

WWW 検索における問題解決事例の利用

問題解決事例に基づく WWW 情報モデリングと検索

牛尼 剛聡[†] 若園 智美^{††} 渡邊 豊英^{†††}

[†]九州大学大学院芸術工学研究院 〒815-8540 福岡市南区塩原 4-9-1

^{††}名古屋大学大学院工学研究科 〒464-8603 名古屋市千種区不老町

^{†††}名古屋大学大学院情報科学研究科 〒464-8603 名古屋市千種区不老町

E-mail: [†]ushiana@design.kyushu-u.ac.jp, ^{††}wakazono@watanabe.nuie.nagoya-u.ac.jp,

^{†††}watanabe@is.nagoya-u.ac.jp

あらまし インターネットが普及し、WWW 上で様々な情報やサービスが提供されるようになった。人が生活し、業務を遂行する上で、目的とする活動に有用な Web ページを効率的に発見し活用する重要性が高まっている。現在、Web ページを検索するために様々な検索サービスが提供され、多くの研究が行われている。しかし、それらの多くは Web ページ自身が有するテキスト情報に基づいて検索を行うものであり、利用者の目的に対する意義や重要性を十分に反映した検索結果が得られているとは言い難い。我々は、この問題を解決するためには、Web ページを実世界上で人間が行う活動の文脈上に位置づけることが重要であると考え、本研究では問題解決活動を対象に、利用者の問題を解決するために必要な Web ページを効果的に検索可能とすることを目的とする WWW 検索手法の概要を述べる。本手法では電子メールや BBS 等のテキストコミュニケーションログから問題解決事例を抽出する。問題解決に利した Web ページの抽出には、メッセージ作成者の WWW アクセスログを利用する。

キーワード Web ページ検索, 問題解決事例, WWW アクセスログ, テキストメッセージ, 推論ネットワーク

Usage of Problem Solving Cases for WWW Search

– Information Modeling and Search on WWW based on Problem Solving Cases –

Taketoshi USHIAMA[†], Tomomi WAKAZONO^{††}, and Toyohide WATANABE^{†††}

[†] Faculty of Design, Kyushu University 9-1 Shiobaru 4-chome, Minami-ku, Fukuoka, 815-8540 Japan

^{††} Graduate School of Engineering, Nagoya University Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan

^{†††} Graduate School of Information Science, Nagoya University Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan

E-mail: [†]ushiana@design.kyushu-u.ac.jp, ^{††}wakazono@watanabe.nuie.nagoya-u.ac.jp,

^{†††}watanabe@is.nagoya-u.ac.jp

Abstract In recent years, variety of information and services are available on the Web. It is important to find the web pages which provide useful information to carry out activities of users. In this paper, we introduce a framework of web page search for problem solving of users. Our main idea for searching is to use contextual information of web pages on real world problem solving activities. Problem solving cases are extracted from text communication logs such as e-mail and BBS, and web pages that are used for problem solving are extracted from web page access logs.

Key words Web page search, problem solving case, WWW access log, text messages, inference network

1. ま え が き

インターネットの普及に伴い、日常生活からビジネスに至るまで様々な種類の情報とサービスが WWW 上で提供されるよ

うになった。こうした背景の下、生活や業務を行う際に、利用者が目的とする活動に有用な Web ページを効率的に検索し活用することの重要性が増大している。本研究では、WWW を利用した問題解決活動を支援するシステムの実現を目標とする。

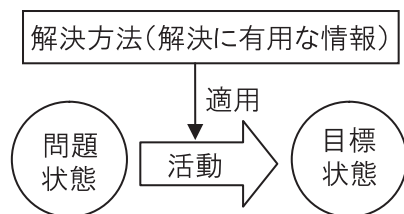


図1 問題解決活動の概念図

本システムが提供するものは、利用者の問題を解決するための具体的な情報が記された Web ページや、問題を解決するためのサービスを提供するサイトのトップページである。本論文では、問題解決活動を支援するための WWW 情報のモデル化手法と検索手法を提案する。

現在、Web ページ検索のために様々なサービスが提供され、Web ページ検索の高度化のために活発に研究が行われている。これらは、Google や Yahoo に代表されるようなキーワードベースの検索を想定したものが多い。Cui はキーワードベースの検索において利用者が検索時に与えるキーワードと、期待される検索結果が表すキーワードとが異なっているにもかかわらず、多くの検索エンジンではそれらを同一なものとして扱っている問題を指摘している [1]。

Web ページ検索で問題解決を行う場合にも、利用者が検索時に与えるキーワードと、期待される結果を表すキーワードとが異なるという問題は重要な意味を持つ。問題解決活動とは、問題状態を目標状態へ変化させる活動である。問題解決活動の概念図を図 1 に示す。問題解決活動をキーワードによって表現する場合、問題状態を説明するキーワード、目標状態を表すキーワード、解決方法を現すキーワードはそれぞれ異なっている事が多い。この場合、利用者はあらかじめ解決に有用な情報を想定して検索を行う必要がある。しかし、この方式では、解決方法の目安がわからなければ有用な検索結果を得ることができない。

例えば、コンピュータ設定の不具合を解決したい場合、問題状態には具体的な状況やエラーメッセージが含まれる。しかし、一般的に解決方法が記述された Web ページには個別の状況に関するキーワードが含まれているとは限らない。このような場合、利用者が直面する問題を説明するための単語を検索キーワードとして与えても満足できる検索結果が得られない。期待する検索結果を得るためには直面している問題そのものではなく、解決方法に関するキーワードを指定しなければならない。

徳永は文献 [2] において、利用者が目的を達成するために現在持っている知識では不十分であると感じている状態の分類として Taylor の 4 階層を紹介している。Taylor の 4 階層を以下に示す。

Q₁:直感的要求

現状に満足していないことは認識しているが、それを具体的に言語化してうまく説明できない状態。

Q₂:意識された要求

頭の中では問題を意識できるが、あいまいな表現やまとまりの

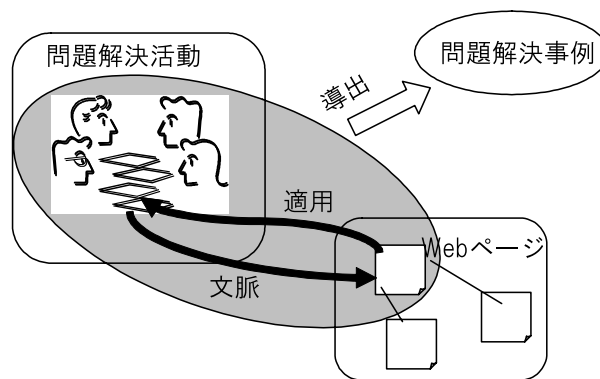


図2 問題解決活動のモデル化の概念図

ない表現でしか言語化できない状態。

Q₃:形式化された要求

問題を具体的な言語表現で言語化することができる状態。

Q₄:調整済みの要求

問題を解決するために必要な情報の情報源が同定できるくらい問題が具体化された状態。

現在の Web 検索システムの多くは Taylor の階層でいえば Q₄ の段階にあるユーザを対象としている。しかし、問題解決を目的とする検索においては、問題自体は明確になっているもの、問題を解決する方法はわからないという Q₃ の段階にいる場合が多い。本研究では Q₃ の段階にいる利用者が WWW を利用して問題解決に有用な Web ページを検索可能とすることを目指す。

我々は、上記の目的を実現するために、Web ページ上に存在する情報が実世界上の実際の問題解決の文脈に対応づける必要があると考えた。つまり、WWW 上のリソースが実世界の活動においてどのように役立っているかという視点から検索可能とすることを重要視する。この考えに基づく問題解決活動のモデル化の概念図を図 2 に示す。問題解決活動において Web ページに記述された情報が適用された場合、問題解決活動は Web ページの実世界における活用文脈を付与すると考えられる。本研究では問題解決活動とそこで活用された Web ページの対応関係を問題解決事例として捉える。

本論文では、Web ページを問題解決という文脈の元に位置づけ、利用者の要求を直接表現できる Web ページ検索システムの基本的な枠組みについて述べる。

本研究で提案するシステムの概念図を図 3 に示す。システムは、Web ページが過去にどのような問題解決に利用されたかを問題解決事例として蓄積する。利用者は解決したい問題に関する記述を検索要求として与え、問題の解決に有効な Web ページを検索結果として得る。この処理の概要は以下の通りである。

- (1) 利用者が解決したい問題に関するキーワードを検索要求として与える。
- (2) 検索要求に基づいて類似事例を検索する。
- (3) 検索された類似事例から、利用者が解決したい問題の解決に有効な Web ページを導出する。
- (4) 導出された Web ページを検索結果として返す。

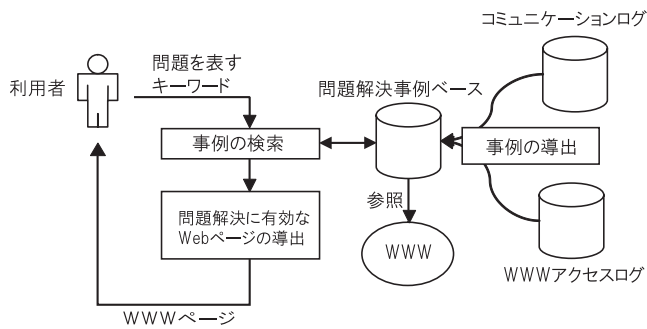


図 3 システムの概要

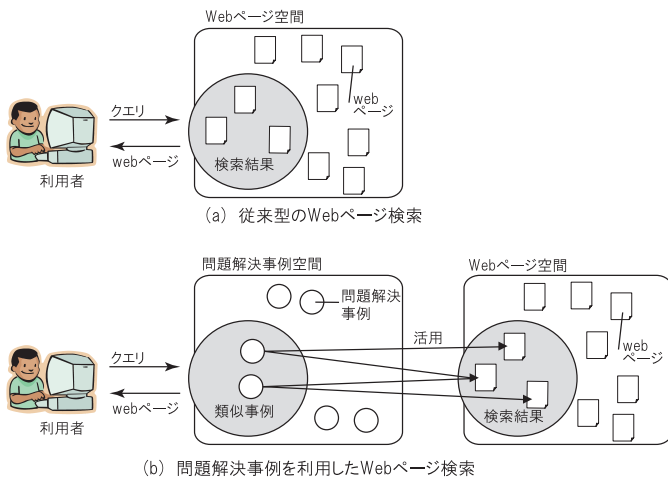


図 4 Web ページ検索アプローチの違い

本論文で提案する Web ページ検索のアプローチの特徴を図 4 に示す．図中の (a) は従来型の Web ページ検索のアプローチを表し，(b) は本論文で提案する Web ページ検索のアプローチを表している．従来型の検索手法では，Web ページを直接検索していたのに対して，本手法では，問題解決事例を検索することにより間接的に Web ページを検索する．問題解決事例は Web ページが問題解決活動において活用される文脈を表している．

問題解決事例は電子メールや BBS 等のテキストコミュニケーションのログを利用して導出する．導出された問題解決事例において問題解決に利用された Web ページを発見するために，メッセージ内に記述された URL とメッセージ作成者の WWW アクセスログを利用する．

2. 関連する研究

従来から，Web ページ検索における文脈利用の重要性は認識され，Web ページ検索で文脈を利用する多くの提案がなされている [3]．Web ページ検索における文脈は様々な視点から考えることができる．

Cui ら [1] は検索時の利用者の問い合わせに対して文脈情報を表すキーワードを付加することによって，検索結果の絞り込みを行う手法している．IntelliZip [4] は，利用者が文書中の単語を選択することにより検索キーワードを指定し，指定した単語の前後の文章から，キーワードの文脈を推定し検索に利用し

ている．これらは利用者の問い合わせ表現の文脈を考慮するものである．一方，我々が対象とするのは Web ページが実世界上でどのように役に立つかという，検索対象に関する文脈である．

Web ページ自体の文脈を与える代表的な方法はアノテーションを利用するものである．赤星ら [5] は携帯電話による電子メールの交換内容から関連情報を推測し，その情報を含む Web ページをユーザに提示する手法を提案する論文の中で，メールを Web ページのアノテーションとして検索に利用する可能性を示している．コミュニケーションログを Web ページ検索における付加情報として利用するアプローチは我々と同一である．しかし，Web ページのアノテーションとして捉えた場合は，Web ページを中心としたモデル化であるのに対して，我々は実世界上の問題解決活動を基準に WWW 情報のモデリングを行う．本研究では，実世界の問題解決活動を基準としたモデル化を行うことにより，未知の活動に関しても利用者に有用な情報を提供可能とすることを旨とする．また，Web ページとの対応付けにメッセージ作成者の WWW アクセスログを利用する点も赤星らの手法と異なっている．

中島ら [6] は利用者の WWW アクセスログを利用して，ブックマークとして登録された Web ページの文脈を導出する方法を提案している．ここで文脈とは利用者の検索アクティビティの事であるのに対し，我々は，Web ページが与える情報が実世界において適用される活動を文脈と捉えている点が異なる．

3. 問題解決事例のモデル化と導出

3.1 問題解決事例のモデル化

蓄積された事例に基づいて，利用者が与えた問題に有用な知識を利用者に提示するためには以下の 2 種類の処理が必要である．

- (1) 利用者が与えた問題に類似した事例を検索する．
- (2) 類似された事例に基づいて，利用者が与えた問題の解決に有用な情報を推定する．

類似した事例を検索するためには，問題解決に関する表現が必要である．また，問題解決に有用な知識を推定するためには，問題解決において活用された知識や情報がわからなければならない．

問題解決事例 c を，識別子 id ，重み付きの単語集合（特徴ベクトル） F ，Web ページ集合 P から構成される 3 項関係として定義する．

$$c = (id, F, P)$$

問題解決事例を構成する F を問題解決記述と呼び， P を問題解決知識と呼ぶ．問題解決事例の概念図を図 5 に示す．

問題解決記述は対象とする問題解決活動の性質を表現する．本手法では，問題解決記述は単語の重みを表す特徴ベクトルによって表現する．問題解決記述は，利用者が与えたキーワードに基づいて，利用者が解決を希望する問題に類似した問題解決事例を検索するために利用する．

問題解決知識は問題を解決するために有用な情報を表し，問題解決に有用な Web ページの集合として実現する．

問題解決活動において適用される知識を獲得する手段には，

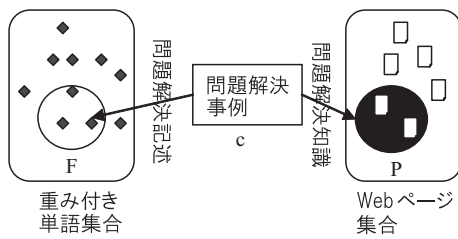


図 5 問題解決事例の構造

WWW の他に、書籍、経験、伝聞など様々なものが考えられる。問題解決に利用した知識が全て Web ページとして存在しているとは限らない。しかし、本研究の目的は、問題解決に有効な Web ページを提供することであるため、Web ページとして存在する問題解決知識のみを対象にする。

本研究では、ページの内部構造は考えず、Web ページを知識表現の最小単位とする。知識は異なる問題解決に利用される可能性があるため、同一の Web ページが複数の事例から参照される場合がある。

3.2 問題解決事例の導出

3.2.1 問題解決事例の発見

問題解決は人間が実世界で行う活動である。コンピュータが実世界上の活動を直接抽出することは困難である。そこで、問題解決事例の導出のためにコミュニケーションログを利用する。

インターネット上でコミュニケーションが行われるコミュニティには様々な種類がある。石川 [7] はインターネット上でのコミュニティを 6 つのパターンに分類し、その一つとして問題解決を目的としたコミュニティがあることを述べている。本研究では問題解決を目的としたコミュニティで行われるテキストコミュニケーションログを利用して問題解決事例を導出する。テキストコミュニケーションを利用するのは、音や画像などを利用したコミュニケーションと比較して、計算機を利用して比較的簡単に内容を抽出可能であることが理由である。

テキストコミュニケーションは同期型と非同期型に分類できる。非同期型の代表例には電子メールや BBS があり、同期型の代表例にはインスタントメッセージがある。非同期型テキストコミュニケーションには以下のような特徴がある。

- メッセージの参照関係を保持している
- 参照メッセージの特定の部分を引用することにより、部分的な話題の関係を表現する習慣がある。

本研究では非同期型のコミュニケーションログを対象とし、メッセージの参照関係によって単一の木構造に組織化されるメッセージの集合を、一つの問題解決事例に対応づける。問題解決事例を導出するために利用したメッセージ集合を導出元メッセージと呼ぶ。問題解決事例を導出する際に利用するデータとの関係を図 6 に示す。

3.2.2 問題解決知識の導出

問題解決知識の導出とは、問題解決に有用であった Web ページと事例とを対応づけることである。対応付けには参照元メッセージに記載された URL とメッセージ作成者の WWW アク

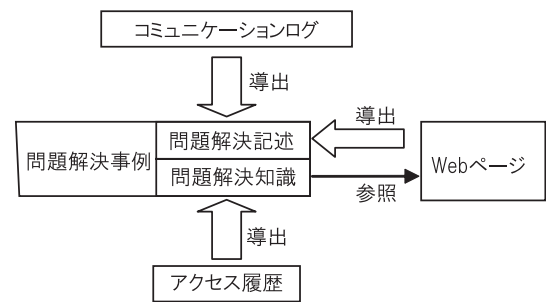


図 6 問題解決事例の導出

セスログを利用する。

問題解決事例 c の参照元メッセージ集合 $M = \{m_1, \dots, m_n\}$ をとする場合、参照元メッセージ集合に含まれる全てのメッセージに対して以下の処理を行うことにより、問題解決知識を決定する。

(1) 対象とするメッセージ m_i 内に Web ページの URL が記載されているかを調査する。記載されている場合は、参照元メッセージに記載されている URL が示す Web ページ p_j が問題解決に有用な情報であると見なし問題解決知識に追加する。

(2) 対象とするメッセージ m_i を送信した利用者の WWW アクセスログ A_i に含まれる Web ページ p_j が、メッセージ m_j との類似度 $\text{sim}(m_j, p_j)$ が閾値 T 以上である場合、 p_j を問題解決知識に追加する。

メッセージ m と Web ページ p との類似度 $\text{sim}(m, p)$ は、 m の特徴ベクトル \mathbf{m} と p の特徴ベクトル \mathbf{p} のコサイン相関値を利用し、以下のように定義する。

$$\text{sim}(m, p) = \frac{\mathbf{m} \cdot \mathbf{d}}{|\mathbf{m}| |\mathbf{d}|}$$

特徴ベクトルの重みは TF*IDF を利用する。具体的には、メッセージ m における単語 t_i の重み $w(m, t_i)$ を以下のように定義する。

$$w(m, t_i) = mtf_i \times \left(\log \frac{N_m}{mf_i} + 1 \right)$$

ここで、 mtf_i は m における t_i の出現頻度、 N_m は全メッセージ集合に含まれるメッセージ数、 mf_i はメッセージ集合中で t_i が出現するメッセージの数である。

メッセージと同様に、Web ページ p における単語 t_i の重み $w(p, t_i)$ を以下のように定義する。

$$w(p, t_i) = ptf_i \times \left(\log \frac{N_p}{pf_i} + 1 \right)$$

ここで、 ptf_i は Web ページ p における t_i の出現頻度、 N_p は全 Web ページ集合に含まれる Web ページ数、 pf_i は Web ページ集合中で t_i が出現する Web ページの数である。

3.2.3 問題解決記述の導出

問題解決記述は問題解決事例の内容を表現するための属性であり、特徴ベクトルとして表現される。問題解決記述は利用者が与えたクエリに基づいて類似した問題解決事例を検索するために利用する。

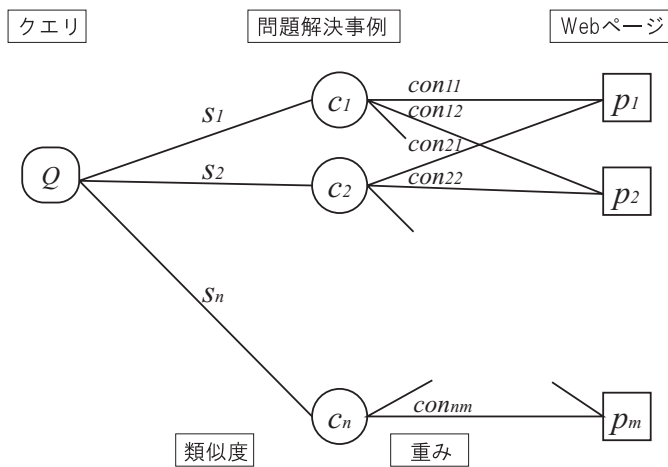


図 7 問題解決事例の類似度と Web ページの貢献度

問題解決事例 c の問題解決記述となる特徴ベクトル F は、 c の参照元メッセージ集合 M に含まれる全てのメッセージの特徴ベクトルと、問題解決知識 p に含まれる Web ページの特徴ベクトルの総和として以下のように定義する。

$$F = \sum_{m_i \in M} \mathbf{m}_i + \sum_{p_j \in P} \mathbf{p}_j$$

4. 問題解決事例を用いた Web ページ検索

本研究で開発するシステムは、解決したい問題を表現するキーワード集合を利用者の入力とし、問題解決に有用な Web ページをランク付けして出力する。システムが行う処理は以下の 2 ステップから構成される。

(1) 利用者が与えたキーワード集合に基づいて類似した事例を検索する。

(2) 検索された事例に基づいて、利用者が与えた問題に有用かどうかという視点から Web ページの重み付けを行う。

事例を検索する際に利用する類似度は、利用者が与えたキーワードからなる特徴ベクトルと記述事例の特徴ベクトルのコサイン相関値を利用する。

1 個の問題解決事例は問題解決知識として複数の Web ページを参照できる。また、複数の問題解決事例が同一の Web ページを参照することもある。そこで、Web ページの重みを、クエリと問題解決事例との類似度と問題解決事例と問題解決知識の類似度の積の総和として実現する。

利用者の与えた問い合わせを Q とする。 Q と問題解決事例 c_i との類似度を $s_i = \text{sim}(Q, c_i)$ とする。Web ページ p_j が c_i によって与えられる重み con_{ij} を c_i と p_j の類似度に基づいて以下のように定義する。

$$\text{con}_{ij} = \text{sim}(c_i, p_j)$$

なお、 p_j が c_i の問題解決知識に含まれない場合 $\text{con}_{ij} = 0$ とする。 con_{ij} を問題解決事例 c_i に関する Web ページ p_j の貢献度と呼ぶ。クエリと問題解決事例の類似度および Web ページの貢献度を表す概念図を図 7 に示す。

問い合わせ Q に対する p_j の重み $\text{rank}(Q, p_j)$ を以下のよう

に定義する。

$$\text{rank}(Q, p_j) = \sum_i (s_i \times \text{con}_{ij})$$

$\text{rank}(Q, p_j)$ は利用者の問題解決に対して p_j が有効な情報を提供できる度合いを表していると考えられる。 $\text{rank}(p_j)$ が大きいほど有用な情報を提供できる可能性が高い。

5. 実験

提案手法の有効性を評価するため、以下の 2 種類の実験を行った。

- (1) 問題解決事例構成手法の有効性
- (2) 問題解決事例検索手法の有効性

5.1 問題解決事例導出手法の有効性

5.1.1 実験の概要

本論文で提案する問題解決事例導出手法のうち、特に Web アクセスログに基づいて抽出された問題解決知識の有効性を評価する実験を行った。本実験では被験者に WWW 上の BBS 上で問題解決を目的としたコミュニケーションを行ってもらい、コミュニケーションログと Web のアクセスログに基づいて事例を導出した。

BBS 上のコミュニケーションに関して、被験者には問題解決を目的としたメッセージを投稿する以外に制約を与えず、被験者は BBS 上に自由にメッセージを作成可能とした。メッセージを作成する際には、すでに投稿されたメッセージに対する返信であることを指定可能である。メッセージには参考となる Web ページのアドレスを記載することができる。システムは投稿されたメッセージに関して、投稿したホスト名と投稿時刻を記録する。本実験で利用した BBS の実行画面を図 8 に示す。

Web アクセスログを取得するために、被験者には Web アクセスの際にプロキシサーバを利用するように依頼した。プロキシサーバでは、アクセス元のホスト名とアクセスした Web ページの URL とアクセス時刻を記録する。

問題解決事例の導出の際、投稿されたメッセージとアクセスログに含まれる Web ページの類似度が 0.05 以上の場合に、Web ページを問題解決知識の対象とした。なお、事例を導出する際には、メッセージ投稿時刻から 1 時間以内にアクセスした Web ページのみを問題解決知識の対象とした。

被験者は 6 名であり、内訳は大学院学生 5 名と大学教員 1 名であった。実験期間中に投稿されたメッセージ数は、29 件であった。取得した被験者の Web アクセスログに含まれるページ数は、のべ 6491 件であった。投稿されたメッセージから 10 件の問題解決事例が導出された。10 件の問題解決事例において問題解決知識として抽出された Web ページの総数は 44 件であった。

5.1.2 実験結果と考察

実験結果から問題解決知識を抽出する手法の有効性を評価する。URL が記載された Web ページは明らかに問題解決に有用な知識であると考え、類似度を利用して抽出できれば、URL が記載されていない場合でも有用な知識が抽出できると予想でき

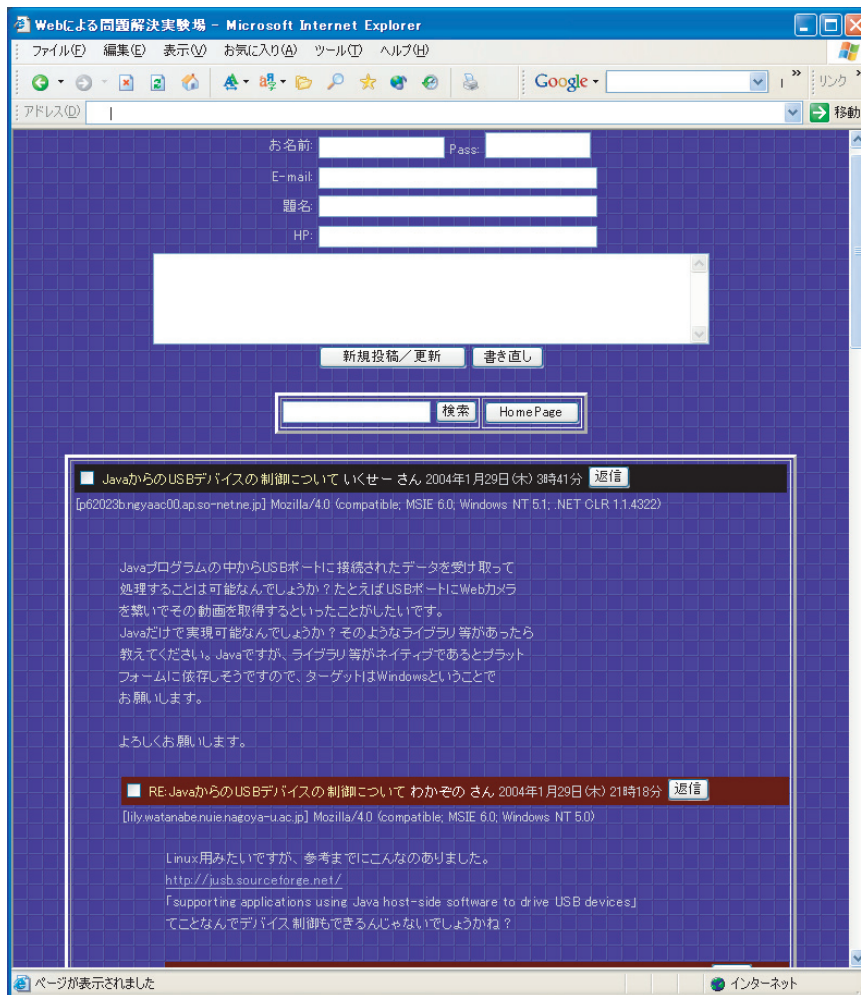


図 8 BBSのスナップショット

表 1 問題解決知識抽出の再現性

事例数	URL が明示 された知識	抽出された 知識	再現率
(件)	(件)	(件)	(%)
10	11	7	63.6

表 2 問題解決知識抽出の適合性

事例数	問題解決 知識		適合知識		部分適合 知識		不適合 知識	
	(件)	(%)	(件)	(%)	(件)	(%)	(件)	(%)
10	44	52.3	23	52.3	12	27.3	9	20.5

る。この考えに基づき、抽出された問題事例に含まれる問題解決知識の再現性を、メッセージと Web ページの類似度によってメッセージ中に URL が明示的に記述された Web ページを抽出できた割合として評価する。実験結果を表 1 に示す。

問題解決事例 2 件に関する 4 件の Web ページが問題解決知識として抽出できなかった。一方の事例においては、回答メッセージに記述が含まれず URL のみが示されていたメッセージであったため、回答メッセージに対応付けることができなかった。もう一方の事例については、メッセージ作成時刻から 1 時間以上前に参照したページであったため、対応付けることができなかった。これらのことから、作成したメッセージと閲覧した Web ページとの類似度は、問題解決に有用な知識を再現することができるのに有効であると考えられる。

問題解決知識を抽出した結果の適合性を評価する。ここで適合性とは、抽出した知識に含まれている知識の正当性を表す。

抽出された問題解決知識を問題に対して、適合するページ、部分適に適合するページ、適合しないページの 3 種類に分類した。分類は実験者が行った。実験の結果を表 2 に示す。実験結果により、抽出された知識の約 8 割は問題の解決に有効なページであった。

Web アクセスログに基づいた問題解決知識抽出の再現性と適合性の評価から、本論文で提案するメッセージと Web ページの類似度に基づいた問題解決知識の抽出は、問題解決事例において有用な知識を抽出するために有効であると考えられる。

5.2 問題解決事例検索の有効性

5.2.1 実験の概要

利用者の与えたキーワードに基づいて問題解決事例を検索する手法の有効性評価を目的とする実験を行った。

実験ではメーリングリストのログから問題解決事例を導出する。導出された問題解決事例集合から、実験者が複数の問題解

検索数	正解事例	検索された	再現率
	正解事例		
(件)	(件)	(件)	(%)
16	36	14	43.8

検索数	検索された	適合事例	部分適合	不適合
	事例	(件)	事例	事例
(件)	(件)	(%)	(件)	(%)
16	80	24 30.0	17 21.3	39 48.7

決事例を任意抽出し、抽出した事例において解決された問題を自然言語文として記述する。被験者に自然言語文として記述された問題を与え、その問題を表すキーワードによりクエリの作成を依頼する。作成されたクエリを用いて問題解決事例を検索し、検索結果を評価する。

WWW 上で一般に過去に投稿されたメッセージが公開されている Java House Mailng List [8] を利用して問題解決事例を抽出した。メーリングリストにメッセージを投稿した人物の Web アクセスログは、メッセージ中に明示的に記述されている URL にある Web ページを問題解決知識とし、Web アクセスログに基づいた問題解決知識の抽出は行わないものとした。

実験では 1641 件のメッセージから 215 件の問題解決事例が導出された。導出された事例の問題解決知識には合計 838 件の Web ページが含まれている。自然言語文を作成した問題は 8 件である。被験者は 2 名である。検索結果はクエリと類似度が高いものから上位 5 事例とした。

5.2.2 実験結果と考察

検索結果にあらかじめ問題に対して類似する問題解決事例を選定し、検索結果に類似する問題解決事例が含まれる割合を表 3 に示す。また、検索結果が問題文と類似していたかについて、類似する、部分的に類似する、類似しないという 3 段階の評価を行った結果を表 4 に示す。

今回の実験では、再現性、適合性共に有効であると判断できる結果を得ることはできなかった。今後、問題解決記述の構成方法を変化させて、有効な問題解決記述の構成方法を比較検討する必要がある。また、Web ページに含まれる索引語のみを利用した検索結果と比較して評価する必要がある。

6. リンクナビゲーション履歴を利用した拡張

本論文で提案した Web ページ検索手法は推論ネットワーク [9] に基づいた文書検索の一種と考えることができる。

今回の実験では複数の問題解決事例の問題解決知識として同一の Web ページを参照する例は少なかった。この原因として、具体的な情報が記載された Web ページが問題解決知識として選択される事が多い事が考えられる。問題解決においては特定の分野のオーソリティサイトのトップページや検索エンジンといった、問題解決に関する具体的な情報は提供しないが問題解決には有効であるページが存在する。これらの Web ページは多くの問題解決事例において共通に参照されることが多いと予

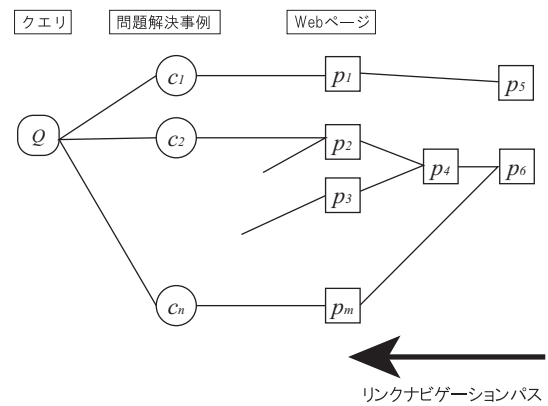


図 9 リンクナビゲーション履歴を用いた拡張

想される。

このようなページを検索可能とするために、問題解決知識に至るまでの利用者の探索アクティビティを考慮することが考えられる。具体的には図 9 に示すように問題解決知識に至るウェブページのリンクナビゲーション履歴を逆向きにたどるように推論ネットワークを成長させることにより、具体的な情報を提供しないが問題解決に有用な Web ページを検索できることが期待できる。

今回提案した手法では、基本的には、既に解決された問題に関する情報しか検索結果として提供されない。また、問い合わせが抽象度の高い問題を表現している場合、検索される事例は複数存在する。このとき、事例が参照する個別の問題解決知識を検索結果として返しても、必要以上に個別化された具体的な知識や、利用者の問題とは直接関係のない情報が含まれている可能性が高い。上記のリンクナビゲーション履歴を利用した拡張により、未知の問題や抽象度の高い問題に対応可能とするために、類似した問題の問題解決知識を導出するプロセスを利用し、利用者が WWW 上で有用な情報を探索するのを補助するための情報を提供できる可能性がある。

7. まとめと今後の課題

本論文では、WWW 上の情報に実世界の文脈を与えるという視点から問題解決事例に基づいた Web ページの構造化を行い、問題解決事例を利用した Web ページ検索のアプローチを示した。

今回の実験では、事例抽出と事例検索に関するものであった。今後、提案手法に基づいた Web ページ検索に関する実験を行う提案手法全般に関する評価を行う予定である。

本研究では、電子メールや WWW アクセスログが利用可能であることを前提とした。しかし、プライバシーの観点からこれらの情報が自由に利用可能な状況は現実には考えられない。実現のためには、個人データを管理し統計的な情報のみを外部に提供するエージェント型のアーキテクチャを持つシステムを設計することが必要である。

提案手法ではコミュニケーションログは問題解決事例を抽出するために利用し、問題解決知識になることは考えていない。しかし、現実の問題解決型のコミュニケーションではコミュニ

ケーションログ自体が有用な知識であることが多い。コミュニケーションログ自体を問題解決知識とすることを検討する予定である。

提案手法では、コミュニケーションを構成するメッセージ全てに関して参照したと思われる Web ページを全て問題解決知識として利用した。しかし、コミュニケーションにおいては結果的に問題解決に有効でなかった Web ページが存在する場合がある。この問題に対処するため、メッセージの内容の理解や対話の構造を解釈し、メッセージの役割を推定し、問題解決知識の精錬を行う手法を開発することが考えられる。

問題解決記述を「事例記述のキーワード集合」を「問題を表すキーワード集合」と「解決方法を表すキーワード集合」に分割することや、問題が発生した時刻、問題が発生した状況、問題を解決した人物等の比較的簡単に抽出可能と思われる特性を個別の属性とすることにより、高度な検索が実現できるように拡張する予定である。

謝 辞

日頃より御指導いただいている九州大学大学院芸術工学研究院藤村直美教授に感謝いたします。また、本研究に関して熱心に御討論頂き、実験に協力して頂いた名古屋大学大学院情報科学研究科渡邊研究室の皆様感謝いたします。

文 献

- [1] H. Cui, J.-R. Wen, J.-Y. Nie and W.-Y. Ma: “Query expansion by mining user logs”, IEEE Trans. on Knowledge and Data Eng., **15**, 4, pp. 829–839 (2003).
- [2] 徳永: “情報検索と言語処理”, 東京大学出版会 (1999).
- [3] S. Lawrence: “Context in web search”, IEEE Data Engineering Bulletin, **23**, 3, pp. 25–32 (2000).
- [4] L. Finkelstein, E. Gabrilovich, Y. M. E. Rivlin, Z. Solan, G. Wolfman and E. Ruppin: “Placing search in context: The concept revisited”, Proc. of WWW10, Hong Kong, pp. 406–414 (2001).
- [5] 赤星, 小山, 角谷, 田中: “携帯端末による電子メール交換に基づく web 検索”, DEWS2003 (2003).
- [6] 中島, 黒田, 田中: “閲覧履歴を反映したコンテキスト依存型 web ブックマーク”, 情報処理学会論文誌: データベース, **43**, SIG5(TOD 14), pp. 23–36 (2002).
- [7] 石川: “インターネットコミュニティ戦略”, ソフトウェアパブリッシング (2001).
- [8] <http://java-house.jp/ml/>.
- [9] H. Turtle and W. B. Croft: “Evaluation of an inference network-based retrieval model”, ACM Transactions of Information Systems, **9**, 3, pp. 187–222 (1991).