知能と情報（日本知能情報ファジィ学会誌）Vol.18, No.5, pp.682-688（2006）

解説

情報抽出を利用した複数文書要約†

難波 英嗣*

1．はじめに

コンピュータを用いてある文書の要約を自動的に生成するという研究は，自然言語処理分野の中でも歴史が古く，既に1950年代にはその最初の試みが報告されている[Luhn 1958]．このLuhnらによる報告では，与えられた文書の中から，まず文書中の頻度に基づいてその文書の重要語(キーワード)を特定，次に文書中の各文の重要度を，重要語を含む割合で算出し，最後にある一定の重要度以上の文を出力するという方法で要約を作成する「重要文抽出」と呼ばれる方法が提案されている．今日ではLuhnらが行なったような，ひとつの文書を要約システムの入力とし，ひとつの要約を出力しようとする研究は「単一文書要約」と呼ばれる．これに対し，あるトピックに関する複数の文書をシステムの入力とし，それらからひとつの要約を作成する「複数文書要約」という研究も1990年代から始まっており，近年では，むしろ単一文書要約より盛んに研究されていると言っても過言ではない．

もっとも，一般的な複数文書要約の作成手順は，Luhnらの単一文書要約手順と多くの点で共通している．具体的には，まず与えられた複数の文書の中から重要語を特定し，次に文書中の各文の重要度を，重要語を含む割合で算出し，ある一定の重要度以上の文を出力するという方法で作成される．

単一文書要約との違いは，「文書間の共通個所の特定」と「抽出された重要文の出力順序の決定」を考慮する必要がある点である．前者に関して，入力される複数の文書間で内容が重複する場合，上述の手法だけで複数文書の要約を作成すると，要約中の記述が重複し，冗長な文書が出力されてしまう可能性がある．このため，冗長な個所(文書間の共通個所)をどのように検出し削除するかが問題となる．後者は，「文書から断片的に抽出した個所(文)を並べただけでは読みにくい」という問題を，抽出された重要文の文間のつながりを考慮して解決する，という課題である．この課題は，単一文書要約においても無関係ではないが，複数文書要約においてこの問題はより一層深刻であり，単一文書要約以上に活発に研究が行われている[Barzilay, et al., 2002；Lapata, 2003；Bollegala et al., 2006]．

複数文書要約研究においては，以上で述べた手順に「情報抽出」と呼ばれる手法を取り入れたり，情報抽出結果を用いるという文抽出とは別のアプローチを採用したりする要約作成研究が行われるようになっている．本稿では，この情報抽出に基づく複数文書要約について述べる．

本稿の構成は以下のとおりである．2節では，情報抽出について簡単に紹介する．続く3節では，複数文書要約を用いた情報抽出手法について述べ，4節で，いくつか研究事例を紹介し，5節で本稿をまとめる．

2．情報抽出

2.1 情報抽出とは

情報抽出とは，あらかじめ指定された情報を自然言語で記述された文書中から抽出することをいう．ここで「あらかじめ指定された情報」とは，対象とする文書によっても異なるが，一般には人名，組織名，日時などの固有表現(Named Entity)のことを指す．

以下に，情報抽出とはどのような課題であるのか，具体例を用いて説明する．例えば，代表的な形態素解析器のひとつである茶筌[1]を用いて「太郎は山谷工業高校の学生です。」という文を解析すると，図1のような解析結果が得られる．

図の各行は，左から順に「見出し(出現形)」「読み」「見出し(基本形)」「品詞」を示している．この結果を見ると，「太郎」の品詞は名詞で，人名を表していることがわかる．一方，「山谷工業高校」は「山谷」と「工業」と「高校」という3つの名詞に分割されている．このような結果になるのは，茶筌が用いているIPADICという辞書の中に「太郎」の品詞は名詞であり，かつ人名であるという情報は登録されているが，「山谷工業高校」と

† Multi-document Summarization using Information Extraction
Hidetugu NANBA

* 広島市立大学　情報科学部
Faculty of Information Sciences, Hiroshima City University

1 http://chasen.naist.jp/hiki/ChaSen/

太郎	タロウ	太郎	名詞-固有名詞-人名-名
は	ハ	は	助詞-係助詞
山谷	サンヤ	山谷	名詞-固有名詞-地域-一般
工業	コウギョウ	工業	名詞-一般
高校	コウコウ	高校	名詞-一般
の	ノ	の	助詞-連体化
学生	ガクセイ	学生	名詞-一般
です	デス	です	助動詞　特殊・デス　基本形
。	。	。	記号-句点

図1　茶筌を用いた「太郎は山谷工業高校の学生です。」の解析結果

という単語は登録されていないためである．もし，辞書に「山谷工業高校」という単語を「名詞-固有名詞-組織」として登録すれば正しく解析できる．しかし，世の中に実在する組織や人物，映画や小説などに出てくる架空の組織や人物なども含め，あらゆる固有表現を網羅するのは現実的にはほとんど不可能であると言ってよい．そこで，辞書に登録されていない未知の固有表現でも，例えば組織名のようなひとつの情報のまとまりを自動的に認識することが，情報抽出の課題のひとつであると言える．そしてこのように認識された固有表現に基づいて，例えば，「太郎」という人物が「山谷工業高校」という組織に所属しているという情報を得ることが情報抽出の目的である．

ちなみに，図1で例に用いた「太郎は山谷工業高校の学生です．」という文を奈良先端科学技術大学院大学の工藤(現，グーグル株式会社)らが開発した南瓜(CaboCha)[2]というツールを用いて解析すると，図2のような結果が得られ，「山谷工業高校」はひとつの固有表現として扱われている．南瓜とは，文の係り受け構造を解析するツールで，同時に組織名や人名などの固有表現の抽出も自動的に行う．図2において，「太郎」の前後には人名を示す〈PERSON〉というタグが，「山谷工業高校」の前後には組織名を示す〈ORGANIZATION〉というタグが付与されており，このシステムでは固有表現が正しく認識されていることが分かる．

```
<PERSON>太郎</PERSON>は---D
<ORGANIZATION>山谷工業高校</ORGANIZATION>の-D
学生です。
```

図2　南瓜を用いた「太郎は山谷工業高校の学生です。」の解析結果

2.2 情報抽出の歴史

情報抽出という研究課題が広く知れ渡るようになったきっかけは，1987年から米国で始まったMUC(Message Understanding Conference)と呼ばれる会議である．これは，複数の研究グループが同一タスクについて競い合う評価型の会議で，1987年から1998年まで7回開催された．MUCでは，海軍の作戦指令やテロ事件など「国防」に関する情報や，マイクロエレクトロニクスの合弁事業や経営トップの交代など「企業活動」に関する情報を新聞記事などから抽出するという課題が設定された．ここで抽出する情報とは，例えばテロ事件の場合，「事件発生日」，「事件発生場所」，「死傷者数」などである．これらの情報の抽出にはパターンマッチング技術が広く使われ，一定の成果を挙げたが，対象となる文書の領域が変わるたびに新たにパターンを一から手作業で作り直さなければならないため，領域を超えた技術の蓄積がほとんどできないという問題があった．この問題に対処するため，情報抽出を領域に依存するものとそうでないものに分け，領域に依存しない情報の抽出を情報抽出分野の課題のひとつとして研究するようになった．具体的には，人名，組織名，場所の名前，日時，数量表現などを文書から抽出するというものである．この課題は，1998年から1999年にかけて国内で開催されたIREX(Information Retrieval and Extraction eXercises)という評価型の会議でも採用され，日本語文書を対象にした情報抽出システムの開発に，国内の多くの研究グループが取り組んだ．現在では，三重大学の桝井らが開発しているNeXT[3]や，2.1節で紹介した南瓜などが日本語の情報抽出器として利用可能である．なお，情報抽出に関する詳しい解説は，[関根，2004，関根，1999，徳永，2005]などの解説記事を参照されたい．

2 http://chasen.org/~taku/software/cabocha/
3 http://www.ai.info.mie-u.ac.jp/~next/next.html

3．複数文書要約における情報抽出の利用

2節で述べた情報抽出の技術を複数文書要約で利用する研究はいくつかあるが，それらは大きく2つに分けることができる．ひとつは，従来の複数文書要約の手順に沿って要約を作成し，その過程で要素技術のひとつとして情報抽出を利用するという方法である．もうひとつは，情報抽出技術を用いて抽出された固有表現を用いて文書を生成したり，表や図にまとめたりするといった方法で，重要文や重要個所抽出に基づく前者のアプローチと異なる．以下，3.1節と3.2節で，これらの2つの手法について説明する．

3.1 従来の複数文書要約手法における情報抽出の利用

一般的な(従来の)複数文書要約の大まかな作成手順[奥村・難波，2005]については，すでに1節でも述べたが，改めて，以下に手順を示す．

(1)関連する文書の自動収集(システム入力)
(2)重要文抽出
(3)文書間の共通点と相違点の抽出
(4)重要個所抽出
(5)重要個所の出力順序の決定
(6)書き換え
(7)要約結果の提示

この手順の中で情報抽出が利用されるのは手順(2)である．その利用方法は，対象となる文書集合のトピックにより異なるが，例えば，ある人物に関する複数の文書から要約を作成する場合[Schiffman et al., 2001]，その人物に関する文を中心に(重要文と考えて)抽出する．その際，文書中の“he”のような代名詞や“the leader”のような定名詞句が具体的に誰を指しているのかを解析しておく必要がある．このような処理を参照関係の同定と呼ぶ．参照関係の同定は，情報抽出において重要な要素技術のひとつと考えられており，2.2節で述べたMUCでも，サブタスクのひとつとして設定されていた．

人物に関する要約以外では，文書集合中に出現する複数のイベントの情報に基づく要約作成でも情報抽出が利用されている[Li et al., 2006]．この手法では，固有表現レベルで重要度を計算しておき，その結果を重要文抽出に反映させる点が，直接文の重要度を計算する従来手法と異なる．その手順は，まず文書集合から複数のイベントに関する固有表現(日時，場所，人名)を抽出し，次に固有表現間の関係を考慮して重要な固有表現を特定，最後にその結果に基づいて重要文を抽出する．なお，イベント情報に基づく要約については，4.1節で詳しく紹介する．

3.2 情報抽出の結果を利用した複数文書要約

情報抽出の結果を利用した複数文書要約では，まず要約対象となる文書集合から情報抽出器を用いて固有表現を抽出し，次にそれらを関連付け，最後に文生成器を用いて文書を生成する[Radev and McKeown, 1998]．あるいは文書生成の代わりに図[難波他，2005]や表[Shinyama and Sekine，2006]を生成し，それらを要約結果としてユーザに提示する．

複数文書の内容をまとめる際，ある種の情報は文書としてよりも図として出力したほうが分かりやすい場合がある．例えば，「日経平均株価」や「失業率」に関する新聞記事集合は，株価や失業率といった数値情報を抽出して，その推移をグラフとして提示したほうが，文書として提示するよりもユーザにとって直感的に理解しやすい．近年，複数の文書から動向情報を抽出・可視化するという研究課題に対する研究者の関心が集まりつつある．また，この課題に関する評価ワークショップMuSTが国立情報学研究所主催のNTCIRワークショップのパイロットタスクとして，2005年から始まっている[加藤他，2004]．本特集号で，掲載されている論文のいくつかはMuSTに関連するものである．

4．研究事例

本節では，3節で述べた要約手法に関して，いくつか研究事例を紹介する．4.1節では，従来の複数文書要約手法において情報抽出を利用した研究[Li et al., 2006]について，4.2節では，情報抽出の結果を利用した複数文書要約研究[Radev and McKeown, 1998; Shinyama and Sekine, 2006；難波他, 2005]について，それぞれ紹介する．

4.1 従来の複数文書要約手法において情報抽出を利用した研究

Liら[Li et al., 2006]は，情報抽出器を用いて，まず要約対象となる複数の文書から固有表現(人名，組織名，時間，場所)と動詞や動作名詞(以後，イベント用語)を抽出してイベントマップという図式表現に変換する．次にGoogle社のWeb検索ランキングアルゴリズムであるPageRankをイベントマップに適用して重要な固有表現を抽出し，最後にこれらの固有表現を多く含んだ文を重要文としてさらに抽出することで要約の作成を行っている．

図3は，ある一文を図式表現に変換した例である．

<Organization> America Online </Organization> was to buy <Organization> Netscape </Organization> and forge a partnership with <Organization> Sun </Organization>, benefiting all three and giving technological independence from <Organization> Microsoft </Organization>

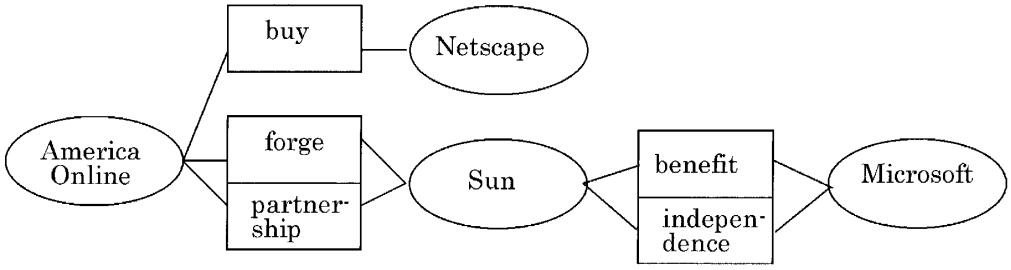

図3 文とその図式表現［Lin et al., 2005］

文中の組織名は情報抽出器を用いて抽出し，これらがイベントマップ中では楕円で示されている．一方，文中で2つの固有表現の間に出現する動詞や動作名詞はイベント用語として抽出し，2つの固有表現を結ぶ形でイベントマップ中に配置する．イベントマップ中ではイベント用語は長方形で示されている．

4.2 情報抽出の結果を利用した複数文書要約研究

●文生成器を用いた要約作成［Radev and McKeown, 1998］

Radevらは，複数のニュース記事から情報抽出を行い，その結果を用いて文生成器で要約を作成するシステムを構築している．まず，テロに関する各記事から情報抽出器を用いて，犯人，犠牲者，事件のタイプなど計25の情報を抽出する．この結果，例えば，表1のような情報が抽出される．次に，これらの情報を，7種類のオペレータを用いて統合する．一般に，古い記事では不完全であった情報が続報記事中で明らかになった場合，要約作成には新しい情報を優先させる必要がある．また，同じイベントが異なる情報源でレポートされ，それらが互いに不完全な情報であるならば，組み合わせることで，より完全な情報が得られる場合がある．7種類のオペレータは，このような考えに基づいて抽出された情報を統合するためのものである．

7種類のオペレータを用いて情報を統合し，パラグラフ・プランナに受け渡して組織化した後，FUF / SURGE［Elhadad, 1993；Robin, 1994］という自然言語生成システムに受け渡し，要約を出力する．例えば，表1の例の場合，次のような要約が作成される．

表1 4つの記事から抽出された情報［Radev and McKeown 1998］

記事ID	TST-REU-0001	TST-REU-0002	TST-REU-0003	TST-REU-0004
情報源	Reuters	Reuters	Reuters	Reuters
記事の日付	March 3, 1996 11:30	March 4 1996 07:20	March 4, 1996 14:20	March 4, 1996 14:30
二次情報源		Israel Radio		
事件発生日	March 3, 1996	March 4, 1996	March 4, 1996	March 4, 1996
事件発生場所	Jerusalem	Tel Aviv	Tel Aviv	Tel Aviv
事件タイプ	Bombing	Bombing	Bombing	Bombing
死傷者	死者:18名 負傷者:10名	死者:少なくとも10名 負傷者:30名	死者:少なくとも13名 負傷者:100名以上	死者：少なくとも12名 負傷者:105名
組織名			Hamas	Hamas

Reuters reported that 18 people were killed in a Jerusalem bombing Sunday. The next day, a bomb in Tel Aviv killed at least 10 people and wounded 30 according to Israel radio. Reuters reported that at least 12 people were killed and 105 wounded. Later the same day, Reuters reported that the radical Muslim group Hamas had claimed responsibility for the act.

●表形式の要約作成[Shinyama and Sekine, 2006]

2.2節で述べた評価型の会議MUCでは，抽出する情報(固有表現の種類)および固有表現間の関係があらかじめ決められていた．例えば，「企業活動」に関する課題では，「誰」が「どの企業」に「何のポスト」で雇われたのか，といったように，決められた関係を持つ情報(固有表現)の組を文書集合から抽出する．これに対し，Shinyamaらは，固有表現間の関係そのものも自動的に検出し，同一の関係にある固有表現の組を新聞記事集合から自動的に収集する手法を提案している．以下，表2は，Shinyamaらのシステムを用いて自動的に作成された表の例である．

この表は，1列目が新聞記事の日付を，2列目がハリケーンの名前を，3列目がハリケーンの襲った場所をそれぞれ示しており，表形式の一種の複数文書要約であると考えることができる．Shinyamaらは，ハリケーンの名前と襲った場所だけでなく，2列目と3列目の関係そのものも自動的に抽出している点がこれまでの他の情報抽出研究と異なる．

このような表を自動生成するための基本的なアイディアを以下に述べる．例えば，ハリケーンに関する2つの新聞記事AとBがあるとする．記事Aには，“Katrina”と“New Orleans”が，記事Bには“Longwang”と“Taiwan”が，それぞれ含まれているとする．これらの固有表現は，あらかじめ固有表現抽出器を用いて抽出しておく．次に記事A中で“Katrina”と“New Orleans”に結びつく表現のパターンをそれぞれ収集する．例えば，記事Aに“Katrina headed”や“Katrina threatened”という表現があれば，“headed”や“threatened”という表現を収集する．同様に，記事Aの“New Orleans”，記事Bの“Longwang”と“Taiwan”についても表現を収集する．こうして収集されたパターンが，例えば図4のようであったとする．図において，記事Aと記事Bの固有表現ごとにパターンを比較すると，“Katrina”と“Longwang”に“headed”というパターンが，“New Orleans”と“Taiwan”に“was hit”というパターンが，それぞれ共通していることが分かる．この結果から，“Katrina”と“New Orleans”，“Longwang”と“Taiwan”の間には何らかの共通の関係があると判断する．このような手法を他の記事にも適用していくことで，同じ関係にある固有表現対を収集し，最終的に表2を得る．

表2 自動的に作成された表の例［Shinyama and Sekine, 2006］

Article	Dump	be-hit
2005-09-23	Katrina	New Orleans
2005-10-02	Longwang	Taiwan
2005-11-20	Gamma	Florida

Keyword: storm, evacuate, coast, rain, hurricane

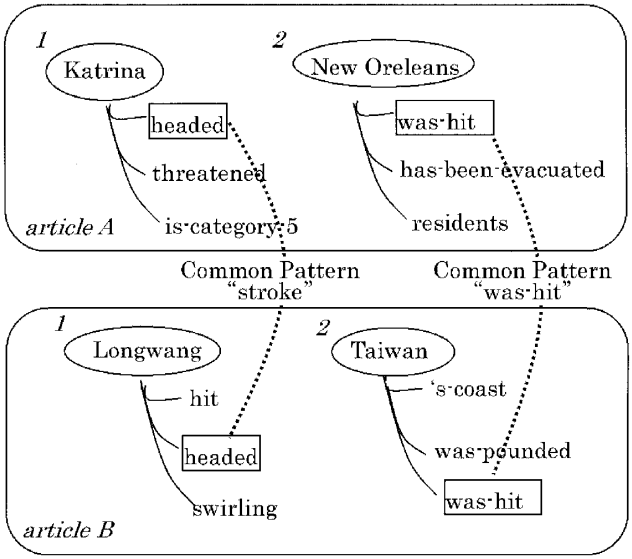

図4 2記事からの類似関係の検出

●グラフ形式の要約作成[難波他，2005]

難波らは，ある期間の日経平均株価や失業率に関する新聞記事集合から，株価や失業率の数値情報を抽出し，それらの数値の推移をグラフとして出力する手法を提案している．グラフを生成するには，数値情報の他に，個々の数値情報に対応する時間情報もあわせて抽出する必要がある．しかし，記事中には，グラフを生成するのに必要な情報とそうでないものが混在しているため，両者を区別する必要がある．この処理を行うため，難波らは文書横断文間関係理論(CST)に着目している．CSTとは，Radevらが提唱している理論[Radev et al., 2000]で，文書中の各文の機能を特定し，文間の依存関係を特定する修辞構造理論(RST)を，文書間関係に拡張したものである．難波らは，CSTの一部を計算機上で実現し，それを用いてグラフ化に必要な数値情報と時間情報の抽出を行っている．

具体的には，以下の2文のように，日経平均株価などの数値の推移に関する2文の対を新聞記事集合から抽出し，さらにそこから数値情報を抽出する．また，情報抽出器を用いて数値情報に対応する時間情報も併せて抽出する．新聞記事中の「昨日」や「三十一日」といった時間表現は，新聞記事の日付から具体的に何年何月何日を指すのか推定する．その情報と先に抽出した情報を対にして蓄積しておき，最終的に新聞記事集合から抽出されたすべての数値情報と時間情報の対をグラフ上にプロットする．

(1) さらに円高の進行や三井グループによるさくら銀行支援を好感し，日経平均株価は前週末終値比192円26銭高の1万4107円89銭と4営業日ぶりに反発，1万4000円の大台を回復した．[毎日新聞 98.09.01]
(2) 日経平均株価は前日終値比218円33銭安と続落し，1万4000円割れ寸前の1万4042円91銭で取引を終えた．[毎日新聞 98.09.05]

図5は難波らの手法を用いて，実際に1992年3月1日から1992年3月31日の新聞記事から抽出された円相場の推移に関するグラフの例である．図のX軸は年月日，Y軸は円を表す．このようにグラフ化することで，文章で提示する方法よりも円相場の推移が直感的に分かる．

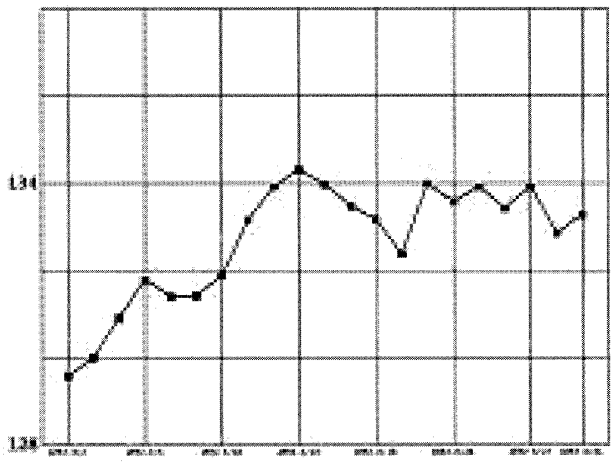

図5 円相場を表すグラフ(システムの出力例)

5．おわりに

本稿では，情報抽出を利用した複数文書要約に焦点を当て，この分野の研究を概観してきた．この分野の研究は，まだ始まったばかりの段階とも言え，今後は様々な方向で研究が進められていくと考えられる．例えば，今回4節で紹介した研究はいずれも新聞記事を対象にしたものであるが，新聞以外のジャンルの文書への適用はそのひとつであろう．例えば，学術論文や特許などの技術文書を対象に情報抽出を利用して要約を作成することは，実用面から考えて，今後ますます重要になってくるものと思われる．学術論文を対象にした情報抽出研究は，専門用語抽出として古くから行われてきており，現在でも多くの研究者がこの問題に取り組んでいる．また，近年，実際に利用可能な専門用語抽出器[4]も出てきている[Nakagawa and Mori, 2002]．この他，Webを対象にした情報抽出および複数Web文書の要約も方向性のひとつとして考えられる．[乾・奥村，2006]および本特集号の乾らの解説記事でも述べられているとおり，近年，自然言語処理の研究分野では，Webから製品，映画，本などに対する評判情報を抽出するという課題に多くの研究者が取り組んでいる．今後は，このような主観情報の抽出およびそれらを要約としてまとめる技術の開発も重要になってくるであろう．なお，この課題に関して，現在2つの評価プロジェクトが始まっている[奥村他，2005，NTCIR，2006]．

参 考 文 献

[Barzilay et al., 2002] R. Barzilay, N. Elhadad and K. McKeown, "Inferring Strategies for Sentence Ordering in Multidocument News Summarization," Journal of Artificial Intelligence Research, Vol. 17, pp.35-55, 2002.

[Bollegala et al., 2006] D. Bollegala, N. Okazaki and M. Ishizuka, "A Bottom-Up Approach to Sentence Ordering for Multi-Document Summarization," Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp.377-384, 2006.

[Elhadad, 1993] M. Elhadad, "Using Argumentation to Control Lexical Choice：A Functional Unification-based Implementation," Ph.D. thesis, Computer Science Department, Columbia University, 1993.

[加藤他，2004] 加藤恒昭，松下光範，平尾努，"動向情報の要約と可視化に関するワークショップの提案，" 情報処理学会 自然言語処理研究会，NL-164，pp.89-94，2004.

[乾・奥村，2006] 乾孝司，奥村学，"テキストを対象とした評価情報の分析に関する研究動向，" 自然言語処理，Vol.13，No.3，pp.201-241，2006.

[Lapata, 2003] M. Lapata, "Probabilistic Text Structuring：Experiments with Sentence Ordering," Proceedings of the 41st Meeting of the Association of Computational Linguistics, pp.545-552, 2003.

[Li et al., 2006] W. Li, M. Wu, Q. Lu, W. Xu and C. Yuan, "Extractive Summarization using Inter-and Intra-

4 http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html

Event Relevance," Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp.369-376, 2006.

[Luhn, 1958] H. Luhn, "The Automatic Creation of Literature Abstracts," IBM Journal of Research and Development, Vol.2, No.2, pp.159-165, 1958.

[Nakagawa and Mori, 2002] H. Nakagawa and T. Mori "A Simple but Powerful Automatic Term Extraction Method," Computerm2：2nd International Workshop on Computational Terminology, COLING-2002 Workshop, pp.29-35, 2002.

[難波他, 2005] 難波英嗣, 国政美伸, 福島志穂, 相沢輝昭, 奥村学, "文書横断文間関係を考慮した動向情報の抽出と可視化," 情報処理学会 自然言語処理研究会, NL-168, pp.67-74, 2005.

[NTCIR, 2006] NTCIR Workshop6, Opinion Analysis Pilot Task, http://research.nii.ac.jp/ntcir/ntcir-ws6/opinion/index-en.html

[奥村・難波, 2005] 奥村学, 難波英嗣, "テキスト自動要約," オーム社, 2005.

[奥村他, 2005] 奥村学, 平尾努, 難波英嗣, "TSC4：意見要約コーパスとそれを用いたワークショップ," 言語処理学会第11回年次大会, 2005.

[Radev and McKeown, 1998] D. Radev and K. McKeown, "Generating Natural Language Summaries from Multiple On-Line Sources," Computational Linguistics, Vol. 24, No. 3, pp.469-500, 1998.

[Radev, 2000] D. Radev, "A Common Theory of Information Fusion from Multiple Text Sources, Step One：Cross-document Structure," Proceedings of the 1st ACL SIGDIAL Workshop on Discourse and Dialogue, pp.74-83, 2000.

[Robin, 1994] J. Robin, "Revision-Based Generation of Natural Language Summaries Providing Historical Background," Ph.D. thesis, Computer Science Department, Columbia University, 1994.

[Schiffman et al., 2001] B. Schiffman, I. Mani and K.J. Concepcion, "Producing Biographical Summaries：Combining Linguistic Knowledge with Corpus Statistics," Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, pp.450-457, 2001.

[関根, 2004] 関根聡, "固有表現から専門用語," 言語処理学会 第10回年次大会 併設ワークショップ「固有表現と専門用語」発表論文集, pp.1-4, 2004.

[関根, 1999] 関根聡, "テキストからの情報抽出 ―文書から特定の情報を抜き出す―," 情報処理, Vol.40, No.4, pp.370-373, 1999.

[Shinyama and Sekine, 2006] Y. Shinyama and S. Sekine, "Preemptive Information Extraction using Unrestricted Relation Discovery," Proceedings of the Human Language Technology-North American Chapter of the Association for Computational Linguistics, pp.304-311, 2006.

[徳永, 2005] 徳永健伸, "情報抽出," 人工知能学事典, 共立出版, pp.406-407, 2005.

(2006年8月24日　受付)

[問い合わせ先]
〒731-3194　広島市安佐南区大塚東3-4-1
広島市立大学情報科学部
TEL：082-830-1584
FAX：082-830-1584
E-mail：nanba@its.hiroshima-cu.ac.jp

著者紹介

難波（なんば）　英嗣（ひでつぐ） [非会員]

2001年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了. 2002年より広島市立大学情報科学部講師. 博士（情報科学）. 自然言語処理, テキストマイニングに関する研究に従事. 言語処理学会, 情報処理学会, 人工知能学会, ACL, ACM各会員.