもベイズ推定を活用した映画推薦システム [21] がある。

いずれにおいても嗜好に基づく推薦では全文検索で問題に挙げた、ユーザのキーワード選択の困難さを解決する。コンテンツに対する重み付けさえ可能であれば、安定した性能を確保できるが、サービス利用開始時は評価情報が少なく精度が悪い（コールドスタート問題）。そのため本来のサービスに対して、付加的に利用する形態をとらない限り、安定した運用は困難となる。

2.1.4 メタデータによる情報検索

メタデータは、「データに関するデータ」といったように、情報の内容を表示すると同時に、情報を構造化する付加的なデータである。

2.1.1, 2.1.3 に述べたように、従来の情報検索では機械処理可能なデータがキーワードのパターンマッチングもしくは評価値の比較など単純な動作に限られていたため、情報検索・推薦結果の並び替えが困難であった。メタデータによる情報検索はデータの意味を考慮したマッチングを行うことから、マッチング性能の向上が期待できる上、メタデータの項目が増えるほど評価軸が増えることになり情報検索・推薦結果の並び替えも自由に行えるようになる。

従来は書籍や CD、映画情報など事前に分類が十分されているデータの検索に扱われることが多かったが、Semantic Web[22] の提唱に伴い RSS (RDF Site Summary) といった RDF (Resource Description Framework)[23] によるメタデータを活用した Web 情報全般に対する情報検索が提案された。

また、コンテンツの処理結果を他のプログラムと交換しあい流通する情報の価値を高める手法 [24] や、CDN (Contents Delivery Network) における接続先の選択の同定 [25] など適用範囲はコンテンツ以外にも広がっている。

検索クエリに含む項目数が多いほど検索精度は向上 [26] するが、ユーザにとってはクエリの項目入力の大きな負担となる。また自らの要求を細かく指定しすぎた場合に個人が特定されるプライバシーリスクが発生する場合がある。これに対して、プルダウンメニューによるクエリ入力の簡易化 [27] や、プライバシーとサービス品質のトレードオフを考慮した個人情報制御機構の提案 [28] がなされているが、依然これを活用したサービスが実運用されていないため、ユーザの許容性に対して解を得ていない。

また、メタデータは RDF などの XML (eXtensible Markup Language) によって記述されるが、検索における動作速度が従来のリレーショナルデータベースエンジンより遅いことが実運用においての問題となっている。また、コンテンツメタデータを生成する手法も確立されていない。コンテンツの中身を解析し内容を理解するに至っていないため、自動で生成できるメタデータは限られている。多くはコンテンツ提供者が手作業によって作成することになり、コンテンツ量の増加

が見込めないのが現状である。

それに対して、コンピュータがコンテンツを解析し、意味付けを行う手法が検討されている。しかし、ほとんどが研究段階のもので実運用化されていない。自然言語処理技術の応用がほとんどで、文書中から住所文字列を特定する手法 [29, 30] や、文書中に記述された出来事の生起時間帯を判定する手法 [31, 32, 33, 34] などがある。

2.1.5 コンテキストウェアサービス

コンテキストウェアサービスは、ユビキタスコンピューティング環境に設置されているセンサや人が携帯するデバイスの出力をサービスへの入力とし、その入力に応じて出力を変えるサービスである [35, 36, 37]。

ほとんどの既存研究が位置に応じた動的なサービスを提供することを目的としている。

GPS や屋内センサによってユーザの位置を常に捕捉しユーザが携帯情報端末に書き込んだ情報との対応付けを自動的に行い、他者に対して位置情報を提供するもの [38] にはじまり、特定の位置にユーザが訪れたときにその位置にあったコンテンツを推薦するシステム [39] がある。実証実験を通して位置コンテキストの正確性に課題を見出すといった知見の抽出 [40] も現れてきた。

さらに提示する位置情報や位置の活用に工夫を凝らしたものとして、ユーザの座標位置を取得して、そのユーザが視認可能なコンテンツのみを推薦する [41] ものや、広告などの展示物にユーザが近づくと、そのユーザの属するコミュニティとのチャットを開始し、より適切なユーザにその広告を配信する [42] システムがある。

近年においてはユーザの要求の多様化に合わせて、コンテキストの次元が増加しつつある。公共端末での情報表示など利用している端末をコンテキストとして扱うもの [43] や、ユーザの動作（立つ、座る、歩くなど）に合わせて適切なサービスを提供するもの [7] がある。

さらに、コンテキストの2次利用を考慮してコンテキストを抽象化する手法 [44]、ユーザの時間経過を考慮したコンテキストの変化を再現したシミュレーション [45]

を通してコンテキストの体系化が進められている。

いずれにおいてもユーザの状況にあったサービスを提供する上で非常に有用だが、サービス高度化のためにコンテキストの次元を増加させるにあたって、センサで自動的に取得できずにユーザに手入力を求めることがあったり、サービスを受けるためのルールを設定をユーザが行わなければならないことがあったりするため、ユーザのサービス参加の敷居がまだ高い。

2.1.6 情報検索手法まとめ

以上の考察により、全文検索にはマッチング・スコアリングの柔軟さに欠ける問題があり、嗜好に基づく推薦ではある程度のユーザの行動の蓄積が必要となるためユーザのその場の状況にあったコンテンツを提供することが困難となるため適用は難しい。メタデータによる検索は様々な属性に対して的確な検索を可能とするが、検索クエリの作成労力がモバイルシーンでの利用において現実的な解を得ていない。コンテキストアウェアはユーザの状況を把握するための十分な議論がされているが、これもメタデータによる検索と同様、複数の属性が存在した場合のユーザの作業量に対する考慮がなされていない。したがって、今後課題となる検索手法はコンテキストアウェアサービスの状況対応性とメタデータによる属性対応性を併せ持ち、ユーザのクエリ生成作業量が限りなく低いものとなる。

2.2. リアルタイム性を考慮した情報流通基盤の分類・考察

第 1 章で述べた行動支援における第 2 の課題であった瞬時に検索結果が得られることを満たすには、コンテンツ増加に対して検索等の演算負荷のスケラビリティを確保する必要がある。現状の情報流通基盤について分類し、考察し本研究の目的に合った情報流通基盤を見いだす。

2.2.1 サーバ集中型

サービス受付およびサービス処理を1のサーバあるいは1つのURLにて実現する。データの管理が楽な一方、コンテンツが膨大に増加した場合に処理遅延が発生する。バックエンドに複数のサーバを用意しドメイン内ロードバランシングなどを図る手法があるが、データの置き場を複数台のコンピュータに分散するパーティショニングの処理が難しい。

2.2.2 エージェント型

エージェントは自立性、協調性、適応性、頑健性、移動性をもつソフトウェアである。サービス構成要素を各々エージェントとして実装することによって、多様なサービスをエージェント基盤内に展開できる。

センサエージェントによってユーザ位置を捕捉し、特定の位置にたどり着いたときなど、条件が発火したときにサービスエージェントから情報を推薦するなどの活用ができ、コンテキストウェアサービスとの親和性が高い。条件発火まで手元の携帯情報端末にエージェントを保持しておくことができるため、ネットワークの接続性がない場合にもサービスを受けることができる。

メモリ使用量の削減を考慮したフレームワークの実装[46]や、Semantic WebにおけるWebサービス部分の実現[47]、コンテキストウェアサービスに特化した実証実験など[48, 49]によって実現可能性が確認されている。

しかし、実験規模がユーザ数十人において処理時間が秒オーダーであるものがほとんどで、莫大なWeb情報をインターネットユーザ全般に展開することは困難であると予想できる。

2.2.3 P2P型

クライアントとサーバ機能を同時に持ったノードを分散させる形態である。情報検索方式について大別すると、ハイブリッド型とピュア型に別れる。

ハイブリッド型はノードのインデックスを集中的に管理するため、P2P ノードの検索が容易であるがコンテンツ増加に対してスケーラビリティがない。Napster[50] が代表的である。ピア型は P2P ノードが自律・分散・協調し経路制御を行い、オーバーレイネットワークを構築し全ての動作において分散する。ピア型からさらに細分化すると、非構造型および構造型に別れる。

Gnutella[51] に代表される非構造型はコンテンツ検索においてクエリをフラッディングする。そのため、発見効率を確保するためにフラッディング範囲を広げる (TTL を長くする) と、ユーザの増加に対してスケーラビリティの問題が発生する。キーワードの関連性を考慮してフラッディングすることによって短い TTL においても発見効率を上昇させるなどの工夫 [52] が提案されているが、ユビキタス環境のユーザ規模を想定するとスケーラビリティの問題は解決されているとはいえない。LL-Net[53] では、位置情報に基づいて P2P ネットワークを構築し検索に必要な最小なクエリを送信する試みがなされ、実空間情報を扱うサービスにおいては非常に有効になると期待できる。

非構造型は、検索が高速な特徴を持つハッシュテーブルを分散管理する DHT (Distributed Hash Table)[54, 55, 56] とよばれる形態によって、数 10 億ノードの規模を想定した P2P ネットワークを実現する。一般的に $O(\log N)$ の検索速度の特徴を持つが、より速度を向上するために DHT をクラスタリングしたもの [57] や階層構造で要約するもの [58]、Bloom フィルタの適用とクエリのキャッシングによって実運用性を考慮したもの [59] がある。しかし、ハッシュを用いる制限として完全一致検索のみしかできないという問題点がある。これは例えばセンサノードを P2P ネットワークに展開する [60] などサービス要素の名前付けに厳格なルールが適用できる場合は問題ないが、範囲検索や部分一致などの要求があった場合に適用できない。

また、複数のオーバーレイネットワークにエージェント単位のピアが参加し、目的に応じてそれらを使い分けるという手法も PIAX[61] と呼ばれるアーキテクチャによって実装がすすめられている。これによって、検索要求を満たす限り構造型 P2P を用い、そうでないときに非構造型 P2P を活用することによって、検索効率と検索速度の両立を可能とする。

2.2.4 情報流通基盤まとめ

以上の議論より、爆発的なコンテンツの増加に対して従来のサーバ集中型およびエージェント型ではスケーラビリティが確保できない。P2P型を活用することが本研究の課題の解決として有望となる。しかし、P2P型は一属性によってノードを配置するため、行動支援情報が持つ属性のうち最も広い範囲で分散する属性を見出す必要がある。さらに、行動支援情報は複数の属性から成るため、その他の状況・属性・嗜好に対するマッチングに対応する必要がある。また、利用シーンによって様々なアルゴリズムの使い分けが必要になる。したがって、P2P型アーキテクチャの適用のみでは本研究の目的は解決できず、複数属性のマッチングに対応し、かつ爆発的なコンテンツの増加に対応できる新たなアーキテクチャを見出す必要がある。

2.3. 行動支援情報の情報源の分類・考察

第1章で述べた行動支援における第3の要件となるユーザのこれからとる行動を示唆する情報をどこから取得することが有効かについて議論する。

2.3.1 商用サイトと個人サイトの比較

まずWeb情報源は大まかに分けて商用サイトと個人サイトに別れる。

飲食店情報を例に挙げると、商用サイトは初期のぐるナビ[62]やYahoo!グルメ[63]といった、コンテンツ提供者が取材を元に集めた情報をWeb上に公開したものが主となる。時間と位置によって飲食店情報を構造化し、検索可能としたサービスに“*What's Open*”[64]がある。これは図2.1に示すように、画面上の地図範囲に存在する飲食店のうち、検索時刻に開店しているもののみを表示するサービスである。状況にあったコンテンツを提示する点においては行動支援情報提供手段として有効である。しかし、未だに電話帳等を元にした飲食店の電話番号、開店時間、場所といった状況に対応する情報の整理に留まる。ぐるナビ等のサイトでは、商用サイトのコンテンツ提供者は取材料を取材先から徴収しコンテンツを掲載す

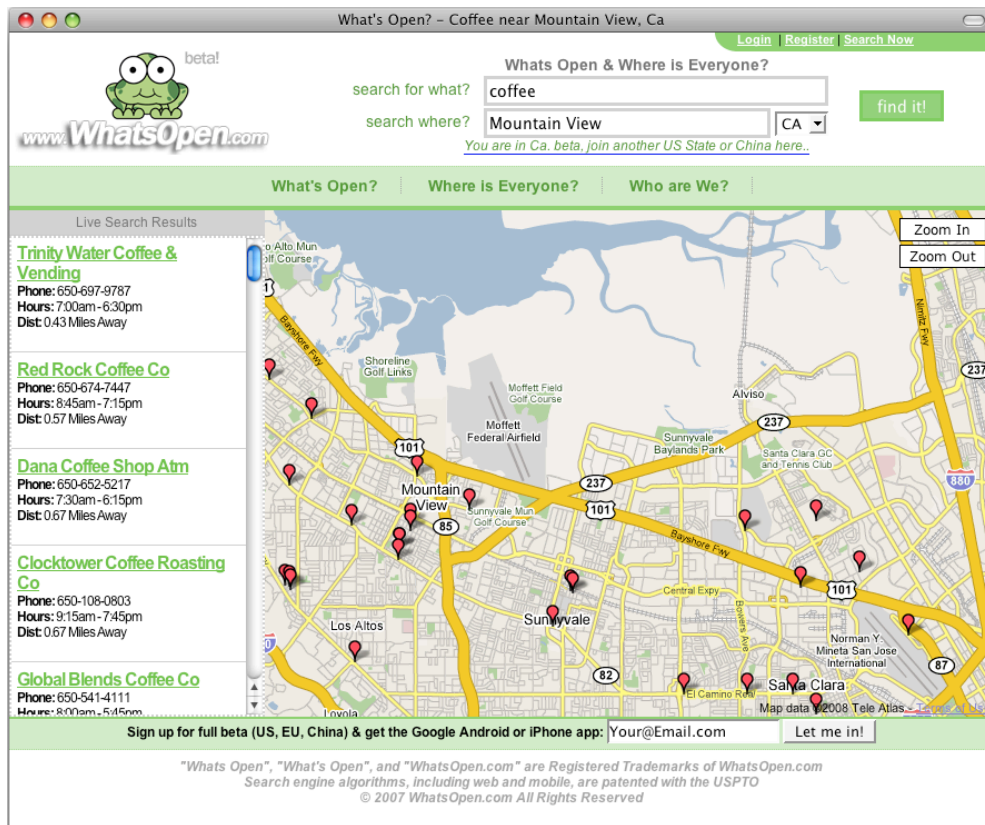


図 2.1 “What’s Open” での検索例

るため、基本的にその飲食店を褒める情報しか記述されない。コンテンツ全般がこのような偏った評価の記述になると、ユーザにとってはありきたりな情報となる恐れがある。

これに対して個人サイトに集まる情報は、商業者と逆の立場である消費者によって作成されたものとなる。個人サイトは、個人がHTML(Hyper Text Markup Language)を記述して公開するものや、すでにWebアプリケーションとして簡易に記事投稿を可能とするブログなどが例として挙げられる。消費者という立場にちなんで、個人サイトに集まる情報をCGM(Consumer Generated Media)と呼ぶ。同じ立場のユーザが報酬を目的とせずに情報を発信していることから、率直な意見を記述している。他のユーザが信頼するといった、口コミ要素が強い。このような評価の平等性の高い情報の中で、ユーザが共感する情報があった場合には飲食



図 2.2 “SHOOTI” での検索例

店情報の場合は、実際にその飲食店に訪れる可能性が高まり、行動支援情報として価値が高いものとなるため、本研究の扱う情報源として有力となる。実際にブログ中の評価表現を自然言語処理によって抽出し提示する手法 [65] がある。そして、これを応用した SHOOTI [66] といったサービスが実現され期待がかかっている。図 2.2 に SHOOTI での検索結果を例示する。通常の全文検索結果に加え、そのブログ中に含まれる評価表現を抽出・提示し、口コミ要素を強調している。ユーザはこの評価表現を参照して、検索対象が自身にとって価値のあるものであるか否かを判断する。

2.3.2 投稿型ポータルサイトとブログサイトの比較

次にこのような情報源が集合するサイトについて分類すると、投稿型ポータルサイトとブログサイトに別れる。

投稿型ポータルサイトでは、現在のぐるナビや Yahoo!グルメ、食べログ [67] といったものが挙げられる。従来の商用サイトを拡張して、ユーザ参加型の評判情報の書き込み機能が加わっている。店舗情報については商用サイトの特徴を持ち、行動支援に必要な店舗の位置（地図のリンクおよび緯度・経度情報）や開店時間が構造化されているためコンテンツの検索においては有利となる。GPS を内蔵した携帯電話からの検索においては近隣の店舗のみを提供するといった応用が可能となる。しかし、口コミ投稿数を調査すると、ぐるナビ (2008 年 2 月現在) において、首都圏など人口密度の高い地域では 1 店舗に対して最も多いもので 331 件近くの投稿があるが、奈良県では口コミ投稿されている店舗が 92 店舗しかなく、投稿数については最も多いもので 12 件、82 店舗が 1 件の口コミ情報しか投稿されていないため、満足な情報量とは言えない。これは、ポータルサイトを目的毎に使い分ける必要があり、そのポータルサイト毎に投稿の方法を身につけなければならない煩わしさが影響していると考えられる。

一方ブログサイトはユーザの投稿する手法は一通りしかないので、気軽に情報のアップロードができる。しかし、簡易な投稿方法であるが故に、情報の構造化がほとんどなされておらず、全文検索以外の検索手法が見出せない問題がある。

以上を踏まえると、ブログサイトに投稿された文書の内容をコンピュータが理解し、コンテンツメタデータを自動抽出することが可能となれば、コンテンツ量および情報構造化の要件が同時に満たされると考えられる。

2.4. 本研究の目的・課題

以上の議論より、行動支援情報をユーザに提供するにあたって、情報源を個人サイトの特徴を持つ CGM とし、情報取得手法はメタデータ検索とコンテキストウェアの要素を両立する手法を取り入れ、P2P 型アーキテクチャを採用することが望ましい。

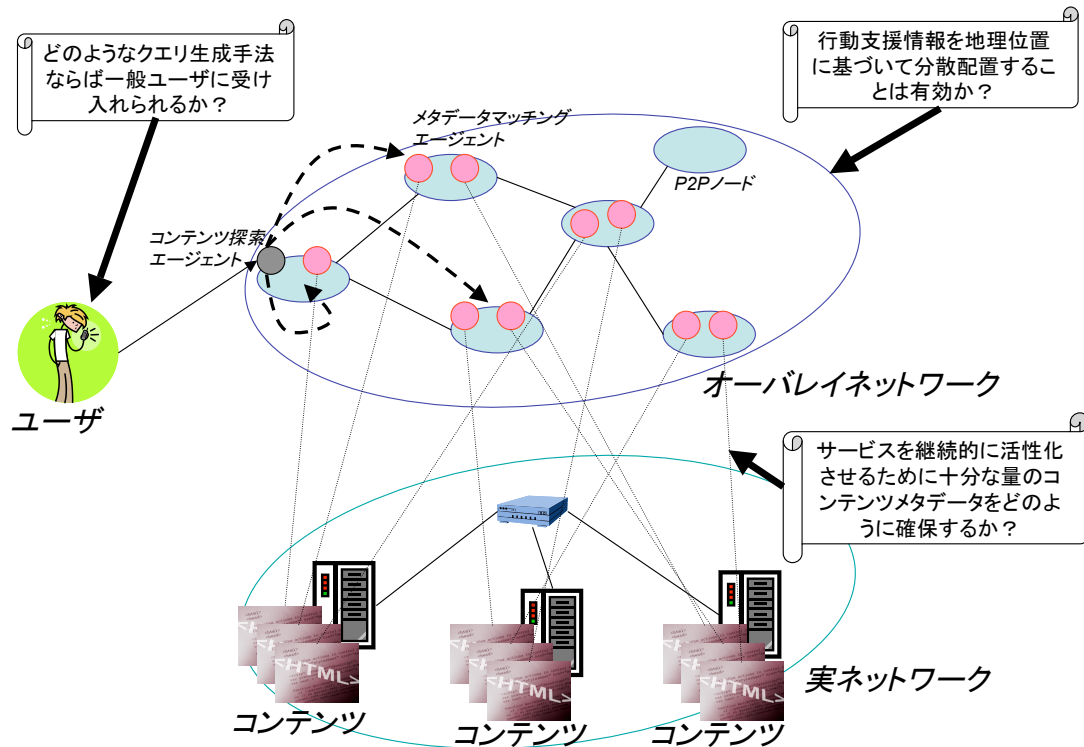


図 2.3 行動支援情報提供モデル

目的とするシステムの全体像を図 2.3 に示す。コンテンツメタデータを保持しメタデータマッチング機能を持つエージェントを P2P ネットワーク上に配置し、コンテンツ増加に対する演算負荷の分散を図る。エージェントが保持するコンテンツメタデータは実ネットワーク上の Web サイトに CGM 作成時点で付加される、あるいはクローラによって自動解析・抽出される。

図中に示す 3 つの課題の詳細を以下にまとめる。

- ユーザーのサービス利用における許容性の課題

ユーザーにサービスを提供する上で最も重要な点となるのはユーザーが使いやすいと思うことである。本研究ではユーザーの状況・嗜好を反映した情報を的確にユーザーに提供するために、メタデータマッチングを行う。このクエリを生成するユーザーの作業量は限りなく少ないことが望ましい。ユーザープロファ

イル相当の入力の手間, プライバシリスク, デジタルデバイドの防止を考慮したサービス利用方法について検討する必要がある. さらに, そのような限られた量のクエリ要素を最大限利用して, ユーザが満足するコンテンツを提示しなければならない. この課題を解決する提案および有効性の検証を第 3 章で述べる.

- システムのリアルタイム性と柔軟性の課題

本研究において扱うメタデータマッチングによるコンテンツ推薦では, 各メタデータがどの程度合致していれば推薦するかを決定するアルゴリズムは必ずしも 1 つに決める必要はない. 最適なアルゴリズムは, 運用主やコンテンツの傾向などに応じて変化することが考えられ, 適切なものを運用中に選択・更新可能であることが望まれる. ここでは, コンテンツの更新が複数の箇所で行われる P2P 型アーキテクチャを想定しているため, P2P 環境においてアルゴリズムの選択や更新が柔軟に行える構成とする必要がある. また, 従来, P2P 型アーキテクチャによりこのようなアプリケーションを実際に運用した例は少ない. 従来の P2P ネットワーク構成技術の研究では, 理想的なシミュレーション評価が主であり, 実フィールドのネットワーク規模でユーザの挙動やコンテンツに偏りが生じる状況のもと, リアルタイム性を確保し, 実用的なアプリケーションを実現可能であるかどうかは, 運用してみなければ分からない. この課題を解決する提案および有効性の検証を第 4 章で述べる.

- 正確なコンテンツメタデータの維持性の課題

メタデータマッチングに利用するコンテンツメタデータが正確でなければ, ユーザの要求にあったコンテンツを推薦できない. しかし, 完全に正確なコンテンツメタデータを作成するには従来の手法ではコンテンツ提供者の手作業に頼ることになり, 今後さらに増加する膨大なコンテンツに対処できなくなる. 今までいくらかのメタデータ入力支援ツールが開発されてきたが, より省作業なツールを作成する, あるいはコンテンツの意味をコンピュータが解析し内容を理解し, 自動的にメタデータを付与することによって, 正確なコ

ンテンツメタデータを大量に生成するといったメタデータ処理機能が必要となる。この課題を解決する提案および有効性の検証を第5章で述べる。

第3章 行動支援情報提供を考慮した メタデータマッチング手法

本章では2.4に挙げた「ユーザのサービス利用における許容性の課題」について主に取り組む。課題を満たすコンテンツ推薦アルゴリズムを提案および設計・実装し、ショッピングモールでの実証実験を通して得られた知見について述べる。

3.1. ユーザの許容性を考慮したメタデータマッチング 手法

第2章でも述べたように、移動中のユーザが検索時に入力できる情報は限られているため、少量のクエリから最大限ユーザの要求に見合った情報を提供する必要がある。コンテキストウェアサービスとメタデータマッチングの長所を取り入れることによってユーザに行動支援情報を的確に提供することが期待できるが、一般ユーザを対象として検証した例は少なく、どれだけのユーザがこの検索手法を受け入れ利用するか、またその検索結果に満足するかといった、ユーザにとっての有用性がまだ明らかにされていない。

また、ユーザの要求を満たす十分な量と質のコンテンツを提供するために、現場のコンテンツ提供者がシステムを運用できる必要があると考えられるが、どのようなシステムであれば実際に運用ができるのか、また、運用上どのような問題があるのかといった知見やノウハウはこれまでに得られていない。

上記の要求を満たすべく、現場の要求に合わせ推薦に必要なメタデータおよび推薦アルゴリズム設計を行った上で、青森県五所川原市のショッピングモール「エ

ルムの街」¹で1ヶ月の長期にわたる実証実験を行った。本章では、本実証実験システムの主要素となるメタデータマッチング部の設計および検証結果について述べる。

3.2. 行動支援情報提供メタデータマッチングの提案

ユーザの許容性を考慮した省作業な推薦方式について述べ、ユーザプロフィールおよびコンテンツメタデータの設計し、次にそれらを活用するコンテンツ推薦アルゴリズムを提案する。

3.2.1 RFID カードを活用したゼロクリック推薦

近年、Suica²、ICOCA³といった無線ICカードを用いた認証・課金手法が普及しつつある。従来の切符を買うために販売所に列をなして待つ必要がなくなる、無線のためカードをケースに入れたまま認識することが可能など、ユーザの利便性に定評がある。また関西私鉄用のICカードPiTaPa⁴を利用したグーパス [68] では、改札機にカードをかざすイベントに連動して、事前に登録しておいた携帯電話のメールアドレスに、占いやニュースなどの情報を送信するサービスを行なっている。その時間に有効なコンテンツを提供するだけといった、ユーザの細かい要求を考慮しない単純な推薦ではあるが、ユーザのクエリ作成労力は限りなく低い。クエリ要素の主となるユーザプロフィールをあらかじめ登録しておき、その後はカードをかざす認証作業のみで推薦システムがユーザプロフィールを取得することが可能であれば、ユーザの作業量を抑えつつユーザが満足する行動支援情報を取得できると期待できる。したがって、本研究ではRFIDカードを端末にかざすのみでユーザを識別し、あらかじめ登録したユーザの属性・嗜好をシステムが自動取得し、適切なコンテンツを提供するサービス形態の実現に取り組む。これによってユーザのサービス利用時の省作業性を確保する。

¹<http://www.elm-no-machi.jp/>

²<http://www.jreast.co.jp/suica/>

³<http://www.jr-odekake.net/guide/icoca/>

⁴<http://www.pitapa.com/>

```
<?xml version="1.0">
<up:UserProfile
  xmlns:pp="urn:liberty:id-sis-pp:2003-8"
  xmlns:up="urn:x-ubiq-profile:2005-9">
  <up:UserId>5058</up:UserId>
  <up:PreferenceProfiles>
    <up:item>和食</up:item>
    <up:item>海外旅行</up:item>
  </up:PreferenceProfiles>
  <pp:LegalIdentity>
    <pp:Gender>urn:liberty:id-sis-pp:gender:m</pp:Gender>
    <pp:LJobTitle>学生</pp:LJobTitle>
    <pp:DOB>1985-04-16</pp:DOB>
  </pp:LegalIdentity>
</up:UserProfile>
```

図 3.1 ユーザプロファイルの記述例

3.2.2 ユーザプロファイルの設計

ユーザが情報検索時に作成するクエリ要素には、時間・場所といったユーザの状況と、属性（年齢・性別・職業等）、嗜好（好みのジャンル、好みの商品等）がある。このようなクエリの元となる情報をユーザが検索毎に手作業で設定することなく機械処理によって自動的に見出せる事が望ましい。ユーザの状況を示す情報は時計による現在時間の取得、GPSあるいは固定位置に設置された端末の事前位置設定による現在地の取得などによって自動化できる。ユーザの属性・嗜好については適切なセンサの実現が難しく自動抽出が困難であるが状況情報に比べて時間に対する変化が緩やかなため、あらかじめ手作業によって設定をしておき、前節で述べたようにRFIDカードの認証によって再利用することでユーザのクエリ入力の手間を削減できる。したがって、本システムではユーザのプロファイルをサービス利用前にあらかじめ設定しておき、そのプロファイルを検索時にクエリとして送信するとともに、検索者の状況を自動抽出し反映する。

本章で行う実証実験はショッピングモールでの商品推薦サービスを提供するた

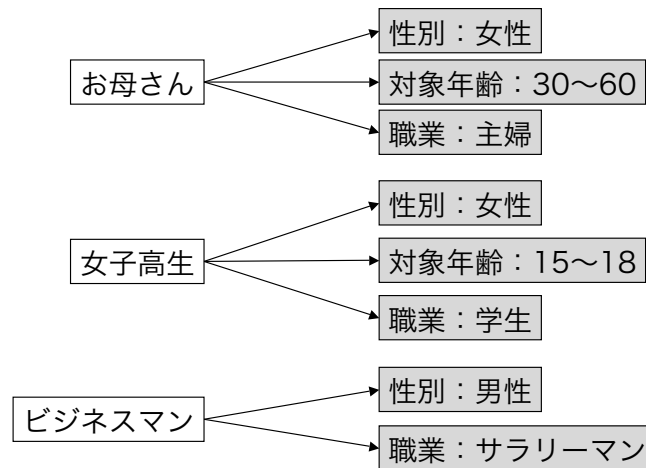


図 3.2 cp:keyword から対象ユーザ特性を連想するリスト

め、商品によって異なる対象者の属性を考慮してユーザプロファイルの設計を行った。ユーザプロファイルの記述例を図 3.1 に示す。コンテンツ推薦に有効となる値は、年齢 (pp:DOB：誕生日から計算)、性別 (pp:Gender)、職業 (pp:LJobTitle)、嗜好キーワード (up:PreferenceProfile) である。記述内容はユーザの特徴を表層的に表す程度の情報量で、一般的なユーザ登録を必要とする Web アプリケーションでの設定項目と同等の量であるため、簡単な選択式のアンケートで十分収集可能で、ユーザにとって入力に負担はかからないと考える。これらの属性からは個人を特定することは困難で、プライバシー的にもユーザにとっての負担は小さい。したがって、2.4 で掲げたユーザのサービス利用に対する許容性の課題を解決できると期待できる。

3.2.3 コンテンツメタデータの設計

ユーザにとって的確な情報検索を実現するにはユーザプロファイル中の属性・嗜好に対応した情報や情報表示に必要な説明情報を含むコンテンツメタデータが必要となる。ユーザプロファイルの設定量を少なく抑えたことによって、それに対応する情報の設定量も削減できているが、その他の有効期間や存在位置などの

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:cp="http://www.ais.cmc.osaka-u.ac.jp/Contents-Profile#"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
  <rdf:Description about="urn:x-ubiq-content-id:0001">
    <dc:title>デニムジャケット</dc:title>
    <dc:publisher>Paul Star</dc:publisher>
    <dc:description>黒のデニムジャケット</dc:description>
    <cp:arealD>1</cp:arealD>
    <cp:price>25000</cp:price>
    <cp:figure>
      <rdf:Seq>
        <rdf:li rdf:resource="http://www.ais.cmc.osaka-u.ac.jp/Contents/0001.jpg"/>
        <rdf:li rdf:resource="http://www.ais.cmc.osaka-u.ac.jp/Contents/0002.jpg"/>
      </rdf:Seq>
    </cp:figure>
    <cp:pr>ファッション雑誌で紹介されたイチオシの商品です!!</cp:pr>
    <cp:flash>http://www.ais.cmc.osaka-u.ac.jp/Contents/0001.swf</cp:flash>
    <cp:shopID>0001</cp:shopID>
    <cp:majorDivision>卸・小売り業</cp:majorDivision>
    <cp:middleDivision>織物・衣服・身の回り品小売業</cp:middleDivision>
    <cp:minorDivision>衣服</cp:minorDivision>
    <cp:keyword>
      <rdf:Bag>
        <rdf:li rdf:resource="urn:x-keyword:ファッション雑誌">
        <rdf:li rdf:resource="urn:x-keyword:メンズヤング">
      </rdf:Bag>
    </cp:keyword>
    <geo:Point>
      <geo:lat>40.758059</geo:lat>
      <geo:long>140.455069</geo:long>
    </geo:Point>
  </rdf:Description>
</rdf:RDF>
```

図 3.3 コンテンツメタデータの記述例

ユーザの状況に対応する情報や画面表示に必要な情報などがあるためコンテンツ提供者のコンテンツメタデータ設定負荷は依然高い。

コンテンツメタデータの継続性については主に第 5 章にて取り組むが、本章での実証実験においても推薦するコンテンツの十分な量の確保が必要となる。元となる電子データが存在しないため、一からコンテンツを作成する必要があることを考慮して実運用上の工夫を行う。

図 3.2 に示すような連想リストを用いて、1つのキーワードを設定することによってユーザプロファイルの対応情報を設定できるようにした。例えば、「リップク

リーム」といった商品をコンテンツとして登録する場合に、コンテンツメタデータに「女子高生」というキーワードを登録するだけで、推薦時に必要となる対象性別（女性）、対象年齢（15～18歳）、対象職業（学生）といった属性を補完できる。このように、キーワードを複数のユーザプロフィール項目に対応させることによってメタデータ入力の省作業化が実現できる。

コンテンツメタデータの記述例を図3.3に示す。コンテンツ作成者のメタデータ作成の労力を考慮して、手入力が必要な情報を必要最小限にしている。cp:majorDivision, cp:middleDivision, cp:minorDivision は大まかな商品分類となる大分類、中分類、小分類の値を保持する。大分類から小分類に向かってツリー構造になっているため、コンテンツ提供者はそのツリーを辿り、商品に該当する小分類を選択肢から選ぶだけでよい。分類の選択肢数は大分類において20、そのあとに辿る中分類・小分類については平均3程度となる。これらの値が商品に設定されていることにより、ユーザが明示的に嗜好に合っていると指定する商品があったときに、その商品とツリー上の位置関係が近い商品を推薦することによってユーザの嗜好にあった商品を推薦することが可能となる。cp:keyword に記述するキーワードはタイムセールスなどのタイムリーなイベントが起こったときに重要度を操作するといったフラグの役割を持たせたり、プロフィールマッチング時には、上述の通り、図3.2に示すようなリストからユーザプロフィールに対応する値を引き出したりするために用いられる。また、これらの情報は基本的にチェックボックスで値を選択するのみで簡易に値を決定できるようにしている。

3.2.4 コンテンツ推薦アルゴリズムの設計

ユーザのサービス参加の敷居を下げるために限られたクエリ入力量で、最大限ユーザが満足するコンテンツを推薦しなければならない。ショッピングモールでの商品推薦に特化した、ユーザおよびコンテンツ提供者の要求を満たす推薦アルゴリズムを設計する。推薦アルゴリズムの全体像を図3.4に示す。

まず、ユーザの状況を考慮した推薦手法に従って、リクエストに合致するコンテンツを保持しているかどうかを判断する。具体的には利用者の近隣にある商品であるかという判断と、コンテンツメタデータに設定されている有効期間内に現

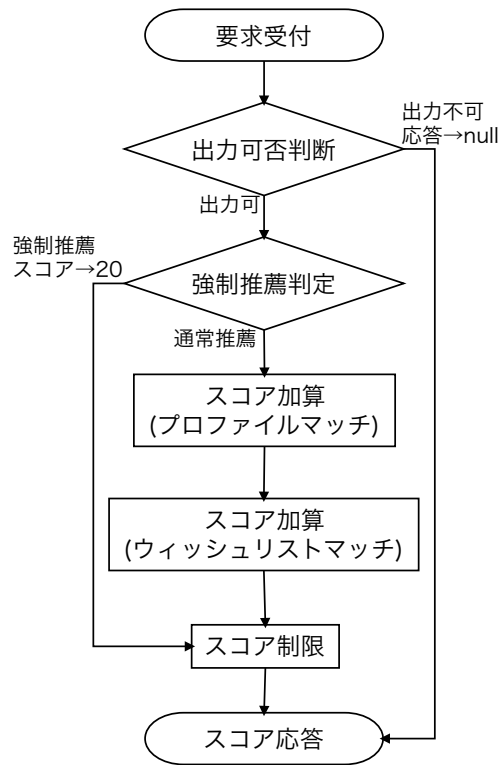


図 3.4 コンテンツ推薦アルゴリズム

在時間が該当しているかという判断である。位置による検索については、詳細は第 4 章にて述べるが、位置に基づいたオーバレイネットワークである LL-Net を活用し、ユーザの周辺のみノードにクエリを送信することによって実現する。

強制推薦判定は、コンテンツメタデータの cp:keyword に「強制」というキーワードが存在するか否かを判定し、存在した場合は最高スコアを加算し、推薦処理は終了する。これはショッピングモール運営者との打ち合わせにより強く勧められた推薦手法である。従来の検索手法からは見出せないような意外性のあるコンテンツを推薦することによってユーザの新たな嗜好を生むのに有効である。

プロフィールマッチは、ユーザプロフィールに含まれる年齢・性別・職業とコンテンツメタデータに含まれる嗜好キーワードから推測される対象年齢・対象性別・対象職業を照合してその合致度をスコアとして算出する。

ウィッシュリストマッチはユーザが注目したコンテンツの履歴を保存しておき、ウィッシュリスト内にあるコンテンツとコンテンツメタデータに記述されている分類キーワードを比較して、類似するコンテンツほど高いスコアを計上する。

スコア制限では、ユーザ毎に推薦履歴を保存しておき、該当するコンテンツがある場合は減点する。実証実験では画面のデザイン上、同時に表示できるコンテンツが5つであった。推薦の度に表示される5コンテンツで、以前推薦されたものが繰り返し現れるのは望ましくないため、この数に合わせて検索履歴キュー長を5とした。最高で90%、最低で10%の減点が行われる。推薦されて間もないコンテンツほど多く減点される。

ウィッシュリストマッチに利用する分類キーワードの体系は頻繁に検索され、ユーザの推薦履歴は頻繁に更新されるため、これらの処理を簡易に高速に行うために、DHTを適用しストレージとして活用した。

3.3. ショッピングモールでの実証実験による評価と考察

実装したシステムを「エルムの街」にて1ヶ月間運用した。以下、実装内容および実験環境について述べる。

3.3.1 実験仕様

ショッピングモールは屋内環境のため端末位置同定にGPSを利用するといった簡易な手法が適用できないことと、PDAの貸出し手続きが大変である理由から本実証実験では店内に情報端末を設置し、その端末周辺の商品をユーザの嗜好・状況に合わせて推薦するシステムを実装した。図3.5に、ユーザが店内設置端末にRFIDカードをかざした後に表示される画面例を示す。左上に推薦された商品のプレビューと簡単な宣伝メッセージが表示される。最もスコアの高かった商品が最初に表示され、5秒おきに下位順位のものに入れ替わり表示される。入れ替わり候補は左下に表示されていて、5秒待たずとも下位順位の商品を見たい場合には左下のプレビューを押すと左上の枠に表示されるようになる。「くわしく見る」ボタンを押すと商品の詳細の画面に遷移する。また、情報提供端末の表示



図 3.5 コンテンツ推薦システムの画面イメージ

画面の右上部分には、操作中の端末の向きを反映した周辺の地図が表示される。店舗毎にクリックが可能となっていて、店舗をクリックすると、その店舗で扱っている商品のみが推薦対象となり、左の画面が更新される。右下は、ユーザプロフィールを反映した星占いなどの情報を提示し、ユーザ毎に表示されるコンテンツが異なることを確認させている。

実装したシステムを「エルムの街」にて実際に販売されている商品の推薦に適用し、一般顧客を対象に実証実験を行った。

ユーザはサービス利用の申し込み時にアンケートを記述し、そのアンケート結果を受け付け担当スタッフがシステム投入し、RFID カードを発行した。また、コンテンツの登録については、コンテンツ入力のための Web アプリケーションを用意し、スタッフルームより PC から店舗スタッフが入力した。

3.3 ショッピングモールでの実証実験による評価と考察

表 3.1 ショッピングモールでの実証実験の詳細

名目	内容
実験期間	1ヶ月間 (2006/2/4 ~ 2006/3/5)
店舗数	98 店舗
設置端末数	20 台
総アイテム数	880 コンテンツ
参加者数	2252 人
利用者アンケート回答数	552 人
テナントアンケート回答数	91 店舗
CPU	Pentium 4 2.56GHz
Memory	512MB
OS	Windows XP SP2
開発言語	J2SDK 1.4.2
Network	100BASE-TX
Passive RFID リーダ	通信距離 5cm, 2.45GHz

実証実験フィールドの詳細を表 3.1 に示す。ショッピングモール全体にわたっての 20 カ所に情報提供端末を設置し、100Base-TX で接続した。本提案のコンテンツ推薦システムには 6 台の端末を用意し、そのうちの 5 台で分散エージェントを構成し、残りの 1 台はコンテンツを保存する Web サーバとして稼働する。分散エージェントの詳細は第 4 章に関わるため、本章では割愛する。実証実験の参加者の平均年齢は 31 歳、最低年齢は 1 歳、最高年齢は 88 歳と、幅広い世代にわたって推薦システムが利用された。また、評価のためにユーザおよび店舗スタッフから以下の事後アンケートを収集した。

- – このサービスが普及したら使いたいですか？
- 端末は何が使いやすいですか？
- 推薦された商品はあなたの嗜好に合っていましたか？
- コンテンツを表示するまでの時間は問題ありませんでしたか？

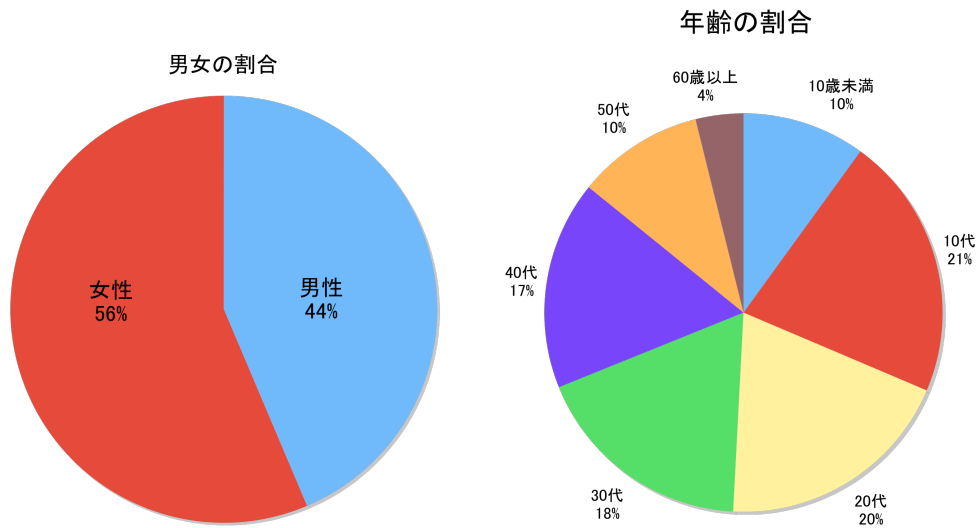


図 3.6 ユーザの性別・年齢別分布

- – コンテンツの商品情報は何個が適切か？
- どのくらいの頻度で更新可能か？
- コンテンツの提供に関して、原稿提出に関する手間は？

ユーザである利用者から集まったアンケートは552人分で回収率は24.5%，コンテンツ提供者である店舗から集まったアンケートは91店舗分で回収率は92.8%となった。

3.3.2 許容性の評価・考察

第一の課題はユーザの許容性を考慮し、どのようなユーザでも利用しやすい情報推薦システムを実現することである。そして、限られたユーザの操作量から最大限満足できるコンテンツを推薦することである。

本実証実験におけるユーザの分布を確認したところ、図 3.6 に示すように、男女の割合・年齢の割合ともに特定の範囲に偏ることはなかった。実証実験ではサービス利用前に実験システムの概要を説明し、サービス利用に意欲的なユーザのみ

3.3 ショッピングモールでの実証実験による評価と考察

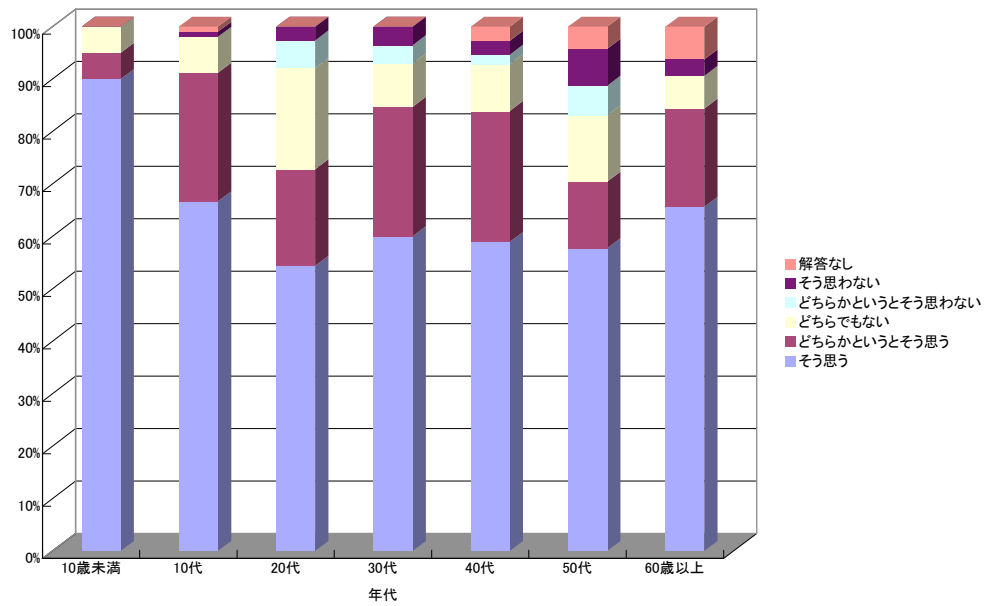


図 3.7 「このサービスが普及したら使いたいですか？」に対する回答 (年齢別)

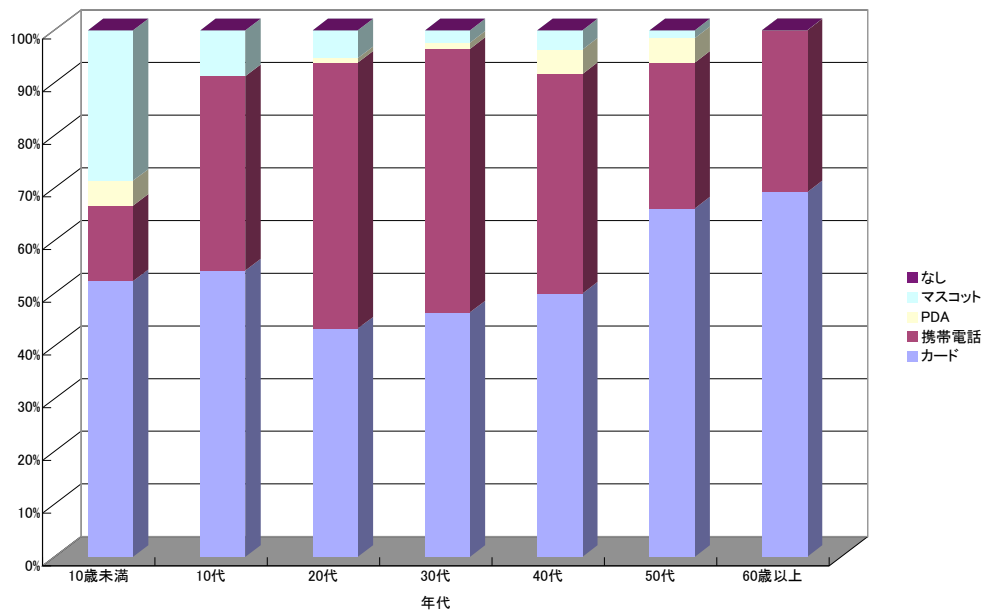


図 3.8 「端末は何が使いやすいですか？」に対する回答 (年齢別)

が利用するため、従来の情報技術の利用に抵抗のないユーザに分布が偏ることを懸念したが、結果として提案するカードをかざすのみの情報取得手法は多くの広いユーザそうに利用されることが確認できた。

また、そのような環境下で、「このサービスが普及したら使いたいですか？」といったユーザに対する質問の回答を年代別に評価したところ、図3.7のようになった。10歳未満の年代は9割近くとかなり多く、それ以外の年代でも概ね6割の人が利用したいと捉えている。

図3.8に示す通り、半数のユーザからカードによる端末操作が使いやすいと回答した。本実証実験のユーザのほとんどがRFIDカードを利用したコンテンツ推薦を初めて受けたにも関わらず、ユーザから十分な支持を得たことから、提案手法はユーザのサービス利用において許容性が高いことが確認できた。

以上、実装したシステムは、サービス利用開始時にユーザがアンケートに答え、その情報を元にプロフィールを事前に設定し、検索時にはカードをかざすのみで情報を得ることによってユーザ認証およびユーザプロフィール読み出し作業を1度に済ませることができた。カードを利用したサービス形態についても半数以上のユーザに指示された。サービス参加者が2252人と多く、ユーザの分布が性別・年齢において概ね均等となったことと、ほとんどの年代から今後普及したら利用したいと回答を得ていることから、どのユーザでもサービス参加しやすい、許容性のある情報推薦システムが実現できたといえる。

3.3.3 推薦の影響

次に、ユーザの要求を満たす推薦の実現具合を、詳細ボタンの押下状況、閲覧時間および事後のアンケートを元に分析する。そして、推薦コンテンツはWebサーバ上に蓄積されているため、そのアクセスログを元に分析する。

コンテンツ推薦手法はユーザの状況と事前アンケートによるプロフィール情報を単純にマッチングするアルゴリズムを用いたが事後のアンケート(図3.9)に示す通り、「おすすめ情報は嗜好にあったか」という質問に対して、「合っていた」、「どちらかといえば合っていた」を合わせると約半数となった。

3.3 ショッピングモールでの実証実験による評価と考察

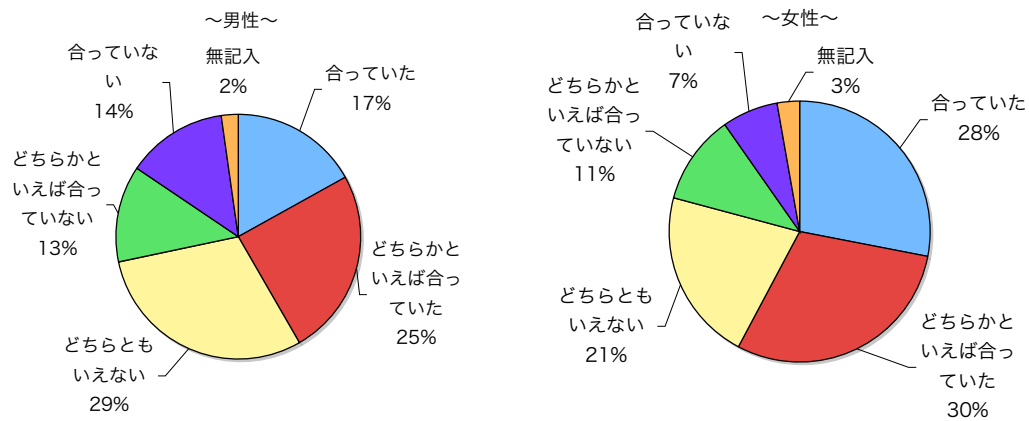


図 3.9 「おすすめ情報は嗜好にあったか」に対する回答 (男性・女性)

表 3.2 コンテンツの閲覧状況

項目	数
コンテンツ閲覧者総数	980
コンテンツアクセス総数	9655
一人あたりの閲覧コンテンツ数	653
「詳しく見る」が押された総数	2504
「詳しく見る」を最初に押した総数	1218
一人当たりの「詳しく見る」を押した総数	126
店舗ボタンが押された総数	7151
店舗ボタンを最初に押された総数	1310
一人あたりの店舗ボタンを押した総数	525

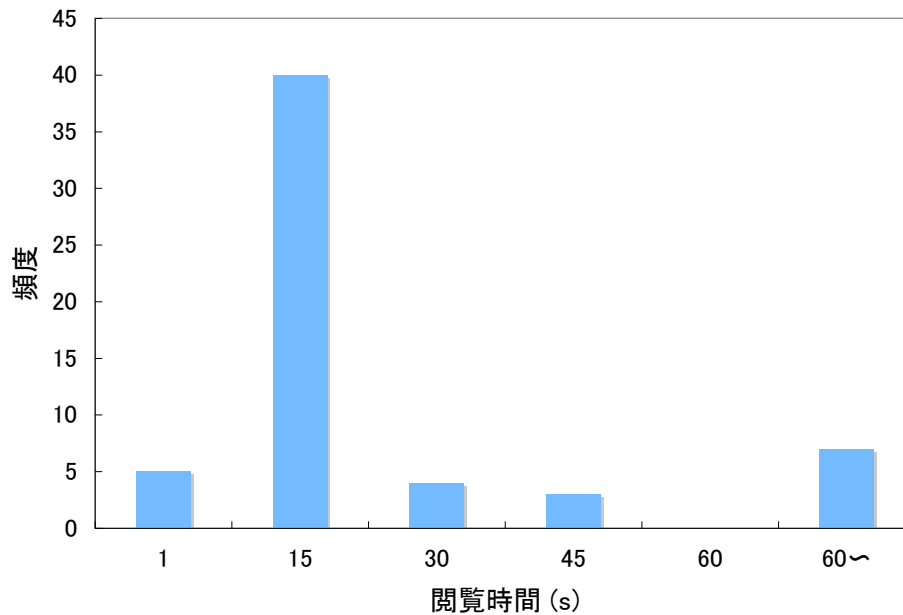


図 3.10 コンテンツ閲覧時間のヒストグラム

表 3.2 に Web サーバのアクセスログから得られたコンテンツの閲覧状況を示す。コンテンツ閲覧者総数が 980 人で、表 3.1 の参加者延べ人数が 2,252 人であったことから 1 人当たり 2 回は期間中に推薦システムを利用している。コンテンツアクセス総数が 9,655 回であったことから一人当たり約 10 回はコンテンツを閲覧している。この表で重要となるのは「詳しく見る」が最初に押された総数である。これは推薦されたコンテンツの詳細を確認するボタンであるため、推薦内容がユーザにとって適していれば押されることになる。期間中に「詳しく見る」を最初に押した総数は 1,218 回であった。約 13% の割合で最初に押されている。ゼロクリック推薦は Web 上の広告と捉えることができるため、WWW 広告の一般的なクリック率が 1% 程度と言われていること [69] と比べてよい数値であるといえる。

コンテンツの閲覧時間はコンテンツの有用さと相関がある [70] ため、その分布を調査した。残念ながら Web アクセスログはコンテンツ閲覧開始時間については容易に取得できるがコンテンツ閲覧終了時間が特定できないため、明らかにコンテンツ閲覧が連続していてコンテンツ閲覧終了時間が特定できるものだけを採取

した。

取得できた 57 件の閲覧時間の分布を図 3.10 に示す。15～30 秒の区間にデータが集中している。初めて推薦コンテンツを閲覧するユーザが大半を占めるため一般的な Web の閲覧時間よりは長くなる予想があるが、既存研究報告では概ね 10 秒以上が有用なコンテンツと考慮されていることからこれは注目されたコンテンツであると判断できる。今後は全てのアクセスに対して閲覧時間を取得できるようにして、ユーザがシステムに慣れた状態で計測することが課題となる。

以上、おすすめ情報が「嗜好に合っていた」、「どちらかといえば合っていた」をあわせて約半数となることや、クリック率および閲覧時間の分析が従来の研究報告などと比較して良好なことからコンテンツの内容にも満足できる推薦が出来たことが確認できた。

しかし、今回はショッピングモールでの商品推薦といった限られた目的に合わせたため、ユーザプロファイルの設計や推薦アルゴリズムの実装を比較的容易に行えたと考えられる。今後はシステムの有効範囲を街や観光地全体、地域全体と広げて展開する際にもユーザの要求を満たすサービスについて検討する必要がある。

3.3.4 コンテンツ提供者のメタデータ作成負荷に対する考察

コンテンツ提供者のコンテンツメタデータの作成労力は、連想リストの活用によって削減された。コンテンツメタデータに必要な情報は以下のものがある。

- 推薦処理に必要なもの
 - － 商品分類 (大分類, 中分類, 小分類)
 - － ジャンルキーワード
 - － 対象年齢
 - － 対象性別
 - － 対象職業
 - － 位置 (緯度・経度)

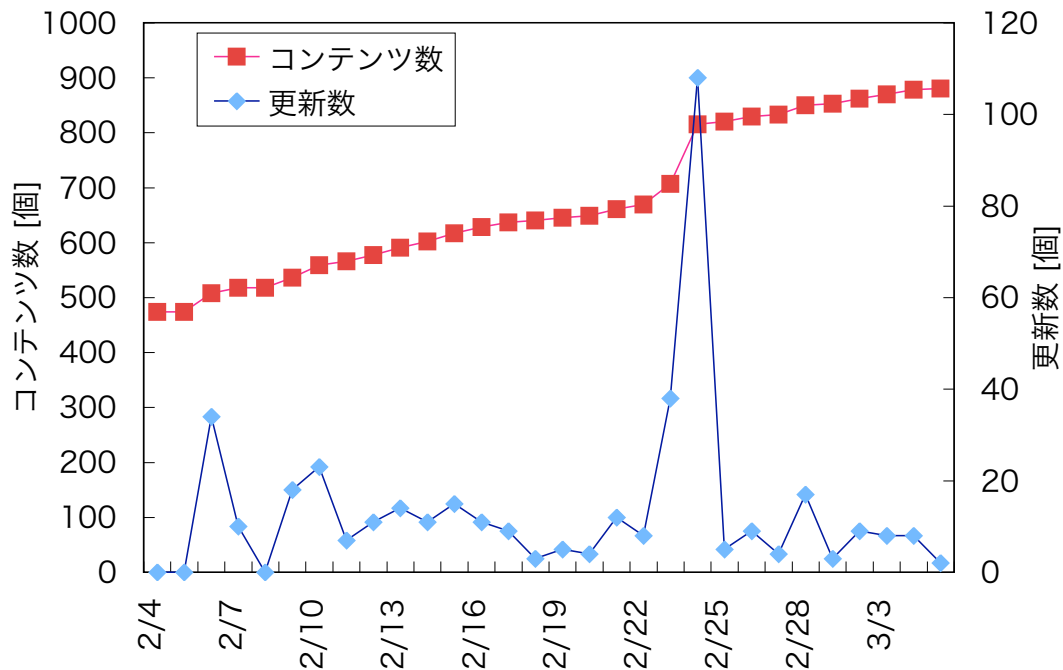


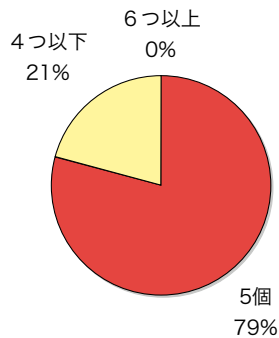
図 3.11 コンテンツ数の推移と更新頻度

- 強制推薦フラグ
- 推薦結果表示に必要なもの
 - 商品名
 - 商品の詳細
 - キャッチフレーズ
 - 写真のリンク
 - 価格

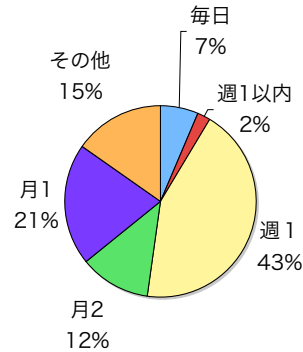
これらのうち、対象年齢、対象性別、対象職業については表 3.2 で設定した連想リストによって cp:keyword の 1 項目に集約されている。必要最小限のユーザープロファイルの設定により設定項目数を少なく抑えたうえで、連想リストによりさらに設定項目を削減できた。またジャンルキーワード、強制推薦フラグは推薦処

3.3 ショッピングモールでの実証実験による評価と考察

1. コンテンツの商品情報は何個が適切か？



2. どのくらいの頻度で更新可能か？



3. コンテンツの提供に関して、原稿提出に関する手間は？

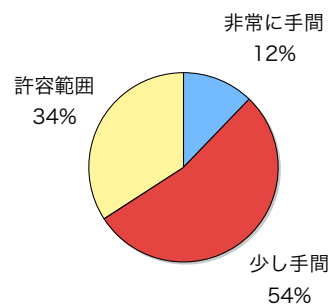


図 3.12 コンテンツ提供者に対するアンケート結果

理中に特定の値として分岐処理できるため cp:keyword に含めることによって設定項目を 14 項目から 10 項目に削減できた。

またコンテンツ登録数は図 3.11 に示す通り、ほとんどの日に更新され増え続けた。2/24 は改装に伴うリニューアルオープン店舗の影響でコンテンツ更新数が大幅に増加している。コンテンツメタデータの設定項目が少ない上で、実験期間中にコンテンツが一定の割合で増加していることから、コンテンツ提供者にとって継続運用が可能な設計であったことが確認できた。

しかし、図 3.12 を見ると、「どのくらいの頻度で更新可能か？」といった質問への回答は週 1 が 43% 近く、「コンテンツの提供に関して、原稿提出に関する手間は？」という質問に対して、許容範囲と回答したテナントは 34% しかいなかった

ことから、手作業によるコンテンツ入力には本来コンテンツ提供者が公開したい情報量を満足に提供しているとは言えないことが分かった。

3.4. まとめと今後の課題

本章ではユーザのサービス利用における許容性の課題について主に取り組んだ。ユーザの検索開始時にユーザの位置や時間を自動的に計算機環境が読み取り、またカードをかざすといった認証作業を通してあらかじめ登録しておいたユーザプロフィールを再利用することによって、移動中のユーザの限定された情報入力量に対して最大限の要素を持ったクエリ生成を実現した。

カードをかざすのみで推薦が受けられるといったサービス形態はショッピングモールでの商品推薦の実証実験を通して得られたアンケートや利用者の分布によると、広い層の多くのユーザに受け入れられたことが確認できた。また、推薦内容についてもアンケート、閲覧時間、クリック率などの統計より良好な結果が得られた。

しかし、ショッピングモールによる実証実験では個別の推薦について明示的な評価をユーザから得ることができなかった。暗黙的な評価についても閲覧時間は有効な指標であるが、個別の推薦毎の閲覧時間を測定する環境が用意できなかったため測定できなかった。今後はこれらの指標を取得するためのユーザインターフェイスおよびセンサについて調査し、実装した上で推薦精度の評価をとりたい。

第4章 行動支援情報をリアルタイムに提供するシステムアーキテクチャ

本章では2.4に挙げた「システムのリアルタイム性と柔軟性の課題」について主に取り組む。課題を満たすコンテンツ推薦システムを実装し、ショッピングモールでの実証実験を通して得られた知見について述べる。

4.1. リアルタイム性と柔軟性を考慮したアーキテクチャ

P2Pネットワーク上のノードにコンテンツを分散保持させることによって、ネットワーク構成の制約を受けず柔軟に、処理時間に影響するコンテンツ増加に適応したシステム拡張が可能となると期待される。しかし、実空間コンテンツのどのような属性に基づいてノードを分散させるかについての議論は少なく、またこのようなシステムを実際に一般のユーザを対象として検証した例は少ない。したがって、行動支援情報を求めるユーザにとっての有用性は十分に明らかにされていない。

さらに、P2Pネットワークによるアーキテクチャを想定したときにコンテンツの増加や、それらの絞り込みに必要な属性の変化およびそのアルゴリズムの変更といった仕様の変更に耐え、無停止で十分な処理速度を達成できる見通しを得られる実フィールドでの検証結果はこれまでに得られていない。

以降、リアルタイム性と柔軟性に対してそれぞれ検討し、コンテンツ推薦システムを提案した後、第3章と同じく「エルムの街」での実証実験を通して得た結果・考察について述べる。

4.2. 分散エージェントによるコンテンツ推薦システムの提案

本節では、リアルタイム性および柔軟性に対して、予想されるコンテンツの分布を考慮し最適なシステムアーキテクチャを見出す議論をする。

4.2.1 位置に基づくオーバレイネットワークの有効性

2.2 でも述べ、2.4 の要件にも挙げたように、今後の爆発的なコンテンツの増加に対応するには、P2P 型の情報流通基盤が適していると考えられる。オーバレイネットワークはある 1 つの属性に基づいて、ノードを分散配置することによって、ノードのサーバ機能としての接続負荷を分散する特徴を持つ。従って行動支援情報を提供するという目的に合わせて、最も広い範囲で分散可能な属性を見出し、適切なオーバレイネットワークを選択しなければならない。

行動支援情報は第 1 章でも述べたように、「実空間との対応がとられた情報を提供すること」が重要である。また、同章の商品情報を取得の例で示したように、時間、位置、目的、ユーザの嗜好、コンテンツ提供者の要望といった属性がコンテンツ側に付随することが想定される。また、その他にも価格、対象年齢、対象性別、対象職業、商品のジャンルなどの属性が考えられる。どのようなコンテンツにおいても扱うことのできる広い範囲に分散する可能性のある実空間情報の属性はこのうち時間と位置が期待できる。時間については秒単位で記録し、時間に基づいたオーバレイネットワークが構築できれば、ノードを分散できるが、実空間でユーザが時間情報を扱う場合は時刻が重要であっても日付が重要でないことがある。ある一定の周期性で情報が集約される可能性があるからである。例えば、現在地周辺の飲食店を検索する場合、コンテンツ中で注目される時間情報は開店時間であり、日付は無視される。

したがって、行動支援情報は第一に地理位置に基づいたオーバレイネットワークを活用することによって、膨大なインターネット上の情報から、最も効率よくフィルタリングされると期待できる。その他の時間や嗜好の一致具合についてはノードに対して個別に問い合わせると、よりの確なフィルタリング結果をユーザ

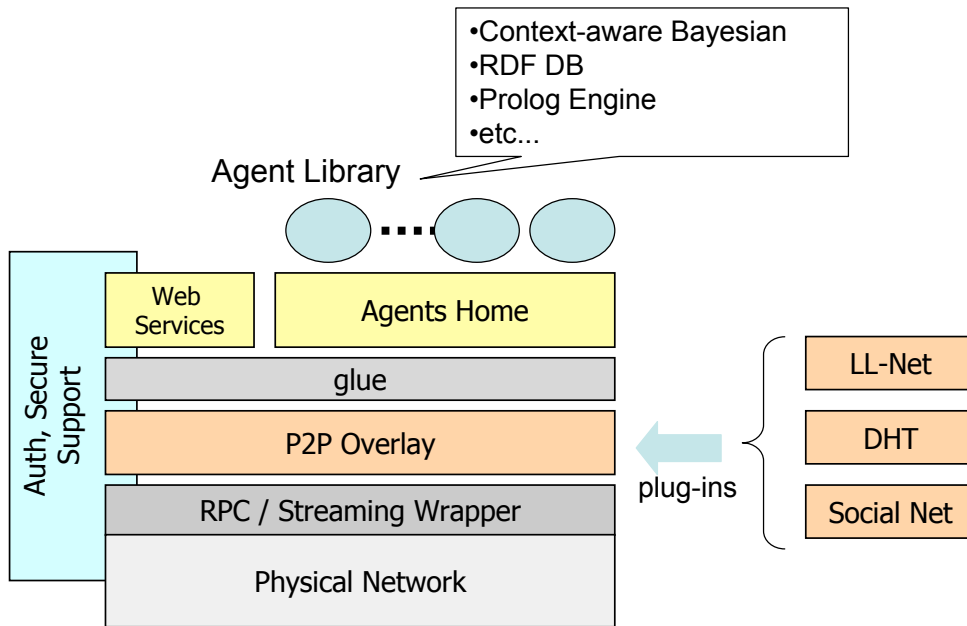


図 4.1 PIAX のアーキテクチャ

が最終的に得ることができる。このようなその他の属性のマッチング演算における負荷分散の考慮は次節の柔軟性の考慮とともに述べる。

4.2.2 柔軟性を考慮した、分散エージェントによる機能構成

2.4 の要件ではさらに P2P 環境での推薦アルゴリズムの柔軟な選択・更新が可能である必要があることを述べた。そして、前節で述べたように位置以外のその他の属性の演算について負荷を分散する必要がある。前者の柔軟性の解決には、オーバーレイネットワークとその上で稼働するアプリケーション（複数属性の演算をこのアプリケーション内で処理）を独立して構築できる環境が必要となる。これらを独立して構築することによって、複数の推薦アルゴリズムを同一オーバーレイネットワークに実装することが可能となり、またアプリケーションの変更が必要な場合もオーバーレイネットワーク自体は停止・更新する必要がなくなる。後者の負荷分散の解決については、先程述べたノード上のアプリケーションをコンテン

ツとその演算機能によって構成することによって、コンテンツを保持する各ノード上に演算処理を分散させることができる。

2.2で紹介した分散エージェントプラットフォーム PIAX はこの機能を満たすことから本研究の目的に合致し活用すべきプラットフォームであると考えられる。PIAX のアーキテクチャを図 4.1 に示す。IP による任意のネットワーク上に選択可能なオーバーレイネットワークを構築し、その上でエージェントが稼働する。現時点で PIAX は LL-Net と DHT, Social Net の 3 つのオーバーレイネットワークを実装している。PIAX におけるエージェントは Aglets[71] を基盤とした特殊な Java クラスとして実装される。エージェントの所在に関係なく、同期・非同期でのメッセージング機能を提供するとともに、マルチキャスト非同期型のメッセージングによる特定のリソースを有するエージェントを発見する手段を提供している。これらのメッセージング機能を用いることで、状況に応じたエージェントを発見し、連携させ、行動支援情報を持つノードを探索することができる。LL-Net によって位置の絞り込みを行った上で、そのノード上のエージェントが保持するコンテンツの残りの他の属性を各ノード上で演算することによって、全ての属性における分散演算を実現できる。

4.3. 分散エージェントによるコンテンツ推薦システムの設計

コンテンツ推薦サービスの概要を図 4.2 に示す。提案するアーキテクチャを実現するプラットフォームは前節の議論に基づいて PIAX を活用する。他のシステムとの連携を考慮して Web サービスとして稼働し、外部システムからリクエストの受付と、コンテンツ探索結果の応答を行う。Web サービスの内部は P2P ネットワーク上でエージェントを分散配置している。前節でも議論したように行動支援情報は地理位置に基づいて分散されることが期待されることから、LL-Net をオーバーレイネットワークに適用した。

SearchAgent はユーザに代わってコンテンツを検索するエージェント、ItemAgent は 1 コンテンツをオブジェクト化したもので検索クエリがコンテンツと適合する

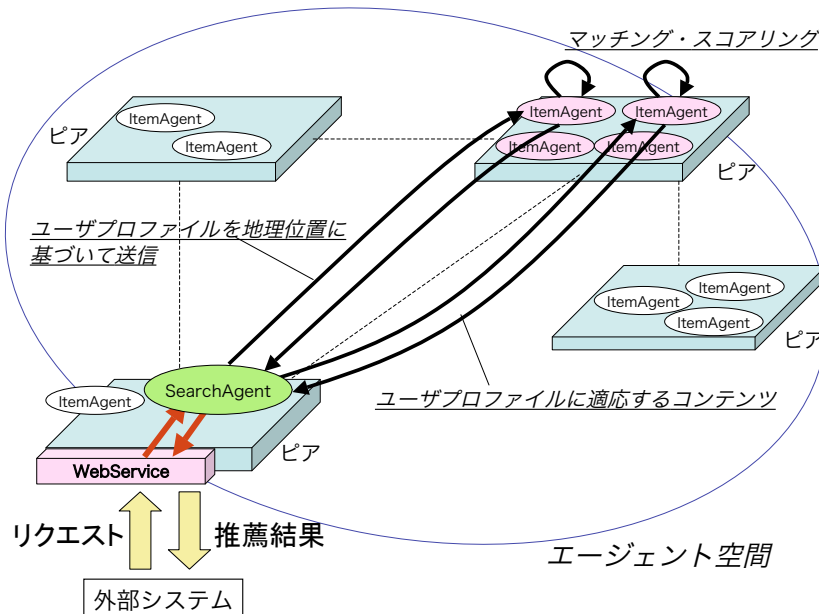


図 4.2 コンテンツ推薦サービスの概要

か否かを判断する。よって、コンテンツ提供者がコンテンツを1つ追加する度に、オーバーレイネットワーク上にItemAgentが1つ増える。

推薦リクエストを受けたノード内の動作を図 4.3 に示す。図中のスコア算出部分は第3章にて提案した図 3.4 と同様である。推薦リクエストが発生すると、SearchAgentはWebサービスのリクエストによりユーザプロフィールを受け取ると同時に、ユーザの状況を示す時間（現在時間）・場所（現在地）といった状況情報をアクセス時間やアクセス端末の設置位置によってコンテキストを取得し、リアルタイムに決定する。LL-Netを通して現在地周辺のItemAgentに対してユーザプロフィールを送信し、ItemAgentがマッチングおよびスコアリング処理を行い、その結果をSearchAgentに返し、Webサービスの返答に伴い推薦は終了する。

コンテンツの増加に応じてItemAgentの数を増やし、検索作業を分散処理することによってコンテンツ増加に対するスケーラビリティを確保する。推薦アルゴリズムは個々のItemAgentが保持し、要求時に自律的に実行する。これにより、

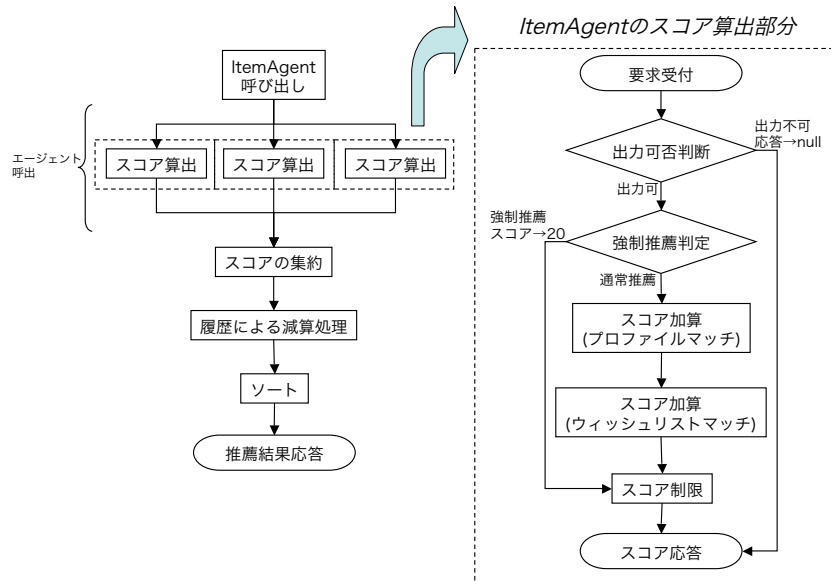


図 4.3 コンテンツ推薦アルゴリズム

コンテンツごとに異なる推薦アルゴリズムを指定できる。また、運用主の要望やコンテンツの傾向などにより推薦アルゴリズムを変更したい場合、新しいアルゴリズムを実装した ItemAgent を起動しなせばよい。

4.4. ショッピングモールでの実証実験による評価と考察

提案・設計したシステムが課題とするリアルタイム性と柔軟性を解決するか否かを、ショッピングモールでの商品推薦による実証実験によって確認する。実証実験は第 3 章で行ったものと同じである。

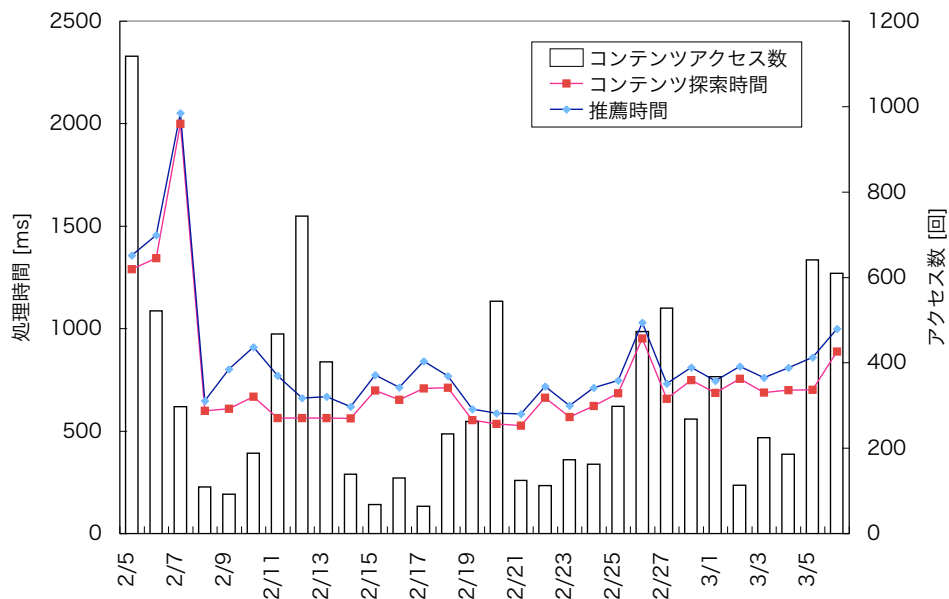


図 4.4 平均処理時間およびコンテンツアクセス数

4.4.1 柔軟性の考察

本実証実験を開始する前日に動作チェックをしている際に、コンテンツ推薦の内容が予想する内容とは異なり、推薦アルゴリズムおよびチューニング用のパラメータを変更する必要があった。稼働中のコンテンツ推薦システム上で、新しく実装した ItemAgent を以前のものと入れ替えるのみで、他の連携要素に影響することなくシステムの仕様変更ができた。従来のコンテンツ推薦システムでは仕様の変更があった場合にシステム全体を再起動する必要があったが、本実証実験システムでは推薦アルゴリズムが PIAX 内で完結するためシステムの再起動は必要ない。したがって、本システムは仕様変更に対して柔軟な設計がなされたといえる。

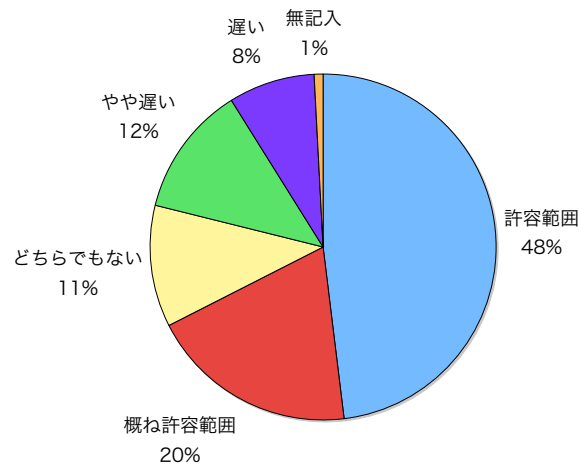


図 4.5 処理時間についてのアンケート結果

4.4.2 処理時間の考察

日ごとのリクエストの平均処理時間とコンテンツアクセス数の推移を図 4.4 に示す。なお、他の連携要素との兼ね合いでシステム過負荷が起こり返答できなかった、あるいは基幹スイッチの不調でネットワークが通じないといった理由で失敗したリクエストは集計対象から省いている。リクエストの受付から応答までの時間を推薦時間、エージェントによりマッチング・スコアリングなどの推薦処理を行っている時間をコンテンツ探索処理時間とした。推薦時間はプロフィール取得時間、ウィッシュリスト取得処理時間、コンテンツ探索推薦処理時間の和となるが全体の 80%~90%がコンテンツ探索推薦処理時間である。2/6 までの過渡的なネットワーク不調が生じていた運用初期以外は、概ね 1 秒以内に処理が完了しており、安定した時間で処理が行えている。コンテンツ数が実験開始時に 474 コンテンツ、実験終了時に 880 コンテンツとほぼ 2 倍の増加を持っていることに対して、平均処理時間は緩やかな増加があるが概ね 1 秒以内であった。また、コンテンツアクセス数が 2/12、2/20 において前後の日より大幅に増加しているが、平

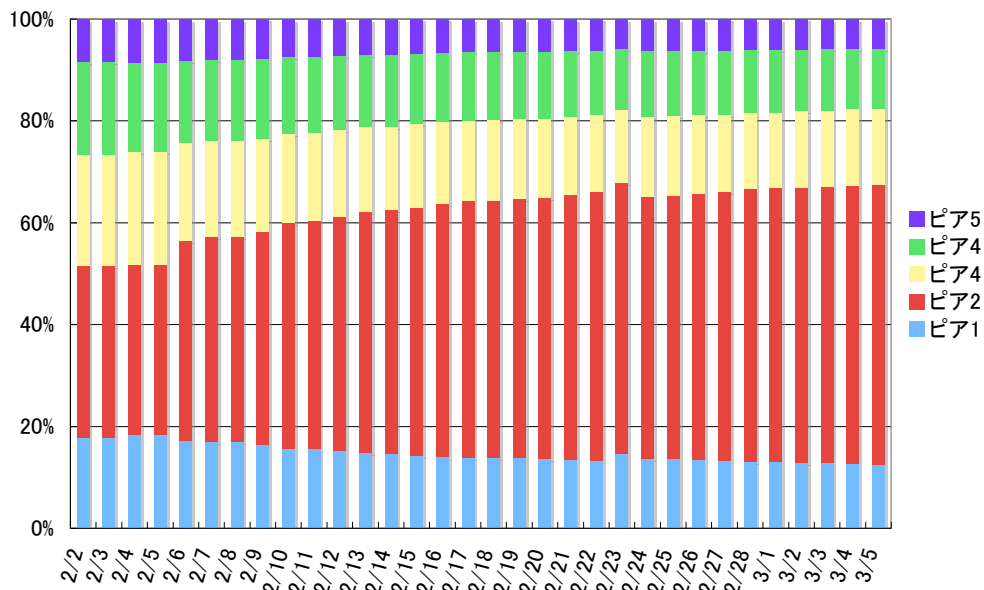


図 4.6 各ピアが保持するコンテンツ割合の推移

均処理時間には同様の特徴が見られないことからコンテンツアクセス数に影響を受けていないことが分かる。これに加えて、利用者が体感した処理時間に対するアンケート結果（図 4.5）では、半数のユーザが許容範囲内と答え、概ね許容範囲内という回答を含めると7割のユーザが処理時間に対して満足していることから、コンテンツ推薦システムにおけるリアルタイム性の要求を満たせたといえる。

4.4.3 コンテンツ分布の考察

平均値には現れないがコンテンツの偏りの影響で局所的に問題が起きたことがあった。図 4.6 に各ピアが保持していたコンテンツ割合の日ごとの推移を示す。このピアが担当するエリアには飲食店エリアが含まれていて、店舗の密度が高かったこともあり、実証実験開始時にすでにコンテンツ量の割合が大きかったが、日を経つにつれさらに割合が増加した。また、図 4.7 に日ごとのコンテンツの分散をグラフで示す。コンテンツ量が偏っているピア2を除外すると、分散に変化が少なくなることが分かる。実環境において位置に基づくオーバーレイネットワークを応用す

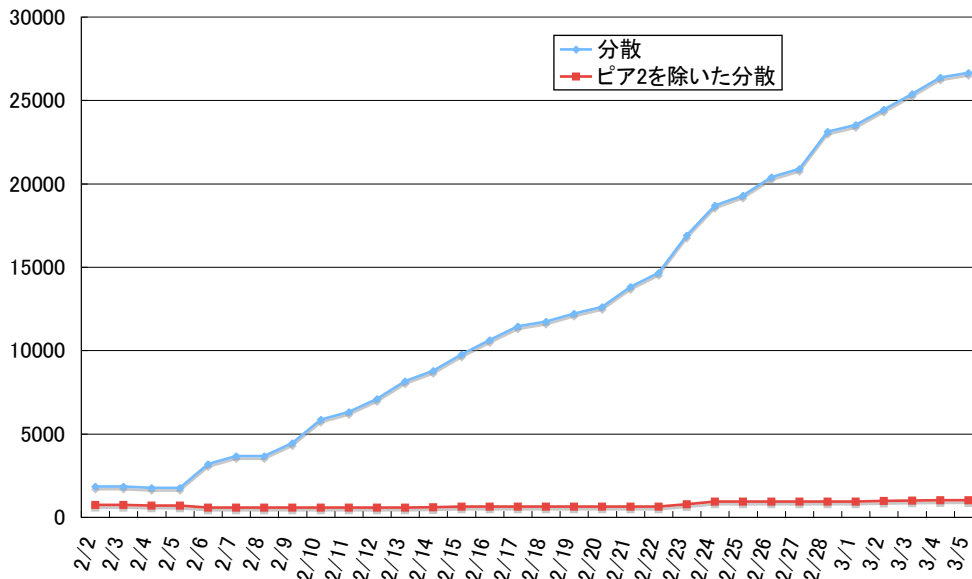


図 4.7 各ピアが保持するコンテンツ量の分散の推移

る際には、事前に実環境のコンテンツ濃度の予測を入念に行わなければ目論見通りの負荷分散は期待できない恐れがあることが明らかになった。

また、本システムでは、いつ返答が得られるか分からない分散環境において一定数の返答を得るために、最初の返答があった後、一定の返答待ち時間を設ける方法をとっていた。実験中に一部のピア内に大量のコンテンツが登録され、そのピアに対してリクエストが送信された場合に、非同期通信を開始して最初の返答に時間がかかってしまったことがあった。これはコンテンツの分布が地理的に一様であると想定してシステム設計をしたことが原因として起こった問題で、実証実験を通じて得られた経験となった。非常に多くの返答が来る場合には例えば先着 20 件の返答を受け取って、後に続く返答については処理を打ち切らせることにより無駄な待機時間を縮小する。あるいは、低スコアであることをスコアリング時に判断した場合に返答をしないことにより無駄な返答数を削減することが今後の課題となる。

4.5. まとめと今後の課題

本章ではシステムのリアルタイム性と柔軟性の課題について取り組んだ。リアルタイム性の解決のために、行動支援情報において最も分散する可能性のある属性が位置であることを見出し、LL-Net を活用したシステム設計を行った。柔軟性の解決のために、位置以外の属性の処理を柔軟に組み替え可能にするために、P2P ノード上のアプリケーションをエージェントにより構成することを提案し、最も適切なプラットフォームとして PIAX を活用した。提案に基づいて設計したコンテンツ推薦システムを実証実験を通して評価した結果、全期間中にわたって概ね1秒以内の推薦応答を確認し、ユーザからのアンケートからも推薦処理時間に満足を得ることができた。しかし、1ヶ月間という固定された期間およびコンテンツの手入力の必要性がある環境では、実際の Web 情報量と比較すると明らかにコンテンツが不足し、設計したコンテンツ推薦システムのコンテンツ増加における演算負荷のスケーラビリティに対する有効性が満足に証明できなかった。実証実験を通して得られた、実ユーザおよび実コンテンツの分布を元に大規模なシミュレーション環境を構築し、著者の研究を含め、他の P2P 型の情報流通基盤に取り組む研究に貢献したい。

第5章 個人サイトからのメタデータ 抽出モデルの提案

本章では最後の課題である、「正確なコンテンツメタデータの維持性の課題」について取り組む。行動支援情報源として価値が高いと期待される個人サイトを構造化する手法を提案し、有効性を検証する。

5.1. 個人サイトに含まれる評価情報

2.3 では個人サイトにはユーザの行動支援情報となる評価情報が含まれていると述べたが、具体的に商用サイトと比べてどの程度の違いがあるか調査する。

情報検索には Google を活用し、飲食店検索を想定して料理名を主キーワードにして検索した。

検索語：「ラーメン 住所 -ぐるナビ -Yahoo!グルメ -グルメウォーカー -all about -MSN グルメ -livedoor グルメ -ラーメンバンク -タウン -NAVITIME」

飲食店の店舗情報が検索結果に含まれやすいように、「住所」というキーワードを追加した。これだけでは検索結果のほとんどが商用サイトとなってしまったため、著名なポータルサイトを除外キーワードに追加した。ラーメンの他にも焼き鳥、うどん、パスタ、すき焼き、餃子、カレー、焼肉、ケーキの8つのキーワードについて調査した。このキーワードの選定は Yahoo!グルメに記載されている料理ジャンルの主要な料理名となっている。

各々のキーワードにおいて、個人サイト・商用サイトのそれぞれ上位10件の詳細を集計した結果を表5.1に示す。飲食店情報とは文書中にその料理を提供する飲食店の店名、住所などといった、飲食店を特定しユーザが実店舗にたどり着くための情報が記述されている情報である。評価情報はその店舗の料理について多店

表 5.1 収集された Web サイトの記述内容

検索語	飲食店情報		評価情報	
	個人サイト	商用サイト	個人サイト	商用サイト
ラーメン	10	10	10	2
焼き鳥	10	10	6	0
うどん	10	10	8	1
パスタ	10	9	6	1
すき焼き	8	10	5	2
餃子	10	10	8	3
カレー	10	10	9	2
焼肉	10	10	6	1
ケーキ	10	10	7	0
平均値	9.8	9.9	7.2	1.3

舗との違いを明確に表記しているもの（★の数、数値の大小（4.5 点，8.0 点），言葉の強弱（普通，旨い，激旨）など）である。

飲食店について記述されている Web ページの 10 件中の平均数は個人サイトが 9.8 件，商用サイトが 9.9 件となり，同程度の割合である。「すき焼き」というキーワードによって誤って収集された個人サイトの 2 件はすき焼き用の牛肉といった食材について説明されている Web ページであった。

評価情報が記述されている Web ページの収集数については個人サイトが平均 7.2 件となり，商用サイトについては収集数が平均 1.3 件となった。これより個人サイトでの飲食店・料理に関する文書にはユーザの行動支援情報の元となる情報が商用サイトよりも多く含まれていることが確認できた。

しかし，2.3 でも述べたように，商用サイトはある程度の構造化がされているため，メタデータマッチングへの適用性は高いが，個人サイトはコンテンツ作成時点で行動支援情報として活用される意識が乏しいためほとんど構造化されていない。今後，行動支援情報の重要性が認知され，個人のコンテンツ提供者がブログな

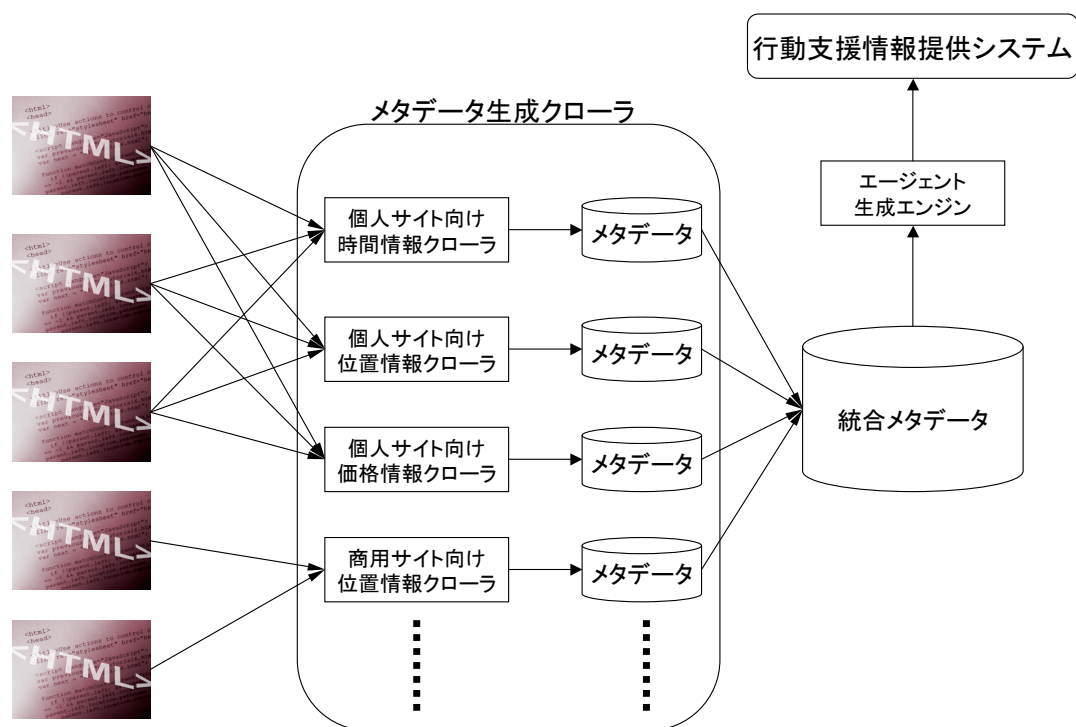


図 5.1 メタデータ抽出モデル

どを作成した時点でコンテンツメタデータを入力してくれれば上記の問題は解決されるが、第 3 章での実証実験からの知見の通り、手作業によるコンテンツメタデータの付加はコンテンツ提供者にとって大きな負担となり、コンテンツ量の増加は全体的に期待できず、サービス不活性の要因となる。

以降、Web ページからのメタデータ抽出モデルを設計し、性能の要であるメタデータ生成クローラを実装・評価することによってメタデータ抽出可能性を検証する。

5.2. Web ページからのメタデータ抽出モデル

提案モデルの全体像を図 5.1 に示す。ブログを収集・解析し最終的にマッチング機能をもつエージェントを生成することで、行動支援情報提供システムの情報

生成基盤として機能することが期待できる。

メタデータ生成クローラは生成するメタデータの性質に合わせて個別に実装される。

例えば、個人サイト向け位置情報クローラは、個人サイトの文章を解析し位置を記述する部分を判定し、その住所文字列をジオコードによって緯度・経度に変換し、メタデータリポジトリにそのコンテンツの URL と緯度・経度、その他加工に必要な情報を持つメタデータを保存する。その他のメタデータ生成クローラによって、生成されたメタデータを統合し、1つの個人サイトに対して位置や時間、その他の属性を記述するメタデータがコンテンツの統合メタデータリポジトリに保存される。最後にその統合メタデータを元にメタデータマッチングエージェントを生成する。

本章ではブログ等の個人サイトを対象とするクローラを実装するが、商用サイトにおいても情報収集 API が用意されていない等の理由で、マッシュアップサイトが独自に商用サイトの情報を構造化する場合も想定されるため、メタデータ生成クローラの情報源は個人サイトに限らない。

このモデルにおいて個人サイトの収集については従来の IP ベースの Web 巡回技術の利用または RSS による更新確認をすることによって収集可能である。メタデータの統合についても、メタデータを XML によって記述した場合には簡易な XSLT の準備によって実現できる。エージェントの生成についても第 4 章と同様の手法で実現ができる。したがって、このモデルの善し悪しを決定する要素はメタデータ生成クローラの解析性能に依存する。

以降、行動支援情報の状況情報として最も重要としてきた時間情報と位置情報についてメタデータ生成クローラの主要素を実装し、実際に個人サイトからどれだけ信頼できるメタデータが生成できるかを検証する。

5.3. Geocrawler の設計と評価

本節では個人サイトに記述された住所情報を抽出する手法について取り組む。住所はジオコードにより緯度・経度に変換可能なため、個人サイトから住所情報

を抽出することが可能ならば、そのサイトを位置座標によってインデックス化可能となる。

5.3.1 Web からの位置座標抽出の現状

Google Map[72] などの地図表示 Web アプリケーションを基盤にし、その地図上にユーザが任意にコンテンツを追加することでコンテンツの位置情報のタグ付けを可能にする手法が増えている。しかし、Google Maps を自身のブログ上に組み込むことの敷居の高さや全てのコンテンツに対して位置情報を付加する意義がコンテンツ提供者に周知されていないことなどから、未だに多くの Web 文書とその中に含まれている位置記述の構造化が実現できていない。前節の時間情報と同じく、コンテンツ提供者が意識せずに位置情報を自動的に抽出することができることが望ましい。

既存の手法に自然言語処理による住所文字列の解析手法 [29] があり、適合度も良好であるが、市町村レベルの粒度にとどまり、番地・号の粒度は実現できていない。外出時にユーザの徒歩圏内の情報を提示するといったことを想定した場合、市町村レベルでは粒度が荒くなってしまうため、より細かい粒度での位置の抽出が必要となる。上記の既存手法を拡張し、粒度を細かくする場合には機械学習の対象を大幅に広げることになり、サービス構築コストが高くなる恐れがあり、また適合度の低下が懸念される。

そこで、本節では形態素解析結果をルールベースのパターンマッチングによって番地・号までを取得する手法を提案する。さらに抽出した住所文字列をジオコードによって緯度・経度に変換することにより GEO タグ [73] を付与する

5.3.2 形態素解析結果を活用した位置座標抽出

住所情報を抽出する機能を実現するために、形態素解析ツールの一つである茶筌 [8] を利用する。解析を行う前に HTML ファイル内のタグは完全に除去する。HTML ファイル内のタグ構造 (`<td>奈良県奈良市三碓 2-1-10 </td>` など) を利用することによって、文書構造の解釈を補うことができる可能性はあるが、今

表 5.2 茶笥を使った解析例

形態素	分類
全国	名詞-一般
店	名詞-接尾-一般
)	名詞-サ変接続
八	名詞-数
ちゃん	名詞-接尾-人名
ラーメン	名詞-一般
住所	名詞-一般
福岡	名詞-固有名詞-地域-一般
市	名詞-接尾-地域
中央	名詞-固有名詞-地域-一般
区	名詞-接尾-地域
白金	名詞-固有名詞-地域-一般
1-1-27	名詞-サ変接続
うどん	名詞-一般
そば	名詞-一般
TEL	名詞-サ変接続
092-521-1834	名詞-サ変接続

回は個人サイトの HTML ファイルを住所抽出対象にしている。これら個人サイトの HTML ファイルは、商用検索サイトのものと違い、作成者のスキル・価値観によって構造が様々であり、タグの使用方法、使用箇所も様々となるため、タグ構造を利用しないこととした。

HTML からテキストに変換したファイルに対し、形態素解析した結果の例を表 5.2 に示す。本研究では、形態素解析により「地域」に分類された形態素を住所の主要部分とする。まず予備実験を行った。HTML ファイル収集コンポーネントで収集した 178 ページの HTML ファイルをテキストファイルに変換し、形態素解析を行った結果を目視により確認した。その結果、「地域」形態素が 3 つ連続して

出現した時点でその周辺の文字列は住所である可能性が高いことを確認した。「地域」形態素が3つ出現した後は「地域」形態素が続く限り形態素を抽出する。残りの丁目・番地等を抽出するためにさらに余分に5つの形態素を抽出した後、住所記述のパターンを網羅する正規表現リストとのパターンマッチングを行い、住所文字列を抽出する。

次に抽出した文字列をジオコードによって緯度・経度に変換し、メタデータを生成する。ジオコードには Yahoo! Maps[74] を活用する。Yahoo! Maps に住所を入力し検索を行うと、検索結果の URL 内に入力した住所に対応する緯度・経度が含まれている。検索結果ページの URL 部分の緯度・経度部分を自動抽出するスクリプトを作成することによって、緯度・経度変換を実現する。入力した住所が Yahoo! Maps の住所データベース内のデータと一致する場合は緯度・経度変換作業を行い、正しい緯度・経度を含む HTML ファイルを返す。しかし、入力した住所文字列が長すぎたり、短すぎたりする場合は、Yahoo! Maps が自動的に近似 (最長一致) の住所に対して緯度・経度変換される。したがって、この変換の精度は住所抽出の精度に依存している。

5.3.3 抽出精度の評価・考察

前節で提案した住所文字列抽出の精度を評価する。

実験した時期の都合上、解析対象となる Web ページ群は 5.1 で収集したラーメンに関する検索結果である。検索結果に含まれる個人サイトの HTML ファイル 178 件を目視で住所確認し、正しい住所数を調査する。正しい住所の定義は、収集した個人サイトの HTML ファイル内に記述された住所が番地以降 (「生駒市高山町 8916-5」, 「三碓 1-5-10」 など) まで記述された住所である。正しい住所の定義に従って、収集した個人サイトの HTML ファイルを目視した結果、769 個の正しい住所を確認できた。提案手法によって抽出された住所は 775 個で、そのうち正しい住所は 457 個であった。再現率および適合率を以下のように定義し、評価した結果、再現率は 59.0%、適合率は 59.0% になった。

$$\text{再現率} = \frac{A \cap B}{A} \quad (5.1)$$

$$\text{適合率} = \frac{A \cap B}{B} \quad (5.2)$$

A : 目視により確認した正しい住所数

B : 住所抽出コンポーネントで抽出した住所数

住所抽出に失敗した主な要因には以下の4つが挙げられる。

- 住所以外を抽出した例 (95 件, 49%)

Web 文書中の住所の番地以降の記述と、電話番号の記述が連続するような状況などで、正規表現による文字パターンマッチングの照合箇所を間違えていることがあった。これは形態素解析によって解決できない問題のため、今後は数値文字列に対する意味づけを解決する手法について検討する必要がある。

- 大字、字、条が含まれた抽出例 (40 件, 20%)

大字等の単語が「地域」形態素として分類されないことがあった。大字は住所で使われることが大半のため、茶笥の辞書に「地域」として新たに登録し、隣接单語の重み付けを調整することによって改善が図れる。

- 茶笥の辞書 (20 件, 10%)

「さいたま市」など、茶笥の辞書に存在しない地域名がいくらかあった。これについても、総務省の都市開発地域データベース等を参照することによって、新住所キーワードを取得することができるため、茶笥に対して新たな辞書登録を行い、隣接单語の重み付けを調整することによって改善が期待できる。

- その他 (41 件, 21%)

主に目立ったのは HTML ファイルのタグ除去作業に失敗していることであった。改行コードを含まない Web 文書などでタグを除去することによって、想定しない形態素が生まれてしまった場合に間違いの原因となる。タグ除去後に、タグ内のキーワードの末尾に改行を付加することによって改善が期待できる。

以上の実験から既存手法に必要であった比較用の辞書を生成することなく、ある程度の再現率を確保した住所抽出手法が実現できたことを確認した。比較用の丁目まで記載された住所を登録した辞書は既存研究の実験当時で約 95 万件の事前登録が必要であった。本実験で明らかになったように最新の茶釜用の辞書を用いてもさいたま市などの地域が登録されていなかったことから Web サービスを構築・運用するにあたっては細かな辞書管理が必要となる。さいたま市には 10 の区があり、それに続く町字・丁目の組み合わせは膨大であることから既存手法のサービス構築コストは提案手法に比べて膨大であることが容易に想像できる。さらに番地・号の辞書登録をすることは困難であるため本提案は既存手法と比較してより低コストで実サービス化が可能な手法であるといえる。

位置情報の抽出では番地・号のパターンマッチングに失敗するものが多くを占めた。これは提案手法が形態素解析出力の地域要素の活用を主提案としていたため、番地・号のパターンマッチングへの高精度化の配慮がかけていたことが原因である。SVM (Support Vector Machine) などの学習型フィルタを適用することによって抽出する数字列が番地・号か電話番号かを分類することができれば、適合率および再現率の向上が期待できる。

5.4. Time-crawler の設計と評価

ユーザが屋外で望む情報を簡易に取得するには位置と時間をキーにして情報検索できることが必要である。位置情報の検索についてはいくらか実サービスが展開されているが時間をベースに検索をするシステムは未だにない。本節ではプロ

グ中に写真がよく写真が掲載されていることに注目して、その写真の Exif 情報に含まれている撮影時間を収集する Time-crawler を提案する。

5.4.1 イベント生起時間の重要性

PDA や携帯電話の発達によって屋外で情報検索をすることが増えた。このような状況下では現在地周辺のレストラン検索や出張先の天気予報の検索など、位置と時間による情報検索の実現が必要となる。つまり、Web 上のコンテンツを位置と時間に基づいてインデキシングすることが課題となる。位置によるコンテンツの構造化には、コンテンツ中に緯度や経度などのメタデータを付加させる手法が有効となる。ユーザは GPS などによって現在地を知り、その位置情報との距離を計算することによってより行きやすい場所をネットワーク上から見出すことができる。最近では Google Maps API などの公開が積極的に行われることによってブログ上に地図を埋め込んで簡単に位置情報を登録できるようになり、たくさんの Web コンテンツが有効利用されつつある。

しかし、時間情報についてはまだ実用例が限られている。位置で情報を絞り込んでも、お昼ご飯を食べたいときにワインバーを紹介されるなど、その時間にそぐわない情報がでるとサービスの有効性は低下する。RSS の自動発行などの工夫がなされ最新のコンテンツを提供する仕組みは成功している。

しかし、屋外で昼食を提供するレストランを検索する場合に必要な時間情報は、記事生成時間ではなくイベント生起時間である。例えば誰かがあるレストランで昼食をとり、その料理に非常に満足した場合にその記事を作成するのは帰宅後の夜となる。RSS を利用してしまった場合、このコンテンツは夜のコンテンツとなってしまう無駄な情報となってしまう。コンテンツ中に含まれる時間を構造化して記述する手段に TimeML[75] といったフレームワークがある。しかし、どのようにこの情報の自動生成手法は確立されていない。手作業によってコンテンツ提供者が時間情報を記述することは現実的ではない。現在、自然言語処理技術によって Web 文書中の時間情報を抽出するアプローチが盛んに研究されている。ニュース記事といったイベント生起時間が明記されているものを対象とした場合に最もよい精度の報告でも適合率が 80 %、再現率が 80%、F-値が 80%となっている [31]。

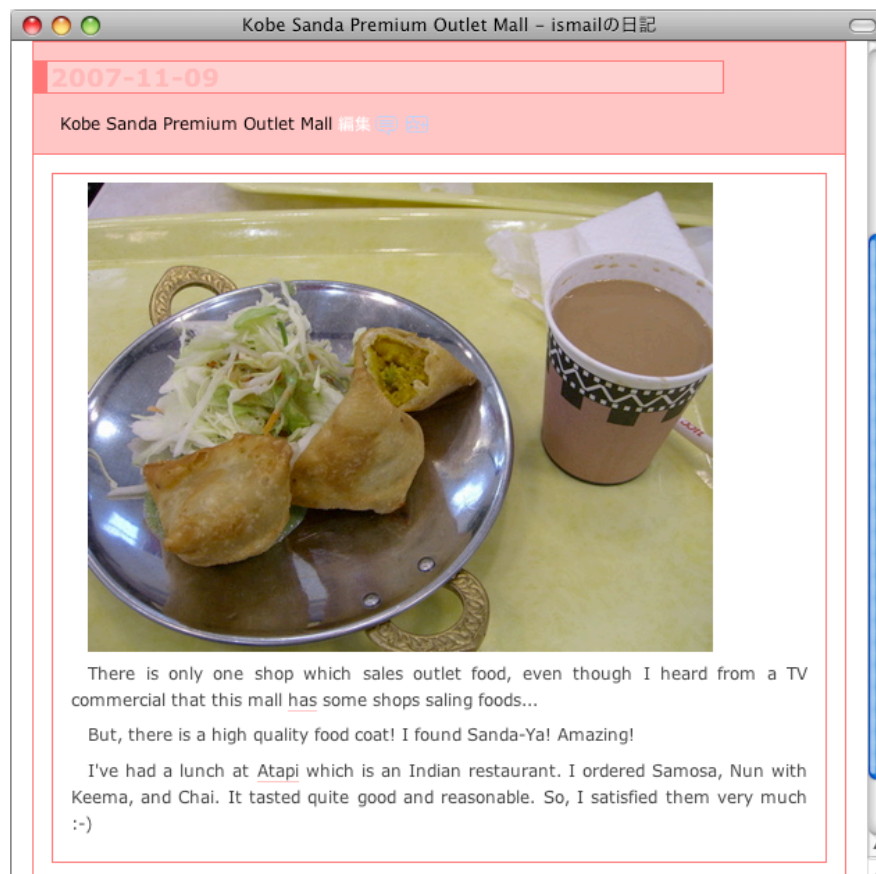


図 5.2 ブログの例

ブログを対象としたものでは時刻ではなく時間帯の判定に留まっている [33, 34]. これはブログ記事中にそもそも時刻が判断できる記述がないことが多く、人間が読んでもイベント生起時刻が見出せないからである. したがって, 自然言語処理とは異なるアプローチによる, ブログからのイベント生起時刻抽出手法を検討する必要がある.

5.4.2 Exifに着目したイベント生起時間抽出

ブログにはよくその日に行った場所や食べたものについて写真とともに文書を記述することが多い. 図 5.2 に示すように, 最初に写真を貼り, その後に文章を

表 5.3 Exif 情報の例

項目名	値
カメラ名	EX-Z750
撮影時刻	2007/11/04 15:00:20
F 値	2.8
シャッタースピード	1/125
画像サイズ	450x338

記述する．ほとんどの場合は写真と文書の内容が一致する可能性が高い．

この画像は個人のデジタルカメラによって撮影される．この画像には Exif 情報が含まれている可能性が高い．Exif 情報の例を表 5.3 に示す．撮影時の情報が事細かに記録されていて，専門的な知識のある人はこれらを活用して最適な加工をしたり，より美しく印刷したりする際にこれらの値を活用する．本研究の目的に合わせて注目するのは撮影時間である．これには抽出目的であるイベント生起時間が秒単位で記録されている．写真と文書の内容が一致すれば，抽出したイベント生起時間は秒単位の正確さをもつため，自然言語処理によるどの既存技術よりも精度が高くなることが期待できる．そこで本研究ではブログに含まれるイベント生起時間を Exif 情報を参照することによって抽出する Time-crawler を提案する．処理の流れは以下の通りとなる．

1. ブログを記事単位で収集する
2. 記事中に写真があることを判定する
3. 写真がある場合には Exif 情報が含まれているかを判定する．
4. Exif 情報から抽出されたイベント生起時間をメタデータに記述する
5. 生起時間と記事の URL を対として持つデータベースを構築する

インデキシングされた時間情報を元に情報検索サービスを構築することによって，よりの確にユーザの要求に応えることが期待できる．

表 5.4 ブログ中に掲載される写真の内訳

	Yahoo!ブログ	アメーバブログ	楽天ブログ	計
撮影時間をもつ写真あり	35	34	22	91
その他の写真・画像あり	55	57	33	145
画像なし	110	109	145	364
撮影時間をもつブログの割合	17.5%	17.0%	11.0%	15.2%

表 5.5 写真と本文の内容の関連

	関連あり	関連なし	写真と本文に関連がある割合
撮影時間あり	78	13	85.7%
撮影時間なし	109	36	75.2%

5.4.3 抽出精度の評価・考察

提案システムの有効性はブログ上に前提通りの写真と文書があるかどうかによって依存している。まずブログを収集し、写真と本文の関係に基づいて分類調査する。その後、適合率の評価を行い、本提案の有効性を検証する。

ブログの分類調査

無作為に集めたブログを観察して、抽出対象となる exif の出現割合と仮説の有効性を確認する。無作為に情報を集めるために、Yahoo!ブログ [76]、アメーバブログ [77]、楽天ブログ [78] の各サイトの新着ブログ 200 件を収集した。記事中に含まれる写真を分析し、ブログを以下の 3 つに分類した。以降、画像ファイルを総じて画像と呼び、そのうち Exif 情報を含む Jpeg ファイルのことを写真と呼ぶ。

- 撮影時間を含んだ写真が掲載されたブログ

ブログが掲載されている画像がデジタルカメラで撮影された Jpeg ファイルの場合、撮影時に記録された Exif 情報が埋め込まれている可能性が高い。そ

の Exif 情報の中に、撮影時刻が記録されている場合は本研究で目的とするイベントの生起時間として再利用する価値がある。

- その他の画像が掲載されたブログ

写真に Exif 情報が含まれている場合でも、ブログに掲載されるまでの過程で加工が施され撮影時間を含まない場合がある。また、イラストやグラフなどの画像も本研究の目的にそぐわないデータとなるが、これには撮影時間が記録されていることはないので収集対象にはならない。

- 写真が掲載されていないブログ

文章によって十分に情報が伝わる場合には写真は掲載しない場合が多い。

表 5.4 に収集したブログの内訳を示す。平均して約 15%と低い割合のブログ記事が撮影時間をもつ写真を持つことが分かった。

しかし、写真と本文の関連性を、本文を読んだ後に写真の内容と一致するか否かを著者の目視によって分類したところ、表 5.5 に示すように画像と本文の関連は、撮影時間を持つ写真の場合には 85.7%と高い結果となった。従って、撮影時間を含む写真が掲載されたブログを提案手法によってインデックス化すると高い適合率が期待できる。

グルメ検索シナリオに適合率の評価

行動支援情報取得の実シナリオとして、グルメ検索を想定して適合率を評価した。Google ブログ検索にて 9 個のキーワードを入力した場合の検索結果においての適合率の評価を行った。Google ランクによる検索結果は変化が少ないため、更新日順によってスコアリングを検索オプションにした上で、上位 100 件の結果を収集した。9 個のキーワードは一般的なグルメポータルサイトである Yahoo!グルメ [63] にて最初にカテゴライズされる料理のジャンルの数に合わせた。写真と文書の関連を判断しやすいように各ジャンルで代表的な料理名をキーワードに全文検索を行った。

表 5.6 グルメキーワード毎の適合率

キーワード	正解数	不正解数	適合率
ラーメン	20	2	0.91
焼き鳥	7	6	0.54
うどん	11	6	0.65
パスタ	24	6	0.80
すき焼き	11	6	0.65
餃子	12	4	0.75
カレー	11	6	0.65
焼肉	11	2	0.85
ケーキ	29	4	0.88
計	109	42	0.72

提案手法によって収集したブログから抽出できた時刻情報と、そのブログ中の文書の内容が一致しているか否かを目視によって判断し、正解と不正解に振り分けた。評価結果を表 5.6 に示す。適合率の平均は 0.72 となった。「ラーメン」、「焼肉」、「ケーキ」といったキーワードは適合率が高く、「うどん」、「すき焼き」、「カフェオレ」、「カレー」では適合率が低い結果となった。料理を作ったことを報告する記事よりも、外食の報告では適合率が高い傾向となった。本実験では検索を全文検索により簡易化したため、同じ料理において調理報告か外食報告かの区別がされていないため、各キーワードにおける適合率の分布にばらつきが生じたと考えられる。

失敗の原因を分析した結果、以下のように分類された。

- 写真と本文の内容が異なる (30 件, 68%)

事前調査の表 5.5 と概ね一致する割合で発生した。しかし、失敗の中での割合を大きく占めるため今後は対処が必要である。コンピュータが画像を解析し、その内容を理解する技術が実用化できれば解決策として適応できる可能性がある。

- ブログサイトのタイトルに検索キーワードが含まれる (8 件, 18%)

本実験は単純な検索手法を用いたためこの失敗が発生してしまった。本文のみを検索することによってこの失敗の割合は削減できると考えられる。

- 検索キーワードが意図と違ってマッチする (6 件, 14%)

キーワードのパターンマッチングをする限りは発生してしまう間違いである。メタデータによる検索を実現すれば失敗の割合は削減できると考えられる。

時刻という細かい粒度を考慮すると 0.72 という適合率は従来の時間情報抽出手法よりも良好であるが、本文と関連性のない写真の抽出によるエラーが割合を多く占めたため、今後の課題としたい。画像のキーワード検索の精度向上を目的とした画像知識データベースの研究 [79, 80] 等を深く調査し、有用なものを本研究に取り入れ適合率向上を目指す。

また、Time-crawler の本来の役割は既に公開されているブログからのイベント生起時間抽出であったが、ユーザがブログをアップロードする時点でその内容を解析しユーザに提示することによって、より正確なメタデータが生成できると期待できる。コンテンツのアップロード確認画面が出た時点で、「自動的に抽出したイベント生起時刻はこれで正しいですか？」などといった確認をすると、もし抽出結果が間違っているときにユーザが積極的に時刻情報を訂正することを期待できる。

5.5. まとめと今後の課題

本章では行動支援情報提供システムを継続的に活性させるためにコンテンツメタデータを既存の Web 文書から抽出するモデルを提案した。特に個人サイトには本音の情報が記入されていることから、ユーザが今後の行動に参考になる情報が多く含まれていると期待して、個人サイトからのメタデータの抽出に取り組んだ。提案したモデルの善し悪しはメタデータ抽出クローラの性能に依存するため、実際に行動支援情報において最も重要な状況に関わる情報となる時間および位置に

ついて実装を行い評価した。その結果、精度に対する課題はいくらか残るものの、抽出した情報の粒度に対して既存の抽出手法よりも優れた精度で、十分な量のコンテンツメタデータを確保できることが確認できた。本提案手法を取り入れることによって、より広範囲のコンテンツを大量に集めることができるため、このようなサービスへの応用には非常に効果があると考えられる。

今後の展開として、それぞれの要素に対して残る課題を解決し、より精度の高いメタデータ抽出手法を検討する。また、行動支援情報検索には上記の時間や位置に加え、ユーザの嗜好や状況に適したコンテンツが何であるかを把握し、機械可読な状態にしなければならない。これらメタデータ自動抽出手法の適合率を 100% とすることは困難なため、メタデータを抽出するタイミングをブログのアップロード時にするユーザインターフェイスについても実験を行い有効性を評価したい。

しかし、ブログから時間および位置情報を抽出し、ブログの作者毎にデータを構造化して表示をすると、そのブログ作者の現実世界での行動を不特定多数の人に示してしまうことになる。もちろん、ブログに公開した時点で、目視によってその人の行動は推測することは可能であるので、提案システムに問題点がある訳ではないが、将来的に上述したブログのアップロード時にメタデータを抽出しユーザに確認を促すユーザインターフェイスにより、公開するメタデータを制御する。あるいは、ブログの本文のみを推薦するなどしてブログの作者を特定できない状態で行動支援情報を推薦する等の工夫も必要になると考えられる。

第6章 おわりに

本論文ではユーザがいつでもどこでも求める情報を省作業で的確に取得するための要件を整理し、実証実験や基礎実験によって検証した。

第1章では本研究で取り扱う行動支援情報のあり方について整理し、「実空間との対応がとられた情報を提供すること」、「瞬時に検索結果が得られること」、「ユーザのこれからとる行動を示唆するもの」の三要件にまとめ、それぞれの重要性について述べた。

第2章では、この三要件を軸に関連研究について調査し、本研究の目的と照らし合わせ取り組むべき課題が何かを考察した。主に情報取得手法、情報流通基盤、情報源の観点において関連研究を分類・考察し、情報取得手法についてはメタデータによる情報検索とコンテキストウェアサービスを元に、情報流通基盤はP2P型を元に課題を設定した。情報源については個人サイトの文書に価値があると判断し、それらを行動支援情報として検索可能にするための構造化の必要性について述べた。関連研究の考察を踏まえて、「ユーザのサービス利用における許容性」、「システムのリアルタイム性と柔軟性」、「正確なコンテンツメタデータの維持性」の課題を見出した。

第3章では、「ユーザのサービス利用における許容性」の課題について取り組み、RFIDカードを活用したゼロクリック推薦を提案し、その許容性について確認した。ユーザから得られる限りのあるクエリ要素を最大限に活用する、状況判断およびメタデータマッチングを主とする推薦アルゴリズムから得られるコンテンツはユーザから一定の満足を得ることが確認できた。

第4章では、「システムのリアルタイム性と柔軟性」の課題について取り組み、システムの柔軟性についてはエージェントによる機能の分割により実運用前の試験期間中およびパラメータチューニング時にシステム全体を停止することなく可

能なことを確認した。リアルタイム性については、コンテンツが位置に基づいて分散されることを期待し、位置に基づいたオーバーレイネットワークである LL-Net を活用し検証したが、実際のコンテンツ分布を確認したところ大きな偏りがあり、実運用時には店舗の密度および人の密度について綿密な分析をした上で、ノードを配置する必要があることが明らかになった。

第 5 章では、「正確なコンテンツメタデータの維持性」の課題について取り組み、個人サイトを巡回しメタデータを抽出する手法を提案した。提案手法はメタデータを抽出する箇所に性能が依存するため、時間情報および位置情報について、どのような粒度・形式でメタデータが抽出できることが望ましいかについて議論し、メタデータの重要性について述べた。実際にイベント生起時間および緯度・経度を抽出する手法について提案し検証したところ、課題はいくらか残ったが、それらを解決することで十分に実運用可能なクローラが構築できることを示した。

今後の課題として、推薦結果に対する明示的な評価の収集による推薦精度の評価、得られた実コンテンツ分布を元にした大規模シミュレーションによるより細かな規模性に対する評価、実験によって得られたエラー原因を克服することによるメタデータ抽出精度の向上について解決する必要がある。

謝辞

本研究全般にわたって熱心な御指導および御支援していただきました, 主指導教官である本学情報科学研究科砂原秀樹教授に深く感謝します.

副指導教官であり, 研究に関する厳しくも暖かい御助言をいただきました, 本学情報科学研究科山口英教授に心から感謝します.

研究活動全般において多大なる御指導および御助言を頂きました, 本学情報科学研究科藤川和利准教授に深く感謝いたします.

本論文をまとめるにあたって, 貴重な時間を割いて丁寧な御指摘を頂き, また本研究の実験環境を平成 17 年度総務省「ユビキタスネットワーク認証・エージェント技術の研究開発」(以下 UAA (Ubiquitous Authentication and Agent) プロジェクト)を通して提供して頂いた, 大阪大学サイバーメディアセンター下條真司教授に心から感謝します.

また, 3 章および 4 章に関わる実験フィールドとして「エルムの街」を提供いただいた, 五所川原街づくり株式会社加西社長および関係者に変感謝します.

UAA プロジェクトを通して本研究についての御助言, お力添えを頂いた本学情報科学研究科奥田剛助手, 大阪大学工学部春本要准教授, 大阪大学情報科学研究科寺西裕一准教授, 竹内亨助教に心から感謝します.

同じく UAA プロジェクトを通して本研究の実験環境構築について御助力頂いた, 大阪大学サイバーメディアセンター特任研究員, 石芳正氏, 株式会社 BBR 吉田幹氏, 貫上秀典氏に心よりお礼を申し上げます.

研究に関する御指導に加え, 大学院生としての生き方, 修了後の生き方についてたくさんの貴重なお話をいただいた九州工業大学中村豊准教授, 本学情報科学研究科島田秀輝特任助教に感謝します.

日頃の大切な研究環境である曼陀羅ネットワーク支えてくれた情報科学セン

ター油谷暁助手, 辻井高浩助手をはじめとする情報科学センターの技術職員の方々に心より感謝します。

株式会社 NTT コミュニケーション川口誠敬君とはグルメ情報に関わる研究を通して熱い議論を交わし, いくらかの学会受賞を通して本研究の価値を確認することができました。非常によい思い出になりました。大変感謝します。

様々な面から研究活動を支えてくださった, 本学情報科学センター事務補佐員の呂悠妃さん, 東京大学工学部江崎研究室事務補佐員の田坂佳苗さん, 大阪大学サイバーメディアセンター事務補佐員の米原夏子さん, 岩田紀子さん, 元事務補佐員の加藤美和さんに心より感謝します。

また, インターネット・アーキテクチャ講座のスタッフ, 学生, 修了生の皆様, 大阪大学下條研究室のスタッフ, 学生, 卒業生, 修了生の皆様, WIDE プロジェクトの皆様とはインターネットの研究に携わる意義を共有させて頂いたことを心より感謝します。

最後に, 博士後期課程の進学に理解いただき陰ながら支えてくれた両親に深く感謝します。

参考文献

- [1] M. Weiser. The computer for the twenty-first century. *Scientific American*, pp. 94–101, September 1991.
- [2] S. Antifakos and B. Schiele. Beyond position awareness. *Personal and Ubiquitous Computing*, Vol. 6, No. 5/6, pp. 313–317, 2002.
- [3] M. Lampe and M. Strassner and E. Fleisch. A ubiquitous computing environment for aircraft maintenance. *SAC'04: Proceedings of 2004 ACM Symposium on Applied computing*, pp. 1586–1592, 2004.
- [4] L.E. Holmquist, J. Falk and J. Wigstrom. Supporting group collaboration with interpersonal awareness devices. *Personal and Ubiquitous Computing*, Vol. 3, No. 1/2, pp. 13–21, 1999.
- [5] L. E. Holmquist and F. Mattern and B. Schiele and P. Alahuhta and M. Beigl and H. -W. Gellersen. Smart-its friends: A technique for users to easily establish connections between smart artefacts. *LNCS*, Vol. 2201, pp. 116–122, 2001.
- [6] J. E. Bardram. Hospital of the future - Ubiquitous computing support for medical work in hospitals. *UbiHealth 2003: The 2nd International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications*, 2003.
- [7] Y. Kawahara and N. Kawanishi and H. Morikawa and T. Aoyama. Top-Down Approach toward Building Ubiquitous Sensor Network Applications. *Proceedings of the 11th Asia-Pacific Software Engineering Conference (APSEC'04)*, Vol. 00, pp. 695–702, 2004.

- [8] 茶筌. <http://chasen-legacy.sourceforge.jp/>.
- [9] MeCab. <http://mecab.sourceforge.net/>.
- [10] Google. <http://www.google.com/>.
- [11] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. *Stanford Digital Library Technologies, Working Paper*, 1998. SIDL-WP-1999-0120.
- [12] 荒谷寛和, 藤田茂, 菅原研次. ウェブページ間相互評価によるウェブ検索手法の提案と実装評価. 情報処理学会論文誌, Vol. 46, No. 2, pp. 337–347, 2005.
- [13] W. Fan and M. D. Gordon and P. Pathak. Discovery of Context-Specific Ranking Functions for Effective Information Retrieval Using Generic Programming. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 4, pp. 523–527, April 2004.
- [14] Yahoo!. <http://www.yahoo.com/>.
- [15] Excite. <http://www.excite.co.jp/>.
- [16] Netscape. <http://netscape.com/>.
- [17] J. L. Herlocker and J. A. Konstan and L. G. Terveen and J. T. Riedl. Evaluating Collaborative Filtering Recommender Systems. *ACM Transaction on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 5–53, January 2004.
- [18] 杉山一成, 波多野賢治, 吉川正俊, 植村俊亮. ユーザからの負担なく構築したプロファイルに基づく適応的 web 情報検索. 電子情報通信学会論文誌 D-I, Vol. J87-D-I, No. 11, pp. 975–990, November 2004.
- [19] Amazon.com. <http://www.amazon.com/>.
- [20] P. Resnick and N. Iacovou and M. Suchak and P. Bergstrom and J. Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews.

Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, pp. 175–186, 1994.

- [21] 小野智弘, 本村陽一, 麻生英樹. 嗜好の個人差と状況依存性を考慮した映画推薦方式の検討. 情報処理学会研究報告 (DPS), Vol. 125, pp. 79–84, 2005.
- [22] SemanticWeb.org. <http://www.semanticweb.org/>.
- [23] Resource Description Framework Model and Syntax. <http://www.w3c.org/RDF/>.
- [24] 幸島明男, 和泉憲明, 車谷幸一, 中島秀之. ユビキタス計算環境におけるコンテンツ流通のためのマルチエージェントアーキテクチャ:consorts. 人工知能学会論文誌, Vol. 19, No. 4, pp. 322–333, 2004.
- [25] 石津健太郎, 岡村耕二. 利用者の属性に適応したコンテンツを提供する cdn の設計と実装. 情報処理学会論文誌, Vol. 45, No. 2, pp. 509–516, Feb. 2004.
- [26] S. Usanavasin and T. Nakamori and S. Takada and N. Doi. A Multi-faceted Approach for Searching Web Applications. *IPSJ Journal*, Vol. 46, No. 3, pp. 816–830, March 2005.
- [27] 角野宏光, 石川憲洋, 小俣栄治, ヨハンイェルム, 宮津和弘, 村上慎吾. メタデータを用いたモバイル向け p2p 検索システムの設計と実装. 情報処理学会論文誌, Vol. 46, No. 2, pp. 365–275, February 2005.
- [28] 宮本崇弘, 竹内亨, 奥田剛, 春本要, 有吉勇介, 下條真司. GrIP: プライバシとサービス品質のトレードオフを考慮した個人情報制御機構. 日本データベース学会 Letters, Vol. 4, No. 1, pp. 145–148, June 2005.
- [29] 横路誠司, 高橋克巳, 三浦信幸, 島健一. 位置指向の情報の収集, 構造化および検索手法. 情報処理学会論文誌, Vol. 41, No. 7, pp. 1987–1998, 2000.

- [30] L. Can, Z. Qian, M. Xiaofeng, and L. Wenyin. Postal Address Detection from Web Documents. *Proceedings of the International Workshop on Challenges in Web Information Retrieval and Integration*, pp. 40–45, April 2005.
- [31] I. Mani and G. Wilson. Robust Temporal Processing of News. *In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)*, pp. 69–76, 2000.
- [32] B. Han, D. Gates, and L. Levin. Understanding Temporal Expressions in Emails. *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*, pp. 136–143, June 2006.
- [33] 野呂太一, 乾孝司, 高村大也, 奥村学. イベントの生起時間帯判定. 情報処理学会自然言語処理研究会報告, Vol. 2005, No. 117, pp. 7–14, 2005.
- [34] T. Noro, T. Inui, H. Takamura, and M. Okumura. Time Period Identification of Events in Text. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pp. 1153–1160, July 2006.
- [35] 山田辰美, 服部正嗣, 平松薫, 柳沢豊, 岡留剛. 実世界指向コンテキストウェアサービスの入力に着目した比較. 電子情報通信学会論文誌 D, Vol. J90-D, No. 3, pp. 820–836, 2007.
- [36] D. Hong, D. K. W. Chiu, and V. Y. Shen. U-commerce & u-business: Requirements elicitation for the design of context-aware applications in a ubiquitous environment. *7th international conference on Electronic commerce ICEC '05*, pp. 590–596, August 2005.
- [37] S. Meyer and A. Rakotonirainy. A survey of research on context-aware homes. *Australasian information security workshop conference on ACSW frontiers*, pp. 159–168, 2003.

- [38] G. D. Abouwd and C. G. Atkeson and J. Hong and S. Long and R. Kooper and M. Pinkerton. Cyberguide: A Mobile Context-Aware Tour Guide. *Special Issue: mobile computing and networking: selected papers from MobiCom '96*, Vol. 3, pp. 421–433, October 1997.
- [39] K. Cheverst and N. Davies and K. Mitchell and A. Friday and C. Efstratiou. Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 17–24, 2000.
- [40] J. E. Bardram and T. R. Hansen and M. Mogensen and M. Soegaard. Experiences from Real-World Deployment of Context-Aware Technologies in a Hospital Environment. *UbiComp 2006: Ubiquitous Computing, LNCS*, Vol. 4206, pp. 369–386, 2006.
- [41] A. Beeharee and A. Steed. Filtering Location-Based Information Using Visibility. *Location and Context-Awareness (LNCS)*, Vol. 3479, pp. 306–315, March 2005.
- [42] 根本博明, 西本一志, 山下邦弘. 広告主・閲覧者間コミュニケーションを促進するコミュニティ向け電子広告システムの提案. 情報処理学会論文誌, Vol. 46, No. 1, pp. 115–126, January 2005.
- [43] 高瀬正明, 五十嵐洋一郎, 掛水光明. 状況依存型サービス提供のための情報流通方式. 信学技報, pp. 77–80, 2004. NS2004-66.
- [44] 藤波香織, 中島達夫. コンテキストウェアなアプリケーション構築のためのフレームワーク. 情報処理学会論文誌 コンピューティングシステム, Vol. 44, No. SIG10, pp. 107–118, July 2003.
- [45] 岡本昌之, 長健太, 岡本雄三, 山崎智弘, 服部正典. Context-Aware エージェントのユーザモデルに基づく適応性検証手法. 電子情報通信学会論文誌 D-I, Vol. J88-D-I, No. 9, pp. 1356–1367, 2005.

- [46] 山崎智弘, 長健太, 岡本昌之, 岡本雄三, 服部正典. 状況依存エージェントフレームワークの実装と評価. 電子情報通信学会論文誌 D-I, Vol. J88-D-I, No. 9, pp. 1331–1343, September 2005.
- [47] H. Chen and T. Finin and A. Joshi and L. Kagal and F. Perich and D. Chakraborty. Intelligent Agents Meet the Semantic Web in Smart Spaces. *IEEE Internet Computing*, pp. 69–79, November 2004.
- [48] 服部正典, 長健太, 大須賀昭彦, 一色正男, 本位田真一. ユビキタス環境における context-aware なパーソナルエージェントの構築とその実証実験. 電子情報通信学会論文誌 D-I, Vol. J86-D-I, No. 8, pp. 543–552, August 2003.
- [49] 竹之内隆夫, 岡本直之, 川村隆浩, 大須賀昭彦, 前川守. ユビキタス環境において動的なコンテキストに応じて知識情報をフィルタリングする推論エージェントの開発. 電子情報通信学会論文誌 D, Vol. J88-D-I, No. 9, pp. 1428–1437, September 2005.
- [50] Napster. <http://www.napster.com/>.
- [51] Gnutella. <http://www.gnutella.com/>.
- [52] K. Nakauchi and H. Morikawa and T. Aoyama. Design and Implementation of a Semantic Peer-to-Peer Network. *High Speed Networks and Multimedia Communications (LNCS)*, Vol. 3079, pp. 961–972, September 2004.
- [53] 金子雄, 春本要, 福村真哉, 下條真司, 西尾章治朗. ユビキタス環境における端末の位置情報に基づく P2P ネットワーク. 情報処理学会 (トランザクション) データベース, Vol. 46, No. SIG18, pp. 1–15, December 2005.
- [54] I. Stoica and R. Morris and D. Liben-Nowell and D. R. Karger and M. F. Kaashoek and F. Dabek and H. Balakrishnan. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. *ACM of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM'01)*, pp. 149–160, August 2001.

- [55] S. Ratnasamy and P. Francis and M. Handley and R. Karp and S. Shenker. A Scalable Content-Addressable Network. *ACM of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM'01)*, pp. 161–172, August 2001.
- [56] A. Rowstron and P. Druschel. Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems. *Proceedings of the 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001)*, pp. 329–350, November 2001.
- [57] 中内清秀, 森川博之, 青山友紀. ユビキタス環境に向けた分散コンテンツ発見. 信学技報 CS2002-110, Vol. 102, No. 352, pp. 7–12, September 2002.
- [58] H. T. Shen and Y. Shu and B. Yu. Efficient Semantic-Based Content Search in P2P Network. *IEEE Transaction on Knowledge and Data Engineering*, Vol. 16, No. 7, pp. 813–826, July 2004.
- [59] P. Reynolds and A. Vahdat. Efficient Peer-to-Peer Keyword Searching. *Middleware 2003 (LNCS)*, Vol. 2672, p. 997, January 2003.
- [60] 神谷英樹, 峰野博史, 石川憲洋, 角野宏光, 水野忠則. 複数センサネットワークを統合管理するアーキテクチャの検討. 情報処理学会 マルチメディア, 分散, 協調とモバイル (DICOMO2007) シンポジウム論文集, pp. 1782–1787, July 2007.
- [61] 吉田幹, 奥田剛, 寺西裕一, 春本要, 下條真司. マルチオーバレイと分散エージェントの機構を統合した P2P プラットフォーム PIAX. 情報処理学会論文誌, Vol. 49, No. 1, pp. 402–413, January 2008.
- [62] グルメ情報検索サイトぐるなび. <http://www.gnavi.co.jp/>.
- [63] Yahoo! グルメ. <http://gourmet.yahoo.co.jp/>.
- [64] What's Open? <http://www.whatsopen.com/>.

- [65] 奥村学, 南野朋之, 藤木稔明, 鈴木泰裕. blog ページの自動収集と監視に基づくテキストマイニング. 人工知能学会, セマンティック Web とオントロジー研究会, SIG-SWO-A401-01, 2004.
- [66] SHOOTI. <http://shooti.jp/>.
- [67] 食べログ. <http://tabelog.com/>.
- [68] グーパス. <http://www.goopas.jp/>.
- [69] 前川徹. 米国インターネット事情：クリック率が低下するバナー広告. 情報処理学会会誌, Vol. 41, No. 9, pp. 1052–1053, September 2000.
- [70] M. Masahiro and Y. Shinoda. Information filtering based on user behavior analysis and best match text retrieval. *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 272–281, 1994.
- [71] Aglets software development kit.
<http://www.trl.ibm.com/aglets/>.
- [72] Google Maps. <http://maps.google.co.jp/>.
- [73] geotags. <http://geotags.com/>.
- [74] Yahoo! Maps. <http://www.google.com/apis/>.
- [75] TimeML. <http://timeml.org/site/>.
- [76] Yahoo! ブログ. <http://blogs.yahoo.co.jp/>.
- [77] Ameba(アメーバブログ). <http://www.ameba.jp/>.
- [78] 楽天ブログ. <http://plaza.rakuten.co.jp/>.

- [79] 獅々堀正幹, 小泉大地, 柘植寛, 北研二. 画像知識データベースを用いた WWW 画像検索システムの開発. 電子情報通信学会論文誌 D-I, Vol. J87-D-I, No. 2, pp. 154–163, February 2004.
- [80] 柳井啓司. キーワードと画像特徴を利用した WWW からの画像収集システム . 情報処理学会論文誌：データベース, Vol. 42, No. SIG10, pp. 79–91, 2001.

研究業績

学術論文

1. 新井 イスマイル, 川口 誠敬, 藤川 和利, 砂原 秀樹, “Geocrawler:個人サイトの評価情報と位置情報に基づいた店舗検索用 Web インデクサの開発,” 情報処理学会 論文誌, Vo.48, No.7, pp.2319–2327, 2007 年 7 月.

国際会議

1. I. Arai, K. Fujikawa and H. Sunahara, “Proposal of Time-crawler which collects an event time by reading Exif data in blogs,” Proceedings of the Annual IEEE Student Paper Conference (AISPC’08), Aalborg, Denmark, February, 2008.
2. Y. Uranishi, M. Sakata, I. Arai, Y. Manabe, H. Sunahara and K. Chihara, “Book Management System Using Various Sensor Network,” Proceedings of International Workshop on Advanced Imaging Technology 2008 (IWAIT 2008), Hsinchu, Taiwan, January, 2008.
3. I. Arai, Y. Uranishi, M. Sakata, K. Fujikawa, H. Sunahara, and K. Chihara, “Next Generation Library: Automatic book management system with ubiquitous sensors,” Demonstration at Fifth annual IEEE International Conference on Pervasive Computing and Communications (PerCom 2007), White Plains, NY, USA, March, 2007.

4. I. Arai, K. Fujikawa and H. Sunahara, “A Proposal of information retrieval method based on TPO metadata,” Proceedings of 2005 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, pp.442–445, Victoria, Canada, August, 2005.

国内研究会

1. 新井 イスマイル, 竹内 亨, 寺西 裕一, 藤川 和利, 砂原 秀樹, 下條 真司, “コンテンツ提供者とユーザの意図を考慮したコンテンツ推薦アルゴリズムの提案と評価,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO2007) シンポジウム論文集, pp.775–783, 2007 年 7 月.
2. 新井 イスマイル, 飯田 龍, 小林 のぞみ, 乾 健太郎, 藤川 和利, 砂原 秀樹, “グルメ情報を含む Web 文書からのユーザ指向型評判情報抽出システムの開発,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO2006) シンポジウム論文集, pp.953–956, 2006 年 7 月. (野口賞 (優秀デモンストレーション賞) 受賞)
3. 川口 誠敬, 新井 イスマイル, 藤川 和利, 砂原 秀樹, “Geocrawler: 位置情報をもとにした個人サイト向け Web インデクサの開発,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO2006) シンポジウム論文集, pp.929–932, 2006 年 7 月. (最優秀論文賞受賞)
4. 石 芳正, 新井 イスマイル, 寺西 裕一, 春本 要, 下條 真司, 武本 充治, 須永 宏, 田中 絵理香, 西木 健哉, “ユビキタス環境における P2P エージェントプラットフォームを用いた情報推薦機構の提案と実装,” 情報処理学会, 第 127 回 「マルチメディア通信と分散処理」 研究会, pp.69–74, 2006 年 6 月.
5. 松浦 知史, 新井 イスマイル, 中村 豊, 藤川 和利, 砂原 秀樹, “P2P ネットワークにおける位置に関連した情報の管理・検索手法の提案と評価,” 情報処理学会, 第 46 回プログラミング・シンポジウム 報告集, pp.83–94, 2005 年 1 月.

6. 益井 賢次, 岡田 行央, 新井 イスマイル, 市川 本浩, 中村 豊, “ネットワークの一時利用を実現するコンポーネント独立で可搬性の高いユーザ管理システムの設計と構築,” 情報処理学会, 分散システム/インターネット運用技術シンポジウム 2004 年度 論文集, pp.19–24, 2004 年 12 月.
7. 新井 イスマイル, 中村 豊, 藤川 和利, 砂原 秀樹, “TPO メタデータに基づく情報検索手法の提案,” 情報処理学会, 第 12 回マルチメディア通信と分散処理ワークショップ, pp.251–256, 2004 年 12 月.
8. 新井 イスマイル, 和泉 順子, 中村 豊, 藤川 和利, 砂原 秀樹, “携帯情報端末における XML 代理サーバによる属性情報管理機構の提案,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO2003) シンポジウム 論文集, pp.769–772, 2003 年 6 月.
9. 市川 本浩, 赤木 永治, 新井 イスマイル, 中村 豊, 砂原 秀樹, “ボランティアによる運用を考慮した簡便で可能性の高い認証管理ゲートウェイシステムの開発,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO 2003) シンポジウム 論文集, pp.33–36, 2003 年 6 月.