

音声認識を用いたマルチメディアコンテンツのインデクシング

大附 克年[†], 別所 克人[†], 水野 理^{†‡}, 松尾 義博[†], 松永 昭一[†], 林 良彦[†]

本稿では、音響処理、音声認識処理、自然言語処理、映像インデクシング処理を統合したマルチメディアコンテンツの自動インデクシングシステムについて報告する。このシステムは、主にニュースコンテンツを対象として、コンテンツの構造や内容を記述するメタデータを自動生成することにより、コンテンツに対する高度なアクセスインタフェースを提供する。音響セグメンテーション、音声認識結果に対するトピックセグメンテーション、映像カット点の情報を統合することによりコンテンツの構造を自動的に抽出する。評価実験により複数の情報の統合がコンテンツの構造抽出の精度向上に貢献することを確認した。

Automatic Multimedia Content Indexing with Continuous Speech Recognition

Katsutoshi Ohtsuki[†], Katsuji Bessho[†], Osamu Mizuno^{†‡},
Yoshihiro Matsuo[†], Shoichi Matsunaga[†], and Yoshihiko Hayashi[†]

This paper describes an automatic multimedia content indexing system, which includes acoustic segmentation, automatic speech recognition, natural language processing, and video indexing features. The system mainly focuses on multimedia news programs. Speech segments, which are extracted from news content, are delivered to the speech recognition module. The speech recognition result sequence is segmented into topics, and those audio indexing results can be integrated with video indexing information to extract news story structure. Experiments show that the integrated results provide better news story structures than any single type of information

1 はじめに

大語彙連続音声認識技術の進展により、読み上げニュース音声など、一定の条件が揃えば高精度な連続音声認識が可能となっている。一方、ネットワークのブロードバンド化が進み、音声・映像コンテンツの流通量が増加する中で、大量のマルチメディアコンテンツに対するアクセシビリティの高度化が求められており、コンテンツに関する情報を記述するメタデータ技術が注目を集めている。しかし、メタデータの作成には膨大な労力が必要とされるため、メタデータ作成の自動化・省力化が望まれている。我々は、大語彙連続音声認識技術と自然言語処理技術を用いることにより音声・映像コンテンツのメタデータを自動生成するインデクシングシステムの検討を進めている[1,2]。

大語彙連続音声認識技術を用いてマルチメディアコンテンツのインデクシングを行い、情報検索へ適用する試みは、いくつかの研究機関において取り組まれている[3-6]。これらは、主にコンテンツの音声を対象としているが、我々のシステムでは、映像インデクシングとの統合を行えるよう柔軟な設計となっている。また、各モジュールの独立性を確保することにより、映像インデクシング以外の各モジュールについても、対象とするコンテンツの特性や用途などによって任意に組み合わせることができる。

本報告では、我々の検討しているインデクシングシステムの全体構成と音響セグメンテーション、音声認識、トピックセグメンテーション、および情報統合の各モジュールについて述べる。また、ニュースコンテンツを対象とした各モジュールの評価実験および評価結果について述べる。

[†] 日本電信電話株式会社 NTT サイバースペース研究所
NTT Cyber Space Laboratories, NTT Corporation

[‡] 現在、東日本電信電話株式会社 通信機器事業部
Currently, Customer Premises Equipment Division, NTT East Corporation

2 システム構成

インデクシングシステムの構成を Figure 1 に示す. 各モジュール間のインタフェースは XML ファイルを介して実現されている. 各モジュールについて以下に述べる.

2.1 音響セグメンテーション

音響セグメンテーション(Acoustic segmentation)部では, 入力コンテンツ(Multimedia content)の音声トラックを分析して得られる特徴量に基づいて, 音声/音楽/雑音の識別を行い, 音声認識部が処理を行うべき音声区間の抽出を行う. このモジュールでは, 信号判別のための特徴量として, 従来手法[7]で用いられているパワー(振幅), 中心周波数, バンド幅, 音源成分の情報に加えて以下のような特徴量を用いている[8].

- ・ 定常性尺度: 時間軸上で隣接する周波数スペクトルの相関値より, 相関値の時間軸上での分散を求め, 信号の定常性を計測する.
- ・ 白色性尺度: 周波数スペクトルを周波数軸上で移動して相関値を求め, 周波数軸上での相関値系列の平均値を求め, 信号の白色性を計測する.
- ・ 形状尺度: 白色性尺度で用いた周波数軸上での相関値系列の一次回帰係数を求め, 信号の周波数スペクトルの大局的な形状を計測する.

音声, 音楽, 雑音(5種類), 無音に対してそれぞれ上記パラメータ7次元の混合正規分布(2混合)を用意して, 約5時間分のニュース番組を用いて学習した. 判別はフレーム単位での尤度比較により行う. 非常に短いセグメントが抽出されるのを抑えるために, フレーム単位の抽出結果を連続する複数のフレームの結果に基づいて平滑化を行っている.

2.2 音声認識

音声認識(Speech recognition)部では, 音響セグメンテーション部で得られた音声区間に対し, 大語彙連続音声認識を行う. 音声認識エンジンは, NTTで開発された VoiceRex[9]を用いている.

音声認識結果は, ポーズで区切られた区間ごとに, 始末端時刻, 音響・言語・信頼度の各種スコアとともに出力される. 信頼度スコアは, 第一位の結果と対立候補とのフレーム当たりのスコア差の平均として求められる.

また音響モデルの駆動方法には以下に挙げる3種類があり, 複数のモデルを用いる場合には, 選択されたモデルの情報が結果に付与される.

- ・ 単一のモデルによる認識
- ・ 複数のモデルによる並列認識: 最もスコアの高い結果を出力する.

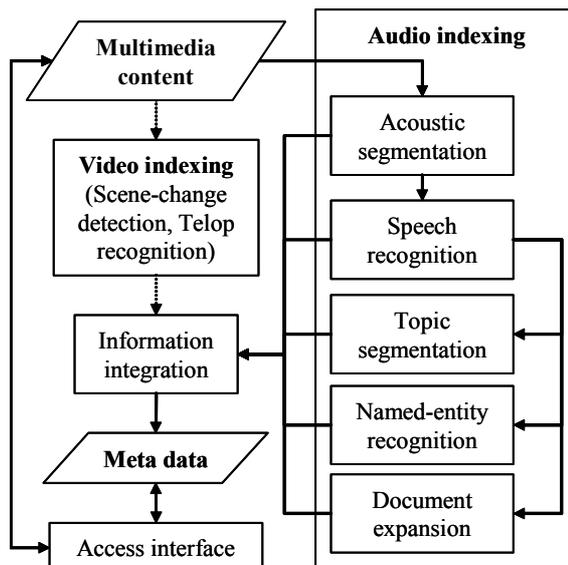


Figure 1: Architecture of the Indexing System.

- ・ 複数モデルから自動選択されたモデルによる認識: 発話の先頭部分を GMM で評価した結果に基づいてモデルを選択する.

また音声認識結果の各単語には, 始末端時刻, スコアに加えて, 読み, 品詞, 終止形(用言の場合)の情報が付与される. これらの情報は, 後段の自然言語処理および検索において用いられる. さらに, 各音声認識区間に対して, 話題抽出[10]により得られた話題語が付与される.

Figure 2 に音声認識部の出力 xml の例を示す.

2.3 トピックセグメンテーション

トピックセグメンテーション(Topic segmentation)部では, 音声認識部で得られた音声認識結果のテキストをトピック単位に分割する. 本システムでは, 単語の意味表現の一つである概念ベクトルを用いたトピックセグメンテーション手法を採用している[11,12]. ニュースなどの特定のトピックに関する区間が複数連続するようなコンテンツでは, このようなトピック境界によりトピックを単位としたアクセスが実現できる.

概念ベクトルとセグメンテーションアルゴリズムについて以下に述べる.

2.3.1 概念ベクトルの生成

概念ベクトルの生成では, まず学習用コーパスを用いて各単語(自立語)間の一文中における共起頻度から単語の共起行列を生成する. 共起行列の各行が各単語に対する共起パターンベクトルとなる. ベクトルの次元数の圧縮とデータスパースネスの解消のために SVD(特異値分解)により行列を変換したのち, 長さ1に正規化したものを概念ベクトルとしてセグメンテーションに用いる.

```

<SR-SEGMENT begin="28902" end="32742" interval="11" segmentId="sr2">
  <NBEST rank="1" speaker="2" a-score="33671" l-score="-7604" c-score="26066">
    <TEXT>アゴヒゲアザラシちゃんも二日ぶりに現れました</TEXT>
    <WORD begin="28916" end="29846" a-score="49240" l-score="-33837" c-score="15402" kana="アゴヒゲアザラシ" yomi="アゴヒゲアザラシ" pos="名詞" can="アゴヒゲアザラシ">アゴヒゲアザラシ</WORD>
    <WORD begin="29916" end="30186" a-score="41488" l-score="-4337" c-score="37151" kana="タマ" yomi="タマ" pos="名詞" can="タマ">タマ</WORD>
    <WORD begin="30186" end="30397" a-score="38546" l-score="-3723" c-score="34823" kana="チャン" yomi="チャン" pos="名詞接尾辞:A 役職敬称:名詞:通用" can="ちゃん">ちゃん</WORD>
    <WORD begin="30397" end="30556" a-score="49486" l-score="-6731" c-score="42754" kana="ガ" yomi="ガ" pos="格助詞:が:通用" can="が">が</WORD>
    <WORD begin="30767" end="31136" a-score="50000" l-score="-8764" c-score="41235" kana="ツカ" yomi="ツカ" pos="名詞:日時:通用" can="二日">二日</WORD>
    <WORD begin="31136" end="31356" a-score="56905" l-score="-9984" c-score="46921" kana="フリ" yomi="フリ" pos="名詞接尾辞:L 日時:名詞" can="ぶり">ぶり</WORD>
    <WORD begin="31356" end="31546" a-score="51710" l-score="-4833" c-score="46876" kana="ニ" yomi="ニ" pos="格助詞:に:通用" can="に">に</WORD>
    <WORD begin="31546" end="31947" a-score="49572" l-score="-12920" c-score="36652" kana="アラワレ" yomi="アラワレ" pos="動詞語幹:A:引用" can="現れる">現れ</WORD>
    <WORD begin="31947" end="32257" a-score="-747" l-score="-9035" c-score="-9782" kana="マシタ" yomi="マシタ" pos="動詞接尾辞:L:終止" can="ました">ました</WORD>
  </SR-SEGMENT>

```

Figure 2: Example of output xml from the Speech recognition module.

2.3.2 セグメンテーションアルゴリズム

音声認識結果の各単語境界において、境界の前後それぞれ一定の単語数に対して単語の概念ベクトルの重心を計算し、境界の前後の重心ベクトルの余弦を求め、境界における結束度とする。セグメンテーションには、境界位置とその前後一定数の境界位置の結束度の平均値を求めることにより平滑化した結束度を用いる。

トピック境界では、結束度が極小となっていると期待される。結束度が極小となる境界位置を i 、 i の左側で結束度が単調減少を開始する位置を l 、右側の右側で結束度が単調増加を終了する位置を r とし、それぞれの結束度を C_l 、 C_i 、 C_r としたとき、 i に対し、結束度の谷の深さ d_i を求める。

$$d_i = (C_l - C_i) + (C_r - C_i) \quad (1)$$

d_i の値が大きいほど、 i がトピック境界らしいと考えられる。

2.4 固有表現抽出

固有表現抽出 (Named-entity extraction) 部では、音声認識結果の単語列から、人名、地名、組織名、日付、時間、金額、割合といった固有表現を、音声認識誤りを考慮して学習した HMM を用いて抽出する[13]。固有表現の情報は検索の際だけでなくトピックの内容を一覧する際にも有効である。

2.5 関連語付与

関連語付与(Document expansion: 文書拡張)部では、音声認識結果を用いて、外部のデータベースから関連文書を検索し、関連語を抽出する。関連語をメタデータに付与することにより、音声認識誤りを補償することができる。また、検索の柔軟性を高めることができる。また、サイドコーパスにおける単語の共起情報を利用することによる音声認識誤りに頑健な重要語抽出[14]や話題表現の抽出[15]について検討している。

2.6 情報統合

情報統合(Information integration)部では、上記の各モジュールが出力した情報を統合し、メタデータ(Meta data)として出力する。コンテンツの構造は、トピックセグメンテーション部の出力だけでなく、音響セグメンテーション部で得られる音楽の開始点や無音区間の情報を用いて抽出し、これをストーリーという単位でメタデータに記述する。また映像インデクシング(Video indexing)[16]を併用する場合には、映像カット点抽出(Scene-change detection)の結果をストーリー境界の推定に用いたり、テロップ認識(Telop recognition)結果をメタデータに含めたりすることも可能となる。

情報統合部におけるストーリー構造の抽出アルゴリズムを以下に示す。

- (1) 時刻 M_i から時刻 $M_i + N$ の区間をフレーム i とする (M : フレームシフト, N : フレーム幅)。
- (2) 各フレームにおけるストーリー境界フレームスコア $S(i)$ を求める。

$$S(i) = \sum_{k \in D} w_k \cdot f(k, i) \quad (2)$$

但し、 D は、各モジュールで抽出されたトピック境界や映像カット点などのイベントであり、 w_k は各イベントに対する重み係数である ($\sum w_k = 1$)。また、 $f(k, i)$ はフレーム i においてイベント k が抽出された場合にそのイベントの有無およびスコアに応じて 0 から 1 の値をとる関数である。

- (3) $S(i)$ がしきい値より大きく、 $i-L$ から $i+L$ までの区間で最大の場合に、 i をストーリー境界フレームとして抽出する。但し L はストーリー長の最小値である。
- (4) 各ストーリー境界フレームにおいて、以下のいずれかに基づいてストーリー境界時刻を抽出する。
 - ・フレーム内でのイベントのスコア
 - ・境界時刻に採用するイベントの優先度 (コンテンツに応じてあらかじめ設定)

以上のアルゴリズムによってストーリー構造を抽出することにより、トピック境界の近傍の映像

カット点や音楽の開始点をストーリー境界とすることが可能となる。

2.7 検索・アクセスインターフェース

生成されたメタデータを用いてコンテンツへアクセスするためのインターフェース (Access interface) を構築した。このインターフェースは、コンテンツのストーリー構造に基づく一覧機能、音声認識結果の表記および読み、固有表現、関連語などでの検索機能、検索結果のコンテンツのストーリー単位、音声認識区間単位での再生機能を提供する。また、XML 形式のメタデータに対する高速な検索を実現するため、XML 検索エンジンを利用している [17,18]。Figure 3 に一覧機能、Figure 4 に検索機能の画面をそれぞれ示す。

3 評価実験

放送ニュースコンテンツを用いて音響セグメンテーション部、音声認識部、トピックセグメンテーション部、情報統合部の評価実験を行った。

3.1 評価データ

評価データとして 5 分間、10 分間、20 分間、30 分間のニュース番組を各 3 本 (3 日分) ずつ合計 12 番組用いた。12 番組中 9 番組には CM が含まれている。評価データ全体の規模は約 1,000 発話 (文)、26,000 単語である。

3.2 音響セグメンテーション

音声 (speech)、音楽 (music)、雑音 (noise)、無音 (silence) の判別性能評価の予備実験の結果を Table 1 に示す。この評価は上記の 12 番組とは異なる 6 番組のデータを用いて行っている。従来のパラメータ (4 parameters) [7] に比べ、2.1 節で述べた 3 つの特徴量を加えた場合 (+3 parameters) にすべての音響信号種別において再現率 (Recall, R.), 適合率 (Precision, P.), F 値 (F-measure, F.) が改善している。また、平滑化 (+smoothing) によりさらに性能が改善している。

予備実験で最も高い性能が得られた条件 (+smoothing) における上記 12 番組のデータに対する評価結果を Table 2 に示す。本システムにおいては、音声認識部で処理する音声区間を高精度で抽出することが必要となるが、音声区間に関しては、F 値 (F-measure) で 95% 以上の判別が達成されている。

Table 2 の評価では、音声と音楽や雑音が重畳している区間を除いているが、そのような区間も含めて音声 (speech) / 非音声 (non-speech) の判別実験を行った結果を Table 3 に示す。重畳区間を含めた場合でも、F 値 95% で音声区間が検出できている。



Figure 3: Access interface screen (browsing news content).



Figure 4: Access interface screen (query input and search result).

Table 1: Acoustic segmentation preliminary evaluation results (R.: Recall [%], P.: Precision [%], F.: F-measure [%])

	4 parameters			+3 parameters			+smoothing		
	R.	P.	F.	R.	P.	F.	R.	P.	F.
speech	62	97	76	79	97	87	94	99	96
music	77	31	44	80	62	70	87	90	88
noise	49	11	28	68	19	30	76	54	63
silence	89	67	76	89	70	78	92	83	87

Table 2: Acoustic segmentation evaluation results

	Recall	Precision	F-measure
speech	94.4%	97.2%	95.7%
music	73.7%	95.9%	83.4%
noise	78.6%	24.0%	36.8%
silence	87.0%	82.1%	84.5%

Table 3: Speech detection evaluation results

	Recall	Precision	F-measure
speech	91.8%	98.4%	95.0%
non-speech	93.6%	73.5%	82.3%

3.3 音声認識

2.2節で述べた3種類の音響モデル駆動方法による音声認識部の評価を行った。言語モデルは、ニュース番組の書き起こしなど約60万文を用いて学習した語彙30,000語のtrigramを用いた。音響モデルは、約300時間の読み上げ音声およびニュース音声をを用いて学習した3状態8混合の状態共有音素triphone HMMを男声、女声、性別非依存について用意した。評価結果をTable 4に示す。男声、女声、性別非依存の3つのモデルを並列に駆動することにより(parallel)、性別非依存のモデルのみを用いる場合(gi)に比べて精度(WER: 単語誤り率)は改善するが、処理時間(RTF)が3倍以上になっている。一方、GMMによるモデル選択を用いる方法(select)では、選択されたモデルのみを駆動するため、処理時間を増加させず精度を改善することができる。

また、GMMによるモデル選択の場合の結果を、評価データの話者(speaker)、雑音(noise)、発話スタイル(style)の属性でそれぞれ分類した結果をTable 5に示す。話者は、アナウンサー(anchor)、レポーター・解説委員(reporter)、その他一般話者(otherwise)の3種類に分類し、雑音および発話スタイルは、雑音のレベル(clean < noisy-low < noisy-high)、発話の自然性(read < spontaneous < free)によってそれぞれ3段階に分類した。分類は検聴者の主観評価による。分類結果をみると、雑音のレベルや発話の自然性が高くなるにつれて、認識精度が低下し処理速度も遅くなっていることがわかる。

Table 4: Speech recognition evaluation result

	WER [%]	RTF
gi	24.9	1.42
parallel	22.8	4.78
select	23.1	1.40

Table 5: SR evaluation result (itemized)

item	#words	WER [%]	RTF
speaker	anchor	16955	15.7
	reporter	5630	28.4
	otherwise	3490	50.6
noise	clean	14286	19.2
	noisy-low	8460	24.9
	noisy-high	3329	35.4
style	read	20411	18.1
	spontaneous	3637	30.4
	free	2027	59.7

3.4 トピックセグメンテーション

評価データの書き起こしテキスト(transcription)および音声認識結果(ASR result)に対するトピック境界抽出実験結果をTable 6に示す。比較

のために単語の出現頻度ベクトルを用いるHearst法[19]による結果(frequency)をあわせて示す。正解の境界は先行ストーリーの終端時刻と後続ストーリーの始端時刻の間の区間であり、抽出時刻が正解区間に含まれる場合(no margin)と3秒までのずれを許容する場合(3s margin)の結果を示す。書き起こしテキストに対する結果をみると、概念ベクトルによる手法(concept)が、Hearst法に比べて高精度にトピック境界を抽出できることが確認できる。音声認識結果に対する評価では認識誤りのために境界抽出精度は低下するが、この場合でも概念ベクトルを用いる方がHearst法よりも高精度で境界が抽出できている。

Table 6: Topic segmentation result
(R.: Recall [%], P.: Precision [%], F.: F-measure [%])

input	method	no margin			3s margin		
		R.	P.	F.	R.	P.	F.
transcription	concept	53	52	52	64	65	64
	frequency	34	35	34	39	40	39
ASR result	concept	45	46	45	50	51	50
	frequency	25	25	25	26	26	26

3.5 情報統合

音響セグメンテーション、音声認識結果に対するトピックセグメンテーション、映像カット点の情報を統合することによるストーリー構造抽出の評価実験を行った。評価結果をTable 7に示す。比較のために各イベントをそれぞれ単独で用いた場合(single)評価結果も示す。情報統合におけるイベントの重み係数は実験によって決めた。

単独のイベントを用いた場合は、トピックセグメンテーション(ts)による境界が音響セグメンテーション(as)、映像カット点(sc)による境界に比べて非常に高い精度を示しており、ニュースコンテンツからの構造抽出における発声内容の重要性が確認できる。また、情報統合を行った場合の結果(intg.)をみると異なるイベントの情報を統合することにより単独の場合に比べて境界抽出精度が向上しており、各イベントがストーリー境界に関する情報を補間しあっていると考えられる。

Table 7: Integration evaluation results
(R.: Recall [%], P.: Precision [%], F.: F-measure [%])

condition		no margin			3s margin		
		R.	P.	F.	R.	P.	F.
single	as	20	12	15	35	22	27
	ts	45	46	45	50	51	50
	sc	27	8	12	53	16	24
intg.	as+ts	46	49	47	50	53	51
	as+sc	16	24	19	18	27	21
	ts+sc	50	47	48	56	53	54
	as+ts+sc	48	48	48	53	53	53

4 まとめ

本稿では、ニュース番組などのマルチメディアコンテンツに対する高度なアクセスを実現するためのメタデータを、大語彙連続音声認識技術を用いて自動生成するシステムについて、その構成と各モジュールの機能について述べ、評価実験結果について報告した。評価実験の結果、音響、音声・言語、映像の情報を統合することがストーリー境界の抽出精度の改善に貢献することが確認された。ストーリー境界の抽出には音声認識結果に対するトピックセグメンテーションの精度が大きく寄与しており、トピックセグメンテーションの精度を改善していくことが重要である。

今後の課題としては、音響セグメンテーション部の音声区間検出の再現率の改善、音声認識結果の信頼度を利用することによるトピックセグメンテーションにおける音声認識誤りの影響の抑制、情報統合のイベントに対する重み係数および境界抽出のしきい値の推定方法の検討、検索・アクセスインタフェースの検索性能の評価などが挙げられる。さらに、対象コンテンツを広げるための検討を進める予定である。

参考文献

- [1] 林良彦, 松尾義博, 大附克年, 池田成宏, 松永昭一, 林実, 水野理, 別所克人, 長谷川隆明, “映像コンテンツのインデクシングのための音声・言語処理,” 情報処理学会全国大会, 2F-2, 2003.
- [2] 大附克年, 松永昭一, 別所克人, 松尾義博, 林良彦, “大語彙連続音声認識を用いた音声・映像コンテンツのインデクシング,” 音講論, 2-4-19, pp.95-96, 2003-3.
- [3] John Makhoul, Francis Kubala, Timothy Leek, Daben Liu, Long Nguyen, Richard Schwartz, and Amit Srivastava, “Speech and Language Technologies for Audio Indexing and Retrieval,” Proceedings of the IEEE, 88(8), pp. 1338-1353, 2000.
- [4] Satya Dharanipragada, Martin Franz, and Salim Roukos, “Audio-Indexing for Broadcast News,” Proc. of TREC7, pp. 115-119, 1998.
- [5] S.E. Johnson, P. Jourlin, K. Spärck Jones, and P.C. Woodland, “Audio Indexing and Retrieval of Complete Broadcast News Shows,” Proc. of RIAO, Vol. 2, pp. 1163-1177, 2000.
- [6] Jean-Manuel Van Thong, Pedro J. Moreno, Beth Logan, Blair Fidler, Katrina Maffey, and Matthew Moores, “SpeechBot: An Experimental Speech-based Search Engine for Multimedia Content of the Web,” IEEE Trans. on Multimedia, Vol. 4, 1, 2002.
- [7] Tong Zhang, Jay Kuo, “Audio Content Analysis for On-line Audiovisual Data Segmentation and Classification,” IEEE Trans., vol.9, No.4, May 2001.
- [8] 水野理, 大附克年, 松永昭一, 林良彦, “ニュースコンテンツにおける音響信号自動判別の検討,” 信学会総合大会, D-14-19, p.186, 2003.
- [9] 野田喜昭, 山口義和, 大附克年, 小川厚徳, 中川聡, 今村明弘, “音声認識エンジン VoiceRex の開発,” 音講論, 2-1-19, pp.91-92, 1999-9.
- [10] Katsutoshi Ohtsuki, Tatsuo Matsuoka, Shoichi Matsunaga, and Sadaoki Furui, “Topic Extraction based on Continuous Speech Recognition in Broadcast News Speech,” IEICE Trans. Vol.E85-D, No.7, pp. 1138-1144, 2002.
- [11] 別所克人, “単語の概念ベクトルを用いたテキストセグメンテーション,” 情報処理学会論文誌, Vol.42, No.11, pp.2650-2662, 2001
- [12] 別所克人, 大附克年, 松永昭一, 林良彦, “概念ベクトルによるトピックセグメンテーションのニュース音声への適用,” FIT2002, F-4, pp.201-202, 2002.
- [13] 長谷川隆明, 林良彦, “隠れマルコフモデルに基づく音声認識結果からの固有表現抽出,” 言語処理学会第9回年次大会, B6-5, 2003.
- [14] 松尾義博, 林良彦, “認識誤りに頑健な重要語抽出,” 言語処理学会第9回年次大会, A3-4, 2003.
- [15] 池田成宏, 松尾義博, 林良彦, “パターンと重要語に基づく関連記事からの話題抽出,” 言語処理学会第9回年次大会, A2-4, 2003.
- [16] Yukinobu Taniguchi, Akihito Akutsu, and Yoshinobu Tonomura, “PanoramaExcerpts: Extracting and Packing Panoramas for Video Browsing,” Proc. of ACM Multimedia, pp.427-436, 1997.
- [17] Yoshihiko Hayashi, Junji Tomita, and Genichiro Kikui, “Searching Text-Rich XML Documents with Relevance Ranking,” ACM SIGIR Workshop on XML and IR, 2000.
- [18] 富田準二, “XML 文書検索システム: LISTA,” NTT R&D, Vol.52, No.2, 2003.
- [19] M.A. Hearst, “TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages,” Computational Linguistics, Vol.23, No.1, pp.33-64, 1997.