手順情報に対する補完情報の検索と統合

尼崎　澄人†　　大島　裕明†　　田中　克己†

† 京都大学大学院情報学研究科社会情報学専攻　〒 606–8501 京都市左京区吉田本町 36-1
E-mail: †{amagasaki,ohshima,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし　本研究では，Web 上に存在する「手順」を説明するページに対して，より詳細な情報を検索する手法と，得られた情報を元のページに統合する手法を提案する．現在，Web 上には，料理レシピやソフトウェアのインストールのやり方についての説明など，手順を記述したページが数多く存在する．しかし，手順についての情報は，その記述の完全性や詳細度などが，ページ毎に異なっており，利用するユーザの立場からは不十分である場合がある．そのような場合，元のページの情報に加えて，それを補完する情報をユーザ自身が検索する必要がある．本研究では，手順情報を記載したページを基に，補完情報を検索し，得られた補完情報を元の手順情報に統合する手法を提案する．補完情報の検索と統合を行うために，本研究では，(1) 補完情報を検索する手法，(2) 手順情報ページに補完情報を挿入・統合する仕組みを提案する．具体的には，(1) では，手順情報の中で選択した手順について，Web 上から補完情報を自動で検索する手法を提案し，(2) では，(1) の手法で得られた補完情報を，手順情報の中に適切に統合する機能を，ブラウザの拡張機能として実現する．

キーワード　情報補完，ページ統合

1. はじめに

日常生活で，自分があまり知らないことをしようとした場合に手順情報が書かれた Web ページを参照することは珍しいことではない．今では，wikiHow(注1) や cookpad(注2) などのような様々な種類の手順情報が集約された Web サービスは増えている．しかし，このような手順情報は，その記述の完全性や詳細度などがページ毎に異なっている．そこで，利用するユーザが情報が不十分であると感じた場合，それを補完する情報をユーザ自身が検索しないといけないようになっているものが多い．また，ユーザが得た補完情報は手順ページとは別々の Web ページであり，ユーザはその補完情報を手順ページと関連付けて見るためには，それぞれページを行き来する必要がある．これには，ユーザが補完されている手順と補完情報を記憶しながら見なければならない．しかし，ユーザが理解していない手順や補完情報については，記憶することは難しく，このページを交互に行き来する動作はユーザの理解の妨げにもなる．

たとえば，「筋肉を鍛えて減量する方法」という wikiHow の手順ページ (注3) がある．このページには，「筋肉を鍛えて減量する方法」として「タンパク質を摂取する」手順が説明されている．それには，赤身肉や鶏肉を食べるように説明されているが，実際には他の方法で摂取したいということがある．このような場合に，「タンパク質 摂取」のようなクエリで Web 検索を行い，他の方法，サプリメントや別の食材でタンパク質を摂取する方法が書かれているページを検索して情報を補完することが考えられる．また，この手順ではタンパク質を摂取できる食材について書かれているが，それらをどのように調理するのかという詳細な情報についての説明はない．このような場合にも，「鶏肉 調理方法」のようなクエリで Web 検索を行い，鶏肉の調理方法について書かれているページを検索して情報を補完することが考えられる．これらのユーザの操作は，ユーザの一貫的な手順ページの閲覧を阻害し，手順の理解の妨げにもなる．

そこで本研究では，任意の手順に対して自動で補完情報を検索することと，手順ページの手順について補完情報を統合することによって，ユーザの手順ページの閲覧の効率を高めることを目的とする．

本論文の構成は，第 2 章で関連研究について述べる．第 3 章では，手順情報に対する補完情報を検索する方法を提案する．第 4 章では，Web 検索エンジンを用いた補完情報の検索手法について説明する．第 5 章では，手順情報と補完情報の統合方法について提案する．第 6 章では，手順情報に対する補完情報の検索の評価を行う．第 7 章では，まとめと今後の課題について述べる．

2. 関連研究

本研究に関連する先行研究がいくつか存在する．それについて以下に述べる．

Eklow らは，補完候補の Web ページそれぞれについて，話題を構造化して補完情報を抽出することによって，Wikipedia のページに対しての Web からの情報補完を行っている [1]．Takata らは，QA サイトの QA ページに対して，回答されている答えとは異なる答えとなる Web ページを補完している．これは，補完候補の Web ページが QA ページの質問の答えとして適しているか，QA ページ内の答えとは異なっているかを考慮する [4]．馬らは，Query-Free 検索機構により，TV 放送の字幕データから話題構造を抽出する．そして，話題構造から検索クエリを生

(注1)：http://www.wikihow.com/
(注2)：http://cookpad.com/
(注3)：http://www.wikihow.jp/筋肉を鍛えて減量する

成して放送内容を補完する Web ページの検索手法を提案している [6]．若宮らは，Web ページ内のシーン記述を動画補完する．そのために，Web 上で共有されている動画内のユーザによるコメントを用いる．手法として，シーン記述を共有動画のユーザコメントの特徴を持つキーワードに変換することで，そのシーンを動画から抽出をする手法を提案している [7]．Okamoto らは，固有表現抽出を用いて，ユーザが見ている HTML 文書中の単語をジャンル分けする．そして，ユーザが見ている Web ページからクエリを抜き出して，検索を行うアプリケーションを開発している [3]．

以上のように，あるコンテンツを Web 上の情報から補完情報を検索して，補完するといった研究はいくつか存在している．本論文では，Web 上のページから手順情報に対する補完情報を検索して，統合することによる補完を提案している．

また，文書間の構造や関係を明らかにする研究がある．戸田らは，文書集合に含まれている話題や各文書へアクセスを効率的に行うために，グラフ分析を用いる．それで，文書集合中に含まれる話題の抽出や，話題間の関係と話題に対する各文書の位置付けを明らかにしている [9]．Zhu らは，Web ページ間の関係をニューラルネットワークによって識別する手法を提案している [5]．Ma らは，Web ページの話題構造を Web ページのタイトルと，ページ内コンテンツから抽出している. [2]

さらには，HTML 文書を自動的に構造化する手法の研究がある．南野らは，HTML 文書中に含まれる繰り返し構造を再帰的に検出することによって，Web ページの自動的な構造化を行っている [8]．砂山らは，HTML 文書を文書中の HTML タグや句点を用いて，テキスト部分を通常の文としてセグメントに分割するシステムを提案している [10]．

補完を行う際には，検索クエリの生成や，文書間の構造を発見することが必要になり，HTML 文書の構造化も行う必要が出てくる．本論文では，補完情報が記載されたページから，補完対象の手順との関係性を見出すために，それぞれの文書の類似度を計っている．

3. 手順情報に対する補完情報検索

3.1 手順情報の定義

本節では，本研究における手順情報について定義する．

Web 上には手順情報が数多く存在し，今でも増加している．たとえば,「筋肉を鍛えて減量する方法」や,「チーズフォンデュの作り方」,「OS のインストール方法」のような手順情報が Web 上に存在している．これらの手順情報は共通して，タイトル，説明，手順群の要素が見られた．そこで本研究では，手順情報をタイトル，説明，手順群から構成される,「やり方」についての情報であると定義する．たとえば，wikiHow の手順情報「筋肉を鍛えて減量する方法」は，次のようになる．

- タイトル
  筋肉を鍛えて減量する
- 説明
  減量をするとともに筋肉を増やすには...（省略）
- 手順群
  - 手順 1
    より多くのタンパク質を摂取する...（省略）
  - 手順 2
    炭水化物を制限する...（省略）
  - 手順 3
    摂取カロリーを調節する...（省略）

  ⋮

  - 手順 15
    リバースグリップ・バーベルカール...（省略）

3.2 手順情報に対する補完情報

本節では，手順情報に対する補完情報について説明する．

ユーザが発見した手順情報は，そのユーザにとって詳細度や具体性が足りなかったり，そもそも情報が欠けていたりする場合がある．補完情報とは，その際に必要とされる情報である．

実際に，どのような補完情報が必要になるのか，wikiHow 内の手順情報をいくつか観察して分析した．その結果，特化補完情報，詳細補完情報，同位補完情報，その他の 4 つに補完情報が分類された．

- 特化補完情報

特化補完情報とは，補完対象の手順に対して，特化ないしは，具体化した記述による補完情報である．たとえば，補完対象の手順が「タンパク質を摂取する」であった場合，特化補完情報としては,「鶏肉を食べる」,「サプリメントを摂る」のような情報が挙げられる．この補完情報は，手順として実行した場合に，補完対象の手順を実行した事にもなる．

- 詳細補完情報

詳細補完情報とは，補完対象の手順に対して，細分化されている記述による補完情報である．たとえば，補完対象の手順が「シャドーボクシングをする」であった場合，詳細補完情報としては,「足を肩幅程度に広げる」,「左ジャブをする」のような情報が挙げられる．

- 同位補完情報

同位補完情報とは，補完対象の手順に対して，その中で並列に列挙されている情報と同位である記述による補完情報である．たとえば，補完対象の手順内で「避けるべき加工食品」がいくつか列挙されていた場合，同位補完情報としては,「ポテトチップ」,「冷凍食品」のような「避けるべき加工食品」のような情報が挙げられる．

- その他

その他とは，補完対象の手順に対して，特化補完情報，詳細補完情報，同位補完情報とは異なる記述による補完情報である．

ここで，以上に挙げたように，分類によって補完は異なる．そこで本研究では,「特化補完情報」と「詳細補完情報」を対象として検索手法と統合手法の提案を行う．

3.3 補完情報検索における問題定義

本節では，手順情報に対する補完情報を，Web 上から検索を行う際の問題を定義する．

補完情報を Web 上から検索するためには，Web 上から補完情報の記述が含まれる Web ページを，検索するためのクエリを生成する必要がある．また，補完情報の記述が含まれる Web ページからその記述の抽出もする必要がある．なぜなら，検索から得られた，補完情報の記述が含まれる Web ページからは，補完情報の記述部分が明示されていない．そのため，補完情報の記述部分を抽出する必要があるからである．

3.4 補完情報検索のアプローチ

節では，3.3 節で述べた問題を解決するためのアプローチを説明する．

補完情報検索のアプローチは，

（1） クエリ生成と Web 検索，

（2） 補完情報抽出

に分けられる．ここで，補完情報検索のアプローチの概要図を図 1 に示す．

では，補完情報検索のアプローチを説明するために，最初に，補完情報検索における入力と出力を示す．まず，補完情報検索における入力である，手順情報，補完対象として選んだ手順を，

$$(H, K)$$

と記述する．ここで，H は手順情報で，K は補完対象として選んだ手順である．また，補完情報検索における出力である，補完情報の記述の候補のリストを，

$$(C, \leqq)$$

と記述する．ここで，C は補完情報の記述の集合で，$\leqq$ は C に対する順序集合である．次に，補完情報検索における入力から，出力を得るまでのアプローチを説明する．まず，クエリ生成と Web 検索では，補完対象の手順 K と，それを含む手順情報 H，それぞれから，補完情報の記述が含まれる候補ページを検索するクエリ集合 Q を生成する．そして，そのクエリ q を用いて Web 検索を行い，候補ページ集合 R を得る．ここで，クエリ生成と Web 検索の概要図を図 2 に示す．それから，補完情報抽出では，検索結果となる候補ページ集合 R の要素である，候補ページ P から，その記述 c を抽出する．ここで，補完情報抽出の概要図を図 3 に示す．

図 1 全体のアプローチの概要図

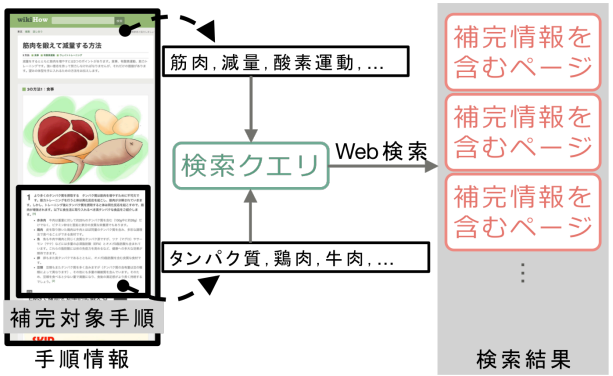

図 2 クエリ生成と Web 検索の概要図

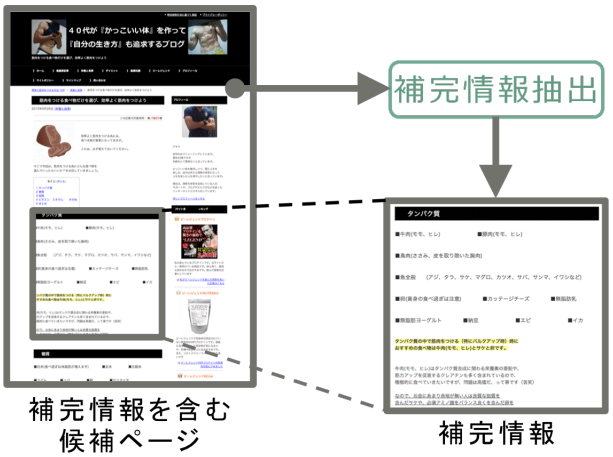

図 3 補完情報抽出の概要図

4. Web 検索エンジンを用いた補完情報の検索手法

本章では，3.4 節で述べた，補完情報のアプローチに基いて Web 検索のためのクエリを生成し，得られた検索結果の Web ページから補完情報を抽出する手法について説明する．

4.1 クエリ生成手法

本節では，3.4 節で述べた，補完情報検索のアプローチのうち，クエリ生成と Web 検索の，クエリ生成手法について説明する．

端的にはクエリ生成手法では，手順情報 H，補完対象として選んだ手順 K を入力として，補完情報の記述が含まれる候補ページ群 R を検索するクエリ集合 Q を生成する．たとえば，wikiHow の手順情報「筋肉を鍛えて減量する」と，その中の補完対象として選んだ手順「タンパク質を摂取する」を入力すると，クエリ「筋肉 タンパク質」を生成する．

次に，クエリ生成手法について，詳しく説明する．まず，入力の手順情報 H は，3.1 節で述べた通り，タイトル，説明，手順群で構成されるものとする．これらを，

$$\left(d^t, d^o, S\right)$$

と記述する．ここで，d^*_* は，「ありうる全ての文」の集合を D と定義した場合に，

$$d^*_* \in D$$

を満たす．また，本研究では，名詞のみに着目し，文章 d_*^* を名詞の多重集合とみなすこととする．次に，S は手順群で，

$$S = \{d_1^s, d_2^s, d_3^s, \ldots, d_n^s\}$$

と表すこととする．そして，もう一方の入力の，補完対象として選んだ手順 K を，

$$K = d_k^s$$

と記述する．

次に，手順情報 H と補完対象として選んだ手順 K で特徴的な名詞を，抽出する．なぜなら，手順情報 H と補完対象として選んだ手順 K で特徴的な名詞は，補完情報の記述 c に含まれると考えられる．そのため，検索クエリ q に用いる語として適していると考えられるからである．まず，特徴的な名詞を抽出するために，文章 d_*^* での名詞 $noun$ の頻度を，

$$tf\,(d_*^*, noun) = |\{w | w \in d_*^*, w = noun\}|$$

と定義する．そして，手順情報 H で特徴的な名詞を求める．手順情報 H で特徴的な名詞は，頻出であると考えられる．ここで，手順情報 H に含まれる名詞 $noun_j^H$ の頻度は以下の式で求められる．

$$\begin{aligned} tf^H\left(noun_j^H\right) &= tf\left(d^t, noun_j^H\right) \\ &+ tf\left(d^o, noun_j^H\right) \\ &+ \sum_i tf\left(d_i^s, noun_j^H\right) \end{aligned}$$

これを，手順情報 H に含まれる全ての名詞について求める．さらに，$tf^H\left(noun_j^H\right)$ が高い順に全ての名詞を並べて，順位 n 位の名詞を $ranked^H(n)$ とする．

また，補完対象として選んだ手順 K で特徴的な名詞を求める．補完対象として選んだ手順 K で特徴的な名詞は，手順 K を含む手順情報内の，他の手順にはあまり含まれていない名詞であり，かつ，手順 K で頻出であると考えられる．ここで，補完対象として選んだ手順 K に含まれる名詞 $noun_j^K$ が出現する，手順の頻度は以下の式で求められる．

$$sf\left(noun_j^K\right) = \left|\left\{d_i^s | tf\left(d_i^s, noun_j^K\right) > 0\right\}\right|$$

さらに，$tf-idf$ の考え方を基にした値を，補完対象として選んだ手順 K に含まれる名詞 $noun_j^K$ について以下の式で求める．

$$tfisf(noun_j^K) = tf\left(d_k^s, noun_j^K\right) * \log\left(\frac{|S|}{sf(noun_j^K)}\right)$$

これを，補完対象として選んだ手順 K に含まれる全ての名詞について求める．さらに，$tfisf(noun_j^K)$ が高い順に全ての名詞を並べて，順位 n の名詞を $ranked^K(n)$ とする．

次に，補完情報の記述が含まれる候補ページ群を検索するための検索クエリ集合 Q を生成する．クエリ生成には，複数名詞手法と intitle-inbody 手法の 2 つの手法を提案する．まず，複数名詞手法では，クエリ $Q = \{q1, q2\}$ を以下のように生成する．

$$\begin{aligned} q1 &= ranked^H(1) \wedge ranked^K(1) \\ q2 &= ranked^H(1) \wedge ranked^K(2) \end{aligned}$$

具体的には，wikiHow の手順情報「筋肉を鍛えて減量する」で複数名詞手法を用いると，

$$\begin{aligned} q1 &= \text{筋肉} \wedge \text{タンパク質} \\ q2 &= \text{筋肉} \wedge \text{鶏肉} \end{aligned}$$

のようなクエリが生成される．このクエリでは，補完対象として選んだ手順 K に含まれる名詞を複数用いているので，多様的な検索結果 Web ページ群を得られると考えられる．次に，intitle-inbody 手法では，Web ページタイトルに $ranked^H(1)$ を含み，Web ページ内に $ranked^K(1)$ または，$ranked^K(2)$ を含む Web ページを検索するクエリ集合 Q を生成する．具体的には，wikiHow の手順情報「筋肉を鍛えて減量する」で intitle-inbody 手法を用いると，

$$q = intitle : \text{筋肉} \wedge (inbody : \text{タンパク質} \vee inbody : \text{鶏肉})$$

のようなクエリが生成される．このクエリでは，Ma らの話題構造を抽出する手法 [2] を参考にしている．

以上の 2 つの手法により得られたクエリの評価は，6.1 節にて行う．

4.2 補完情報抽出手法

本節では，3.4 節で述べた，補完情報検索のアプローチのうち，補完情報抽出について説明する．

端的には補完情報抽出では，補完情報の記述が含まれる候補ページ P と，補完対象の手順 K，それを含む手順情報 H を入力として，その記述 c を抽出する．たとえば，wikiHow の手順情報「筋肉を鍛えて減量する」と，その中の補完対象として選んだ手順「タンパク質を摂取する」，その補完「筋肉をつける食べ物だけを…（省略）」を入力すると，補完情報の記述「無脂肪ヨーグルトや納豆でもタンパク質が摂取できる」が抽出される．

次に，補完情報抽出の手法について，詳しく説明する．まず，候補ページ P は，以下のように段落に分けられる．

$$P = \{d_1^p, d_2^p, d_3^p, \ldots, d_t^p\}$$

この段落について，最も補完情報の記述に適している段落を得る．そのために各段落で，補完情報に適している度合いを示すスコア $score_n$ を計算する．スコア計算手法には，非類似補完手法と弱類似補完手法の 2 つの手法を提案する．しかしその前に，2 つの手法で用いる類似度について説明する．2 つの手法では，補完情報に適している度合いを，「手順情報 H」，「補完対象として選んだ手順 K」，「候補ページ P の段落 d_1^p」のそれぞれの間の類似度を用いて求める．そのため，文章間の類似度を $sim(d_1, d_2)$ で定義する．このとき，手順情報 H の文章を d^h とする．

まず，非類似補完手法では，スコア $score_n$ を以下のように計算する．

$$score_n = (1 - sim(d_k^s, d_n^p)) - \left|sim\left(doc^h, doc_k^s\right) - sim\left(doc^h, doc_n^p\right)\right|$$

非類似補完手法は,「手順情報 H と補完対象として選んだ手順 K」と「手順情報 H と補完情報の記述」は類似の仕方が近いと仮定する．そのため,「手順情報 H と補完対象として選んだ手順 K」の類似度と,「手順情報 H と Web ページ P の段落 d_n^p」の類似度が近いほどスコア $score_n$ が高くなる．また,「補完対象として選んだ手順 K と補完情報の記述」は類似していないと仮定する．そのため,「補完対象として選んだ手順 K と Web ページ P の段落 d_n^p」の類似度が低いほどスコア $score_n$ が高くなる．次に，弱類似補完手法では，スコア $score_n$ を以下のように計算する．

$$score = (|0.5 - sim(doc_k^s, doc_i^p)| \times \left|sim\left(doc^h, doc_k^s\right) - sim\left(doc^h, doc_i^p\right)\right| + 1)^{-1}$$

弱類似補間手法では,「手順情報 H と補完対象として選んだ手順 K」と「手順情報 H と補完情報の記述」は類似の仕方が近いという仮定は非類似補完手法と同じである．しかし,「補完対象として選んだ手順 K と補完情報の記述」は弱く類似しているという仮定が異なる．そのため,「補完対象として選んだ手順 K と Web ページ P の段落 d_n^p」の類似度が 0.5 に近いほどスコア $score_n$ が高くなる．

以上の 2 つの手法の評価は，6.2 節にて行う．どちらかの手法で，Web ページ P の各段落についてスコアを計算する．これにより，最も補完情報の記述に適している段落を得ることが出来る．しかし，求める補完情報の記述は，複数の段落に渡って記述されていることが考えられる．そのため，補完情報に適している度合いを下げずに，前後の段落を結合する．その方法を以下に示す．

（1） スコア $score_n$ が最高である段落を d^c とする．
（2） 段落 d^c とその前の段落を結合した段落を d^b とする．
（3） 段落 d^c とその次の段落を結合した段落を d^n とする．
（4） 段落 d^b のスコアと段落 d^n のスコアがどちらも，最初のスコア $score_n$ 以上ならば，スコアが高い方の段落を d^c として，手順 2 に戻る．そうでないならば，段落 d^c を補完情報の記述 c として終わる．

よって，候補ページ P から，その記述 c が抽出される．

5. 手順情報に対する補完情報の統合

本章では，手順情報に対する補完情報の統合について提案する．

5.1 補完情報統合における問題定義

本節では，手順情報に対する補完情報を統合する際の問題を定義する．

第 4 章では，手順情報に対する補完情報の検索を行った．それにより，手順情報に対する補完情報を得ることが出来た．しかし，手順情報が記載されたページと，第 4 章で得た補完情報の記述が含まれるページは別々のページである．そのため，お互いのページを行き来しながら，ページの補完を頭の中で行う必要がある．これらのページを同時に見ることが可能ならば，煩わしい操作や，情報を記憶しながらページの閲覧を行う必要が無くなる．しかし，現在のブラウザでは，複数のページを同時に見るには，ウィンドウを分割したりしなければならない．しかしそれは，手順情報に対する補完情報の統合に適切であるとは言えない．

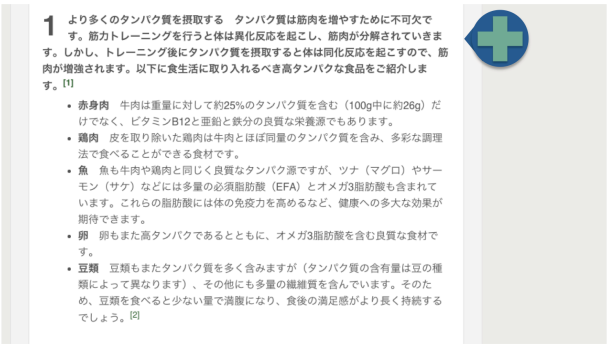

図 4 統合ボタンが手順に対して現れた例

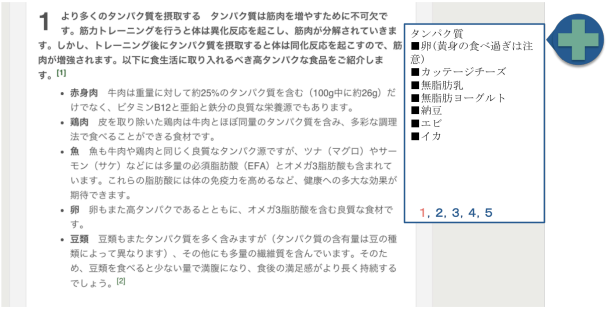

図 5 補完情報が統合された例

5.2 補完情報の統合手法

本節では，手順情報に対する補完情報の統合を行う手法について説明する．

まず，ブラウザ拡張によってどのようにページ群の統合がされるのか説明する．

最初に，手順情報の手順毎に図 4 のような統合ボタンがブラウザ上で表示される．これをクリックすると，図 5 のように第 4 章で得られた補完情報が統合される．また，1 つの手順につき，複数の補完情報が補完されており，補完情報下部の数字をクリックすることで，他の補完情報が図 6 のように表示される．

この手法では，手順毎に補完情報が補完できるため，手順情報の複数の手順についての補完情報の補完を同時にすることが出来る．また，補完情報の候補を複数挙げることにより，補完情報の多様化も行われる．

6. 評 価

本章では，本研究で提案した，手順情報に対する補完情報の検索手法の評価について述べる．

6.1 検索クエリ生成手法の評価

本節では，4.1 節で述べた，2 つの検索クエリ生成手法，複

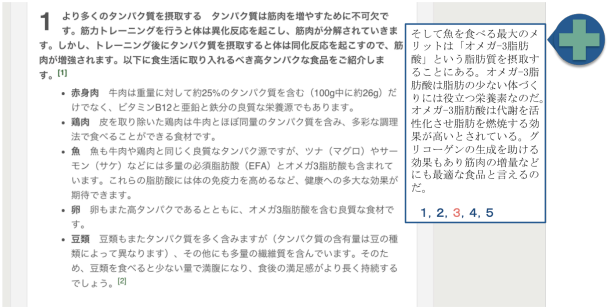

図 6　統合する補完情報を変更した例

表 1　補完情報を含む候補ページ検索の評価に使う手順

wikiHow ページ名	パート
筋肉を鍛えて減量する	1 パートの 1
インフルエンザを治す	3 パートの 3
鶏肉をさく	2 パートの 2
猫を追い払う	1 パートの 5
パソコンのメンテナンスをする	1 パートの 1

表 2　実験結果：特化補完情報適合率

手順情報	複数名詞手法	intitle-inbody 手法
筋肉を鍛えて減量する	0.5	0
インフルエンザを治す	0.3	0.1
鶏肉をさく	0	0
猫を追い払う	0.3	0
パソコンのメンテナンスをする	0	0

数名詞手法と intitle-inbody 手法について評価を行う．

評価方法は，まず，wikiHow の手順情報を無作為に 5 つ選ぶ．そして，それぞれの手順情報について，補完対象とする手順も無作為に選ぶ．次に，選ばれた補完対象とする手順について，2 つの検索クエリ生成手法を用いて，検索クエリを生成し，その検索クエリで Bing で Web 検索を行い，10 件ずつの候補ページを取得する．最後に，それぞれの候補ページについて，特化補完情報の記述が含まれているかどうかを評価，また，詳細補完情報の記述が含まれているかどうかを評価する．

今回設定した 5 つの手順情報と，その補完対象として選んだ手順がどの部分であるかを，表 1 に示す．

選んだ手順に対して，特化補完候補として挙げられることを期待する結果は，たとえば，「筋肉を鍛えて減量する」で選んだ手順では，より多くのタンパク質を摂取することが書かれているため，タンパク質を摂取することの特化であると考えられる，「鶏肉を食べる」，「サプリメントを摂る」などが候補として挙げられることを期待する．

結果として，特化補完情報の記述が含まれていると認められた候補ページが，生成したクエリでの検索結果にどれだけ現れたかを表す適合率を，表 2 に示す．また，詳細補完情報の記述が含まれていると認められた候補ページが，生成したクエリでの検索結果にどれだけ現れたかを表す適合率を，表 3 に示す．

以上の結果より，intitle-inbody 手法に比べると，複数名詞手法が補完情報を含む候補ページを検索するクエリを生成する手法として若干優位であるということが分かる．しかし，「鶏肉をさく」，「パソコンのメンテナンスをする」の手順情報に対しての補完が含まれているページがどの手法からも得られなかった．恐らくこれは，これ以上特化できないレベルの手順を対象にしていたことが原因であると考えられる．

表 3　実験結果：詳細補完情報適合率

手順情報	複数名詞手法	intitle-inbody 手法
筋肉を鍛えて減量する	0.5	0
インフルエンザを治す	0.1	0.1
鶏肉をさく	0	0
猫を追い払う	0.1	0
パソコンのメンテナンスをする	0	0

表 4　評価する補完情報を抽出する手順情報と候補ページの組合せ

手順情報	候補ページ名
筋肉を鍛えて減量する	筋肉をつける食べ物だけを選び、効率よく筋肉をつけよう — 簡単に筋肉をつける方法
インフルエンザを治す	インフルエンザ
猫を追い払う	猫の爪切りの仕方〜爪の構造・爪切りの道具や手順を理解し、猫の爪のケアを学ぶ

表 5　実験結果：正しく補完情報を抽出した数

手法名	正解抽出数
非類似補完手法	1
弱類似補完手法	1

6.2　スコア計算手法の評価

本節では，4.2 節で述べた，補完情報を含む候補ページから補完情報を抽出する際に用いる，2 つのスコア計算手法，非類似補完手法，弱類似補完手法について評価を行う．

評価方法は，まず，6.1 節で得られた，補完情報を含む候補ページのうち，特化補完情報の記述が見られる候補ページ 3 件を選ぶ．それぞれについて，2 つのスコア計算手法を用いて，特化補完情報の記述が抽出できるているか評価する．

今回設定した 3 件の手順情報と候補ページの組合せを，表 4 に示す．

結果として，全組合せについて，2 つの手法それぞれで，抽出したい補完情報が抽出できている数を，表 5 に示す．

以上の結果より，どちらの手法も 3 件中 1 件の抽出しか正解を得ることが出来なかった．結果として得られた抽出文は，短すぎる文が抽出されることがあった．よって，得られる抽出の文量も考慮して抽出を行う必要があると考えられる．これより，スコアの計算手法だけではなく，抽出のアルゴリズムも改善する必要がある．

7.　まとめと今後の課題

本研究では，まずは，手順情報に対する補完情報の検索を行った．まずは，補完情報が記載されたページを Web 上から検索するために，検索クエリの生成を行った．クエリ生成の手法として，複数名詞手法と intitle-inbody 手法を挙げて，実験を行った．その結果，複数名詞手法が若干優位な結果となったが，まだ候補ページを上手く検索できているとは言い難い．次

に，検索した結果の候補ページ内から，補完情報を抽出した．その内，補完情報の抽出に用いるスコアの計算手法についても，非類似補完手法と，弱類似補完手法を挙げて実験を行った．その結果，どちらの手法についても好ましい結果を得ることが出来なかった．手順情報に対する補完情報の検索については，今後改善していく必要がある．

もう一方では，手順情報が記載されたページに補完情報を統合する仕組みの提案を行った．提案手法では，手順毎に補完情報を補完することで，同時に複数の手順に対しての補完情報を補完できた．また，多様な補完情報を提示することができた．今後は，インタフェースの面でより良い統合を行えるようにすることを課題とする．

謝　　辞

本研究の一部は，文部科学省科学研究費補助金（課題番号15H01718，24680008）によるものです．ここに記して謝意を表します．

文　　献

[1] Damien Eklou, Yasuhito Asano, and Masatoshi Yoshikawa. How the web can help wikipedia: A study on information complementation of wikipedia by the web. In *Proc. 6th International Conference on Ubiquitous Information Management and Communication*, pp. 9:1–9:10, 2012.

[2] Qiang Ma and Katsumi Tanaka. Topic-structure-based complementary information retrieval and its application. *ACM Transactions on Asian Language Information Processing*, Vol. 4, No. 4, pp. 475–503, 2005.

[3] Masayuki Okamoto, Nayuko Watanabe, Masaaki Kikuchi, Takayuki Iida, Kenta Sasaki, Kensuke Horiuchi, Tomohiro Yamasaki, Sumi Omura, and Masanori Hattori. First query term extraction from current webpage for mobile applications. In *Proc. 9th International Conference on Mobile and Ubiquitous Multimedia*, pp. 19:1–19:9, 2010.

[4] Natsuki Takata, Hiroaki Ohshima, Satoshi Oyama, and Katsumi Tanaka. Searching the web for alternative answers to questions on webqa sites. In *Proc. 11th International Conference on Web-Age Information Management*, pp. 441–452, 2010.

[5] Zhu Xing, Shen Huang, and Yong Yu. Recognizing the relations between web pages using artificial neural network. In *Proceedings of the 2003 ACM symposium on Applied computing*, pp. 1217–1221, 2003.

[6] 馬強, 田中克己. 話題構造に基づく放送と web コンテンツの統合のための検索機構. 情報処理学会論文誌データベース, Vol. 45, No. 10, pp. 18–36, 2004.

[7] 若宮翔子, 北山大輔, 角谷和俊. Web ページ補完のための共有動画に付与されたユーザコメントを用いたシーン抽出手法. 全国大会講演論文集, pp. 775–776, 2010.

[8] 南野朋之, 齋藤豪, 奥村学. 繰返し構造に基づいた web ページの構造化. 情報処理学会論文誌, Vol. 45, No. 9, pp. 2157–2167, 2004.

[9] 戸田浩之, 北川博之, 藤村考, 片岡良治, 奥雅博. グラフ分析を利用した文書集合からの話題構造マイニング. 電子情報通信学会論文誌, Vol. 90, No. 2, pp. 292–310, 2007.

[10] 砂山渡, 井山晃洋, 谷内田正彦. 重要文抽出による web ページ要約のための html テキスト分割. 電子情報通信学会論文誌, Vol. 87, No. 12, pp. 1089–1097, 2004.