様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成　25　年　6　月　3　日現在

機関番号：13904
研究種目：基盤研究(C)
研究期間：2010～2012
課題番号：22500090
研究課題名（和文）　音声処理と言語処理の融合に基づく大規模音声ドキュメントの内容検索
研究課題名（英文）　Content Retrieval against large-scale spoken documents based on the integration of speech and language processing

研究代表者
秋葉　友良（AKIBA　TOMOYOSI）
豊橋技術科学大学・大学院工学研究科・准教授
研究者番号：00356346

研究成果の概要（和文）：大規模な音声ドキュメントを対象とした内容検索の研究開発を行った。まず、音声ドキュメントの中に検索語がそのまま表れる位置を検出する音声検索語検出の問題に対して、検出閾値を用いずにもっともらしい候補から順番に高速に検出する手法を開発した。最終的に、ベースラインの連続 DP マッチング法とほぼ同等の検出性能を維持したまま約 70 倍高速な検出を達成した。次に、自然言語で表現した検索要求に適合する音声区間を検索する音声内容検索の問題に対して、音声検索語検出を前処理に用いた誤認識や未知語に頑健な手法を開発した。提案法は、クエリに未知語が含まれている場合に効果があること、大語彙連続音声認識を用いた従来法と混合することで相補的に検索性能を向上させること、が分かった。

研究成果の概要（英文）：We conducted the research and the development of spoken content retrieval targeting large-scale spoken documents. Firstly, for the spoken term detection (STD) task, which aimed to detect the position in a spoken document that a given term appeared at, we developed the method that did not require any detection threshold but, instead, outputted the candidates in increasing order of their plausibility. Finally, we achieved about 70 times faster detection at the almost same detection performance than the baseline continuous DP matching. Next, for the spoken content retrieval (SCR) task, which aimed to find the segment in a spoken document that was relevant to a given query topic represented in natural language, we developed the method robust for recognition errors and out-of-vocabularies (OOVs) that made use of STD as its preprocessing. We found that the proposed method was effective for the query including OOVs and worked complementally with the conventional SCR method, which made use of the large vocabulary continuous speech recognition (LVCSR), and that the combination of them improved the retrieval performance.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	1,500,000	450,000	1,950,000
2011 年度	800,000	240,000	1,040,000
2012 年度	1,100,000	330,000	1,430,000
年度			
年度			
総　計	3,400,000	1,020,000	4,420,000

研究分野：総合領域
科研費の分科・細目：情報学・メディア情報学・データベース
キーワード：情報検索，音声情報処理，自然言語処理

１．研究開始当初の背景

音声・画像・ビデオの記録・編集機器の拡大、およびインターネットをはじめとする情報通信網の発展により、誰でも気軽にコンテンツを作成・公開することが可能となり、マルチメディアコンテンツの情報爆発が進行しつつあった。これらのコンテンツには、ファイル名やタイトル以外にはメタデータが付与されていないことが多く、従来のテキストベースの検索技術だけでは、目的のコンテンツにたどり着くことは困難である。一方、人の話し言葉音声を含むコンテンツの場合には、大語彙連続音声認識技術を利用することで、言語情報を利用した検索が可能である。このような音声言語情報を対象とした検索技術は「音声ドキュメント検索」と呼ばれ、マルチメディアコンテンツの情報爆発時代に必要不可欠な技術である。例えば、Web 検索大手の Google も、動画内音声をテキストで検索する Google Audio Indexing のベータ版を、2008 年 9 月に公開した。

２．研究の目的

音声ドキュメントに対する検索の従来研究は、検索クエリとして与えた用語が出現する位置を特定する「キーワード検索」(または、Spoken Term Detection; STD)を対象としてきた。前述の Google Audio Indexing も、このキーワード検索を目指している。キーワード検索は``Known Item Retrieval"とも呼ばれ、検索者が検索の対象(用語)を既に知っている状況(ナビゲーショナルな質問)を想定したタスクである。しかし、人が検索を行なう実際の場面では、知りたい事項に対して漠然としたイメージしか持っていない場合は多く、人の曖昧な情報要求(インフォメーショナルな質問)から関連情報を見つける技術が必要とされる。このような状況における検索タスクは「内容検索」と呼ばれ、知りたい内容を表現した文やキーワードリストなどの検索クエリから、その内容を含む未知の文書を見つけることを目的とする。正解文書は、必ずしも検索クエリ中の表現(語)が含まれているとは限らない。

十分に記述内容を推敲するテキストの場合と異なり、音声ドキュメントは自発性の高い発話音声から構成される。そのため、検索者が想定するようなキーワードは必ずしも発話中に現れないことが多い。したがって、音声ドキュメントを検索の対象とする場合、内容検索は特に重要な技術となる。

内容検索は、情報検索や自然言語処理分野におけるテキストを対象とした検索では中心的な研究課題であり、米国 NIST 主催の評価型ワークショップの TREC、国立情報学研究所主催の NTCIR などで、様々な評価タスクが設定され、活発に研究が進行中であった。一方、音声を対象とした検索では、1997 年から 2000 年の間に前述の TREC において内容検索(アドホック検索)の評価実験が行なわれたが、それ以降、内容検索の研究はほとんど行なわれてきていなかった。これは、音声を対象とした検索の場合、より単純なキーワード検索でも自明ではない困難な課題であったこと、TREC 以降に大規模な内容検索用のテストコレクションが構築されなかったこと、が理由であった。

これに対し研究開始当初は、音声ドキュメントのキーワード検索に関しては、まだ決定的な手法は無いものの、音声データとのマッチング手法および効率的な索引付け手法が種々提案されつつあった。また、申請者らは、情報処理学会音声言語情報処理研究会の「音声ドキュメント処理ワーキンググループ」の活動として、日本語の講演音声を対象とした大規模な内容検索用テストコレクションを構築してきた。このような背景のもとで、本研究では、キーワード検索の技術を利用した、高精度な音声ドキュメントの内容検索法の開発を目的とした。

３．研究の方法

音声ドキュメントの内容検索を実現するアーキテクチャの第 1 近似は、従来の音声キーワード検索技術とテキスト内容検索技術をカスケードに組合せることである。一方、単純な結合では各々の処理誤差(例えば、音声認識誤りの影響や文書検索の精度の影響)が乗算的に拡大し、性能低下を招く。このような状況では、トップダウンの予測情報が役に立つことが知られている。例えば、大語彙連続音声認識においては、認識結果の予想を言語モデル(トップダウン情報)で表現し認識処理中に利用することで高精度の認識を可能にしている。音声キーワード検索においても、検索クエリとして与えるトップダウン予測情報が鍵であった。しかし、本課題研究では未知の情報を音声中に見つけることを要するため、十分なトップダウン予測情報が利用できないことが問題である。

この問題に対し、音声キーワード検索技術

とテキスト内容検索技術を密に結合することで、予測と認識を交互に実行し、相互に情報を共有し相補的な処理を可能とする新しいアーキテクチャの実現を目指した。そのために、言語処理、音声処理の両方向からアプローチを行ない、統合アーキテクチャのための基盤技術を開発するとともに、最終的に統合システムの開発を行なった。

４．研究成果

４．１．音声検索語検出

音声検索語検出の問題に対し、新しい索引付け手法 Metric Subspace Indexing(MSI) 法を開発した。従来の索引付け手法では、閾値をあらかじめ設定しておき，その閾値内の候補を検出結果として出力するのに対し，MSI 法では、閾値を用いずにもっともらしい順に検出結果を出力できる。この特長により、従来では不可能であった音声検索語検出の利用法が可能となった。たとえば、システム全体の処理時間に対して検索に割ける時間が制限されるような応用場面において,その時間内で見つかる候補だけをもっともらしい順(距離順)に出力するということができる。あるいは、一番もっともらしい候補だけを高速に検出する場合にも有用である。また、最適な閾値は対象音声ドキュメントの質、音声認識の性能、アプリケーションの要求、等によって異なるため、閾値の設定は 難しい問題であるが、MSI 法ではこの問題を避けることができる。

MSI 法について、様々な検出性能の改善手法を検討した。検討した手法の検出性能を下図に示す。

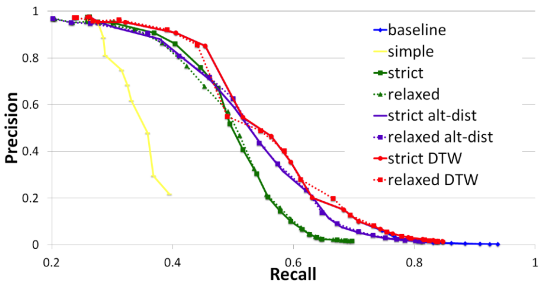

図１：精度-再現率曲線

図の縦軸と横軸は、それぞれ検出性能の精度と再現率を示す。各手法において、検出の閾値(baseline の場合)および検出順の出力数(MSI 法の場合)を変化させたときの精度と再現率をプロットした。ここで、baseline は非索引付け手法である連続 DP マッチング手法であり、この性能を高速に達成することが目標となる。simple は提案手法の最もナイーブな実装、strict は初期の直線検出に基づく MSI 法の厳密な距離順アルゴリズム、relaxed はその距離順を緩和することで検出速度を向上させたアルゴリズムである。strict alt-dist は、音声認識の挿入・脱落誤り対策として直線制約を緩和する代替距離尺度を導入した厳密距離順アルゴリズム、relaxed alt-dist はその距離順を緩和したアルゴリズムである。strict DTW は、直線検出の代わりに Dynamic Time Warping(DTW)による検出を導入した拡張手法の厳密距離順アルゴリズム、relaxed DTW はその距離順を緩和したアルゴリズムである。

図より、ナイーブな実装、距離順直線検出、代替距離尺度の導入、DTW の導入、と MSI 法の改良によって baseline に近づく性能を達成できた。特に、strict DTW は理論上は baseline と同じ性能を高速かつ距離順に達成できることが示されている。

次に、各手法の検出性能と検出速度の関係を下図に示す。

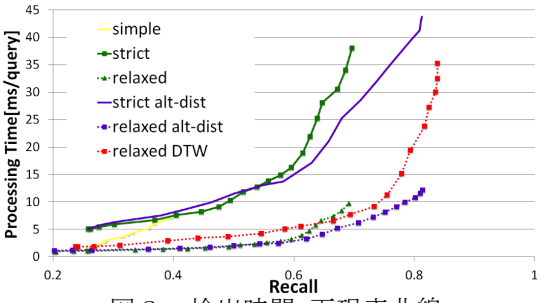

図２：検出時間-再現率曲線

図の縦軸と横軸は、それぞれ検出速度(msec)と再現率を示す。一般に、高い再現率を達成するには、検出結果を多く出力するために時間を要することを示している。

各手法の厳密アルゴリズム(strict)と距離順緩和アルゴリズム(relaxed)を比較すると、relaxed バージョンは検索性能を大きく低下させること無く(図 1)高速な検出を達成している（図２）ことがわかる。relaxed alt-dist は baseline の約 150 倍、relaxed DTW は約 70 倍の検出速度を達成した。

４．２．音声内容検索

音声内容検索の問題に対し、音声ドキュメント中の任意の可変長音声区間を検索する手法、および音声検索語検出を前処理に用いた認識誤りや未知語に頑健な音声内容検索手法の研究開発を行った。以下では、本研究課題の中核を成す後者の研究成果について報告する。

開発した音声内容検索システムの構成図を図 3 に示す。まず、音声データに対し大語彙連続音声認識あるいはサブワードの連続音声認識を適用し、サブワード系列からなる

自動書き起こしテキストを得る。与えられた質問文も単語集合に変換した後、各語をサブワード表現へと変換する。これらの語を検索語とする音声検索語検出を音声データに対して行い、文書毎に各語の出現頻度などの統計情報を得る。この統計情報に基づき，各文書と質問文との関連度を算出し、関連度の高い文書から検索結果として出力する。

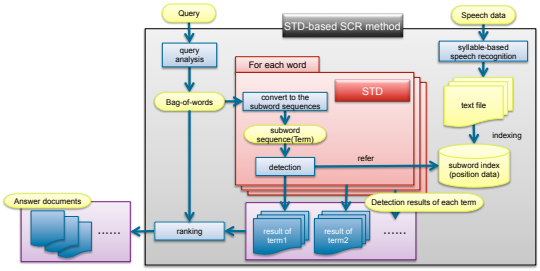

図３：音声検索語検出を前処理に用いた音声内容検索システム

提案法について、検索性能を向上させる改善手法を検討した。第１に、提案法と従来の単語認識結果を用いた典型的な音声内容検索手法との統合を行った。提案法は未知語や認識誤りの語については効果的に検索を行うことが出来るが、正しく認識された語に対しては湧き出し誤りによる悪影響が生じるという問題がある。一方で，従来法は正しく認識された語は有効に利用することができ、提案法とは相補的な関係にある．そこで、従来法と提案法の混合システムを構築した。第２に、音声認識の不確実性を考慮に入れた新しい検索モデルの開発を行なった。音声認識誤りの影響は、本来は文書中に出現しているが認識結果には出現しないという誤り(偽陰性)と、本来は文書中に出現していないが認識結果には出現しているという誤り(偽陽性)の２種類が考えられるが、従来はこれらのうち偽陰性への対処のみが行われて来た。提案手法では音声検索語検出により偽陽性の誤りが増加するため、信頼できる手がかりを重視する単語組合せ素性を用いた検索モデルを新たに開発した。

NTCIR-9 SpokenDoc テストコレクションの dry-run および formal-run の 125 クエリを用いた評価実験結果を表１に示す。評価指標は、Mean Average Precision(MAP)である。音声ドキュメントに対する大語彙連続音声認識の単語正解率は 59.5%、音節正解率は 80.6%、連続音節認識の音節正解率は 75.5%であった。比較した従来手法は、大語彙連続音声認識に対して単語を索引として検索を行う CSCR(単語)、および音節 bi-gram を索引として検索を行う CSCR(音節)である(連続音節認識結果に対して音節 bi-gram を索引とする手法とも比較したが、CSCR(音節)には性能で及ばなかった。)。これらと、連続音節認識結果のみを用いた音声検索語検出を前処理に用いる提案法 STD-SCR、および CSCR(単語)と STD-SCR の混合システム CSCR-STD-SCR と比較を行った。検索モデルは、いずれも TF-IDF 重み付けを行ったベクトル空間法を共通に用いている。

表１：従来検索モデルによる検索性能(MAP)

	IV	OOV	ALL
CSCR(単語)	0.373	0.149	0.307
CSCR(音節)	0.236	0.105	0.194
STD-SCR	0.291	0.237	0.273
CSCR-STD-SCR	0.384	0.277	0.349

表１の列 IV は既知語のみからなる 84 クエリの結果、列 OOV は未知語が含まれる 41 クエリの結果、列 ALL は全クエリ 125 の結果をそれぞれ示している。実験結果より、STD-SCR は全クエリでは単語ベースの従来法 CSCR(単語)の性能に及ばないが、OOV クエリに対して大きな改善を達成している。また、CSCR(音節)に対してはクエリの種類に因らず性能を改善した。さらに、従来法との混合システム CSCR-STD-SCR は、最も高い検索性能を達成した。

表２は、単語組み合わせ素性を用いた検索モデルを用いた場合の実験結果である。表１と比べると、いずれの場合も検索性能が改善しており、提案する検索モデルの有効性が示されている。

表２：単語組み合わせ素性を用いた検索モデルによる検索性能(MAP)

	IV	OOV	ALL
CSCR(単語)	0.393	0.150	0.322
CSCR(音節)	0.238	0.113	0.198
STD-SCR	0.290	0.241	0.274
CSCR-STD-SCR	0.399	0.331	0.376

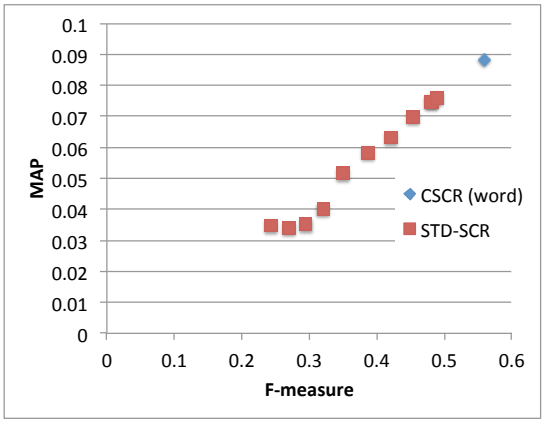

図４：音声検索語検出の性能(F-measure)と音声内容検索の性能(MAP)の関係

図４は、提案法および従来法の様々な手法

における、音声検索語検出の検出性能(F-measure)と音声内容検索の性能(MAP)の関係を調査したものである。この結果から、音声検索語検出の性能を向上させることによって、音声内容検索の性能も向上できることがわかる。すなわち、提案法の前処理における音声検索語検出の検出性能を向上させることによって、誤認識や未知語に頑健な提案法の特徴を保持したまま、提案法単独でも従来法に匹敵する検索性能を達成することが期待できる。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　7 件）

[1] 有賀美明, 秋葉友良, “フレーズベース統計的機械翻訳との統合に基づく言語横断質問応答”, 電子情報通信学会論文誌, 査読有, Vol.J96-D, No.3, pp.713-722, 2013.
http://search.ieice.org/bin/summary.php?id=j96-d_3_713&category=D&year=2013&lang=J&abst=

[2] 瀧上智子, 秋葉友良, “音声検索語検出を前処理に用いた未知語や認識誤りに頑健な音声ドキュメント検索”, 情報処理学会論文誌, 査読有, Vol.54, No.2, pp.506-517, 2013.
http://id.nii.ac.jp/1001/00090255/

[3] 大野哲平, 秋葉友良, “音声継続時間を利用した直線検出に基づく音声検索語検出”, 情報処理学会論文誌, 査読有, Vol.54, No.2, pp.484-494, 2013.
http://id.nii.ac.jp/1001/00090252/

[4] 伊藤慶明, 西崎博光, 中川聖一, 秋葉友良, 河原達也, 胡新輝, 南條浩輝, 松井知子, 山下洋一, 相川清明, “音声中の検索語検出のためのテストコレクションの構築と分析”, 情報処理学会論文誌, 査読有, Vol.54, No.2, pp.471-483, 2013.
http://id.nii.ac.jp/1001/00090251/

[5] 秋葉友良, “音声ドキュメント検索: マルチメディアデータを対象とした音声言語情報検索”, 情報の科学と技術, 査読無, Vol.63, No.1, pp.21-27, 2013.
http://ci.nii.ac.jp/naid/110009562662

[6] Hiromitsu Nishizaki, Tomoyosi Akiba, Kiyoaki Aikawa, Tatsuya Kawahara, Tomoko Matsui, “Evaluation Framework Design of Spoken Term Detection Study at the NTCIR-9 IR for Spoken Documents Task”, Journal of Natural Language Processing, 査読有, Vol.19, No.4, pp.330-350, 2012.
http://dx.doi.org/10.5715/jnlp.19.329

[7] 金子泰輔, 秋葉友良, “部分距離空間上の索引付けに基づく音声中の高速検索語検出法”, 電子情報通信学会論文誌, 査読有, Vol.J95-D, No.3, pp.608-617, 2012.
http://search.ieice.org/bin/summary.php?id=j95-d_3_608&category=D&year=2012&lang=J&abst=

〔学会発表〕（計　40 件）

[1] 大島翔, 秋葉友良, 音声ドキュメント検索のための自発クエリの収録と検索性能評価, 日本音響学会春季研究発表会, 3-P-25a, pp.233-234, 2013 年 3 月 15 日, 東京工科大学.

[2] 加瀬健太, 秋葉友良, 音声ドキュメントのパッセージ検索における適合音声区間の自動決定, 日本音響学会春季研究発表会, 3-P-29a, pp.237-240, 2013 年 3 月 15 日, 東京工科大学.

[3] 瀧上智子, 秋葉友良, STD を前処理に用いた音声ドキュメント検索法のパッセージ検索での評価, 日本音響学会春季研究発表会, 3-9-9, pp.109-112, 2013年3月15日, 東京工科大学.

[4] 大野哲平, 秋葉友良, 部分距離空間上の索引を用いた DTW 距離順の Spoken Term Detection, 日本音響学会春季研究発表会, 3-9-10, pp.113-116, 2013 年 3 月 15 日, 東京工科大学.

[5] Teppei Ohno and Tomoyosi Akiba, Incorporating Syllable Duration into Line-Detection-Based Spoken Term Detection, 2012 IEEE Workshop on Spoken Language Technology, Paper No. TU-AM.8, 2012 年 12 月 4 日, マイアミ・アメリカ.

[6] 秋葉友良, 音声および他言語ドキュメントを対象とした情報アクセス技術, 平成24年度電気関係学会東海支部連合大会, S3-3, 2012 年 9 月 25 日, 豊橋技術科学大学.

[7] 大野哲平, 秋葉友良, 直線検出に基づく音声検索語検出への音節継続時間の導入, 日本音響学会秋季研究発表会, 3-P-29, pp.195-198, 2012 年 9 月 21 日, 信州大学.

[8] Tomoyosi Akiba, Hiromitsu Nishizaki,

Kiyoaki Aikawa, Tatsuya Kawahara, and Tomoko Matsui, Designing an Evaluation Framework for Spoken Term Detection and Spoken Document Retrieval at the NTCIR-9 SpokenDoc Task, In Proceedings of the Eight International Conference on Language Resources and Evaluation, 2012年5月25日, イスタンブール・トルコ.

[9] 酒井哲也, 上保秀夫, 神門典子, 加藤恒昭, 相澤彰子, 秋葉友良, 後藤功雄, 木村文則, 三田村照子, 西崎博光, 嶋秀樹, 吉岡真治, Shlomo Geva, Ling-Xiang Tang, Andrew Trotman, Yue Xu, NTCIR-9 総括と今後の展望, 情報処理学会研究報告, Vol.2012-IFAT-106, No.5, 2012年3月26日, 白百合女子大学.

[10] 金子泰輔, 秋葉友良, 部分距離空間上の索引を用いた音声検索語検出における距離順計算の厳密化, 日本音響学会春季研究発表会, pp.271-274, 2012年3月15日, 神奈川大学.

[11] 瀧上智子, 秋葉友良, STDベース音声ドキュメント検索法の索引付けによる高速化, 日本音響学会春季研究発表会, pp.267-270, 2012年3月15日, 神奈川大学.

[12] Tomoyosi Akiba, Hiromitsu Nishizaki, Kiyoaki Aikawa, Tatsuya Kawahara and Tomoko Matsui, Overview of the IR for Spoken Documents Task in NTCIR-9 Workshop, In Proceedings of the 9th NTCIR Workshop, pp.223-235, 2011年12月8日, 学術総合センター.

[13] 金子泰輔, 秋葉友良, 部分距離空間上の索引とビット並列演算を用いた距離順音声検索語検出手法, 日本音響学会秋季研究発表会, pp.193-196, 2011年9月22日, 島根大学.

[14] 瀧上智子, 秋葉友良, 音声検索語検出結果を用いた音声ドキュメントの内容検索, 日本音響学会秋季研究発表会, pp.187-188, 2011年9月22日, 島根大学.

[15] Tomoyosi Akiba, Koichiro Honda, Effects of Query Expansion for Spoken Document Passage Retrieval, In Proceedings of International Conference on Speech Communication and Technology, pp.2137-2140, 2011年8月29日, フレンツェ, イタリア.

[16] 秋葉友良, 西崎博光, 相川清明, 河原達也, 松井知子, 伊藤慶明, 胡新輝, 中川聖一, 南条浩輝, 山下洋一, NTCIR-9 Spoken Doc: 音声検索語検出と音声ドキュメント検索の評価枠組みの設計, 情報処理学会研究報告, Vol2010-SLP-84 No.18, 2010年12月21日, 国立オリンピック記念青少年総合センター.

[17] Yoshiaki Itoh, Hiromitsu Nishizaki, Xinhui Hu, Hiroaki Nanjo, Tomoyosi Akiba, Tatsuya Kawahara, Seiichi Nakagawa, Tomoko Matsui, Yoichi Yamashita, Kiyoaki Aikawa, Constructing Japanese Test Collections for Spoken Term Detection, In Proceedings of 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010), pp.677-680, 2010年9月28日, 幕張・千葉.

[18] Taisuke Kaneko, Tomoyosi Akiba, Metric Subspace Indexing for Fast Spoken Term Detection, In Proceednge of 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010), pp.689-692, 2010年9月28日, 幕張・千葉.

[19] 秋葉友良,音声ドキュメント検索の現状と課題, 情報処理学会研究報告, Vol.2010-SLP-82 No.10, 2010年7月23日, 仙台・秋保温泉.

[20] Koichiro Honda, Tomoyosi Akiba,Language Modeling Approach for Retrieving Passages in Lecture Audio Data, In Proceedings of International Conference on Language Resources and Evaluation (LREC 2010), 2010年5月21日, マルタ.

〔図書〕(計 1件)

中川聖一編著, 小林聡, 峯松信明, 宇津呂武仁, 秋葉友良, 北岡教英, 山本幹雄, 甲斐充彦, 山本一公, 土屋雅稔 共著, コロナ社,“音声言語処理と自然言語処理”, 2013, pp.120-150, 196-200.

６．研究組織

(1)研究代表者

秋葉 友良 (AKIBA TOMOYOSI)

豊橋技術科学大学・工学研究科・准教授

研究者番号：00356346

(2)研究分担者

中川 聖一 (NAKAGAWA SEIICHI)

豊橋技術科学大学・工学研究科・教授

研究者番号：20115893