

## ラダリング型検索サービスのためのドメイン知識構築, 及び, 実証実験

下畑 さより<sup>†</sup> 北村 美穂子<sup>‡</sup> 介弘 達哉<sup>‡</sup>

池野 篤司<sup>‡</sup> 坂本 仁<sup>‡</sup> 折原 幾夫<sup>†</sup> 村田 稔樹<sup>†</sup>

沖電気工業株式会社 ユビキタスサービスプラットフォームカンパニー

<sup>†</sup> 〒335-8510 埼玉県蕨市中央 1-16-8 OKI システムセンター

<sup>‡</sup> 〒541-0053 大阪市中央区本町 2-5-7 大阪丸紅ビル 4F

E-mail: {shimohata245, kitamura655, sukehiro564, ikeno546, sakamoto535, orihara351, murata656}@oki.com

**あらまし** 対話によってユーザの意図を理解し, 最適な検索結果を提供する「ラダリング型検索サービス」では, 対話を円滑に進め, ユーザの発話内容を理解し, 検索の決め手となる情報を獲得するための知識(ドメイン知識)が不可欠である. 本稿では, 転職ドメインを対象に, 我々が構築したドメイン知識の概要とその開発工程について説明する. また, 本ドメイン知識を用いて行った実証実験の結果についても報告する.

**キーワード** 対話システム, オントロジー, ラダリング

## Development of Domain Knowledge for Laddering Search Service and the Experiment using the Demonstration System

Sayori SHIMOHATA<sup>†</sup> Mihoko KITAMURA<sup>‡</sup> Tatsuya SUKEHIRO<sup>‡</sup>

Atsushi IKENO<sup>‡</sup> Masashi SAKAMOTO<sup>‡</sup> Ikuo ORIHARA<sup>†</sup> and Toshiki MURATA<sup>†</sup>

Ubiquitous Service Platform Company, Oki Electric Industry Co., Ltd.

<sup>†</sup> 1-16-8 Chuou Warabi-shi, Saitama, 335-8510, Japan

<sup>‡</sup> 2-5-7, Honmachi, Chuo-ku, Osaka, 541-0053, Japan

E-mail: {shimohata245, kitamura655, sukehiro564, ikeno546, orihara351, murata656}@oki.com

**Abstract** The "laddering search service" aims to provide the best search results with detailed information derived from intensive user interviews. The domain knowledge is essential for the search service to understand what a user wants to say and to acquire the necessary information. This paper describes its outline and development process of the domain knowledge which we built for the laddering search service, and also contains the experimental results of the demonstration system on the web.

**Keyword** dialogue system, ontology, laddering

### 1. はじめに

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, ネット上に様々なサービスやコンテンツが提供されるようになってきた. 多様化・細分化された情報はユーザの多様なニーズにこたえるという側面があるが, 一方で, 多くの情報が閲覧されることなく埋もれたままになってい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. しかしながら, このような検索サービスでは, ユーザ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, ほしい情報が明確な場合や一般的(mass)な場合には有効な手段であるが, ニッチなニーズや, 漠然としたニーズを探し出す場合には適さない. 例えば, キーワードが思いつかなくなったり, ほ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.

そこで我々は, 対話によってユーザのニーズを浮き彫りにし, ユーザの求めるサービス・コンテンツを提供する「ラダリング型検索サービス」を提案する. 『ラダリング』とは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。ラダリングに基づく対話を計算機上で実現することにより、ユーザが単独では表現することが出来なかった詳細な検索条件を獲得し、真のニーズにマッチしたサービスやコンテンツを検索する。このような検索サービスは、従来の検索サービスと比べて、具体的なイメージが明確でない場合や個別化したニーズを満たしたい場合に適しているといえる。

ユーザの発話内容を理解し、対話によって検索の決め手となる情報を獲得するためには、検索対象ドメインについての広範かつ詳細な知識が不可欠である。本稿ではラダリング対話のための知識に焦点を当て、ラダリング型検索サービスの実現に向けた我々の取り組みを説明する。また、Web上にサービス形態をイメージしたデモサイトを構築し、実際の転職希望者を対象に行った実証実験の結果について報告する。なお、ラダリング型検索サービスのシステム概要については、[6]を参照されたい。

## 2. 対話のための知識とは

### 2.1. ラダリング型検索サービスにおける対話処理

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がそうであるように、ユーザに話しやすい環境を提供し、対話を通じてユーザが考えを整理し、新しい価値観を創造していくプロセスを支援する。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固有の言語知識(ドメイン知識)を記述したドメイン知識部、発話文を解析し、ユーザの意図を理解する意図解析部、ユーザ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、から構成される。

本サービスにおける対話は、以下の3つのプロセスを繰り返すことによって進行する。

- ・ **対話を進める**

対話戦略に従ってどの質問をするか決める、または、発話内容に応じて次の対話アクションを選択する。対話アクションには、もっと掘り下げて聞く(具体的に聞く、質問を変えて別の角度から聞く)、情報を提供する、要約して考えを整理する、次の話題に移るなどがある。

- ・ **発話内容を理解する**

対話が成立しているかを理解する。対話が成立している場合には、ユーザ発話文から検索に必要な情報を取得する。対話が成立していない場合、例えば、ユーザが質問の意図を理解していない、誤って解釈している、真面目に答えていないといった場合には、状況に応じてユーザに

指示や説明を行い、軌道を修正する。

- ・ **発話を促す**

発話内容に対して、的確なレスポンス(確認、要約、共感を示す言葉など)を返すことで、ユーザに安心感や寄り添い感を与え、発話しやすい環境を作る。ユーザが答えに窮した場合に質問を別の言葉で言い換える。

### 2.2. ラダリング対話のための知識

対話をスムーズに進めるためには、システムがドメインに応じた典型的なトピックや対話の流れを知っている必要がある。また、ユーザに安心感を与え、多くの情報を引き出すためには、ユーザの発話内容を正しく理解し、的確な応答をするための知識が不可欠である。このような観点から我々は、ラダリング対話を行うための知識を以下のように定義した。

- ・ **テーマ**

検索対象ドメインにおいて、ユーザのニーズを把握し、最適なコンテンツやサービスを提供するために聞いておくべき話題。

- ・ **シナリオ**

ユーザから得たい情報の種類(以下、属性名と呼ぶ)と、そのための質問文、また、質問をすることによってユーザから得られた情報(以下、属性値と呼ぶ)に応じて、次にどのようなアクションをするかを記述したもの。

- ・ **オントロジー**

検索対象ドメインに関する語彙を網羅的に収集し、体系化したもの。木構造で表現され、第1階層に全体の概念を示すオントロジー名を、第2階層以下に広義語(概念的な表現)から狭義語(より具体的な表現)へと具体的な語彙を階層的に記述する。

次にこれらの知識間の関係について説明する。対話は、複数のテーマから構成される。シナリオはテーマごとに1つ、または、複数用意される。テーマを代表する最も重要な属性に関するシナリオを代表シナリオと呼び、基本的な対話の流れは、代表シナリオに優先度を設定することで決定する。ただし、ユーザの回答によっては、詳細質問や関連質問を優先させる場合もある。

シナリオから生成された質問文に対してユーザから回答が得られると、意図解析部がそれを解析し、属性値と意図を獲得する。ここで、獲得される属性値は、シナリオに定義された属性名の要素として、オントロジーに記述された語彙そのものである。また、ここでの意図とは、発話文に含まれる肯定、否定、分からな

表 1 テーマと属性名, 代表質問, オントロジーの対応の例

テーマ	属性名	代表質問	オントロジー
転職理由	転職理由	どうして転職しようと思ったのですか？	転職理由
希望する仕事の条件	希望の職種	転職先の職種にはこだわりますか。こだわる場合は希望の職種を教えてください。	職種
	希望しない職種	これだけはいやだという業種はありますか？	職種
	希望の勤務地	勤務地にはこだわりますか。こだわる場合は希望の場所を教えてください。	地名
	希望する会社の特徴	どんな会社に勤めてみたいと思いますか？	会社の特徴
	希望しない会社の特徴	こんな会社には入りたくないという会社はありますか？	会社の問題点

表 2 作成したオントロジーの例

オントロジー名	語彙表現	語彙表現	語彙表現
職種	営業職	海外営業 海外セールス 国外営業	
職種	営業職	ルート営業 ルートセールス ルート販売	
職種	営業職	カウンター営業 カウンターセールス	
職種	営業職	MR メディカルリプレゼンティブ プロパー	
制度	休暇	リフレッシュ休暇	
制度	休暇	育児休暇 育児休業 育休 育児休暇 子育て休暇	
制度	休暇	夏季休暇 夏休み 夏期休暇 お盆休み	
制度	休暇	産前・産後休暇	産前休暇
制度	休暇	産前・産後休暇	産後休暇
強み	ビジネススキル	セルフマネジメント力 自己管理能力 自己+管理する	
強み	ビジネススキル	課題解決力 問題解決力 課題+解決する 問題+解決する	
強み	ヒューマンスキル	コミュニケーション能力 コミュニケーションスキル 対人スキル	
強み	指導力	統率力 牽引力 求心力 統率する 牽引する	

といったユーザの心情・状態を示し、獲得した属性値が有効かどうかを判断するために利用される。

抽出された属性値は、属性名と対にして、ユーザの個人データに蓄積される。蓄積された情報は、ユーザが考えを整理するために参照したり、検索条件として利用したりする。また、検索対象のデータに対しても、同様の意図解析処理によって事前に属性情報が取得され、検索対象データに蓄積されている。ラダリング型検索サービスのマッチングは、個人データと検索対象データの属性情報の一致度をもとに算出される。

### 3. ドメイン知識の開発

ラダリング型検索サービスの一形態として、「転職者向け職業紹介サービス」(転職ドメイン)を想定し、ドメイン知識を構築した。転職ドメインを対象としたのは、職歴や転職理由、転職先への希望といったユーザの検索条件が多様でキーワード化が難しい分野だ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。

#### 3.1. テーマと対話の流れの検討

ラダリング対話を行ううえでまず重要なのは、対話によって何を聞き出したいか、どの順序で聞くべきかを明らかにすることである。このプロセスは、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。

我々は、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(属性名)と、その情報を聞き出すための質問文(代表質問)を作成し、ユーザの回答から情報を抽出する際に参照するオントロジーを対応付けた。対話の統一性を保つため、共通する概念で扱える属性では、共通のオントロジーを用いることとした<sup>1</sup>。表 1 にテーマと属性名、代表質問、オントロジーの対応付けの例を示す。

次に、これらの対話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、それぞれの対話テーマに対して、予測されるユーザの発話を分類し、分類ごとに次の対話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。

#### 3.2. オントロジーの構築

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、その体系化を行う(この概念体系をオントロジーと呼ぶ)。ここでのオントロジーは木構造で記述され、上位下位を表す階層構造を持つ。オントロジー内の各要素は、概念の内容を示す値(属性値)と具体的な語彙表現からなり、同じ内容を示す複数の語彙表現(同義表現)は 1 つの属性値に対応付けられる。語彙表現のパタ

<sup>1</sup> 例えば地名オントロジーは都道府県名や市町村名を記載したオントロジーであるが、ユーザの住所、希望する勤務地、会社の所在地など、複数の属性の値を

ーンは、名詞句、動詞句、名詞句と動詞句の組み合わせの3種類とした。

概念の洗い出し、および、階層化の工程は、ドメインエキスパートとの議論と協業を重ね、システム設計者が取りまとめた。転職ドメインに特化したとはいえ、対話に出現する可能性のある語彙を網羅するため、概念数は職種、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、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。作成したオントロジーの例を表2に示す。

### 3.3. 語彙拡張

本対話システムの意図解析処理では、ユーザの発話文に含まれる属性値を捉えることにより、発話内容を理解し、検索に必要な情報を獲得する。そのため、属性を示す表現としてユーザ発話に出現する可能性のある語彙はすべて、オントロジーに登録されている必要がある<sup>2</sup>。そこで、オントロジーの各要素について、同義表現を網羅的に収集し、語彙の追加を行った。

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、基本的に人手で行った。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、最終的に合計11,200件、異なり42,330件の語彙を持つオントロジーが構築された。

### 4. 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、対象者が自ら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。検索対象としてリクナビNEXTの実求人データ約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。

実験の結果を図1にまとめる<sup>3</sup>。

獲得する際に参照される。

<sup>2</sup> 一部の表記ゆれについては、意図解析処理内の形態素解析にて吸収される。

<sup>3</sup> 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が参加した。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。

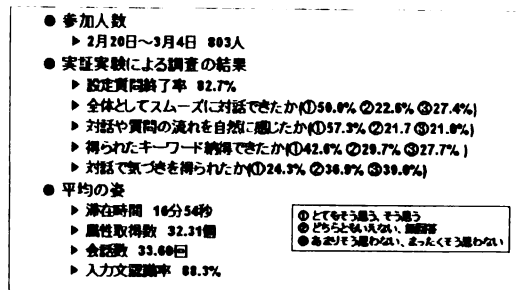


図1 実証実験の結果

まず、平均滞在時間約17分<sup>4</sup>、平均対話数33回と、多くのユーザ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<sup>5</sup>。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、それ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、否定的な回答が肯定的な回答を上回った。対話により得られたキーワードは納得できるものの、それが新たな発見や認識に結びついているユーザは24%にとどまっている。これに対しては、今後、新たな対話戦略や要約、言い換えの技術を導入し、対話履歴や取得情報、検索結果の表示方法を工夫することで、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!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3に示す。

表3 既存の検索サービスとの比較

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%

<sup>4</sup> 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。

<sup>5</sup> 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、一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ーとマッチングしない場合があるので、会話数=取得属性値数とはならない。

ユーザ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、「手間」がかかるものとして、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手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。

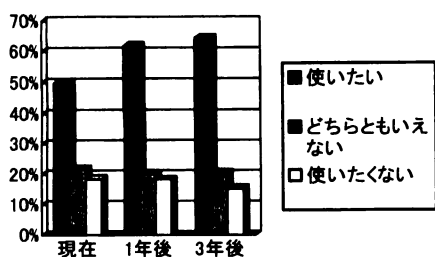


図2 「ラダリング型検索サービスを今後使いたいかな」のアンケート結果

最後に、ラダリング型検索サービスを今後使いたいかなについてのアンケート結果を図2に示す。使いたいと回答したユーザは、「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る。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様々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、よく分からない情報を検索したいという潜在ニーズの現れであり、対話型検索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。

## 5. 関連研究

システムとの対話を通じて、ユーザから情報や要望を引き出すシステムの研究として、ダイアログナビ[7]、対話エージェント[8]、対話型意見収集システム[4][9]などがある。

ダイアログナビ[7]は、大規模テキスト知識ベースに基づく対話的自動質問応答システムである。パーソナルコンピュータの利用者を対象とし、あいまいな質問に対して対話的に聞き返しを行うことにより、ユーザの求める回答を導き出す。聞き返しは、聞き返しの手順を記述した対話カードを用いる手法と、聞き返しの選択肢を自動的に編集して提示する手法とを組み合わせる。想定されるあいまいな質問をあらかじめ調べ上げ、その対応を対話カードで定義する方法は、我々のシナリオによる対話制御と類似している。ユーザ質問文と知識ベースの表現のずれを吸収し、検索精度を上げるために、小規模な同義表現辞書(3,512語217フレーズ)と上位・下位語辞書(200語)を用いている。

鳥澤らは、Webデータから自動的に獲得した知識を用いて一般ユーザと対話し、多様な情報を収集する対話エージェントを実装した[8]。ドメイン依存性の低い対話によってCGM(Consumer Generated Media)的な情報を獲得するというコンセプトは、我々のシステムと異なっているが、常識や語彙の上位下位関係といった知識を自動的に獲得し、対話に利用している点、他のユーザとの対話で獲得した情報を利用して対話を行う点など、今後我々がシステムを改良していく上で参考になる点が多い。

大塚らの対話型意見収集システム[4][9]は、市民参画型事業の初期段階において、市民から意見を収集するための対話支援システムで、確認や質問を繰り返しながら、入力者の意見を、あらかじめ用意した「懸念事項」に分類する。問いかけを繰り返し、入力者に気づきを与えつつ内容の本質に迫るリフレーミング手法はラダリングと類似するが、このシステムでは、意見の収集・分類が目的のためオントロジーは参照せず、形態素解析結果を意見内容分類器と意図・感情分類器にかけることによって、分類先を推定している。

対話システムで利用する知識は、一般的な用語辞書ではなく、より詳細な情報や語彙間の関係を記述できるオントロジーが適している。ドメイン知識としてのオントロジーは、医療やバイオ、製造など様々な分野で開発されている[3][5][10]。アプリケーション依存のオントロジーでは、タスクが限定されていることもあり、オントロジーエディタなどを用いて人手で構築されることが多い。一方、専門辞書や大量の電子化されたテキストが存在する分野では、語義文や定型表現を用いてオントロジーを(半)自動構築する研究が進んで

いる[1][2].

今回我々が開発したオントロジーは、職種や転職理由のようにドメイン依存性の高いものと、性格や趣味のように汎用性の高いものが混在している。そのため、オントロジー全体を統一的な手法により自動構築することは現実的ではない。しかしながら、オントロジーの網羅性やドメイン移植性を高めるためには、構築プロセスの自動化は必須である。今後は、語彙の収集、上位下位概念の抽出、オントロジー内の矛盾の検出など、適用分野の特徴にあわせた効率的なオントロジー構築方法を検討したい。

## 6. まとめ

ラダリング型検索サービスでは、有効な検索をおこなうための情報を対話により獲得する。今回は、自然な対話を行い、ユーザの発話内容を理解するための知識の設計、および、開発を中心に、本サービスを説明した。ラダリング対話は、問いかけを繰り返し行うことにより、ユーザ自身が気づかなかった新たな価値観、ニーズを引き出していく。本システムは未だ開発途中であるが、実証実験では23%のユーザが、システムとの対話を通じて気づきを得られたと回答し、「ラダリング型検索サービス」の実現が可能であることを実証した。

今後は、ドメインエキスパートの知識を効率的にドメイン知識に移転するために、開発ツールの整備や知識の自動獲得技術、システム発話文の自動生成技術の導入を行っていく。さらに、システム全体では、対話履歴の要約や検索結果の効果的な表示など、ユーザに新たな視点や気づきを与えるための工夫を取り入れていきたい。

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殿と共同で行われた。

## 文 献

- [1] Aramaki, E., et al, "Toward Medical Ontology via Natural Language Processing," Proc. of International Joint Conference on Natural Language Processing (IJCNLP) workshop OntoLex2005, pp53-58, 2005.
- [2] Snow, R., Jurafsky, D., and Ng, A. Y., "Semantic Taxonomy Induction from Heterogenous Evidence," Proc. of 44<sup>th</sup> Annual meeting of the ACL, pp.801-808, 2006.
- [3] MeSH. Medical Subject Headings, National Library of Medicine, 2006.
- [4] 大塚裕子, 山本瑞樹, 乾孝司, 丸元聡子, 奥村学, "市民参画型道路計画における対話支援-対話型アンケートシステムプロトタイプ版の実装-,"

言語処理学会第13回年次大会, pp.302-305, 2007.

- [5] 來村徳信, 溝口理一郎, "オントロジー工学に基づく機能的知識体系化の枠組み," 人工知能学会論文誌, Vol.17, No.1, pp.61-72, 2002.
- [6] 北村美穂子他, "ラダリング型検索サービスのための対話エンジンの設計・開発," 情報処理学会研究報告, 2008-NL-185, 2008.
- [7] 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.
- [8] 鳥澤健太郎, "一般ユーザにインタビューする対話エージェント," 情報処理学会研究報告, 2007-NL-180, pp.25-30, 2007.
- [9] 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.
- [10] 森田幸伯, 福重貴雄, 大沼宏行, 野本昌子, 玉利公一, 松平正樹, 豊岡豊, "情報家電オントロジー構築と活用に向けた取り組み," 第21回人工知能学会全国大会, 2007.