

59 対話を用いた情報検索手法の有効性について

鹿島啓吾* 荒木健治 栢内香次
(北大工)[†]

1 はじめに

最近、WWWが急速に普及してきており、計算機による情報検索を行う機会が増大している。また、WWWの情報源は量が膨大であるため、ユーザが自分のほしい情報を探し出すことが困難となってきた。またユーザは自分のほしい情報に対する認識があまり定まっていない状態で検索を行い、結果を見ることにより曖昧であった認識をはっきりさせ、その上でほしい情報を決定することも多い。

現在、Web上の情報を扱っている研究は様々行われている[1]、[2]、[3]。これらの研究はあらかじめ対象とする情報の種類を絞った限られた範囲での情報の検索、抽出などを目的としており、前述した情報検索の時に起こる問題を解決するための研究は少ない。また、対話を通してユーザの要求に答えるシステムの研究は、あらかじめ対象の範囲を絞りユーザの入力に対して応答文を生成する規則が組み込まれている解析的なアプローチでなされている[4]。この手法ではあらかじめ想定しておいたことにしか対処できないなどの問題点がある。

そこで本研究ではユーザが欲しい情報に対する認識がはっきり決まっておらず、どのようなものを探すか曖昧な時に対話を通してユーザの意思決定を支援し、対話の中からキーワードを獲得してそれらをもとにWeb上の情報の検索を行い、そのページを提示するという手法を提案する。本手法は対話を行いながらシステムがユーザの欲しい情報に関する曖昧な認識を具体化させる上で有効な知識を獲得して行く学習型の手法である。ここでいう対話とは主にシステムがユーザに質問をし、それに対する返答から欲しい情報を具体化してゆくという、目的をもった対話である。システムが知識を獲得、蓄えて行くことによりユーザの意思決定を支援する対話を行うことができるようになる。

一方でシステムの発話に対しユーザが返答をするために必要な知識を持たないために答えることができない状況が起こってくると考えられる。この問題を解決するために、本手法ではシステムが蓄えている知識を元にWeb上の情報を検索して、ユーザが返答するとき必

要となる知識、情報が獲得できないかということも検討した。このようにWeb上の情報を利用することにより、ユーザの返答を支援することができるようになると考えられる。

本稿では以上のことを踏まえ、Web上の情報を参照しながら対話を行い知識を獲得してユーザの欲しい情報に関する曖昧な認識を具体化しその情報をWeb上で検索、獲得する手法について2章では処理の概要、3章では各処理部の説明、4章では実験とその結果、考察について述べる。5章はまとめである。

なお今回キーワードを決定した後の検索エンジンとして“Yahoo!” [5]を使用した。

2 概要

まず本手法はシステムの「何を探していますか?」という発話より対話が始まる。その発話に対するユーザの返答から知識獲得部であらかじめ与えておいた規則に基づいて知識を獲得する。また、この時ユーザがシステムに蓄えられている概念の知識をもとにWeb上の情報を検索し、返答するときに参考となるようなページを表示させる。その後は応答文生成部でこの発話に続くシステムの発話が生成され、この処理を繰り返して行う。このような流れに沿って本手法は処理を行いユーザの欲しい情報に対する曖昧な認識を具体化して、その情報をWeb上で提示できるようにする。

3 処理過程

3.1 知識獲得部

ここでは対話からの知識獲得方法について述べる。まずシステムにあらかじめ文のひな形を複数与えておく。これらのひな形はユーザの要求する情報を絞る上で有効である概念の階層構造を作成するためのひな形を用意した。文生成にはスロット法を用いている[6]。この手法は限られた場面、目的の発話では限られた型の文型を用意しておき、そのスロットに入れるべきデータをうまく用意することで目的を達することができる。本システムの対話は目的をもった対話であるのでスロット法を利用した。この文のひな形とひな形のスロットに入る単語からその発話に対するユーザの返答により獲得できる

*kashima@media.eng.hokudai.ac.jp

[†]札幌市北区北13条西8丁目北海道大学工学部

知識がどのようなものであるかの規則をひな形と共に与えておく。今回はシステムがまだ未完成であるためユーザーの発話は単語で行っている。将来的にはユーザーの発話を自然言語を用いた文で行えるようにするべきであると考えている。このような規則をシステムに与えておく事によりユーザーとシステムの対話から知識を獲得する。対話自体がユーザーの欲しい情報を具体化する方向であるので得られる知識は対象を分類する時の概念の階層構造である。図1に知識獲得の例を示す。図1のようにひな形とともに規則を与えておくことにより知識を蓄えて行く。

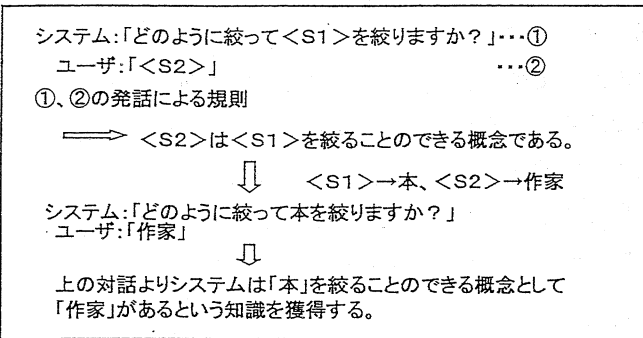


図1 知識獲得の例

3.2 応答文生成部

あらかじめシステムに対話の流れを与えておく。この流れはシステムの問いに対するユーザーの返答で制御される。この流れに沿ってひな形を決定しさらに獲得している概念の階層構造の知識を利用して応答文を生成する。例として「さらに<S1>を絞りますか?」というシステムの問いに対してユーザーの返答がYESなら<S1>をさらに絞るための文のひな形、NOなら<S1>に関する質問を終わらせるひな形というようにひな形を決定して対話を行う。このように知識として獲得している概念の階層構造を利用することによってユーザーに対して情報の絞り方の例を示したり一覧を示したりなどユーザーの曖昧な認識を具体化することを助ける発話を行うことができる。図2に蓄えられた知識とひな形による例を示す。この例では本に関する知識とひな形によってユーザーが何か本を探している時にどんな本を探しているのかを具体化する発話を行う例を示している。図2の発話例が示すように本を絞る上で有効である発話が行うことができる。

3.3 Web上の情報の検索について

本手法ではユーザーのほしい情報に関する曖昧な認識を具体化するための対話でシステムの問いに答えると

きにWeb上の情報を表示することによりユーザーが返答するとき参考ができるようにしている。この時の検索のキーワードには対話での話題のことおよびシステムに蓄えられている概念の階層構造を利用している。よって対話によって蓄えられた知識が多いほどこの検索時のキーワードの選択の幅が広がるのでユーザーの返答の参考となるページを提示できるキーワードを含む確率が大きくなる。なお、今回はこの処理は手動で行った。

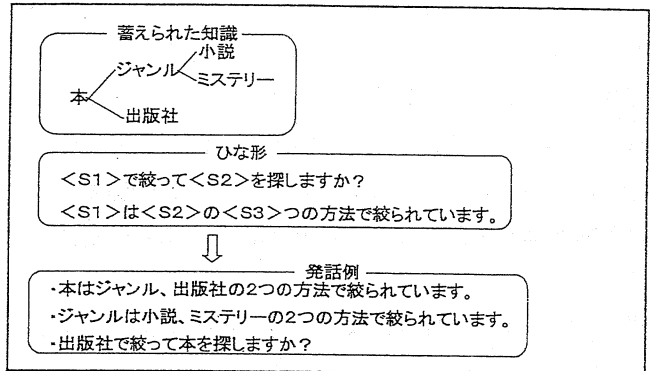


図2 応答文生成の例

例として図2で示される知識がシステムに蓄えられていたとする。このときユーザーが本を出版社で絞りたいが出版社名がわからないとする。この時にWeb上で本の出版社に関する情報の獲得を目指す。この時蓄えられている知識である「本」、「出版社」をキーワードとして検索すると図3の結果がWeb上から獲得できる。

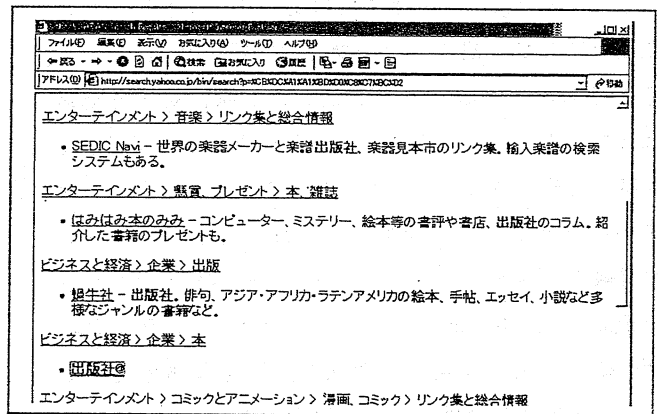


図3 「本」「出版社」で検索した結果

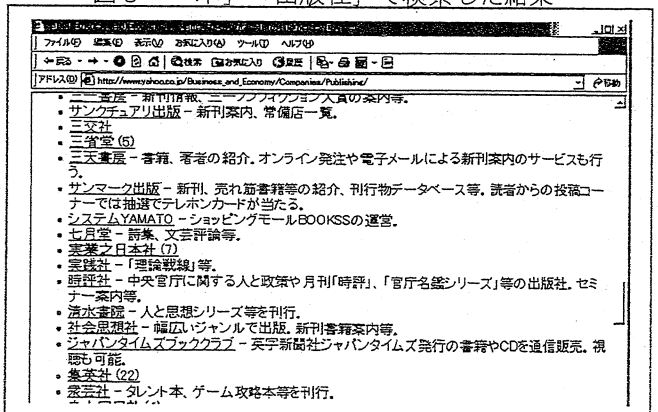


図4 「出版社@」で示されるページ

さらにこのページから「出版社@」という箇所をクリックしてページを獲得すると図4のような結果が得られ出版社の名称が掲載されているページが獲得できる。このページを見ることによってユーザは本の出版社の名称を答えることができる。今回どのページを見るかという処理は手動で行った。一つ一つこのページの内容を見て自分の欲しい情報が掲載されているページが「出版社@」であった。今後はこの処理を自動化するためページ選択の知識、ルールを考慮する必要がある。このようにWeb上の情報を利用することによりシステムの発話に返答することができユーザの意思決定を支援できる。

以上のことを踏まえて対話を行いユーザの曖昧な認識を具体化してその情報が掲載されているWeb上のページを検索、表示させる。

4 実験

4.1 実験の評価について

本手法の性能をどのような方法で評価すべきかはそれほど明白ではない。それは獲得した情報の有用性、無用性に関してはユーザー一人一人の主観によることやWeb上の情報は日々更新されているのでWebページの取得の結果には再現性がないことによる。

今回検索エンジンとして用いた“Yahoo!”はカテゴリが存在しすべてのページはどこかのカテゴリに属するという特徴を持つ検索エンジンである。

4.2 実験方法

以上のことを踏まえて今回はほしい情報が漠然としすぎていてそれだけの情報では満足な情報が得られないユーザの要求を想定して本システムの流れに沿って処理を行い満足するページが得られたかを“Yahoo!”のカテゴリに存在するものとし無いものを比較して評価した。

4.3 実験結果

表1 実験結果

	成功	失敗
カテゴリとして存在するもの(「映画」など)	8回	4回
カテゴリとして存在しないし無いもの(「本」など)	3回	9回

実験結果は上の表1のようになった。結果が示すように、はじめの漠然とした要求で求められている対象が今回検索エンジンとして用いた“Yahoo!”のカテゴリとして存在するかしないかということで結果が大きく異なった。カテゴリに存在するときにはユーザの漠然とした要求を具体化して満足できるページを獲得する

成功率は高いがカテゴリに存在しない場合には成功率が低かった。

4.4 考察

このような結果になった理由を考察する。これはカテゴリに存在する概念を具体化してゆこうとした場合はシステムに知識が蓄えられていない状態でも検索のキーワードとしてこれを入力すると必ず一つ以上のカテゴリに一致するので具体化するための例などが参照できることが多い。このことによりユーザに知識がなくてもシステムの質問に答えて対話を続けることができ最終的に欲しい情報を具体化することに成功するからであると考えられる。

S* : →システムの発話
U* : →ユーザの発話
S1 : 何を探していますか?
U1 : 映画
S2 : 映画の絞り方の一覧を示しましょうか?
U2 : はい
S3 : 映画はジャンル 主演俳優 製作国の3つの方法で絞ることができます。
S4 : どのように絞って映画を探しますか?
U3 : ジャンル
S5 : さらにジャンルを絞ってみますか?
U4 : はい
S6 : どんなジャンルですか?
U5 : アニメ
S7 : さらにアニメを絞ってみますか?
U6 : いいえ
S8 : 他の方法で絞りますか?
U7 : はい
S9 : どのように絞って探しますか?
U8 : 監督
S10 : さらに監督を絞りますか?
U9 : はい
S11 : 宮崎 駿

図5 実際の対話例の一部

逆にカテゴリが存在しない場合にはユーザがシステムの質問に答える時に参照できるページが表示されることが少ない。このことにより対話を続けて行くことが困難になりキーワードとなりうるものの候補が少ない中でユーザの要求を満たす情報を検索しなければならぬ状況に陥るからだと考えられる。また、今回の実験においてページ獲得の時の誤りも起こった。一番起こった誤りの種類は例えば「本」をキーワードとして検索した時に「日本」や「本当」といった「本」という文字そのものが入った自分がまったく意図していないページが提示されてしまうことであった。これらのことを解決するために同義語、関連語などの情報を導入して検索の幅を広げることや「本」と共起するような言葉、例えば「読む」などをキーワードとして加えて自分が全く意図していないページを提示させないようにするなどの必

要があると考えられる。実験量が少ないので推測ではあるが、対話を繰り返し行うことでシステムに知識が蓄えられるのでカテゴリに無い事柄についても情報の獲得率は上がってゆくと考えられる。図5に実際に行われた対話例の一部を示す。

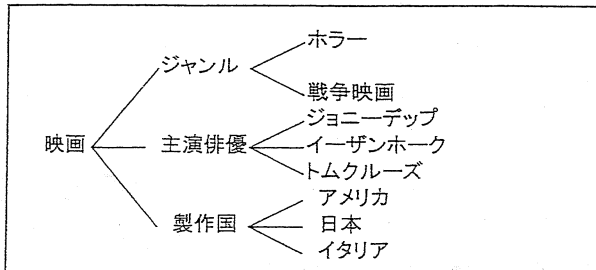


図6 図5で示される対話が行われる前にシステムに蓄えられていた映画に関する知識

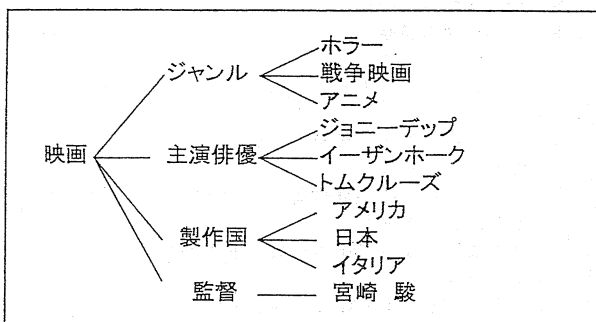


図7 図5で示される対話が行われた後にシステムに蓄えられている映画の知識

図5で示される対話例ではユーザの漠然と映画を探して対話を通して曖昧であった対象が具体的にたて行く対話である。この対話が行われる前にシステムに蓄えられている知識は図6で示されるものである。蓄えられている知識を利用することにより図5の中のS2、S3の発話が行われている。これらの発話はユーザが欲しい映画の情報を絞る時に有効な発話である。またS6に対するU5の返答から映画のジャンルにはアニメという概念があること、S9に対するU8の返答から映画は監督から絞ることができるという知識を獲得している。この結果この対話が行われる前のシステムの知識は図6で示されていたが対話の後には図7で示されるようになりこの対話で獲得したものが加わった新しい知識を持つことになる。ユーザは初めは漠然と映画を探していたがこの対話を通してジャンルはアニメで監督は宮崎駿である映画を探すというように自分の欲しい情報を具体化させている。

5 まとめ

本稿では対話を用いてユーザの欲しい情報に対する曖昧な知識を具体化してそのページをWeb上で検索、

提示するシステムについて述べた。最終的な情報の提示に関しては結果が示すように“Yahoo!”のカテゴリに存在するものに対して本手法を用いた時の成功率は12回行って8回成功であった。これは手動で行った処理もあるのでシステムとしては不完全であるが、システムと対話を行うと知識が蓄えられてゆくことから考えて本手法は有効であると考えられる。カテゴリに存在しないものに対しては12回行って3回成功と結果が良くなかった。この結果はカテゴリとして存在しないと具体化するための情報が得にくいことや実験量不足のため蓄えられている知識が少ないこと原因と思われる。実験の量を増やしてゆけば蓄積される知識が増加するので、推測ではあるがカテゴリに存在しないものに対してもユーザが満足するようなページを提示できるのではないかと考えている。また最終的なページの獲得の成功、失敗により概念の階層構造の知識に対してフィードバック処理を行う必要があると考えられる。この処理でユーザの欲しい情報が獲得できるキーワードを残してゆく処理を行うことにより、情報検索に有効な概念の階層構造ができてゆくと考えられるからである。また、今回手動で行ったWeb上の情報獲得のため処理の自動化やユーザの発話を自然言語の文でできるようにすること、さらに他の検索エンジンを利用した時の結果についても考慮して行く必要があると考えられる。

参考文献

- [1] 佐藤理史：ワールドワイドウェブを利用した住所探索，情報処理学会論文誌，Vol42，No. 1，pp. 59-67，2001
- [2] 山田洋志，福島俊一，松田勝志：Webページからのタイプ別情報抽出・分類方式，情報処理学会研究報告，2000-FI-57，2000-NL-136，pp. 143-150，2000
- [3] 岡部正幸，山田誠二：関係学習を用いた対話型Webページ検索，人工知能学会研究会資料，SIG-FAI-A003-13，pp. 65-70，2000
- [4] 日笠亘，藤井綱貴，黒橋貞夫：入力質問と知識表現の柔軟なマッチングによる対話的ヘルプシステムの構築，情報処理学会研究報告，99-NL-134，pp. 101-108，1999
- [5] “Yahoo!” <http://www.yahoo.co.jp>
- [6] 長尾真：自然言語処理，岩波書店，1996