

ラダリング型検索サービス 「ラダサーチ®」

下畑 さより
池野 篤司

北村 美穂子
折原 幾夫

介弘 達哉
村田 稔樹

インターネットが普及し、そのインフラコストが小さくなったことにより、ネット上にさまざまなサービスやコンテンツが提供されるようになってきている。多様化・細分化された情報はユーザーの多様なニーズに応えるという側面があるが、一方で、情報の氾濫を招き、多くの情報が閲覧されることなく埋もれたままになるという問題をはらんでいる。

サービスやコンテンツを探し出す手段としては、GoogleやYahoo!に代表されるキーワード検索が一般的である。しかし、このような検索サービスでは、ユーザー側に適切なキーワードを選択するスキルと膨大な検索結果から求める情報を探し出す労力が要求される。したがって、キーワードが明確な場合や情報が一般的な場合には有効な手段であるが、条件が複雑な場合や、漠然としたニーズを探し出す場合には適さない。たとえば、キーワードが思いつかなかったり、欲しい情報の具体的なイメージがなかったりする場合には、求める情報にたどり着くことは困難である。

そこで著者らは、対話によってユーザーのニーズを明確にし、多くの情報の中からユーザーの求めるサービス・コンテンツを探し出して提供するラダリング型検索サービス「ラダサーチ®」*1)を提案する(注1)。ユーザーとの対話を行うシステムとしてダイアログナビ¹⁾、対話型意見収集システム²⁾などがある。これらのシステムと比べて「ラダサーチ®」では、ユーザー回答およびマッチング状況に応じて対話の流れを制御し、ユーザーのニーズを多角的に引き出していくという特徴がある。以下では、「ラダサーチ®」の概要を紹介し、実際にシステムを構築して行った実証実験の結果について紹介する。

システム概要

ラダリング型検索サービスは、カウンセラーやファシリテーターと同様に、ユーザーに話しやすい環境を提供し、対話を通じてユーザーが考えを整理し、新しい価値観を創造していくプロセスを支援する。そして、そのプロセスを通して得られた情報を用いて検索を行い、ユー

*1) ラダサーチは、沖電気工業株式会社の登録商標です。

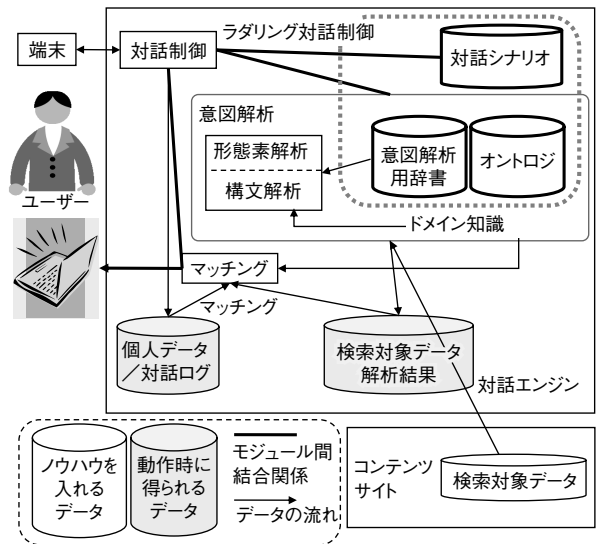


図1 ラダリング型検索サービスの概観

ザーが単独では獲得することのできなかつた、真のニーズにマッチした情報を提供することを目的とする。

本サービスの概観を図1に示す。本サービスで用いる対話エンジンは、検索対象となるドメインの言語知識を記述したドメイン知識部、発話文を解析しユーザーの意図を理解する意図解析部、ユーザー意図と対話シナリオに基づいて対話アクションを決定する対話制御部、そして、対話によって獲得した情報にマッチしたデータを抽出するマッチング部から構成される。

TIP

【基本用語解説】

ラダリング

マーケティング調査などにおいて活用される対話手法で、質問を投げかけ、その回答に応じてさらに問いかけを繰り返すことにより、ユーザー自身が気づいていないニーズや価値観を引き出していくことを特徴とする。

(1) ドメイン知識

ユーザーの発話内容を理解し、スムーズに対話を進めるためには、対象となるドメインを網羅する詳細な言語知識が必須である。「ラダサーチ[®]」では対話および検索に必要なドメイン知識を以下のように定義し、データベース上に構築する。

a) テーマ

検索対象ドメインにおいて、聞いておくべき典型的なトピックとその順序を定義したもの。

b) シナリオ

ユーザーとシステムとの対話のルールを記述したもので、一般シナリオと特別シナリオの2種類がある。前者は、ユーザーから獲得すべき情報の種類と、それを聞き出すための質問文、および、ユーザー回答に応じたアクションを定義したもので、後者は、意図解析結果によって起動されるアクションを定義したものである。

c) オントロジ

検索対象ドメインに関する概念クラスを網羅的に収集し、体系化したもの。木構造で表現され、第1階層に概念全体を示すクラスを、第2階層以下に広範な概念から詳細な概念へとサブクラスを階層的に記述する。各クラスはその概念を示す具体的な語彙表現と対応付けられている。

(2) 意図解析

ユーザー発話文を解析し、システムの質問文に対する回答値（クラス名）、および、ユーザーの発話意図を獲得する。本システムにおける発話意図とは、「はい（肯定）」、「いいえ（否定）」、「わかりません（曖昧）」といったユーザーの状態や心情である。発話意図を理解するために、意図解析では、一般的な形態素解析や構文解析に加えて、表現の強度や極性判定といった意味表現の解析処理も行う。

(3) ラダリング対話制御

予め設定したテーマの流れに沿ったシナリオ選択を基本として、シナリオ内のルールに基づいて対話を進行させる。具体的には、ユーザー発話文を意図解析した結果に基づき、次の5つのいずれかの処理を行う。

- a) ある属性の一般シナリオを継続する。
- b) 異なる属性の一般シナリオに遷移する。
- c) 特別シナリオを呼び出す。
- d) ある属性に関する一連の質問は終了し、シナリオ優先順位に従って、次の属性のシナリオを継続する。
- e) ある属性に関する一連の質問は終了し、マッチングから新しく情報要求（属性）を受け取る。

さらに対話制御モジュールは、1つのシナリオが終了する度に、対話から獲得した情報を個人データに書き込む。

(4) マッチング

検索対象のデータ群と個人データとのマッチングを行い、検索結果を出力する。さらに、検索結果のデータ数を最適にするため、次に聞くべき属性を対話制御に与え、対話の流れを指示することができる。また、対話の終了を判断するのもマッチングモジュールの役割である。

ドメイン知識の開発

ラダリング型検索サービスの適用事例として、「転職者向け職業紹介サービス」を対象とするシステムの構築を行った。「転職」を対象としたのは、転職先を決定する要因が、職歴や転職理由、転職先への希望など多岐に渡っており、キーワード化が難しいドメインであると考えたからである。以下では、ドメイン知識構築の過程について、具体的な事例を用いて説明する。

(1) テーマとシナリオの検討

ラダリングを行ううえでまず重要なのは、自然な対話の流れで、状況に応じた適切な質問を繰り返し、ユーザーの価値観創造プロセスを支援することである。このプロセスの構築には、ドメインエキスパートによる知識の表出、および、その体系化が不可欠である。

著者らは、キャリアカウンセラーとの議論を繰り返し、そのノウハウに基づいて転職ドメインにおける12の主要テーマを設定した。さらに、それぞれのテーマの代表シナリオに対して、聞き出したい情報と、その情報を聞き出すための質問文を作成し、ユーザーの回答から情報を抽出する際に参照するオントロジを対応付けた。次に、これらのテーマに優先度を設定することで大局的な対話の流れを定義した。さらに、それぞれのテーマに対して、予測されるユーザーの発話を分類し、分類ごとに次のアクションを記述することで、細かな対話の流れを制御できるようにした。

(2) オントロジの体系化

ドメインに関する対話を行う上で必要な概念を洗い出し、その体系化を行った。概念の洗い出し、および、体系化の工程は、ドメインエキスパートとの議論を重ね、システム設計者が取りまとめた。転職ドメインでの対話に出現する可能性のある語彙を網羅するため、概念は職種、制度、転職理由、趣味、将来のビジョンなど31種類となった。また、概念ごとに階層の深さ、網羅する範囲を定め、

表1 実証実験における対話例

シナリオ名	システム発話文	ユーザ回答	取得属性
仕事の経験	今どういったお仕事をされていますか?	書籍販売、図書館司書をしていました。	現在の職種:図書館司書 所有する技能:書籍
仕事の実績	わかりました。書籍、音楽、販売の経験をお持ちなのですね。現在あるいは過去で、仕事で評価されたことや褒められたことを教えてください。	まじめな勤務。接客態度。	仕事上でのあなたの強み:まじめさ 所有する技能:接客
:	:	:	:
希望する仕事	どのような仕事を希望しますか?	まだよくわからない	希望する仕事:不明
仕事上で強みが発揮される時	よくわからない、ということですね。では、仕事をする上での強みや何か得意分野について、すでにお話しいただいています。ではそうしたあなたの強みが発揮された結果、何がどう変わりましたか?	お客が親しみを持って来店して下さるようになった。	仕事上で強みが発揮される時:お客が……
:	:	:	:
仕事上での課題	あなたが仕事をする上で苦手と思っていたり、自分自身の課題だと考えていることがありますか?	売上など、実際のデータを分析し仕事に活かすこと。	仕事上での課題:売上など、実際の……
課題を指摘する人	そうですか。あなた自身の苦手な分野や課題を、あなたの周りで指摘してくれる人はいますか?	いない。	課題を指摘する人:なし
課題が生じる理由	ではその苦手分野や課題は、どういった理由で生じていると思いますか?	やり方がわからないから。	課題が生じる理由:やり方が……
課題を解決したいか	苦手分野や課題を、あなた自身は解決したと思いますか?	思う。	課題を解決したいか:思う
課題の改善方法	わかりました。それでは苦手分野や課題に対して、どうしたら克服、あるいは改善していけると思いますか?	情報収集する	課題の改善方法:勉強する
	勉強するということですね。		

基本的な語彙のピックアップとその上位下位関係の定義を行った。

(3) 概念の語彙化

オントロジの各クラスについて、概念を示す表現としてユーザー発話に出現する可能性のある語彙を網羅的に収集し、クラスに対応付けた。語彙の収集は、既存のデータの利用（資格表、類義語辞典など）、テキストデータからの自動抽出、予備実験および実証実験の対話ログからの不足表現の抽出などにより行った。語彙の追加は、基本的に作業者が行った。これは、一般的には同義語とみなされない表現でも、特定の分野や文脈によっては同じ意味に解釈される場合があり、自動化が困難だったからである。また、不適切な語彙の削除や上位・下位関係の修正は、対話ログを参照して随時実施した。

これらの作業により、最終的に11,200クラス、42,330語彙を持つオントロジを構築した。

実証実験

ラダリング型検索サービスの基本機能、および、ドメイン知識を構築し、実際の転職希望者を対象に、実証実験を行った。実験は、実証実験サイトにアクセスし、システムとの対話を行う形式を取った。検索対象として実求人データ約3,000件を使用し、サービス利用前と利用後にアンケート調査を実施した。実証実験における対話例を表1に、実験の結果を図2にまとめる（注1）。

まず、平均滞在時間約17分（注2）、平均対話数33.6回と、多くのユーザーが時間をかけてシステムとの対話に取り組んだことが分かる。また、獲得された属性値数が

- 参加人数
 - ▶ 2月20日～3月4日 803人

- 実証実験による調査の結果
 - ▶ 設定質問終了率 82.7%
 - ▶ 全体としてスムーズに対話できたか (①50.0% ②22.6% ③27.4%)
 - ▶ 対話や質問の流れを自然に感じたか (①57.3% ②21.7% ③21.0%)
 - ▶ 得られたキーワードに納得できたか (①42.6% ②29.7% ③27.7%)
 - ▶ 対話で気づきを得られたか (①24.3% ②36.9% ③39.0%)

①とてもそう思う、そう思う
②どちらともいえない、無回答
③あまりそう思わない、まったくそう思わない

- 平均の姿
 - ▶ 滞在時間 16分54秒
 - ▶ 属性取得数 32.31個
 - ▶ 対話数 33.60回
 - ▶ 入力文認識率 88.3%

図2 実証実験の結果

対話数にほぼ匹敵していることから、システムとの対話が十分に機能していたと考えられる（注3）。アンケート調査の結果からも、スムーズな対話、自然な対話の流れ、キーワードの納得性などの項目で、肯定的な評価が否定的な評価を上回り、有効な対話が行われたことを裏付けるものとなっている。

その反面、対話により気づきを得られたかという設問については評価が低く、対話が新たな発見や認識に結びついているユーザーは24%にとどまっている。この評価は、シナリオの質および量が十分ではなく、ユーザー回答に対する反応が単調であったため、ドメイン知識の拡充が必要であると考えている。さらに、対話履歴や取得した情報、検索結果の表示方法を工夫することで、気づきを与える効果が期待できるのではないかと考える。

次に、ラダリング型検索サービスをGoogleやYahoo!

表2 既存の検索サービスとの比較による
ラダリング型検索サービスの特徴

項目	割合
時間がかかる	48%
面倒だ	36%
楽しい	31%
非効率的だ	31%
意外性がある	24%
親しみを感じる	22%
可能性を感じる	20%
気づきがある	16%
興味深い	16%

のような既存の検索サービスと比較してどう感じるかを複数選択形式で尋ねた結果を表2に示す。

ユーザーは既存の検索サービスと比較して、「楽しい」「可能性がある」といった内容で肯定的に捉えながらも、「時間がかかる」「面倒な」ものとして、本対話エンジンをイメージしていることが分かる。

冒頭にも述べたとおり、検索の目的や対象が明確である場合には、従来型の検索サービスにより迅速に情報を入力する方法が適している。職業紹介サービスにおいても、すでに転職先の職種や条件が明らかな場合は、既存の転職サイトにある条件検索などのインターフェースで十分であろう。しかし、希望の転職先のイメージがあいまいであったり、転職を悩んでいる段階であったりする場合には、対話によって考えを整理し、ニーズや価値観を具体化し、その結果を用いて検索していく提案の方法が適している。従来型の検索サービスとラダリング型検索サービスは相反するものではなく、シームレスに連携し、ユーザーの状況や検索対象に応じて使い分けられるようにしていくべきであろう。

最後に、ラダリング型検索サービスを今後使いたいについてのアンケートについては、使いたいと回答したユーザーは、「現在使いたい」が49.8%、「1年後」が61.8%、「3年後」が64.4%と、年を追うごとに伸びており、ラダリング型検索サービスへの期待が伺えた。また、使いたいシーンについての自由回答では、グルメやファッションなどのレコメンドから、年金や法律のコンサルティングまでさまざまなシーンが挙げられた。1年後、3年後の期待が高いことや、具体的な利用シーンが多く意見されたことは、従来のキーワード検索では未充足な、言語化できない情報を検索したいという潜在ニーズの現れであり、ラダリング型検索サービスがそうしたニーズに応えるものとして期待されていることを示唆している。

注1：実験は約1,000人の転職希望者に参加を呼びかけ、821人のユニークユーザーが参加した。本論文では821人のうち、1回以上の対話データが得られた803人について分析を行う。

注2：滞在時間は対話の開始時間から最後の対話を行った時間の差分を計算した。なお、5分以上の中断時間（入力のなかった時間）は、滞在時間から削除している。

注3：基本的に本システムでは、対話と属性値取得は1対1で行われるが、1対話で複数属性値が獲得されたり、逆に、発話内容がオントロジとマッチングしない場合があるので、対話数＝取得属性値数とはならない。

まとめ

本稿では、ドメイン知識の設計、および、開発を中心に、ラダリング対話を実現する検索サービス「ラダサーチ®」を紹介した。また、実証実験の結果から、「ラダサーチ®」が、対話によってユーザーが求める情報を探し出すためのキーワードを取得できることを示した。

現在「ラダサーチ®」は開発の第2段階に入り、実証実験で明らかになった課題の解決、および、さらなる対話品質の向上に取り組んでいる。今後は、ユーザーの求める情報を探し出すだけでなく、ユーザーに新しい視点や価値観を提示できる検索サービスを目指して、システムの改良を続けていく予定である。

本研究は、経済産業省平成19年度情報大航海プロジェクト「ラダリング型検索サービスの研究開発」の一環として、株式会社リクルートと共同で行われた。◆◆

参考文献

- 1) 清田陽司, 黒橋禎夫, 木戸冬子: “大規模テキスト知識ベースに基づく自動質問応答 –ダイアログナビ–” 自然言語処理, Vol.10, No.4, pp.145–175, 2003年
- 2) 丸元聡子, 鈴木泰山, 大塚裕子, 伊藤裕美, 乾孝司, 奥村学: “空港計画における対話型意見収集システムの実装と課題”, 言語処理学会第14回年次大会, pp.5–8, 2008年

筆者紹介

- 下畑さより: Sayori Shimohata. システムプラットフォームセンタ 技術第二部
 北村美穂子: Mihoko Kitamura. システムプラットフォームセンタ 技術第二部
 介弘達哉: Tatsuya Sukehiro. システムプラットフォームセンタ 技術第二部
 池野篤司: Atsushi Ikeno. システムプラットフォームセンタ 技術第二部
 折原幾夫: Ikuo Orihara. システムプラットフォームセンタ 技術第二部
 村田隼樹: Toshiki Murata. システムプラットフォームセンタ 技術第二部