

# 検索意図を考慮したナビゲーション支援システム

宋 曦媚

高久 雅生

筑波大学図書館情報メディア研究科

筑波大学図書館情報メディア系

songximei@slis.tsukuba.ac.jp

masao@slis.tsukuba.ac.jp

**概要** ユーザーの検索意図を考慮して、ユーザーの意図が曖昧な場合であっても具体的な場合であっても、ユーザーを満足させることを目指して開発したナビゲーション支援(IntentNAVI)システムについて報告する。IntentNAVIシステムは4つのサブシステムと意図推定モジュールから構成され、サブシステムのうち、ウェブサブシステムは常に提示され、他の3つのサブシステムは個別意図(News, Shopping, QA)に対応し、クエリ意図に近いサブシステムをシステムが推定して表示する。本稿では、IntentNAVIシステムの意図推定手法に関する評価とユーザー実験の結果を示す。評価実験の結果では4つのサブシステムの全体から見ると、Webのサブシステムの結果が参加者の主観満足度が高いと言え、個別意図の3つのサブシステムのうち、Newsの課題において、Newsのサブシステムに対してユーザーの満足度が高いと言える。このシステムを通じて、ユーザー意図の把握と2画面インタフェースがユーザーに与える影響を考察する。

**キーワード** ナビゲーション支援システム, ユーザーの意図, 意図推定, 多様化検索

## 1 はじめに

現在のウェブ検索エンジンに入力されるクエリは、ユーザーの情報要求を十分に表現できないことがある。これはウェブ検索の発展とともに、ウェブ検索における曖昧もしくは不明快なクエリが多くなったためである。このようなタイプのクエリに対する技術として、最近、多様化検索の研究が盛んである。多様化検索の目標は結果リストで多くの意図を含むようにすることである[1]。一方、単一の意図を持つユーザーにとって、サーチエンジンで希望するカテゴリ(News, Shopping など)をクリックして、検索を切り替える場合がある。この検索対象を切り替えることを本稿ではナビゲーションと呼ぶ。多様化検索では多くの意図を含んでいるものが良い結果だが、単一の意図しか持たないユーザーにとって、良い結果かどうか疑問が残る。本稿ではユーザーの意図の曖昧性にかかわらず、ユーザーを満足させるためのナビゲーション支援システム(IntentNAVIシステム)を構築する。本研究ではユーザー意図推定の手法とIntentNAVIシステムを開発しユーザー実験を通じて、その有効性を評価する。評価においては以下のResearch Questionを設定し、それらに答える。

Research Question:

RQ1: IntentNAVI システムの意図推定はどの程度正確か。

RQ2: IntentNAVI のサブシステム単体はどの程度有効か。

RQ3: IntentNAVI のサブシステム単体の有効性はシステム全体の有効性にどの程度影響するか。

## 2 関連研究

本研究に関連する研究には意図推定に関する研究とインタフェースに関する研究がある。

ユーザー意図を推定する研究には、Nanba ら[2]のNTCIR12-IMine2<sup>1</sup>多様化検索のタスクで与えられたクエリに対して、サブピックのパーティカルインテントを取得する研究が存在する。この研究では特徴語(154語)を使った分類器を用いて、サブピックのパーティカルインテントを推定した。推定したパーティカルインテントはNTCIR12-IMine2 多様化検索タスクの評価結果で一番良い結果であった[3]。本稿では意図推定手法にNanbaらの手法を用いる。

多様化検索に関するインタフェースに関する研究にはIwataらの不明確なクエリに対してユーザーができるだけ多くの適切な側面を見つかることをサポートするインタフェース(AspecTiles)[4]に関する研究が存在する。このインタフェースは2画面のインタフェースから構成され、左側に各ドキュメントのクエリの各側面に関する適合性を判断し、適合の程度を色付きタイトルとして表示する。右側にBingのAPIを利用した検索結果を提示した。評価実験の結果でAspecTilesはBingより検索性能とユーザーの満足度においてよい結果を示した。本研究との違いはユーザー意図の推定や切り替えるインタフェースを持たない点がある。

<sup>1</sup> <http://www.dl.kuis.kyoto-u.ac.jp/imine2/>

### 3 IntentNAVI システムの構成

IntentNAVI システム全体の構成として、ユーザーインタフェース、サブシステム、意図推定の 3 つからなる。システム全体の概要図は図 1 のようになる。

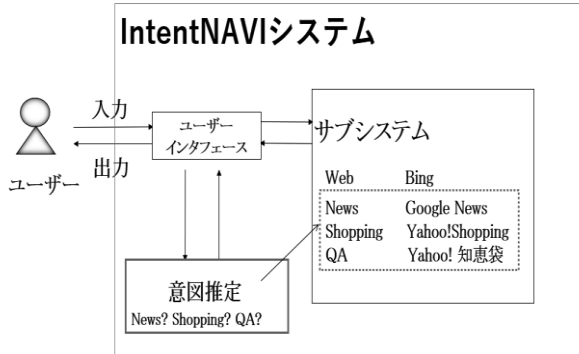


図 1 システム概要図

全体としては、サブシステムは 4 つあり、多様な検索をサポートする。サブシステムのうちウェブ検索は常に提示する。他の 3 つのサブシステムは個別意図 (News, Shopping, QA) に対応し、クエリ意図に近い 3 つのサブシステムを切り替える。利用したサブシステムは Google news<sup>2</sup>, Yahoo!Shopping<sup>3</sup>, Yahoo!知恵袋<sup>4</sup>, Bing<sup>5</sup> それぞれのサブシステムである。

4 つサブシステムそれぞれの API を通じ、クエリに関するタイトルや URL や Snippet などを結果として提示する。

意図推定モジュールはクエリに応じて、その意図を推定して 3 つのサブシステムを自動的に切り替える。

IntentNAVI システムのインタフェースの検索結果画面例を図 2 に示す。



図 2 システムの 2 画面のインタフェース

IntentNAVI システムのインタフェースは 2 画面であり、左側に 3 つのサブシステムのいずれかを表示し、News, Shopping, QA それぞれの検索エンジンを用いた。ユーザーが検索する時、ユーザーの意図を推定して、3 つのサブシステムから意図に対応するサブシステムの結果のみを提示する。ユーザー意図は News, Shopping, QA 以外の場合をサポートするために、右側に通常のサーチエンジン Bing の検索結果を常に提示している。そして、ユーザーの検索意図により IntentNAVI システムにある News, Shopping, QA それぞれのボタンをクリックして、対応するサブシステムに切り替えることもできる。システムの右側に Web ボタンもあって、クリックしてから、Web サブシステムの検索結果が更新できる。

### 4 ユーザー意図推定手法

#### 4.1 ユーザー意図推定手法の概要

Nanba ら[3]の研究では TinySVM 分類器<sup>6</sup>を用い、News, Shopping, QA 意図に対する分類結果が良いと判断されており、本研究でも Nanba らの手法を参考にする。NTCIR-12 の IMine2 多様な検索タスクで与えられた日本語の Web Corpus<sup>7</sup>から抽出した 500 ファイルを元のデータセットとして利用した。TinySVM 分類器は二値分類器であり、News, Shopping, QA の 3 つのモデルが必要である。News, Shopping, QA に関するそれぞれの特徴語を集め、解析した 500 ファイルで各特徴語の頻度を抽出した。抽出した頻度を分類器で使ったトレーニングデータとして用いて、News, Shopping, QA モデルを作成した。

ユーザーが現実のクエリを入力した後に、このクエリに対応するウェブ (Bing) の上位 10 件の検索結果を用い、各特徴語の頻度の合計を分類器の入力データとする。分類器を用いて、入力データをどのモデルに属するかを判断し、判断した結果を IntentNAVI システムに渡し、対応するサブシステムを呼び出して結果を出力する。

#### 4.2 分類器の評価

TinySVM 分類器を用いて、クエリに応じてユーザー意図の分類を行った。ユーザー意図を判断することを通じて、ナビゲーション支援システムでユーザーに支援する。分類器がよいかどうかを評価し、評価に関する結果を説明する。

ここでは、NTCIR12 の IMine2 タスク (多様な検索タスク) から与えられたクエリを現実のクエリとして評価する。NTCIR12 の IMine2 タスクで使われたクエリに関するサブトピックのパーティカルintent毎の重要度の合計から最も大きいジャンルをそのクエリの正解ジャンルとみなして評価する。評価結果は表 1 のようになる。

<sup>2</sup> <https://news.google.co.jp/>

<sup>3</sup> <http://shopping.yahoo.co.jp/>

<sup>4</sup> <http://chiebukuro.yahoo.co.jp/>

<sup>5</sup> <https://www.bing.com/>

<sup>6</sup> <http://chasen.org/~taku/software/TinySVM/>

<sup>7</sup> <http://www.dl.kuis.kyoto-u.ac.jp/imine2/dataset/>

表1 分類した評価結果

ページジャンル	Topic(個)	精度	再現率
News	15	0.359	0.933
Shopping	40	0.727	0.600
QA	45	0.679	0.422

NTCIR12のIMine2タスクの評価結果で、100件の中で15件のクエリの意図がNews、40件のクエリの意図がShopping、45件のクエリの意図がQAと判断した。TinySVM分類器で分類した結果は、100件のクエリの中で39件のクエリの意図がNews、33件のクエリの意図がShopping、28件のクエリの意図がQAと判断した。TinySVM分類器で分類した結果はNTCIR12のIMine2タスクの評価結果と比較し、分類器の結果で正解したものはNewsにおいて14クエリ、Shoppingにおいて24クエリ、QAにおいて19クエリであった。

各ジャンルのクエリ数を通じ、分類器の精度と再現率を計算した。Shoppingジャンルの精度が一番高く、Newsジャンルの再現率が一番高いと評価される。

## 5 ユーザー実験

IntentNAVIシステムの有効性を評価するため、ユーザー実験を行った。

### 5.1 実験参加者

実験参加者は筑波大学に在籍する学生10名とした。男性が6名で女性が4名で、平均年齢は23.8歳である。参加者のうちで9名が情報学、1名が数学を専攻している。

### 5.2 実験課題と方法

ユーザー実験として3つの課題を配り、News(N),Shopping(S),QA(Q)それぞれに関する3つの課題を設定し、それぞれに対する必要な情報を集めるために、IntentNAVIシステムで検索を行った。Newsの課題は熊本地震の復興に関する情報を集めることとし、Shoppingの課題は家電製品を買いたいこと、家電製品に関する情報を探ること、QAの課題は就職活動における良いエントリーシートを書くための見本などの情報を集めることとした。課題の詳細を付録に示す。

各課題の遂行時間は10分で検索の目的と合うページを見つけた際はブックマークをし、検索を行った後で、システムの有効性やユーザー満足度などを評価するアンケートを行った。各課題の遂行順序は実験参加者間でカウンターバランスを取った。

### 5.3 実験の結果

ユーザー実験中に記録したアクセスログとサーバログとスクリーンキャプチャを用いて、ユーザー行動を分析

する。分析した結果は表2のようになる。

表2 課題遂行中のユーザー行動

課題	クエリ数(件)	サブシステムクリック数				
		全体	N	S	Q	Web
News	4.1	17.1	6.8	0	0.8	9.5
Shopping	6.5	17.6	0.4	7.4	0.2	9.6
QA	3.5	16.4	2.8	0.4	3.8	9.5

表2により、全体の結果から見ると、どの課題にしても、参加者がウェブのサブシステムの結果をクリックした数は全体クリック数の半分を占めた。

News課題で参加者の平均クエリ数は4.1件(標準偏差1.85)で平均クリック数は17.1回(標準偏差6.85)である。Newsのサブシステムのクリック数は全体のクリック数で40%を占めている。Shopping課題で参加者の平均クエリ数は6.5件(標準偏差3.84)で平均クリック数は17.6回(標準偏差6.40)である。Shoppingのサブシステムのクリック数は全体のクリック数で42%を占めている。QA課題において参加者の平均クエリ数は3.5件(標準偏差1.51)で平均クリック数は16.4回(標準偏差5.50)である。QAのサブシステムのクリック数は全体のクリック数で23%を占めている。

結果のクリック数により、インタフェースの右側のウェブサブシステムの結果以外でみると、Shopping課題でShoppingのサブシステムの結果のクリック数が多く、Newsの課題でもNewsのサブシステムのクリック数も多かった。この結果により、NewsとShoppingの課題においては他のサブシステムに比べて、それぞれの意図に対応するサブシステムがよく使われると分かる。

実験のログファイルを用いて、ユーザーが検索を行った際に、システムが自動的にユーザー意図と推定した結果を表3に示す。

表3 システムが自動推定した意図の回数

課題	検索回数(平均)	システム自動判断した意図(平均回数)		
		News	Shopping	QA
News	8.1	7.1	0.2	0.8
Shopping	12.7	2	8.0	2.7
QA	10.9	3.4	1.1	6.4

表3の検索を行った回数とシステムが自動推定した意図により、News課題で平均8.1回検索のうちで7.1回Newsの意図と推定し、Shopping課題で平均12.7回検索のうちで8回Shoppingの意図と推定し、QA課題で平均10.9回検索のうちで6.4回QAの意図と推定した。

課題後のアンケートで参加者の主観評価を記入するので、問6[検索時に左側のサブシステムに示した情報

はあなたの検索意図と一致しましたか。1:まったく一致しなかった 2:あまり一致しなかった 3:少し一致した 4:完全に一致した]と問 8[左側のサブシステムでの切り替えるボタンを利用しましたか]のような質問があり、IntentNAVI システムの意図推定はどの程度正確かというRQ1に対する回答が得られた。各課題で10名の参加者の答えにより、問 6 に対するスコアの平均値を計算した。表 4 に示す。また問 8 の結果は表 5 に示す。

表 4 アンケートによる検索意図推定の一致度スコアの平均値

課題	平均値
News	3.1
Shopping	2.5
QA	2.5

表 4 の結果からみると、News の課題で News サブシステムの結果は一致度が高い。一方で、Shopping と QA の課題に対する推定した意図とユーザー意図の一致度がさほど高くない。

表 5 アンケートによるサブシステムを切り替える機能を利用した実験参加者数

課題	利用した	利用しなかった
News	3	7
Shopping	7	3
QA	7	3

表 5 の結果からみると、News 課題の方は切り替えるボタンを利用した参加者がほかの課題より少ないため、システム自動意図推定の正解率がユーザー主観評価に影響を与えた可能性がある。

RQ2 でサブシステム単体はどの程度有効かに対して、問7[左側のサブシステムを切り替える前で左側に示した情報に満足しましたか。1:非常に不満である 2:少し不満である 3:少し満足している 4:とても満足している]と問11[右側のサブシステムから示した情報に満足しましたか。1:非常に不満である 2:少し不満である 3:少し満足している 4:とても満足している]というアンケートを行った。問 7 と問 11 の結果の平均値を表 6 に示した。

表 6 サブシステムに対する満足度スコア平均値

課題	個別サブシステムの満足度	Webサブシステムの満足度
News	3.2	3.2
Shopping	2.6	3.7
QA	2.4	3.2

表 6 の結果により、参加者は左側の結果より右側の結果の満足度が高いといえる。News, Shopping の課題では個別意図サブシステムに対する満足度がやや高い。

RQ3 でサブシステム単体の有効性はシステム全体の有効性にどの程度影響するかに対する回答を得るために、問9[左側のサブシステムの中で、あなたはどのサブシステム(News, Shopping, QA)から示した情報に満足しましたか。(複数選択可能)]と問 12[ナビゲーション支援システムの全体で、あなたはどのサブシステム(News, Shopping, QA, Web)から示した情報にもっとも満足しましたか]というアンケートを行った。

表 7 個別意図サブシステムに対して満足した実験参加者数

課題	News	Shopping	QA	どれも満足しない
News	10	0	0	0
Shopping	0	8	1	2
QA	4	0	3	3

表 7 の結果により、左側のサブシステムのうち News サブシステムの結果は参加者に満足度が高いと言える。QA の課題で QA のサブシステムの結果に対する満足度があまり高くないと言える。

表 8 サブシステム全体で最も満足したサブシステム

課題	News	Shopping	QA	Web
News	5	0	0	5
Shopping	0	4	0	6
QA	2	0	1	7

表 8 の結果により、全体から見ると、ウェブサブシステムの結果が参加者の主観満足度が高いと言える。News の課題において、News の単体サブシステムの結果に対する満足度がやや高く、全体から見ると、News サブシステムの結果と Web サブシステムの結果に満足する実験参加者の人数が同じである。QA の課題において、QA のサブシステムの結果に満足する人が少ないため、右側の Web のサブシステムの結果を利用して満足する人が多い。

システム全体に対して満足度を測るため、問 13[示された検索結果は全体として見て、満足できるものでしたか。1:非常に不満である 2:少し不満である 3:少し満足している 4:とても満足している]というアンケートがある。問 13 の平均値は表 9 のようになった。

表 9 検索結果に対する満足度

課題	満足度
News	3.4
Shopping	3.0
QA	3.2

表 9 により、参加者全体の満足度は[少し満足している]と[とても満足している]の間である。右側のウェブサブシステムの結果にある程度満足したことにより、個別意図のサブシステムへの不満がある程度カバーされていることがわかる。また、News の課題のように意図推定と検索結果が良かった課題では、全体の満足度もやや向上している。

## 6 考察と今後の課題

IntentNAVI システムに対するユーザー実験から得られた結果は Web のサブシステムの結果に満足度が高く、News の課題における News のサブシステムの結果も満足度が高かった。

News のサブシステムで利用した結果は Google News の API を用いて取得した結果である。この結果にはタイトル、URL、Snippet 以外にも写真やサブタイトルなどもあり、参加者にとって、News のサブシステムのような表示が見やすかった可能性がある。

ユーザー実験の課題後のアンケートにおける自由記述コメントがいくつかあり、コメントに対して、IntentNAVI システムの改善点を考え、ナビゲーションを更に支援することを目指す。

QA の課題に対する結果があまり高くなかった理由はエントリーシートに対する質問が Yahoo!知恵袋において対応する回答がさほど多くないと考えられる。そのため、QA 課題に対して、Yahoo!知恵袋よりウェブの検索結果が良いと判断された。

Shopping の課題において、買い物を探す場合は、ユーザーにとって、買い物の値段の安い順などの順番を付ける結果リストが欲しいという参加者からのコメントがあった。商品のイメージを返したら、便利だというコメントもあった。本研究では Yahoo!Shopping の API を直接利用し、結果を取り出しているが、今後の課題として、Shopping のサブシステムの結果に商品の写真などを付け、安い順や更新順などにより、結果リストをリランクすることが考えられる。

また、左側では 3 つサブシステムがあるが、ユーザーの意図に対応するにはサブシステムの数が不十分だと考えられる。もっと多くの種類のサブシステムを入れることと、各サブシステムから得られた結果リストをユーザーのニーズに合う順番で表示すること、見やすいインタフェースに取り組むことも今回の課題として存在する。

## 7 おわりに

本研究ではユーザーの意図の曖昧性にかかわらず、ユーザーを満足させるためのナビゲーション支援システムを構築した。ナビゲーション支援システムを開発手法と IntentNAVI システムの有効性に評価するためのユーザー実験の実施を説明した。

IntentNAVI システムの意図推定の手法を利用して、予備調査で分類器の評価を行って、Shopping のジャンルの結果の精度が一番高く、News のジャンルの結果の再現率が一番高い。

ユーザー実験を通じて、IntentNAVI システムが News の課題に対して、自動的にユーザーの意図を推定した正確率が一番高いと言える。4 つのサブシステムの全体から見ると、ウェブのサブシステムの結果が参加者の主観満足度が高いと言え、左側の 3 つのサブシステムのうち、News 課題においては News のサブシステムに対して、ユーザーの満足度が高い。一方で、QA のサブシステムの満足度とクエリ意図との一致度が高くない。システム全体として、参加者は個別意図のサブシステムの結果に満足する場合は個別意図のサブシステムの結果とウェブのサブシステムの結果を両方利用する。個別意図のサブシステムの結果に満足しない場合は切り替える機能を利用する一方で、個別意図のサブシステムの結果に満足しないで、ウェブのサブシステムの結果を利用するようになる。

今後の課題として、個別意図の要求に合わせて検索結果を並べ替えたり、追加の情報を表示するなどの工夫が考えられる。

## 参考文献

- [1] D. Rafiei, K. Bharat, and A. Shukla. Diversifying web search results. In Proceedings of the 19th International Conference on World Wide Web, WWW '10, pages 781–790, 2010.
- [2] Hidetsugu Nanba, Tetsuya Sakai, Noriko Kando, Atsushi Keyaki, Koji Eguchi, Kenji Hatano, Toshiyuki Shimizu, Yu Hirate, and Atsushi Fujii. NEXTI at NTCIR-12 Imine-2 Task. IN Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, pages 27-30, June 7-10, 2016.
- [3] Takehiro Yamamoto, Yiqun Liu, Min Zhang, Zhicheng Dou, Ke Zhou, Ilya Markov, P.Kato, Hiroaki Ohshima, Sumio Fujita. Overview of the NTCIR-12 Imine-2 Task. IN Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, pages 8-26, June 7-10, 2016.
- [4] Mayu Iwata, Tetsuya Sakai, Takehiro Yamamoto, Yu Chen, Yi Liu, Ji-Rong Wen. AspectTiles: Tile-based Visualization of Diversified Web Search Results.

Proceedings of SIGIR '12, Pages 85-94. 2012.

## 付録

### A:News の検索課題

2016年4月14日に、熊本県と大分県で相次いで地震が発生しました。気象庁震度階級では最も大きい震度7を観測する地震が4月14日夜および4月16日未明に発生したほか、最大震度が6強の震度が2回、6弱の震度が3回発生しました。一連の地震回数(M3.5以上)は内陸型地震では1995年以降で最多となりました。熊本地震は被害が拡大し、復興には時間がかかりそうな状況です。被災地が日常を取り戻し、復興を遂げると、被災された地域への観光客が戻ってくることを期待できます。

あなたは、ある出版社でニュース雑誌の編集者として働いています。あなたは編集長から、このニュースを掘り下げる記事を書くための情報を集めるよう指示を受けました。編集長は、1時間後に行われる他の部門の編集長との編集会議で記事について話し合いをする予定です。編集長は、会議でそれぞれの記事で採り上げる内容を示したいと考えています。あなたは、現在まで熊本地震以降5ヶ月にわたって復興の作業はどこまで順調に進んでいるかを示したいと考えています。そのため、熊本地震の被害の様子や復興状況に関する多様な情報を取る必要があります。

### B:Shopping の検索課題

毎年の四月に、大学生が入学する時期になります。出身地と大学の地域が違う場合は、アパートやマンションなどを借りる必要があります。アパートで冷蔵庫やテレビなどの必要な家電製品を提供しない場合が多いです。

あなたは、北海道の出身で四月に筑波大学に入学する予定です。筑波大学の近くでアパートを借りました。アパートへの入居は二日後の予定で、このときまでに冷蔵庫などの家電製品を届けることを保証するために、早く買う必要があります。そのために、インターネットで家電製品の値段やサイズや仕様などに関する情報を多様な探す必要があります。

### C:QA の検索課題

毎年、就職活動が行われています。学生が良い会社に就職するために、会社の要求に応じた書類の審査が必要です。書類選考の課題として提出を求められる採用書類をエントリーシートと呼びます。エントリーシートで自分の長所や人に負けない強みなどをうまく書けば、就職活動の役に立ちます。

あなたは、就職中の大学四年生です。ある会社の説明会に行きたいので、明日までこの会社にエントリーシ

ートを提出しなければなりません。エントリーシートで自己PRや自分の強み、弱みなどを書く必要があります。良いエントリーシートを書くために、内容の書き方や見本、注意点などの多様な情報を探します。