

インターネットにおけるユーザフレンドリーな情報取得

岩田 麻佑[†] 原 隆浩[†] 西尾章治郎[†]

[†] 大阪大学大学院情報科学研究科マルチメディア工学専攻 〒565-0871 大阪府吹田市山田丘 1-5

E-mail: †{iwata.mayu,hara,nishio}@ist.osaka-u.ac.jp

あらまし インターネットから情報を取得することは非常に便利である一方、ユーザによっては必要な情報を探し出すことが困難であったり、情報の取得に伴うプライバシー侵害などの問題が起きる場合も考えられる。このような問題に対し、ユーザが容易に安全にインターネット上の情報を取得できるように、これまでに取り組んできた (i) 子供による Web 検索の支援、(ii) 曖昧な検索クエリによる Web 検索の支援、(iii) 位置情報サービス利用時のプライバシーの保護に関する研究内容について紹介する。

Mayu IWATA[†], Takahiro HARA[†], and Shojiro NISHIO[†]

[†] Department of Multimedia Engineering Graduate School of Information Science and Technology

1-5 Yamadaoka, Suita, Osaka 565-0871, Japan

E-mail: †{iwata.mayu,hara,nishio}@ist.osaka-u.ac.jp

1. はじめに

現在では、いつでも、誰でも、どこでも、様々な端末を用いて、ユーザはインターネット上の情報を取得することが可能となっている。しかし、インターネットから情報を取得することは非常に便利である一方、ユーザによっては情報の取得が困難であったり、情報の取得に伴う問題が起きる場合も考えられ、必ずしもユーザフレンドリーな環境において情報を取得できるとは限らない。例えば、子供や高齢者などは大多数のユーザを占める大人と比較すると情報処理能力が低いいため、検索クエリの入力や検索結果の選択を苦手とし、自分の必要とする情報を探すが難しく、結果的にインターネット上の情報を十分に活用できないという可能性がある。また、インターネット上の情報に対してのユーザの検索要求は多岐に渡っているが、検索ログの大多数を占める短く曖昧な検索クエリでは、ユーザそれぞれの要求に合わせた情報を取得することは難しいと考えられる。さらに、インターネットを利用して情報を取得する際には、検索クエリを始めとして、ユーザのプロファイル情報や現在地などを入力する場合もある。そのような場合、入力した個人情報などが外部に漏れ、プライバシーが侵害されてしまうというような問題が起りうるとも考えられる。このように、現状では、ユーザにとって必ずしもユーザフレンドリーな環境でインターネット上の情報を取得できるわけではない。

そこで本稿では、全てのユーザが簡単に必要な情報を取得できるわけではない、全てのユーザが同じような情報を収集したいわけではない、全てのユーザが情報を得るに当たってプライ

バシ侵害を受ける危険性があるという3つの問題についての取り組みについて紹介する。具体的には、それぞれの問題に対応する研究内容として、(i) 子供による Web 検索の支援、(ii) 曖昧な検索クエリによる Web 検索の支援、(iii) 位置情報サービス利用時のプライバシーの保護について紹介する。

2. 子供による Web 検索の支援：検索結果の子供向けリランク

本章では、全てのユーザが簡単に必要な情報を取得できるわけではないという問題、特に子供による Web 検索の支援に着目した検索エンジンの検索結果の子供向けリランクについて紹介する。

近年、子供がインターネットを利用し、Web 検索を行うことが一般的となってきている^(注1)。子供がインターネットを利用する際の目的は、“学習”に関する Web 検索が 53.9%、“趣味”に関する Web 検索が 53.3%であると報告されており^(注1)、Web 検索を頻繁に使用していることが分かる。

しかし難解な文章を苦手とし、画像を好む [12] というような子供の特徴を考慮した Web 検索エンジンは存在しないため、現状では、子供が Web 検索を快適に利用できる環境が整っていないとはいえない。具体的には、現状の検索エンジンでは、子供が必要とする分かりやすく、興味を持てる情報が上位にランクされるとは限らない。小学生や中学生を対象とした子供の Web 検索行動を調査した既存研究において、子供には、検索結果が

(注1): goo リサーチ: <http://research.goo.ne.jp/>

表 1 スコア化に用いる指標

タイプ	指標	説明
構成	<i>Size</i>	ページの面積 (小さいほど高スコア, 図 1)
	<i>ImageRate</i>	ページ中の画像やアニメーションが占める面積の割合 (大きいほど高スコア, 図 1)
	<i>TextRate</i>	ページ中のテキストが占める面積の割合 (小さいほど高スコア, 図 1)
	<i>AncorRate</i>	ページ中に出現するリンクの数 (小さいほど高スコア)
	<i>Component</i>	文献 [11] の方法により分割したページ中のコンテンツの数 (小さいほど高スコア)
	<i>Color</i>	ページ中で使用されている色の数 (大きいほど高スコア)
文章	<i>KanjiRate</i>	ページ中のテキストの文字数の内、漢字、アルファベットの占める割合 (小さいほど高スコア)
	<i>ChildrenExpression</i>	ページ中のテキストに出現する全単語中の子供向け表現の割合 (大きいほど高スコア)
	<i>DifficultExpression</i>	ページ中のテキストに出現する全単語中の難解表現の割合 (小さいほど高スコア)
	<i>Easy</i>	帯 [13] により推定したページ中のテキストの難易度 (小さいほど高スコア)
	<i>SentenceLength</i>	ページ中のテキストの文章ごとの平均文字数 (小さいほど高スコア)

ら自分が必要とするページを選択するのが苦手とし、検索結果の上位 5 件を超えるページをあまり閲覧しないという傾向が明らかになっているため [3], [5], 検索結果のランキングでは、子供向けのページを上位にランクする必要がある。

そこで筆者らの研究グループでは、小学生の Web 検索を支援するため、子供向けページを上位にランクすることを目的とし、一般的な検索エンジンの検索結果を子供向けにリランクする手法を提案している [6]。提案手法では、子供にとって興味を持つことができ、見た目が見やすく、内容が分かりやすいようなページを子供向けページと定義し、一般向けページと子供向けページが混在したページ集合から子供向けページを判定できるように、ページの構成、文章に関する指標を設定し、ページの子供向け度合いをスコア化する。そして、検索エンジンの検索結果をスコアの降順に並び変えることで、子供向けのリランクを行う。これにより、子供向けのページが検索結果上位にランクされるため、子供はランキング下位まで探す手間なしに、子供向けのページを容易に得ることができる。

2.1 子供向けリランク手法

2.1.1 子供向けページの定義

本研究では、既存研究で得られた知見に基づき、Web ページの構成と内容について、網羅的に以下の特徴量を考慮する。

- Web ページの構成 [3], [9], [12]

ページの大部分が文字で構成され、含まれている情報量が多く、サイズの大きいページは、子供にとって面白みに欠け、どのコンテンツに注目すべきか分からない。さらに、子供にとってスクロール操作は負担である。そこで、子供にとっては、子供の興味をひき、分かりやすい構成が必要と考えられるため、以下の点を満たすページを子供向けと定義する。

- 画像やアニメーションが使用されている。
- テキストは適度な量である。
- リンクはどこをクリックすべきか分かるように把握しやすい量である。
- スクロール操作が少量で済むページサイズである。
- カラフルな色使いである。



図 1 指標の例

- Web ページの文章 [9], [12], [15]

子供向けのページとしては、子供に親しみや興味を持たせる文章であることが重要と考えられる。そのため、以下の点を満たすページを子供向けとする。

- 一般的な文と比べると、長さは短めで、一文に含まれる情報が少ない。
- 漢字、アルファベットや難解な表現が少ない。
- 語りかけの表現などの子供向け特有の表現が含まれている。
- 難解な表現が使用されず、理解が容易な文である。

2.1.2 指標の設計

前項で述べた子供向けページの定義に基づき、表 1 に示す指標を設定した。設定した指標は、構成に関する指標 6 個、文章に関する指標 5 個の計 11 個であり、これらの指標を用いて Web ページのスコア化を行い、スコアが大きいほど子供向けとする。

2.2 リランク手順

表 1 に示す指標に基づき、Web ページのスコアを算出し、リランクを行う。

- (1) ユーザがクエリを入力する．
- (2) クエリに対する検索結果を検索エンジンより取得する．
- (3) 取得した検索結果の Web ページのスコアを各指標に基づき算出する．
- (4) スコアの降順にリランクして表示する．

2.3 評価実験

指標を組み合わせた際の有効性を検証するために行った評価実験について述べる．

2.3.1 データセット

提案手法を評価するためには、実際に子供に Web ページを評価してもらった正解データが必要である．そこで、以下のようにデータセットを作成した．まず、Yahoo!きっず^(注2)における週間検索キーワードランキングの上位より選んだ6個のキーワードを実験に使用するクエリとし、各クエリごとに Yahoo! JAPAN^(注3)、Yahoo!きっずの検索結果のランキングのそれぞれの上位から25の Web ページを取得した．そして、合計300の Web ページを小学生の男女34名に評価してもらい、Yahoo! JAPAN 用、Yahoo!きっず用のデータセットをそれぞれ作成した．具体的には、それぞれの被験者に Web ページを閲覧してもらい、“読みたいと思うか? ”、“見た目が見やすいか? ”、“内容が分かりやすいか? ”の3つの質問に Yes, No の2択で答えてもらい、1ページ当たり4人から7人による評価を得た．そして、Yes を1点、No を0点として各ページごとに点数を平均したものをそのページの子供向けスコアとし、子供向けスコアの降順に並び変えたものを理想ランキングとすることでデータセットを作成した．

ここで、Yahoo! JAPAN、Yahoo!きっずのそれぞれでデータセットを作成したのは、Yahoo! JAPAN と Yahoo!きっずでは検索結果に含まれるページが大きく異なるためである．Yahoo!きっずのランキングでは、クエリに対応する登録サイトがあれば、そのサイトを検索結果上位に表示した後、Yahoo! JAPAN と同様のランキングを表示するが、Yahoo!きっずに登録されていないページは表示されない．そのため、Yahoo!きっずのランキングによるデータセットでは、Yahoo! JAPAN のランキングよりも子供向けページが多く含まれる傾向がある．

2.3.2 評価指標

評価指標には、正解データがクエリへの多段適合度を持つ場合の順位付き検索結果の性能を測る NDCG (Normalized Discounted Cumulative Gain) を用いた．NDCG は適合度の高い順に並べた理想的な結果とのずれを表す指標であり、今回は上位5件における NDCG 値を用いた．

2.3.3 実験結果

教師付き機械学習を用いたランキング関数の学習の代表的な手法である Ranking SVM を用いて、指標を組み合わせて利用した場合の性能を評価した．実装には svm_rank ^(注4)を用い、カーネルには線形カーネルを用いた．Ranking SVM による学

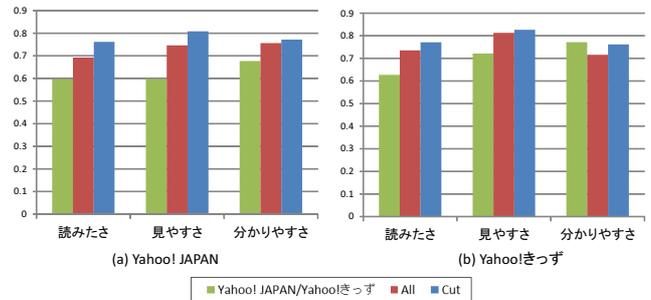


図2 指標を組み合わせた NDCG の平均値

習を行うことで、各指標を適切に組み合わせることが可能となる．ここで、学習は5種類のクエリに対応する250ページで行い、残り1種類のクエリに対応する50ページで評価を行うことを1セットとし、6種類のクエリ全てで評価できるように、6セットの評価を行った．

また、提案手法における指標の組合せ方法として、単純に11の全ての指標を使用する *All* に加え、指標間の相関関係、単独の指標によりリランクを行った際の精度を考慮して、不要な指標を削除し、*TextRate*, *AnchorRate*, *Component*, *Color*, *ChildrenExpression*, *SentenceLength* という6種類の指標を使用する *Cut* を用いた．

Yahoo! JAPAN のデータセットにおけるリランクの NDCG 値の平均を図2(a)に、Yahoo!きっずのデータセットにおけるリランクの NDCG 値の平均を図2(b)に示す．緑色が Yahoo! JAPAN、Yahoo!きっず、赤色の *All*、青色が *Cut* の NDCG 値を示す．

まず、全ての指標を組み合わせた *All* では、図2(a)に示す Yahoo! JAPAN では、全ての質問で5%から15%の NDCG 値の向上が見られた．この結果より、単純に全ての指標を用いても、指標を組み合わせる効果があることが分かる．また、図2(b)に示す、NDCG 値がもともと高い Yahoo!きっずでは、分かりやすさでは Yahoo!きっずの NDCG 値を下回ったが、それ以外の質問で5%から10%の NDCG 値の向上が見られた．分かりやすさで *All* の NDCG 値が下がってしまったのは、高学年の子供にとっては、授業で学んでいると考えられる“地球温暖化”などのクエリについては、単純に画像が多く、文章の少ないページよりも、やや詳しくに内容が記述されているページの方が高評価であったためであると考えられる．

また、不要な指標を削除した *Cut* は、両データセットの全ての質問において、全ての指標を用いる *All* よりも NDCG 値が上回っており、指標間の関係や単独で精度の悪い指標を除く効果が大きかったことが分かる．

2.4 まとめと今後の課題

本章では、大人以上に Web 検索を苦手とする子供が容易に情報を取得できることを目指した検索結果の子供向けにリランク手法について紹介した．

今後は、テキスト量、画像量、文章の難易度などの年齢に応じた適切な値を調査し、学年の特徴に従った各指標の最適な組み合わせ方法を検討する予定である．

(注2): Yahoo!きっず: <http://kids.yahoo.co.jp/>

(注3): Yahoo! JAPAN: <http://www.yahoo.co.jp/>

(注4): svm_rank : http://www.cs.cornell.edu/people/tj/svm.light/svm_rank.html

3. 曖昧な検索クエリによる Web 検索の支援: タイルによる検索結果の可視化インタフェース

本章では、全てのユーザが同じような情報を収集したいわけではないという問題、特に曖昧な検索クエリによる Web 検索の支援に着目したタイルによる検索結果の可視化インタフェースについて紹介する。

検索クエリは短く曖昧な場合が多い。例えば、ハリーポッターというクエリをユーザが入力した場合、ユーザがハリーポッターの本に関する情報を必要としているのか、映画に関する情報を必要としているのかは分からない。そのような曖昧なクエリに対して、検索結果を多様化し、多くのユーザの検索意図を検索結果のランキングリストに含むことで、できるだけ多くのユーザの検索要求を満たすことを目指した研究が多く行われている [1]。このような多様化した検索結果を一般的な検索結果と同じインタフェースを用いて表示してもユーザにとってはどのページがどんな検索意図を示しているのかを把握するのが難しい。

そこで筆者らの研究グループでは、多様化された検索結果の中からユーザが効率的に効果的に自分の必要としている情報を取得できるよう支援することを目的とし、通常の検索エンジンの検索結果をシンプルに拡張するインタフェースである AspecTiles を提案している (図 3)。AspecTiles は、検索結果中の各ページの左側に、色付きのタイルを用いて、どのページがどの検索意図 (アスペクト) にどの程度関連しているのかを視覚化する。通常の検索結果では、どのページがどのアスペクトにどれだけ関連しているかという情報はユーザにとって簡単には分からないが、AspecTiles のような情報が付加された検索結果のインタフェースでは、ユーザにとって欲しい情報を把握しやすく、インターネット上の情報取得の支援になると考えられる。

3.1 提案インタフェース

図 3 は、“elliptical trainer” を検索クエリとした際の AspecTiles のスクリーンショット画像である。通常のインタフェースと同様に、検索結果の各ページのタイトル、スニペット、URL を表示する。さらに、(a) アスペクトラベル: 検索結果上部に入力クエリに関するアスペクトのラベル (“reviews”, “best machines” など) を表示、(b) アスペクトごとのタイル: ページの各アスペクトへの関連度を 5 段階の色で表示 (色が濃いほど高い関連度)、という追加情報を提示する。また、Remove, Add ボタンを押すことで、表示するタイル数はユーザが自由に調節することができる。

追加の情報を表示するために、通常の検索エンジンの検索結果とクエリに対するアスペクトを、検索結果多様化のアルゴリズムに入力し、通常の検索結果のランキングを並び替え多様化した検索結果と検索結果のページの各アスペクトへの関連度を取得する。具体的には、ユーザがクエリを入力すると Bing API^(注5)を用いて、検索結果上位 300 ページとクエリサジェス

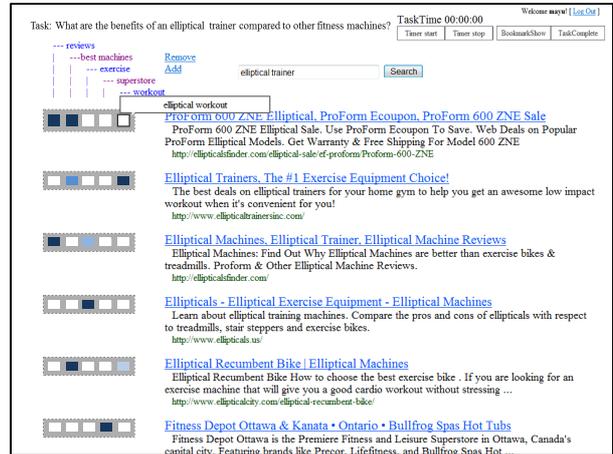


図 3 AspecTiles (elliptical trainer)

ションを取得する。本研究では、取得が容易で、適度な精度が期待されるクエリサジェスションをクエリのアスペクトとして利用する。アスペクトとして取得したクエリサジェスションを利用し、Dou らのアルゴリズム [4] に基づいて、各アスペクトへの関連度を算出し、Bing API で取得した検索結果をリランクし、多様化を行う。

3.2 ユーザ評価

提案インタフェースである AspecTiles の有効性を検証するためにに行った評価実験について述べる。

3.2.1 タスク

クエリは、特定のアスペクトに関する狭い情報要求、もしくは複数のアスペクトに関する広い情報要求を示している 2 種類のケースがありうる。例えば、“elliptical trainer” というクエリでは、ユーザは商品名という特定のアスペクトに関する情報のみ、もしくは効果やレビューという複数のアスペクトに関する幅広い情報を取得しようとしているのかもしれない。そのため、本研究では、シングルアスペクトタスクとマルチアスペクトタスクの 2 種類のタスクを設定し、それぞれのタスクでの提案インタフェースの効果を検証する。

具体的には、TREC 2009 web diversity より 15 種類のトピックを利用し、トピックごとに 1 個のシングルアスペクトタスクと 1 個のマルチアスペクトタスクを作成した。図 4 はタスクとして選択した TREC のトピックの例を示している。各トピックは、クエリ、概要情報、いくつかのサブトピックを含んでおり、概要情報からマルチアスペクトタスク、サブトピックの 1 つからシングルアスペクトタスクを作成した。

3.2.2 実験手順

AspecTiles の有効性を検証するため、提案インタフェースである AspecTiles と比較インタフェースとしてタイルの無い通常の検索結果インタフェースを用いてユーザ評価を行った。両インタフェースで提示する検索結果のランキングは、Dou らの手法 [4] により多様化した結果を用いた。被験者は 32 名で、内訳は 10 代から 30 代の男性が 21 名、女性が 11 名である。

実験では、インタフェースを用いて、被験者に 4 種類のタスク (2 種類のマルチアスペクトタスク + 2 種類のシングルアス

(注5): Bing API: <http://www.bing.com/toolbox/bingdeveloper/>

表 2 実験結果：各指標の平均値（標準偏差）

タスク		比較インタフェース	提案インタフェース
シングルアスペクト	Precision	0.498 (0.387)	0.469 (0.395)
	Recall	0.143 (0.127)	0.152 (0.145)
	Time-FirstRel	187.5 (96.8)	183.8 (100.9)
	DocClick-FirstRel	2.500 (2.502)	2.258 (2.065)
マルチアスペクト	Precision	0.684 (0.248)	0.717 (0.232)
	Recall	0.130 (0.088)	0.214 (0.204)
	Time-FirstRel	166.9 (117.0)	118.0 (120.6)
	DocClick-FirstRel	3.345 (3.199)	1.714 (3.199)

表 3 アンケート結果：各質問の 5 段階の平均スコア（標準偏差）

質問	比較インタフェース	提案インタフェース
Q1: 適合情報を見つけるのにインタフェースは役立ちましたか？	2.875 (0.941)	3.718 (0.581)
Q2: インタフェースを使うのは容易でしたか？	3.156 (0.919)	3.562 (0.800)
Q3: またインタフェースを使いたいですか？	2.875 (0.906)	3.593 (0.837)

```

<topic type="faceted" number="33">
<query>elliptical trainer</query>
<description>
Find information about elliptical trainer machines.
</description>
<subtopic type="inf" number="1">
I'm looking for reviews of elliptical machines.
</subtopic>
<subtopic type="inf" number="2">
Where can I buy a used or discounted elliptical trainer?
</subtopic>
<subtopic type="inf" number="3">
What are the benefits of an elliptical trainer compared to other fitness machines?
</subtopic>
<subtopic type="inf" number="4">
What are the best elliptical trainers for home use?
</subtopic>
</topic>

```

Multi-aspect task created from the description field

Single-aspect task created from one particular subtopic

図 4 TREC 2009 web diversity task のトピックとタスク例

ペクトタスク)を行ってもらった。具体的には、各タスクにおいて、被験者にそれぞれのインタフェースを用いて、制限時間内にできるだけ多くの答え(タスクの内容に適合する文章や画像)を集めてもらった。タスク中に答えを容易に集められるように、検索結果に提示される URL を選択すると、選択したページ中のドキュメント、画像などをドラッグすることで保存できる機能を全てのインタフェースに追加した。

タスク終了後には、各インタフェースの使いやすさなどについてアンケートを行った。

3.3 実験結果

32 人の被験者に評価実験を行った結果、合計で 1,262 個の答えを取得した。インタフェースの効果や性能を測るため、3 人の評価者で取得した答え全てについて適合性判定を行った (Fleiss' kappa:0.666)。具体的には、取得したそれぞれの答えについて、「適合している」、「やや適合している」、「適合していない」から適合度を選んでもらった。そして、3 人の評価者の判定結果のうち一番低い適合度をそれぞれの答えの適合度として ground-truth のデータセットを作成した。

以上のように作成したデータセットを用いて、提案インタフェースである AspecTiles と比較インタフェースの性能を評価した。あるタスクにおいてある被験者が保存した適合している答えの数を n とし、 n /そのタスクにおいてそのユーザが保存した全ての答えの数を Precision、 n /そのタスクにおける全ての適合している答えの数を Recall とした。これらの指標は高いほどユーザが情報を見つけるのに効果的なインタフェースであるといえる。また、最初の適合している答えを取得するまでの時間 (Time-FirstRel)、最初の適合している答えを取得するまでに選択したページ数 (DocClick-FirstRel) についても評価した。これらの指標は低いほどユーザが効率的に情報を見つけれられるインタフェースであるといえる。

表 2 はそれぞれの指標の結果を示している。インタフェースとタスクのトピックによる影響を分析するために分散分析を行った。その結果、シングルアスペクトタスクでは、Precision、Time-FirstRel、DocClick-FirstRel においてタスクの作用のみに有意差が見られた ($p \leq 0.05$)。マルチアスペクトタスクでは、Precision においてタスクの作用、Recall、Time-FirstRel、DocClick-FirstRel においてインタフェースの作用に有意差が見られた ($p \leq 0.05$)。インタフェースの作用に有意差が見られた場合のみ、インタフェース間で Bonferroni 法による多重比較を行った ($p \leq 0.05$)。

マルチアスペクトタスクにおいては、提案インタフェースは比較インタフェースよりも有意に Recall が高く、さらに有意差は見られなかったが Precision も高い傾向となった。また、Time-FirstRel、DocClick-FirstRel については、提案インタフェースの方が比較インタフェースよりも有意に小さい値となった。これらの結果より、マルチアスペクトタスクにおいては、提案インタフェースは、比較インタフェースよりも短い時間で、少ないページ選択で適合情報をユーザが取得できたことを確認した。

シングルアスペクトタスクにおいては、提案インタフェースと比較インタフェースの結果に有意差は見られなかった。これ

は、マルチアスペクトタスクと比べて、シングルアスペクトでは提示されるアスペクトがユーザが必要とするアスペクトと合致する場合が少ないことが主な原因と考えられる。最初に提示されるアスペクトのうち、タスクの答えに適合するアスペクトの数を手で数えたところ、シングルアスペクトでは0.683個、マルチアスペクトタスクでは1.113個となった。

3.4 アンケート結果

表3に各インタフェースにおけるアンケートの結果を示す。各被験者には、Q1からQ3の質問に5段階で答えてもらった。マン・ホイットニ検定を行い($p \leq 0.05$)、比較インタフェースとAspecTilesを比較した。表3の結果より、提案インタフェースは比較インタフェースと比べて、適合情報を探すのに便利であり、使うのが容易であり、また使いたくなるインタフェースであるということが分かった。

3.5 まとめと今後の課題

本章では、ユーザの多様な検索意図に対応するためのインタフェースとして、多様化された検索結果の中からユーザの検索意図に適合する情報を容易に取得できるように、検索結果をタイトルを用いて拡張したインタフェースを紹介した。

今後は、アスペクトごとのリランク機能の付加やクエリに関するアスペクトの決定方法の改善について検討している。

4. 位置情報サービス利用時のプライバシーを保護： ダミーによる位置曖昧化

本章では、全てのユーザが情報を得るに当たってプライバシー侵害を受ける危険性があるという問題、特に位置情報サービス利用時のプライバシーの保護に着目したダミーによる位置曖昧化に関する研究について紹介する。

GPS技術の発展に伴い、ユーザの位置に対応した情報を提供する位置情報サービスが展開されている。位置情報サービスを利用する際には、ユーザは自身の位置をサービスプロバイダへ通知する必要があるが、この位置情報が流出することにより、ユーザの訪問箇所が特定され、住居や勤務先などを第三者に把握される可能性が指摘されている。このようなユーザの位置情報に関するプライバシーを保護するために、筆者らの研究グループでは先行研究において、実環境における制約を考慮し、複数のダミーの位置情報を生成することでユーザの位置を曖昧化する手法を提案している[14]。しかし、この手法はユーザは停止することなく移動すると想定しており、いくつかの地点で停止しながら移動するというより自然なユーザの行動に対応することは難しい。

そこで、筆者らの研究グループでは、さらに、ユーザのユーザがいくつかの地点で停止しながら移動するという環境において、自然なダミーの動きを作成することのできるダミー生成手法を提案している[8]。提案手法では、ユーザの停止地点や停止時間も含めた行動予測が可能な状況を想定し、既知であるユーザの行動に基づいて、実環境における制約条件を考慮したダミーの行動を決定する。このような想定は実環境では必ずしも妥当ではないが、ユーザが事前に行動プランを登録したり、ユーザの過去の行動履歴から予測したりなど、ある程度の精度

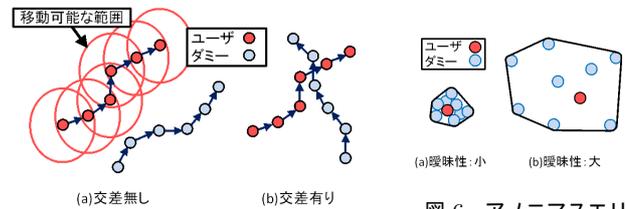


図5 追跡可能性

で予測できる場合も多い。

提案手法では、ダミーが停止すべき時間および地点を決定し、その時間にその地点を経由して移動を行うダミーの行動を決定する。この際、ユーザや他のダミーがあまり存在しない地点にダミーを停止させることで、ユーザの位置が曖昧になるようにし、ユーザや他のダミーと停止する地点を共有して交差をさせることで、ユーザの位置が一時的に特定された場合でも、その曖昧性を短時間で回復できるようにする。

4.1 ユーザの行動に基づいたダミーの移動経路生成手法

4.1.1 ダミー生成時の要求

提案手法では以下のようなダミー生成時の要求を考慮する。

- 追跡可能性

短期間の連続したサービス要求の際には、複数の位置情報が与えられた際に、それらを結合することにより、その軌跡を推測できる性質である追跡可能性を考慮しなければならない。図5のように、ユーザとダミーの経路を定期的に変換することで、追跡可能性を下げることができる。

- アノニマスエリア

ユーザの位置プライバシーを保護するためには、複数の位置情報から一意に特定できないだけでなく、どの程度の大きさの領域に位置情報が曖昧化されているかも重要である。例えば、6(a)のようなダミーの配置は、ダミーの存在範囲が小さいため、ユーザのおおよその位置の予測が可能になってしまう。本研究では、Luら[10]の定義に基づき、ユーザとすべてのダミーを包括する凸多角形をアノニマスエリアと定義し、その大きさをユーザ位置の曖昧度の評価値として用いる。

4.1.2 ダミーの行動決定方法

本研究では、ユーザが所有するモバイル端末上のシステムは地図情報を保持しており、ユーザやダミーが通っても不自然ではない道路、停止しても不自然ではない位置を全て把握しているものとする。また本研究では、ユーザの停止地点、停止時間、移動経路といったユーザの行動がすべて事前に予測できるものと想定する。

以上のような想定環境下で、提案手法では、ユーザが指定したダミーの個数、アノニマスエリアのサイズ、さらに、予測されたユーザの行動に基づいて、ダミーが停止すべき地点(停止地点)、その地点に到着すべき時間(停止地点到着時間)を決定する。そして、停止地点に向かって移動し、停止地点到着時間にそこに到着し、その後停止するというダミーの行動スケジュールを生成する。

具体的には、以下の手順を要求されたダミー数だけ繰り返

図6 アノニマスエリア

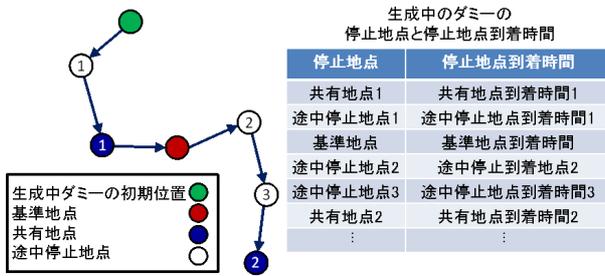


図7 最終的に決定されるダミーの行動

し、ダミーの行動を1つずつ順に決定する。初めのダミーの行動はユーザの行動のみを考慮して決定し、2番目以降のダミーの行動はユーザと生成済みダミーの行動を考慮して決定する。

(1) アノニマスエリア確保のための停止地点(基準地点)および停止地点到着時間(基準地点到着時間)の決定:

アノニマスエリアを十分に確保できるように、ユーザおよび生成済みダミーの存在が少ない領域とその時間を求め、その領域内に生成中ダミーの通るべき地点である基準地点、基準地点到着時間を1つ決定する。

(2) 追跡可能性低下のための停止地点(共有地点)および停止地点到着時間(共有地点到着時間)の決定:

追跡可能性を低下させるため、基準地点および基準地点到着時間をもとに、ユーザや生成済みダミーと共有可能な停止地点である共有地点、その地点に到着すべき時間である共有地点到着時間を決定する。そして、基準地点および決定した共有地点をもとに、共有可能な地点がなくなるまで、順次共有地点を増やし、できる限り多くの停止地点をユーザ、ダミー間で共有させ、交差を発生させる。

(3) 基準地点、共有地点を通るダミーの移動経路の決定: 無理のない移動経路を作成するため、基準地点や共有地点間に途中停止地点を適宜設定し、途中停止地点においてもダミーを停止させる。

生成した最終的なダミーの移動経路の例を図7に示す。

4.2 評価実験

提案手法の有効性を確認するために、京都の街の地図上でユーザの動きをシミュレートし、評価実験を行った。ユーザの動きは、道路上をランダムに停止しながら移動するモデルを利用した。シミュレーションにおけるサービス利用間隔は180[s]、ユーザの歩行速度は1.30[m/s]、要求ダミー数は16, 25[個]、ユーザ、ダミーの最大停止時間は600[s]、最小停止時間は60[s]、要求アノニマスエリアは1000², 1200², …, 2000²[m²]とした。

4.2.1 評価指標

● AAAR-Count (Anonymous Area Achieving Ratio - Count)

要求されたアノニマスエリアを、ダミー配置により実際に達成できた回数のサービス利用(要求)の総数に対する割合。要求アノニマスエリアをどの程度の頻度で達成できたかを示しており、常時達成できた場合には100%となる。

● MTC (Mean Time to Confusion)

ユーザが何かしらの情報で特定された後に、ユーザとダミーが

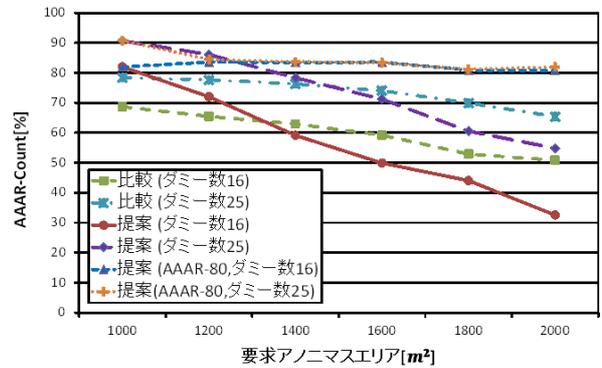


図8 AAAR-Count

交差することによりユーザである確率が一定値以下になる(十分に曖昧化される)までの平均時間。値が小さければ追跡可能性が低いということを表している。

4.2.2 評価手法

(1) 比較手法

ユーザの行動を予測できない状況を想定した、先行研究の提案手法[14]。この手法では、アノニマスエリアを確保し、ユーザの動きの変化に対応させるため、ダミーをユーザの周りにグリッド状に配置し、移動させる。

(2) 提案手法

予測したユーザの行動に基づいて、停止地点で停止しながら自然に移動するダミーを生成する提案手法。

(3) 提案手法 (AAAR-80)

80%以上のAAAR-Countを達成できるように設定アノニマスエリアを拡大した提案手法。準備実験を行った結果、提案手法が非常に低いAAAR-Countとなることがあった。これを防ぐために、提案手法の設定アノニマスエリアを実際にユーザが要求するアノニマスエリアよりも大きく設定し、要求アノニマスエリアを十分に確保できる際の提案手法の有効性を確認する。

4.2.3 実験結果

● AAAR-Count

図8にAAAR-Countの結果を示す。ダミー数に関わらず、提案手法と比較手法を比較すると、要求アノニマスエリアが大きくなると、提案手法は、比較手法に比べ、AAAR-Countが小さくなる。これは、提案手法はアノニマスエリアを確保するための基準地点は1つであるのに対して、できる限り多くの停止地点を共有させることで、ユーザとダミーを積極的に交差させるためである。

そこで提案手法において、各要求アノニマスエリアに対して、80%以上のAAAR-Countを達成するための、設定アノニマスエリアの大きさを調べた。調査した結果の設定アノニマスエリアの値を用いた際の提案手法のAAAR-Countを提案手法(AAAR-80)に示している。

● MTC

図9にMTCの結果を示す。提案手法と比較手法を比較すると、提案手法は、ダミー数が16と25共に、全ての要求アノニマスエリアにおいて、比較手法よりもMTCを低減でき

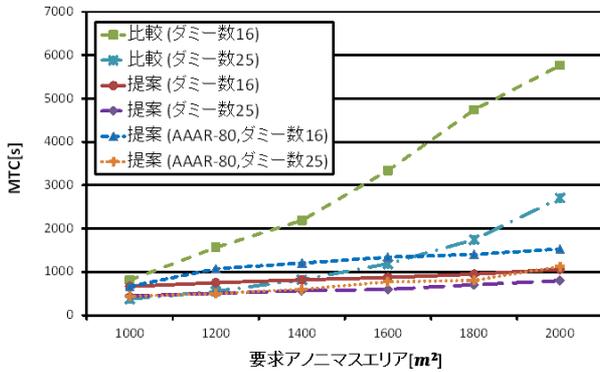


図 9 MTC

ている。提案手法において、ダミー数 16 と 25 共に、80%の AAAR-Count を達成する設定アノニマスエリアを用いても、全ての要求アノニマスエリアにおいて、提案手法 (AAAR-80) の MTC は比較手法よりも小さな値となるこの結果より、80%の AAAR-Count を達成するために、提案手法の設定アノニマスエリアを拡大しても、MTC の増加分はそれほど大きくなく、追跡可能性を十分に低く保てる事が分かる。

4.3 まとめと今後の課題

本章では、位置情報サービス利用におけるユーザの位置プライバシー保護を目的として、停止を伴うユーザの行動を考慮したダミーによる位置プライバシー保護について紹介した。

今後は、実際に人の目でユーザとダミーの動きを見た際に、どの程度ユーザを曖昧化できているか視認性の評価を行う予定である。さらに、ユーザの行動予測が外れてしまった場合でも、ユーザの行動に対応できるように提案手法を拡張することを検討している。

5. おわりに

本稿では、インターネット上の情報を全てのユーザが容易に安全に取得できることを目指した取り組みとして、全てのユーザが簡単に必要な情報を取得できるわけではない、全てのユーザが同じような情報を収集したいわけではない、全てのユーザが情報を得るに当たってプライバシー侵害を受ける危険性があるという問題の解決を目指した研究について紹介した。

具体的には、(i) 子供による Web 検索の支援を目指した Web 検索エンジンの検索結果の子供向けリランク手法、(ii) 曖昧な検索クエリによる Web 検索の支援を目指した多様化された検索結果のタイルによる可視化インタフェース、(iii) 位置情報サービス利用時のプライバシーの保護を目指した停止したユーザの移動経路を考慮したダミーによる位置曖昧化手法について紹介した。これらの取り組みにより、ユーザはますます快適に安全にインターネット上の情報を取得することができるものと考えている。

6. 謝 辞

本研究の一部は JSPS 科研費 (24187) による助成、マイクロソフトリサーチアジアの研究助成 (CORE 7) によるものである。ここに記して謝意を表す。

- [1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Jeong: Diversifying Search Results, *Proc. WSDM 2009*, pp. 5–14 (2009).
- [2] D. Bilal: Children’s Use of the Yahoo!igans! Web Search Engine: I. Cognitive, Physical, and Affective Behaviors on Fact-based Search Tasks, *Journal of the American Society for Information Science*, Vol. 51, No. 7, pp. 646–665 (2000).
- [3] D. Bilal and J. Kirby: Differences and Similarities in Information Seeking: Children and Adults as Web Users, *Information Processing and Management*, Vol. 38, No. 5, pp. 649–670 (2002).
- [4] Z. Dou, S. Hu, K. Chen, R. Song, and J.-R. Wen: Multi-Dimensional Search Result Diversification, *Proc. WSDM 2011*, pp. 475–484 (2011).
- [5] A. Druin, E. Foss, L. Hatley, E. Golub, M. L. Guha, J. Fails, and H. Hutchinson: How Children Search the Internet with Keyword Interfaces, *Proc. IDC 2009*, pp. 89–96 (2009).
- [6] M. Iwata, Y. Arase, T. Hara, and S. Nishio: A Children-Oriented Re-ranking Method for Web Search Engines, *Proc. WISE 2010*, pp. 225–239 (2010).
- [7] M. Iwata, T. Sakai, T. Yamamoto, Y. Chen, Y. Liu, J.-R. Wen, and S. Nishio: AspectTiles: Tile-based Visualization of Diversified Web Search Results, *Proc. SIGIR 2012*, pp. 85–94 (2012).
- [8] R. Kato, M. Iwata, T. Hara, A. Suzuki, Y. Arase, X. Xie, and S. Nishio: A Dummy-based Anonymization Method based on User Trajectory with Pauses, *Proc. GIS 2012*, pp. 249–258 (2012).
- [9] 菊地秀文, 赤堀侃司: 小学校情報教育における児童の Web ブラウジングの特徴分析, *日本教育工学会論文誌*, Vol. 27, No. 2, pp. 143–153 (2003).
- [10] H. Lu, C. S. Jensen, and M. L. Yiu: PAD: Privacy-Area Aware, Dummy-Based Location Privacy in Mobile Services, *Proc. Workshop on Data Engineering for Wireless and Mobile Access 2008*, pp. 16–23 (2008).
- [11] 前川卓也, 原 隆浩, 西尾章治郎: モバイル端末のための Web ページ自動スクロール方式, *日本データベース学会 Letters*, Vol. 4, No. 2, pp. 29–32, 2005.
- [12] J. Nielsen: Usability of Websites for Children: 70 Design Guidelines based on Usability Studies with Kids, *Nielsen Norman Group Report* (2002).
- [13] S. Sato, S. Matsuyoshi, and Y. Kondoh: Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus, *Proc. LREC 2008*, pp. 28–30 (2008).
- [14] A. Suzuki, M. Iwata, Y. Arase, T. Hara, X. Xie, and S. Nishio: A User Location Anonymization Method for Location Based Services in a Real Environment, *Proc. GIS 2010*, pp. 398–401 (2010).
- [15] 湯浅千映子: 子ども向け文章の情報の配列 「小学生新聞」を対象に, *文体論研究*, Vol. 52, No. 52, pp. 41–56 (2006).