Vol.2010-CLE-3 No.4
2010/12/9

授業用教材スライド内の文字の出現頻度と属性に基づく重要語自動抽出法とその評価

下園幸一[†1]　菅沼明[†2]

近年の大学等の授業では，PowerPoint 等のスライドを用いて行われることが多くなってきている．このため学生が授業の内容を理解するためには,「スライドの質」が重要となっている．我々は，スライドの質を向上させるために「学生がどのようにスライドを理解するか」という点に着目し，これを教員に提示するシステムの構築を目指している．このシステムを利用することにより，教員は自分の意図する内容との相違点を認識でき，スライドの修正支援に役立てることができと考える．本稿では，受講生からみたスライド内の「重要語」を受講生からの情報なしに抽出する手法の構築を目指す．まず，スライドより形態素解析を用いて語を抽出し，出現頻度に基にした統計手法を利用して語に順位付けを行い，重要語の候補となる語の推定を行った．さらに，語のスライド中の属性を考慮し，推定精度の向上を図った．

Development and Evaluation of an Extraction Method of Important Words Based on Contextual Frequency and Attributes from the Teaching Materials Slides

Koichi Shimozono[†1] and Akira Suganuma[†2]

Recently, many approaches of effective teaching at universities and colleges have been performed. One of them is presentation of the teaching material slides such as PowerPoint by using a computer. Therefore, “the quality of a slide” is important in order for a student to understand the contents of the class. We aim to develop a system which presents “how a student understands a slide” to the teacher for improvement of its quality. By using our system, the teacher can recognize difference between his thought and understanding of his students about his slide contents and then he can improve his slide. This paper describes an extraction method of words which students consider to be important before his lecture. By using a morphological analysis, words are extracted from text in slides. We rank the extracted words by a statistical procedure based on contextual frequency and regard the ranked words as candidates for important words. Furthermore, the accuracy of the candidates is enhanced in consideration of the attribute of the word in the slides.

1. はじめに

近年，日本の大学では，授業を効果的に行うためにさまざまな取り組みがなされている．その 1 つに ICT の利活用がある．その中でも WebCT や moodle 等の LMS(Learning Management System) に注目が集まっている．しかしながら，システムや設備に多大なコストがかかる事や，教員へ教材作成の負担がかかるため，学内の一部組織でのみ取り組んでいる場合や，教員が個人的に取り組んでいる場合も多い．また，導入コストと比較して，その効果を疑問視する声も存在する．吉田らの調査[1)] でも，2006 年現在で，LMS を利用している教員は 20%程度である．現状では，LMS を導入している場合であっても，授業で利用する教材スライド（PowerPoint ファイル）を授業開始以前に LMS 上で公開するのみにとどまっている場合が多い．

しかしながら，同調査によると，PowerPoint 等プレゼンテーションソフトを利用して授業を行う教員は 72%にのぼっており，実際の授業では多用されていることがわかる．この理由は，第一に “授業準備における教材の蓄積性，編集の容易性，再利用性等”，第二に “授業場面において映像資料を容易に用いることができること，授業のスピードアップなど” が指摘されている．

このような状況の中で， 我々は “授業用教材スライドの質” が重要となっていると考える．教員が授業用教材スライドを作成する際，まず考えることは「学生が授業スライドをどのように理解するか」である．実際に授業を行ってみて，学生の反応を見ることにより，授業スライドに変更を加える場合もある．つまり，“授業用教材スライドの質” とは「どれだけ学生がそのスライドを理解できるか」にかかっている．そのため，授業スライドの質を高めるには，まず，学生がどのように授業用教材スライドの内容を理解するかを知ることが重要である．

先行研究では，学生（受講者）からのフィードバックを利用して，スライドの質の向上および学生の学習支援に役立てるものがある[2)]．しかしながら，先行研究の手法では，講義を

†1 鹿児島大学 Kagoshima University
†2 有明高専 Ariake National College of Technology

Vol.2010-CLE-3 No.4
2010/12/9

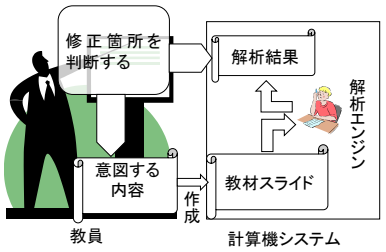

図 1 システムの概要

実際に行わなければフィードバックが得られず，スライドの修正を事前に行うことができない．我々は受講者からのフィードバックを利用せずに「学生がどのように授業スライドの内容を理解するか」を計算機上で推測し，それを教員に提示することによって，教員のスライド修正支援を行うことができるのではないかと考え，スライド修正支援システムの構築を目指す．システムの概要を**図 1** に示す．支援手順としては以下のようになる．

(1) 教員はスライドを作成する
(2) 解析エンジンが，スライドの内容を解析し，学生がどのようにスライドを理解するかを教員に提示する
(3) 提示された内容が教員の意図する内容と異なった場合は，スライドの修正を行う

この（2）（3）を繰り返すことによってスライドの質を高める．システムは解析エンジンに従って解析結果を提示するのみであり，その結果を見て実際にスライドの修正を行うのは教員である．このようにすることにより，スライド修正のための複雑な処理をシステム側で行わなくてもよいと考える．また，(2)（3）を繰り返しながらスライドの改善を行うため，計算機上の解析は，なるべく短時間で終わる方がよい．

本稿では，この解析エンジンの機能の一つとして，講義毎の重要語の抽出法の構築を行う．重要語は，その回の講義を理解する上で重要な要素の一つであり，スライドを修正することにより，受講した学生が考える重要語と教員が考える重要語をなるべく一致させることができれば，スライドの質は向上したと言える．先行研究では，講義情報（スライドのテキスト情報および実際の講義中に収集した発話情報）と学生の Web 検索情報を基に各スライドから重要語句を抽出し，その語句に関連するスライドに重要度を自動的に付与するものがある[3]．しかしながら前述のように実際に講義を行わなければ，重要語句の抽出はできず，また，学生がどのような語を検索したかを得るようなシステムが必要となる．我々は，授業で使用されるスライドから得られる情報のみによる重要語の抽出を目指す．

表 1 学生アンケート結果

順位	指摘語	得点	指摘数	順位	指摘語	得点	指摘数
1	セキュリティインシデント	150	19	11	FireWall	67	17
2	セキュリティ	145	17	12	公開鍵	67	9
3	機密性	119	16	13	アクセス管理	63	10
4	アルゴリズム	107	21	14	アクセス制御	58	13
5	公開鍵暗号	102	21	15	不正アクセス	54	9
6	共通鍵暗号	94	20	16	シーザー暗号	51	9
7	ユーザ認証	89	15	17	秘密鍵	50	7
8	可用性	83	14	18	ファイアーウォール	49	9
9	完全性	79	12	19	パスワード	48	10
10	リモートユーザ認証	73	13	20	コード	40	9

2. アンケート結果による重要語

まず，受講生がどのような語を重要語として認識するかを調査するためにアンケートを行った．アンケートは，実際に筆者が行っている講義形式の授業「情報ネットワーク論」の第 11 回授業 (以後 C.1 No.11 と記載) 終了後に「今回の授業を受けて重要だと思った語を重要な順に 10 語記載しなさい」という形式で行った．なお，この授業は，教科書がなく，PowerPoint のみで授業を行った．PowerPoint ファイルは授業開始以前に Web で公開していた．受講生は 40 名であり，1 名が 8 語までしか記載していなかったため，得られた語数は 398 語である．

得られた語 (以後，指摘語) に対して，指摘数と重要度を考慮した順位付けを行う．各受講者が指摘した語に重要度に応じて 10 点から 1 点までの点数を付け，指摘語毎にこの点数を加算して順位付けを行った．なお，指摘語には「ファイアーウォール」,「ファイアウォール」といった表記の揺れがあったため，表記の揺れに関しては正規化を行った．また，一人のみが指摘した語に関しては，その重要語としての精度が疑われるため指摘語から外した．この結果，全指摘語の違い語数は 43 語である．このうち，上位 20 語を**表 1** に示す．

3. 抽出法の構築

本稿では，スライドにある情報のみから学生が重要語であると思う語の抽出を行いたいため，シソーラス等外部知識は用いない．そのため，まず，語の出現頻度に着目した．スライ

表 2 2×2 分割表

	b の出現数	b 以外の出現数
a の出現数	O_{11}	O_{12}
a 以外の出現数	O_{21}	O_{22}

ド中に繰り返し出現する語を学生が重要であるとみなす可能性が高いためである．しかし，単に頻出する語を抽出しただけでは次のような誤りを犯すと考える．

誤り 1 全体を通して出現頻度は少ないが，ある特定の回で多く出現している語を抽出できない

誤り 2 各回に頻繁に出現する一般的な語を候補として抽出してしまう

そのため，統計的手法を用いて重要語の特定を行う．本稿では統計的手法として，G スコア，TF-IDF の 2 つを試みる．

3.1 G スコア

大規模コーパスからある単語の共起関係を求める統計的手法として，尤度比検定の 1 つである G 検定（G スコア：以下 G 値）[4] が用いられている．G 値を求める一般的な式は

$$G = 2\sum_{i=1}^{n} O_i \ln(\frac{O_i}{E_i}) \tag{1}$$

で表すことができる．O_i が観測値，E_i が期待値である．n 語のコーパスから**表 2** のような 2×2 分割表を用いて単語 a と単語 b の関連性の性検定を行う場合は以下の計算になる．

$$\begin{aligned}
A &= O_{11}\ln O_{11} + O_{12}\ln O_{12} \\
&\quad + O_{21}\ln O_{21} + O_{22}\ln O_{22} \\
B &= (O_{11}+O_{12})\ln(O_{11}+O_{12}) \\
&\quad + (O_{21}+O_{22})\ln(O_{21}+O_{22}) \\
&\quad + (O_{11}+O_{21})\ln(O_{11}+O_{21}) \\
&\quad + (O_{12}+O_{22})\ln(O_{12}+O_{22}) \\
C &= (O_{11}+O_{12}+O_{21}+O_{22}) \\
&\quad \times \ln(O_{11}+O_{12}+O_{21}+O_{22}) \\
G &= 2(A-B+C)
\end{aligned}$$

O_{11} は単語 a と単語 b が共起している回数，$O_{11}+O_{12}$ はコーパス内の単語 a の出現回数，$O_{11}+O_{21}$ はコーパス内の単語 b の出現回数，4 つのセルの合計が n（コーパス内の単語総数）である．G 値が大きければ，単語 a と単語 b の共起関係は強いといえる．

この G 値の計算手法は，語と語の共起関係を数値化するものである．しかしながら，今回は，各講義で使用するスライドとそのスライドに含まれる語の共起関係から G 値を求めることで，“ある語 a は，ある講義とどの程度関係が深いか”を数値で表すことができると考える．

以下のような極端な例を想定して説明する．

- ある科目は 10 回の講義からなり，それぞれの講義に使用するスライドは 500 語からなる
- 語 c は，各講義でまんべんなく 20 回使用されている
- 語 d は，ある特定の講義 e で 10 回使用されており，他の講義では毎回 1 回のみ使用されている

この場合，語 c の G 値は各講義で全て 0 となり，各講義において頻出語ではあるが，抽出の候補に入れるべきではないないと判断できる．従って，“誤り 2” を犯さない．語 d の G 値は，講義 e において 21.8 となり，他の講義では 0.56 となる．講義 e において大きな値を示しているため，“誤り 1” を犯さずに，講義 e における重要語の候補であると判断できる．

3.2 TF-IDF

TF-IDF 法は，TF(term frequency) という指標と IDF(inverse document frequency) という指標の 2 つを用いて，大量の Web ページからの情報検索や文章中の特徴的な語を抽出するために用いられている．TF-IDF 法による語の重みを求める一般的な式は，総文書数を N とし，文書 i における語 w_j の重みを w_j^i とすると，

tf_j^i ：文書 i における w_jの出現頻度 (出現個数)

$idf_j = \log \frac{N}{df_j}$：$N$ は総文書数.df_jは語 j が出現する文書数

$w_j^i = tf_j^i \cdot idf_j$

となる．総文書数 N は，Web ページの解析では総 Web ページ数，大量文章解析の場合は，文書数がよく用いられている．しかしながら，授業で用いられる PowerPoint スライドの場合，文書数は 1 つの講義で 10 数個程度であるので，以下の 2 つの場合で試算してみる．

TFIDF1 N として総スライドファイル数，df_j として w_j が 1 回でも出現するスライドファイル数

TFIDF2 N として総スライドファイル中の総スライドページ数，df_j として w_j が 1 回でも出現するスライドページ数

本稿では，統計的手法による「G 値」,「TFIDF1」,「TFIDF2」を以後，**出現頻度指標値**と

呼ぶ.

3.3 語の抽出

実際にスライドから語を抽出するにはスライドに現れる文に対して形態素解析を行わなければならない．形態素解析エンジンには京都大学で開発された MeCab[5] を，解析辞書には IPA で作成された辞書を用いて語の抽出を行う．スライド中に現れる文は，一般の文章に表れる文とは異なり，完全な文になっていない場合や，改行や記号により文が分断されている場合がある．そのためいくつかの語に関しては，間違った解析結果を出力している．また，名詞が連節している場合，学生は，それらを複合名詞として認識し，重要語であるかどうかを判断すると思われる．例えば,「公開鍵暗号」という語の解析結果は,「公開」「鍵」「暗号」の 3 つの語になるが，学生は「公開鍵暗号」を 1 つの語と認識すると考えられる．そのため，形態素解析結果に対して以下のような修正を行い語の抽出を行う．

- 名詞が連節している場合は，一つの複合名詞として語の抽出を行う．
- カタカナ，漢字，英字からなる未知語は固有名詞と考えることができるため，これらが連節している場合は，一つの複合名詞として語の抽出を行う．
- 英字と記号が連結している場合は，一つの単語としてみなし語の抽出を行う．
- 平仮名のみからなる語は候補から外す.
- 数字のみからなる語は候補から外す.
- 代名詞，非自立名詞は候補から外す.
- 一文字語は候補から外す.

この修正を行った結果，C.1 No.11 からの抽出語は 332 語となる．この抽出語中に含まれる指摘語は 36 語である．網羅率 (全指摘語に対する抽出語中の指摘語数) は 83.7%であり，単純抽出精度 (全抽出語中の指摘語数) は，10.8%である．

3.4 抽出法の評価に用いる指標

今回我々が構築しようとしているシステムでは,「学生が重要であると思う語」を「学生が思う重要度の順」に教員に提示することを目指している．そのため，抽出法の評価尺度として「アンケート調査結果の順位」と「抽出結果の順位」との間でどの程度相関関係があるかを数値化し用いる．順位の相関度を表す指標としては，スピアマンの順位相関係数がある．この相関係数はノンパラメトリックな指標であり，2 つの変数の間の関係が任意の単調関数によってどの程度忠実に表すことができるかを示すものである．−1 から 1 までの値をとり，0 に近い値の場合は「相関関係はない」，1 に近い場合は「正の相関がある」，−1 に近い場合は「負の相関がある」と言うことができる．また，帰無仮説 H_0(2 つの変数には相関がない) を立てて有意性の検定を行うことが可能である．今回は，この係数を利用し,「アンケート調査結果」と 3 つの頻度情報指標値による語の並びとのそれぞれの相関係数を算出する．

表 3 各手法毎の抽出語と出現頻度指標値

G 値			TFIDF1			TFIDF2		
順位	抽出語	値	順位	抽出語	値	順位	抽出語	値
1.5	平文	37.868	1.5	平文	9.713	1	喪失	16.47
1.5	パスワード	37.868	1.5	パスワード	9.713	2	平文	14.948
3.5	暗号文	33.646	3.5	暗号文	8.633	3	暗号文	14.062
3.5	喪失	33.646	3.5	喪失	8.633	4	パスワード	13.633
5	片方	25.213	5	片方	6.475	5	片方	12.353
7	暗号	21.002	7	暗号	5.396	7	本人	9.414
7	可用性	21.002	7	可用性	5.396	7	暗号	9.414
7	推測	21.002	7	推測	5.396	7	推測	9.414
11.5	完全性	16.795	11.5	アルゴリズム	4.317	9	可用性	8.304
11.5	アルゴリズム	16.795	11.5	監視	4.317	11	アルゴリズム	8.235

順位は平均順位を表す

3.5 頻度情報のみによる抽出法の評価

3 つの出現頻度指標値に対し，全ての抽出語に値が大きい順に平均順位を付与する．これにより順位づけられた語のうち上位 10 語のみを例として**表 3** に示す．スピアマンの順位相関係数を計算する場合は，指摘語の 43 語に得点順に平均順位を付与し，それぞれの抽出法により得られた語 (332 語) の平均順位との間で計算する．また，欠損値 (抽出語と指摘語のうち片方にしか現れない語) は計算対象から除外するので，実際には抽出語のうち指摘語と一致している語 (36 語) の順位のみで係数の計算を行っている．計算には統計処理ツールである R を用いた．3 つの出現頻度指標値による平均順位との相関係数を**表 4** に挙げる．全ての場合において帰無仮説 (指摘語の順位と抽出法の順位に相関はない) を棄却できる (5%の危険域内に収まっている) ため，有意な相関関係は存在している．

4. 文字の属性を考慮した重みづけ

実際の授業で用いられるスライドでは，重要語は本文領域だけでなくタイトル領域に出現したり，視覚的効果（フォントサイズ，色，アンダーライン）が施されていたりする．これらの情報を語の重みとして加味することで，相関係数が向上すると考える．本節では，タイトル領域出現語，段落レベル，文字色，フォントサイズに関する重みづけを考察する．

4.1 タイトル領域出現語

タイトル領域に出現する語は，そのスライドページ全体を表す語であるため，学生も重要

Vol.2010-CLE-3 No.4
2010/12/9

表 4 出現頻度指標値のみによる学生アンケート結果との順位相関係数

抽出法	相関係数
G 値	0.409
TFIDF1	0.499
TFIDF2	0.476

表 5 タイトル係数を考慮した場合の相関係数

タイトル係数	抽出法	相関係数
T1	G 値	0.430
	TFIDF1	0.513
	TFIDF2	0.502
T2	G 値	0.470
	TFIDF1	0.509
	TFIDF2	0.501

語として認識する傾向があると考えられる．実際に C.1, No.11 の全抽出語に対するタイトル領域出現語の割合は 14.8%であるのに対して，アンケート結果の 62.8%の語がタイトル領域出現語である．

タイトル領域に出現する語に対しての重みづけ (タイトル係数) として以下の 2 つを考える．スライドファイル i における語 w_i のとし，タイトル領域での出現数を Et_{w_i} とする．

タイトル係数 1 出現頻度指標値とタイトル出現数で乗算を行う

$$T1_{w_i} = (1 + Et_{w_i}) \tag{2}$$

タイトル係数 2 出現頻度指標値と総出現数に対するタイトル出現数で乗算を行う

$$T2_{w_i} = (1 + Et_{w_i} \cdot \frac{Et_{w_i}}{E_{w_i}}) \tag{3}$$

この T1 または T2 を出現頻度指標値に掛けた値での順位と指摘語の順位との相関係数を**表 5** に示す．出現頻度指標値のみの場合より若干相関係数が上がっている．

4.2 段落レベル

PowerPoint スライド本文領域 (箇条書きプレースホルダー) では，一般に階層構造の箇条書きを行う場合が多い．第 1 レベルでそのページで説明する小見出しを提示し，一段下がった第 2 レベルで，その説明を行うような使い方がなされている．また，段落レベルが下がると，それに応じてフォントサイズも小さくなるため，視覚的にも上位レベルにある語を学生が注目すると考えられる．そのため，段落レベルも重要語を推定する手がかりとなる．

スライドファイルで使用されている最大の段落レベルの深さをと n する．前述のように w_i，E_{w_i} を定義し，ある段落レベル j での w_i の出現数を Ev_{w_ij} として，段落レベル係数 I_{w_i} を以下のように設定する．

$$I_{w_i} = \frac{\sum_{j=1}^{n} \frac{1}{k} \cdot Ev_{w_ij}}{E_{w_i}} \tag{4}$$

このように段落レベル係数を設定した場合，例えば，ある語 A が段落レベル 1 にのみ 5 回出現する場合，段落レベル係数は 1 となり，ある語 B が段落レベル 3 に 5 回出現する場合は，段落レベル係数は 0.333 となる．つまり段落レベルが深い場所に多く出現する語の出現頻度指標値を下げることができる．この段落レベル係数を出現頻度指標値に掛けた場合の順位およびタイトル係数も含めて掛けた場合の順位と指摘語の順位との相関係数を**表 6** に示す．段落レベル係数のみを出現頻度指標値に考慮した場合も出現頻度指標値のみの場合より少し相関係数は上がっている．また，タイトル係数も考慮する場合は，タイトル係数 1 よりタイトル係数 2 の方がよい．

4.3 文 字 色

スライド中で，ある語を視覚的に目立たせたい場合文字色を変える事がある．そのため，文字色も重要語を推定する手がかりとなる．PowerPoint では，スライドのデザインに応じて標準の文字色や標準の背景色等，標準色が設定されており，外部プログラムよりある色が標準色であるかどうかを判断可能である．これを用いて「文字色が標準文字色でない場合」を文字色係数と考えることとする．標準の文字色を使用していない w_i の出現数を Cn_{w_i} として，文字色係数 C_{w_i} を以下のように設定する．

$$C_{w_i} = 1 + \frac{Cn_{w_i}}{E_{w_i}} \tag{5}$$

この文字色係数を出現頻度指標値, タイトル係数 2, 段落レベル係数も含めて乗数にした場合のそれぞれの相関係数を**表 7** に示す．相関係数は全体的に下がってしまっている．

4.4 文字サイズ

スライド中で視覚的に語を目立たせるために文字色と同様文字サイズを大きくする場合があると思われる．PowerPoint では，スライド内のオブジェクトに応じて標準の文字サイズが設定されている．今回は，これを用いて「文字サイズが標準文字サイズより大きく設定されている場合」を文字サイズ係数と考えることとする．標準の文字サイズより大きい文字サイズを使用している w_i の出現数を Fs_{w_i} として，文字色係数 F_{w_i} を以下のように設定する．

$$F_{w_i} = 1 + \frac{Fs_{w_i}}{E_{w_i}} \tag{6}$$

この文字サイズ係数と出現頻度指標値, タイトル係数 2, 段落レベル係数，文字色係数も含めて掛けた場合の順位と指摘語の順位との相関係数を**表 8** に示す．相関係数は文字色係数までの場合よりさらに少し下がっている．

表 6 段落レベル係数を考慮した場合の相関係数

係数	抽出法	相関係数
I のみ	G 値	0.429
	TFIDF1	0.506
	TFIDF2	0.506
T1×I	G 値	0.420
	TFIDF1	0.491
	TFIDF2	0.500
T2×I	G 値	0.463
	TFIDF1	0.527
	TFIDF2	0.474

表 7 文字色係数を考慮した場合の相関係数

抽出法	相関係数
G 値	0.436
TFIDF1	0.494
TFIDF2	0.457

表 8 文字サイズ係数を考慮した場合の相関係数

抽出法	相関係数
G 値	0.436
TFIDF1	0.482
TFIDF2	0.455

5. 考　　察

今回の手法では，スライドに存在する語の各属性情報を属性の出現数を基に数値化し，その語の統計的に処理を行った出現頻度情報との間で乗算を行い，重要語の候補としての順位を決定している．さらに，以下の点に関して吟味をしなければならない．

(1) 各属性を考慮した場合の相関係数は約 0.41〜0.52 に間に収まっている．これらの値の差に統計的な有意な差が存在するかを検定してみたが，有意な差とは認められなかった．他資料でも同様な結果となるか調査が必要である．

(2) 文字色および文字サイズを考慮した場合，相関係数が下がっている．一つの原因としては，今回のスライドでは，語を強調させるために文字色の変更や文字サイズの変更を行っていないためである．また，実際には，文字色だけでなく，その文字の背景色も考慮する必要がある．

(3) 考慮しなければならない属性として，他にも語への下線やフォント自体の変更が挙げられる．

また，構築しようとしているシステムでは，抽出語を教員に提示する場合，抽出語の上位何語かを提示するような形での実装を考えている．仮に 20 語提示するとして，最も相関係数が良い「TFIDF1, タイトル係数 2，段落レベル係数との乗数」の上位 20 語と「TFIDF1, タイトル係数 2，段落レベル係数，文字色係数, 文字サイズ係数との乗数」の上位 20 語で，どのような学生指摘語が含まれているかを**表 9** に示す．二重下線の語は学生指摘語の上位 20 語に含まれている語であり，下線の語は，学生指摘語のうち 21 位から 43 位の語である．その他の語は学生指摘語に含まれない語であり，重要語の候補として本来提示すべきでない語である．どちらの場合も 20 語中 11 語が学生指摘語に含まれる語である．

表 9 係数を考慮した抽出語上位 20 語と指摘語上位 20 語の比較

TFIDF1×T2×I	TFIDF1×T2×I×C×F
暗号	暗号
公開鍵暗号	公開鍵暗号
分類	平文
パスワード	分類
セキュリティインシデント	パスワード
平文	暗号文
片方	セキュリティインシデント
リモートユーザ認証	片方
解読	リモートユーザ認証
共通鍵暗号	解読
暗号文	共通鍵暗号
セキュリティ	アクセス管理
第 3 者	持ち物
喪失	知識
アルゴリズム	FireWall
IBM	安全性
可用性	監視
推測	セキュリティ
セキュリティ対策	第 3 者
RSA	喪失

6. ま と め

今回，スライドに存在する語の頻度情報およびその語の視覚的属性情報を元に受講生が重要であると思う語の候補を抽出する手法を提案した．今後，他のサンプルを用いて本抽出法の有効性の評価を行う．また，実際の支援ツールとして使えるようにソフトウェア開発を行い，実用に耐えられるかどうかの検証を行う予定である．さらに，重要語の抽出だけでなく，他のスライド改善手法も考えていく．

参 考 文 献

1) 吉田　文，田口真奈：大学教員の IT 利用実態調査，NIME 研究報告第 38 号，独立行政法人 メディア教育開発センター (2008).
2) 仁木啓司，松浦健二，後藤田中，金西計英，矢野米雄：スライド教材のカスタマイズ環境を用いた教員・学習者支援研究：パワーポイントのレポートを用いた支援例，電子情報通信学会研究会報告 (教育工学)，Vol.106, No.583, pp.139–144 (2007).

Vol.2010-CLE-3 No.4
2010/12/9

3) 山田博文，金子喜俊，松田和彦，桂田浩一，新田恒雄：講義再現システムにおけるスライドへの重要度自動付与法とその評価，電子情報通信学会研究会報告 (教育工学), Vol.101, No.41, pp.25–32 (2001).
4) Sokal, R.R. and Rohlf, F.J.: Introduction to Biostatistics, W.H.Freeman and Company, San Francisco and London (1973). （藤井宏一訳：生物統計学，共立出版株式会社 （1983））.
5) Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), Association for Computational Linguistics, pp.230–237 (2004).