「マルチメディア，分散，協調とモバイル
(DICOMO2019) シンポジウム」 令和元年7月

Wordy：ワードクラウドによる要約と検索を支援する e ラーニングシステム

朱偉[1] 張軍[1] 飛田博章[1]

概要：本論文では，授業動画で教師が話している部分を抽出しワードクラウド化することにより，授業動画のキーワードとダイジェスト閲覧を支援する Wordy システムについて述べる．授業動画から教師の発話内容を音声認識によりテキストに変換し，そのテキストから抽出したキーワードをワードクラウドに要約し，動画のキーワード検索を行う．また，ワードクラウドから動画閲覧者が関心を持つキーワードを選択すると，システムはキーワードについて教師が話している動画のポイントを検索して一覧表示する．一覧表示された動画は教師が個別に再生できることに加え，キーワードに関連する動画をまとめてダイジェスト再生する機能も提供されている．提案手法をクラウドサービスによる音声認識を利用した Web アプリケーションとして実現し，比較実験により有効性を検証した．

Wordy: Word Cloud to Summarize and Browse Online Videos for eLearning

WEI ZHU[1] JUN ZHANG[1] HIROAKI TOBITA[1]

1. はじめに

ネットワークの高速化とモバイルデバイスの発達により，オンライン e ラーニングシステムが広く使われるようになっている．インターネットに接続することで，いつでもどこでもビデオ講義を受講することができる点に特徴があり，自然科学，情報技術，およびビジネススキル等様々なビデオコンテンツが提供されている．しかし，既存のシステムでは，ビデオコンテンツから必要な情報を効率的に検索することは困難である [1, 2, 3, 4].

一般的に， e ラーニングシステムで学ぶための操作は 2 つに大別される．まず，学習テーマに沿って視聴したい授業動画を選択する．この時，学習者は e ラーニングのウェブサイトで授業動画の概要説明を読むことや，お試し版の動画をみることで，自分に適した動画を選択する．次に，選択した動画の再生位置を選択する．全体を学習したい場合は最初から動画を再生し，知りたい部分を中心に学習したい場合はスライダーを操作して再生位置を探す．特に，一度見た動画を復習する際には，スライダー操作を繰り返して復習ポイントを特定して動画を再生する．従って，授業動画による e ラーニングシステムでは，動画の選択方法と学習および復習ポイントを特定する手法が重要であると言える．加えて，こうした一連の操作は単一の動画の操作を対象としているため，複数の動画にまたがる検索や閲覧も重要となる．

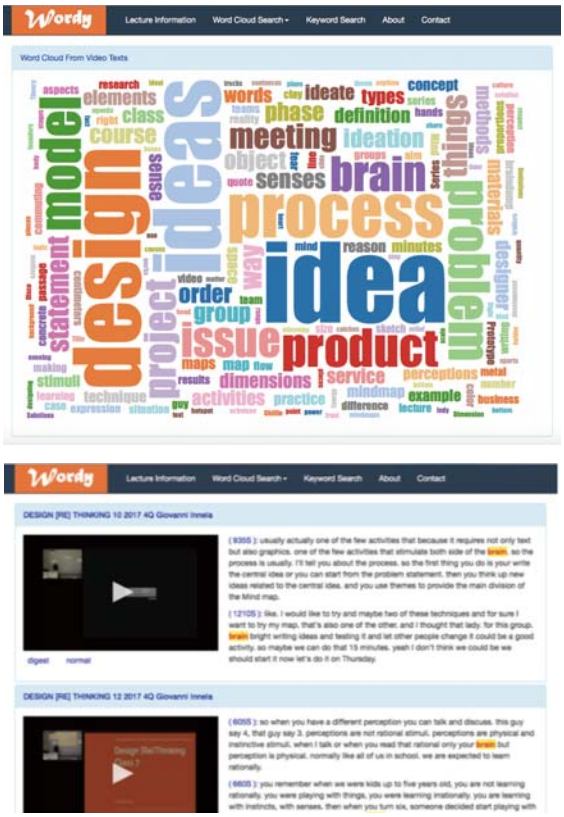


図 1 ワードクラウド（上）と検索結果一覧（下）

しかし，既存の e ラーニングシステムはビデオ内容に対するキーワード検索を提供していないため，検索や閲覧に際し制約が多い．特に，従来の e ラーニングシステムでは，

1 産業技術大学院大学

学習者はビデオを再生せずにビデオコンテンツの内容をチェックすることが難しい．授業動画の検索は，現在のところ，タイトル，概要紹介や，タグ等事前に編集された情報をもとに検索する．しかし，動画の内容にはタイトルまたは，概要や，タグ等に含まれない情報が多いため，内容に関連する情報を検索する際にこうした手法では十分とは言えない．例えば， Google や他の検索エンジンの場合，ユーザはキーワードを入力することで，インターネットに公開されている情報からキーワードに関連する情報を容易に取得できる．また，こうしたキーワード検索を日常的に使っているため多くの人々が操作に慣れているため，授業動画の検索にも有効に作用すると考えられる．

本論文では，授業動画で教師が話している部分を抽出しワードクラウド化することにより，授業動画のキーワードとダイジェスト閲覧を支援する Wordy システムについて述べる．授業動画で教師の話している部分は中心的な役割を果たしている要素の 1 つであり，キーワード検索と連携させることを考えた．そこで，授業動画から発話部分を抽出し，音声認識エンジンを介してテキストに変換した．次に，動画の再生位置とキーワードを関連付けワードクラウドとして表示する．このワードクラウドのキーワードをクリックすることで授業動画の関連する部分を検索が可能となり，検索結果からキーワードをクリックすると該当する部分を再生する機能を実現した．提案手法をクラウドサービスによる音声認識を利用した Web アプリケーションとして実現し，比較実験により有効性を検証した．

2. システム設計

2.1 従来の問題

授業動画を使った既存の e ラーニングシステムの問題点について述べる．特に，授業動画の配信を行うオンライン学習では，授業動画の選択と閲覧で制約があった．

2.1.1 講義選択

オンラインの授業動画による学習では，動画内容に関する情報を十分に伝えることが難しい．内容を入念にチェックして情報を提供することも考えられるが，授業の更新や変更への対応が難しくなる．近年，技術的な進歩の速さは目覚ましく，プログラム言語やクラウド技術は継続的に発展しているため，授業内容は毎年更新や変更されることが一般的である．また，講義全体のランキング情報や講義のサンプリング動画が提供されていればそれらを併用して講義を選択することも可能である．しかし，書籍と異なり学習者が自分に必要かどうかの判断をするためにはこれらの情報では講義選択には十分とは言えない．そのため，講義内容を知るための効果的な可視化方法が必要とされている．

2.1.2 ビデオ学習

学習者が授業動画を選択すると，動画を再生して勉強を進める．対象が動画であるため，単純な UI を操作することで動画を閲覧することができる（例えば，再生，停止や，早送りボタン等）．こうした手法は誰でも扱える点で有効であるが，単純な操作のみサポートされ，学習者は興味持つ部分を効果的に検索することが難しい．また，動画を効果的に視聴するためには，再生速度を制御することも重要である．

2.2 システムデザイン

Wordy は講義選択とビデオ学習の両方において，動画内容の視覚化とその検索により，学習効率を上げることを目的としている．

提案手法は動画内の教師の発話からワードクラウドを生成する．動画内容のキーワードは自動的にワードクラウドにまとめられるため，学習者はワードクラウドの検索ページにアクセスして，ワードクラウドからキーワードを選択することで目的の動画を見つけることができる（図 1（上））．学習者が興味を持つキーワードをクリックすると，対応する動画と発話のテキストがリストに表示され，ヒットしたキーワードがハイパーリンクとして表示される．

また，提案手法は動画コンテンツの視覚化のために 2 種類のキーワードクラウドと，効果的な閲覧を支援するために 3 種類の再生モードを提供している．

3. 実装

Wordy は，Web サービスにオンプレミスサーバーとパブリッククラウドサービスの両方を使用している．従来のシステムと同様に，オンライン e ラーニングをサポートするための Web サービスとして機能する．

3.1 システムアーキテクチャ

Wordy のアーキテクチャは，ウェブサーバとクライアントアプリケーションの 2 つの部分に分割されている（図 2）．システムは，Spring Boot フレームワークを使用してウエブサービスとして動作する．サーバー側では，ffmpeg を使用して動画から音声情報のみを抽出し，およそ 1 分単位のチャンクに分割して wav ファイルに保存する．その後，Google Cloud Speech API を使用して音声認識を行い，認識結果を Elasticsearch にロードする．また，Elasticsearch を組み合わせることで，全文検索が可能となっている．クライアント側では，Spring Boot と相性の良い Thymeleaf テンプレートを使用してウェブページを生成し検索結果を表示する（図 3（右））．動画の再生は JW Player を使用している．

3.2 ワードクラウドの生成

クライアントアプリケーションには，大量の大学講義ビデオの内容を解析して得たキーワードで構成されたワードクラウドが表示される．ワードクラウドは D3 ワードクラウドのプラグインによって生成される．キーワードの出現頻度に応じて文字サイズを決定し，高頻度のキーワードは大きく表示され，低頻度のキーワードは小さく表示される（図 3（左））．

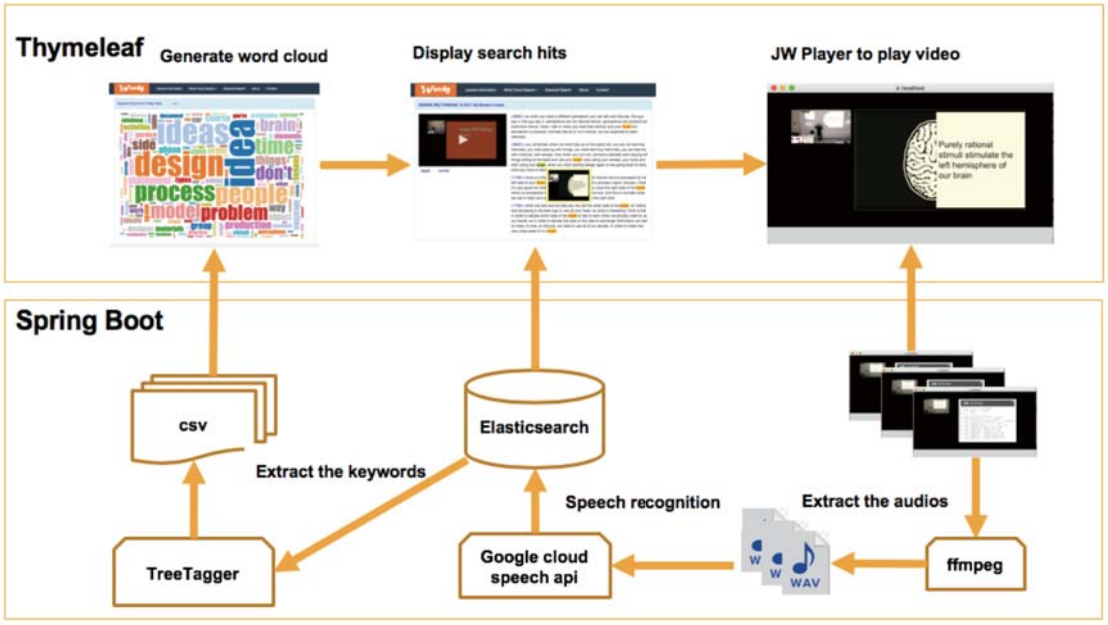


図 2 システムアーキテクチャ：Wordy はウェブサービスとして開発され，ウェブサーバ内のビデオコンテンツを分析し，検索用のユーザインターフェースを提供する．

図 3 ワードクラウド（左）と検索結果一覧（右）

提案手法では 2 種類のワードクラウドの表示が可能となっている．まず，音声認識によって生成されたテキストから作成されるワードクラウドで，TreeTagger という形態素解析ツールを使用して，音声認識によって生成されたすべてのテキストの解析を行い，解析結果から名詞のみを選び出し，出現回数を数え，不要なキーワードを除外した後に csv ファイルに保存してワードクラウドを生成する．また，学習者に検索されたキーワードのランキングからワードクラウドを作成することも可能である．Wordy システムは，学習者がクリックしたキーワードを記憶し，その出現回数をカウントしワードクラウドを生成するため，注目ランキングを反映させるワードクラウドとなる．

3.3 ワードクラウドの編集

作成されたワードクラウドに対して，教師が対話的に編集を行うことが可能であり，不要なキーワードを削除することができる．Wordy システムでは，ワードクラウドは音声認識結果に基づいて自動的に作成されるたま，作成された単語には重要な単語と重要でない単語の両方が含まれている．不要な単語のサイズが大きい場合，またはそれらの数が多い場合，現在のシステムは単語の視覚化を効果的に提供することができない．不要な単語を削除し，必要な単語を残すための編集機能がある．不要な単語を減らすことで，単語スペースを効果的に利用することができる．

ワードクラウドを編集する例を図 4 に示す．この例では，ワードクラウドには，講義のトピックとは関係のない“people”，“don’t”，“time”，“things”などの不要な単語が含まれている．また，“idea”と“ideas”は，同じ単語の単数形と複数形でありながら重複して出現している（図 4

(1)). 特に，“idea”と“ideas”の両方が頻繁に登場するため，両方ともサイズが大きくなってしまう．従って，教師はそのうちの 1 方を選択し削除する（図 4 (2)）．単語を削除すると，単語領域に空白が表示されるので（図 4 (3)），レイアウトを再計算して新しいワードクラウドを作成する（図 4 (4)）．

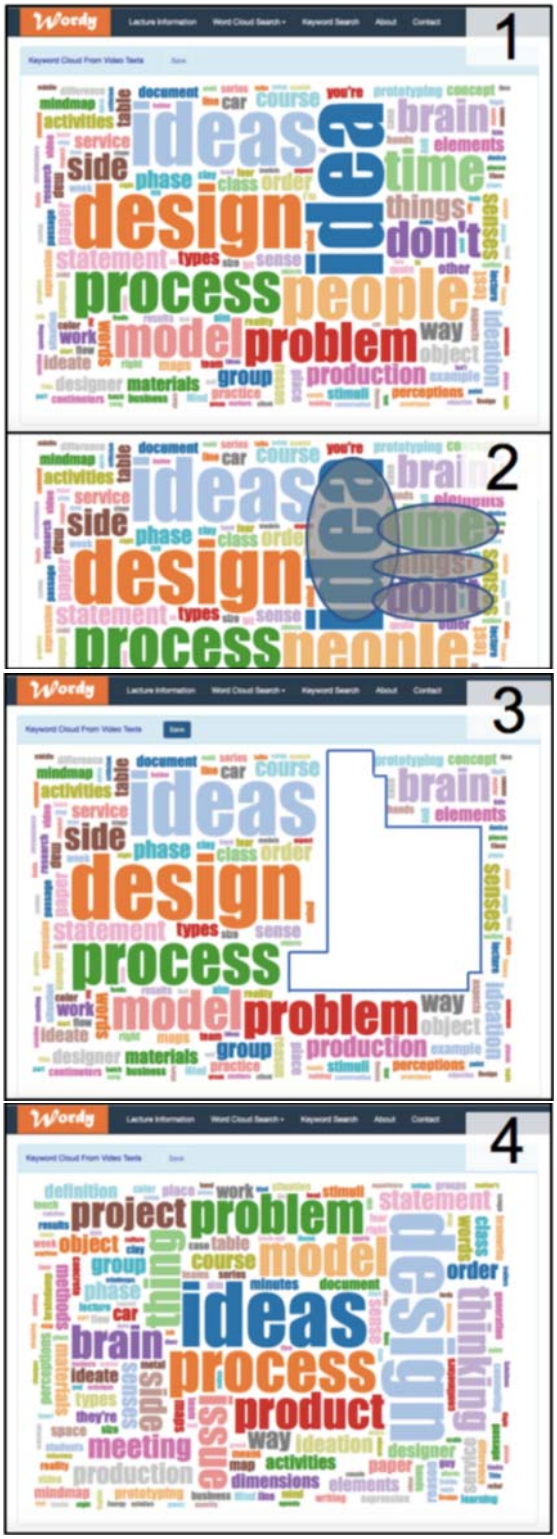
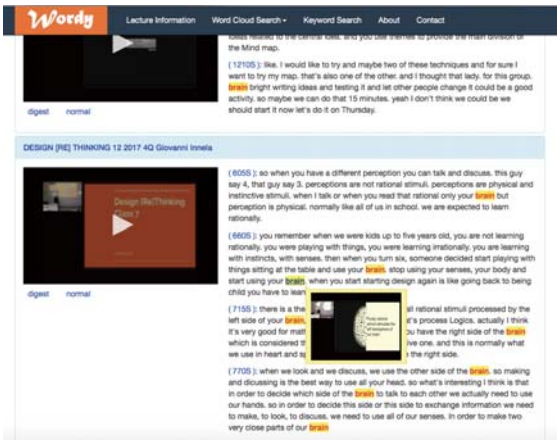


図 4 ワードクラウドの編集



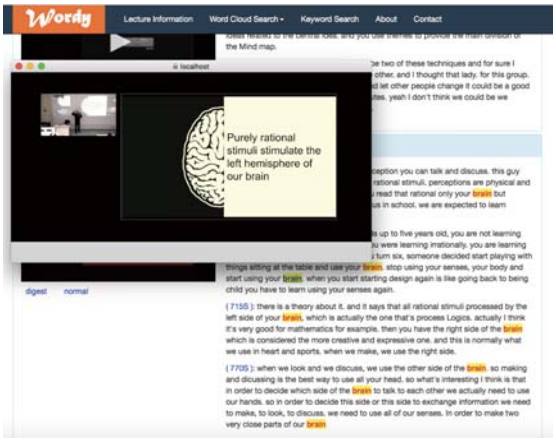


図 5 キーワードのクリックによるプレビュー

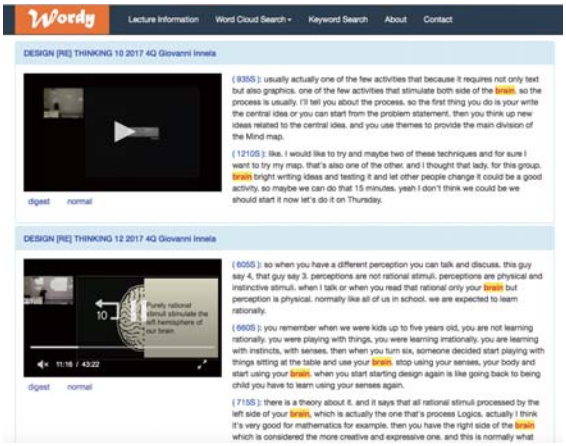


図 6 ダイジェストプレビュー

3.4 動画の操作

動画の内容がワードクラウドにより可視化されるので，学習者はキーワードを選択することで，関心を持つ動画位置を検索し閲覧することができる．検索結果は Web ページとして表示され，動画を再生する領域と，キーワードを含むテキストが表示される．右側には，音声認識によって生成されたビデオのテキストがリストとして表示され，検索でヒットしたキーワードがハイライト表示されていること

に加え，対応する動画の位置にハイパーリンクが貼られている．

3.5 動画のプレビュー

効果的な動画のプレビューはビデオコンテンツを素早く見るために重要である． 提案手法では，複数の授業動画を対象にしてワードクラウドを作成するため，検索結果には複数のビデオが含まれている．動画を効果的に再生し学習に役立てるために，Wordy は 3 種類のプレビューモードを提供している．通常のプレビューモード（図 1（下））に加えて，システムは 2 つのダイジェストプレビューモードを提供している．

キーワードクリックモードでは，検索結果でハイライトされたキーワードのリンクをクリックして再生を開始する．マウスをハイパーリンクの上に置くと，対応するシーンのサムネイルがツールチップとしてポップアップ表示される（図 5（上））．学習者はサムネイルを確認して，ビデオを再生するかどうかを決める．ハイパーリンクをクリックすると，テキストにリンクされているターゲットシーンから再生が開始され，ビデオの最後まで再生される（図 5（下））．

また，ダイジェストモード（図 6）では，ユーザはビデオファイル毎にキーワードを含む部分を連続で見ることができる．Wordy は，JavaScript によるコントロールで関連部分を含むダイジェスト再生を可能にした． これにより，ユーザはキーワードに関する知識をより詳細的に知ることができる．例えば，学習者が“Network”に興味がある場合，大量な授業ビデオの中，教師が“Network”を話しているシーンのみダイジェスト映画を見ることができる．

4. ユーザテスト

Wordy システムの有効性を検証するため，動画検索に関する比較実験を行った．

4.1 実験方法

実際に動画授業の視聴により学習している大学院生 5 名を被験者とし，机の上に配置したノート PC を介して実験を行った．対象とした授業動画は実際に大学院の授業で使われている 3 本の授業動画を使用した．

また，同一の授業動画に対して以下の 3 つの検索手法による比較実験を行った．

- P1 ビデオのみ提供
- P2 ビデオとそのビデオの音声テキストを提供
- P3 提案手法

P3 の提案手法では，ワードクラウドを作成し，被験者に特定のキーワードに関連する動画の部分を特定する検索を行ってもらった．

4.1.1 実験 1

以下の 3 パターンでそれぞれ実験を実施し，ターゲットシーンを見つけるまでにかかった時間を計測した．

- Q1: 1 つのキーワードでターゲットシーンを探す
- Q2: 3 つのキーワードでターゲットシーンを探す
- Q3: 1 つのセンテンスでターゲットシーンを探す

4.1.2 実験 2

実験 1 の後でさらに実験が終わった後にアンケート調査を実施した．アンケートは以下の 3 項目である．

- Q4 該当のシーンを見つけるまで何回操作をしたか
- Q5 ビデオ検索に役立ったか
- Q6 実際に授業で使いたいか

アンケート点数換算方法として，Q4 は回数を以下の 5 段階の中から選択してもらった．

1: 5 回以上, 2: 4 回, 3: 3 回, 4: 2 回, 5: 1 回

Q5 と Q6 に関しても以下の 5 段階の中から選択してもらった．

1: 全くそう思わない, 2: あまりそう思わない, 3: どちらでもない, 4: そう思う, 5: 非常にそう思う

4.2 実験結果

ターゲットシーンを見つけるまでにかかった時間を計測した結果を表 1 に示す．エラーバー（誤差範囲）付き棒グラフ（図 7）から分析すると，Wordy を利用した場合，ターゲットシーンを見つけるまでの時間が普通のビデオのみ提供する場合の約 1/4 であることがわかった．

実験が終わった後に実施したアンケート調査の結果を表 2 に示す．Wordy を利用した場合，他の 2 パターンより高い点数を得た（図 8）．特に，被験者全員が Wordy システムはビデオ検索に非常に役立つという回答を得た（表 2）．

5. 議論

提案手法に関して，プロトタイプ実装や，ユーザテストでの知見を踏まえて議論する．

5.1 Wordy システム

ユーザテストの結果，単一のキーワードと複数のキーワードの場合において，Wordy システムはターゲットシーンを効果的に見つけられることがわかった．まず，検索時間の比較から，Wordy システムはオンライン学習時に勉強したいポイント及び復習したいポイントを他の手法に比べ速やかに見つけられることがわかった．また，実験後のアンケートから，受験者の全員が Wordy システムを実際に授業で使いたいといった評価を得た．こうした結果から，Wordy システムを使用することで，既存の授業動画を使った e ラーニングシステムよりも，学習効率が上げることが可能であると考える．

Wordy では，ビデオのスピーチテキストに基づくワードクラウドと学習者の選択ランキングに基づくワードクラウドの 2 種類のワードクラウドを実装した．学習者の目的や好みに応じてそれぞれのワードクラウドを選択して講義ビデオを受講することが可能である．例えば，キーワードを検索して目的のビデオを見つけるには，1 つ目のワードクラウドが効果的である．他人の利用ランキングを参考にし

てビデオを選択して学習するには，2 つ目のワードクラウドが効果的であると考える．ワードクラウドに関する有効性は今後調べていきたい．

		被験者1	被験者2	被験者3	被験者4	被験者5
P1	Q1	233	91	229	132	400
	Q2	312	136	310	168	176
	Q3	237	51	335	392	396
P2	Q1	103	143	169	142	126
	Q2	87	109	47	122	160
	Q3	142	66	60	137	70
P3	Q1	80	90	70	89	86
	Q2	54	37	42	77	45
	Q3	60	63	67	58	66

表 1 ターゲットシーン検出時間（単位：秒）

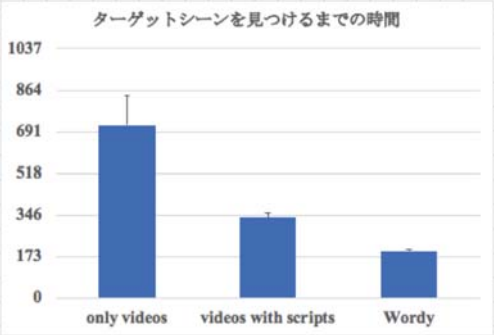


図 7 ターゲットシーン検出時間の比較

		被験者1	被験者2	被験者3	被験者4	被験者5
P1	Q4	1	1	1	1	1
	Q5	1	2	5	1	1
	Q6	2	1	1	1	2
P2	Q4	2	3	5	3	4
	Q5	5	4	5	4	4
	Q6	5	4	3	4	4
P3	Q4	3	5	4	4	5
	Q5	5	5	5	5	5
	Q6	5	4	4	5	5

表 2 アンケート結果

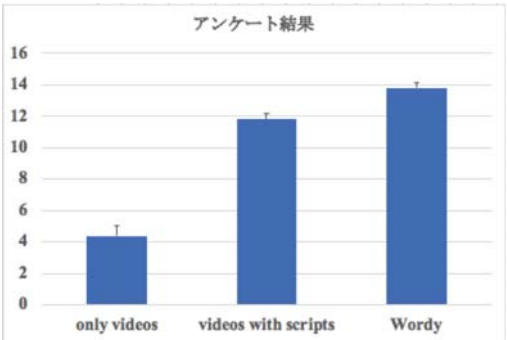


図 8 アンケート結果の比較

5.2 Wordy の課題と今後

まず，ワードクラウドを生成するために音声認識技術を利用しているため，音声認識の精度が視覚化に大きな影響を与える．今回のプロトタイプ実装では，現在認識精度が高い Google Cloud Speech API を使用して音声認識を行なっている．Google Cloud Speech API は普通の新聞など発音が良い音声の場合には，認識精度が 90%以上に達することがある．しかし，講義ビデオの場合，教師の発音によって認識精度が大きく左右される．現在のプロトタイプでは，音声認識結果に含まれるエラーを手動で訂正しているので，教師の発音が悪い場合は，教師による手動訂正の負担が重いと想定される．

また，Wordy はビデオコンテンツの音声情報を使ってワードクラウドを作成するため，ターゲットビデオに大量な音声が含まれている場合は，この方法が効果的であるが，次の 2 つの場合では，音声テキストだけでビデオを検索することが困難であった．最初のケースは，スピーチの少ないビデオである．この場合，ワードクラウドは小さくてビデオ検索にあまり効果がない．もう 1 つのケースは，バックグラウンドノイズが多い場合，または複数の人が同時に話す場合である．この場合，システムはスピーチ部分を正しく認識することが困難である．これらの問題を回避するため，講演スライドのような視覚的な情報を提供することが効果的である．スライド情報を含むビデオコンテンツの場合は，光学式文字認識（OCR）などの画像処理テクノロジを使用してビデオからキーワードを抽出し，ワードクラウドを作成することも検討する必要がある．

6. 結論

本稿では，授業動画に基づくワードクラウドを作成し，ユーザがインタラクティブにコンテンツをまとめて検索できる Wordy システムについて述べた．また，大学のオンライン講義ビデオを使用して，システムの有効性を比較実験により示した．

音声認識の精度を高め，使いやすくするために，システムの改良を続け，実際の授業動画を使った e ラーニングシステムとして運用することを計画している．

参考文献

[1] Yao, T., Mei, T., Ngo, C. W., and Li, S. P. Annotation for free: video tagging by mining user search behavior, In Proceedings of MM'13, pp. 977-986, 2013.

[2] Morris, M. J. and Kender, J. R. VastMM-Tag: a semantic tagging browser for unstructured videos, In Proceedings of MM'11, pp. 957-960, 2011.

[3] Sebastine, S. C., Thuraisingham, B. and Prabhakaran, B.. Semantic web for content-based video retrieval. In Proceedings of ICSC'09, pp. 103-108, 2009.

[4] Law-To, J., Grefenstette, G. and Gauvain, J. L. News: robust automatic segmentation of video into browsable content, In Proceedings of MM'09, pp. 1119-1120, 2009.