

プレゼンテーションコンテンツのための 概念的構造と表層的構造に基づくスライドの関係判定方式

王 元元[†] 北山 大輔^{††} 角谷 和俊^{††}

[†] 兵庫県立大学大学院環境人間学研究科 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

^{††} 兵庫県立大学環境人間学部 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

E-mail: [†]{nd09s005,ne07p001}@stshse.u-hyogo.ac.jp, ^{††}sumiya@shse.u-hyogo.ac.jp

あらまし プレゼンテーションコンテンツを閲覧する際、コンテンツを部分的に閲覧したり、ある部分に補足的な説明をすることが必要になる場合がある。しかし、コンテンツ内の各スライドが複雑に関係しているため、必要な内容のみを抽出して閲覧することが困難である。すなわち、単にコンテンツの内容を要約したり、検索トピックを補足している説明を抽出するだけでは、ユーザが興味を持ったトピックに関する文脈が不明であるという問題がある。そこで、スライド間の関係に基づく要約や補足を実現するために、意味的構造と表層的構造を用いたスライド間の関係判定方式を提案する。具体的には、概念辞書を用いてキーワードの意味階層とスライドのインデントによる表示階層を定義し、それに基づくスライドの概念的・表層的关系を抽出することによりスライド間の関係を判定する。

キーワード プレゼンテーションコンテンツ, マルチメディア, e-ラーニング

Determination of Relations between Slides based on Conceptual Structure and Visible Structure for Presentation Content

Yuanyuan WANG[†], Daisuke KITAYAMA^{††}, and Kazutoshi SUMIYA^{††}

[†] Graduate School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

^{††} School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

E-mail: [†]{nd09s005,ne07p001}@stshse.u-hyogo.ac.jp, ^{††}sumiya@shse.u-hyogo.ac.jp

Abstract It is easier for many users than before to view presentation content which consists of slides. In this context, when a user views whole presentation content, many slides are not necessary for the user. The user is hard to view the necessary content only in short time, and it is useful to search for the relevant slides based on the relation that can help the user's grasp the necessary content well by summarizing and supplying content. In this paper, we propose a method to determine relations between slides based on conceptual structure that semantic relation of keywords and visible structure that layer structure of indents in slide text of keywords. We describe here our method by using keywords' semantic relations and keyword's indent position in slides.

Key words presentation content, multimedia, e-learning

1. はじめに

プレゼンテーションコンテンツ（以降、単にコンテンツと記す）は、発表に使用されたスライド資料により構成される。このようなコンテンツは e-ラーニングや会議などの場面で利用されている。たとえば、会議や講義に参加できなかった者が、後日にプレゼンテーション資料を閲覧する場合や、発表者や聴講者が内容を知った上で再度参考にする場合に利用することがで

きる。最近では SlideShare [1] や MPMeister [2] のような、インターネット上でコンテンツを公開するサービスも出現し、ますます利用が促進されると考えられる。しかし、コンテンツにおいて情報を減らさずに短時間で必要なスライドを閲覧することは困難である。たとえば、重要なスライドのみを検索して抜き出すだけでは、プレゼンテーションの文脈が失われるため、抽出されたスライドのみを閲覧しても不十分な場合がある。また、ユーザが興味を持ったトピックを検索した場合に、トピッ

クを補足している説明を抽出することは可能であるが、それが文脈に従った説明であるとは限らず、効率的でない。そのため、ユーザの興味がある内容を、文脈を失わずスライドを抽出するためにスライド間の関係を判定する手法を提案する。

コンテンツは、テキストを含んだ複数枚のスライドによって構成される。そして、あるスライドからみて他のスライドが概念的な内容を詳細に説明しているというような、スライド間の詳細化/要約関係が存在する。我々は、詳細化/要約関係に当たる他のスライドを共に提示することで「コンテンツの内容に関する要約」や「指定スライドの内容を補足説明する」といった効率的な閲覧を支援できると考えた。このとき、スライド間に存在する概念的な関係と表層的な位置づけの構造を分析することによりスライド間の関係を求めることができると考えた。本稿では、概念辞書の WordNet [3] からスライド内のテキストにおけるキーワードの意味階層を用い、スライド間の概念的関係を抽出し、スライド間の概念的構造と定義する。また、スライド内のインデントによるキーワードの表示階層と意味関係を用いて表層的关系を抽出し、スライド間の表層的構造と定義する。

本研究では、概念的関係と表層的关系を組み合わせることで6種類のスライド間の関係を定義する。これを用いることでコンテンツを効率的に閲覧するためのスライド間の関係を明示することが可能となる。

以下、2節では本研究の概要と関連研究について述べる。3節では、スライドの概念的関係と表層的关系について説明する。4節では、スライドの関係判定について述べ、5節では、評価実験とプロトタイプシステムについて述べ、最後に6節でまとめを述べる。

2. 本研究のアプローチ

2.1 本研究の概要

ユーザがコンテンツを閲覧する際、あるスライドの内容に対して、情報を減らさず、かつ文脈が失われないように必要なスライドを閲覧したいという要求が発生する場合がある。このとき、コンテンツ内のスライド間の関係を提示し、指定されたスライドと他のスライドの関係を適切に抽出することで要求に応えることが可能である。しかし、時間的に前後して出現するスライドや、指定したスライドに類似したスライドが、適するスライドであるとは限らない。さらにスライド内のキーワードそのものの間に意味的な関係が存在し、さらに、作成者によってスライド内でのキーワードの使い方が異なる。そのため、意味階層と表示階層を利用してスライド間の関係を判定する必要がある。図1に本手法の概念図を示す。

コンテンツ中のスライドは、他のスライドと詳細化/要約関係を持つ場合がある。たとえば、概念的関係で、スライドXとYにIS-A関係がある($Y \text{ IS-A } X$)。スライドXとYは共通キーワード「根菜」が表層的关系で「根菜」の位置づけが上位と判定され、スライドXの内容に対してスライドYの内容が特化した内容を説明している。これらのスライド間にはSpecializationという関係があると考えた(図2)。このように、あるスライドに対し、スライド間の概念的関係と表層的关系

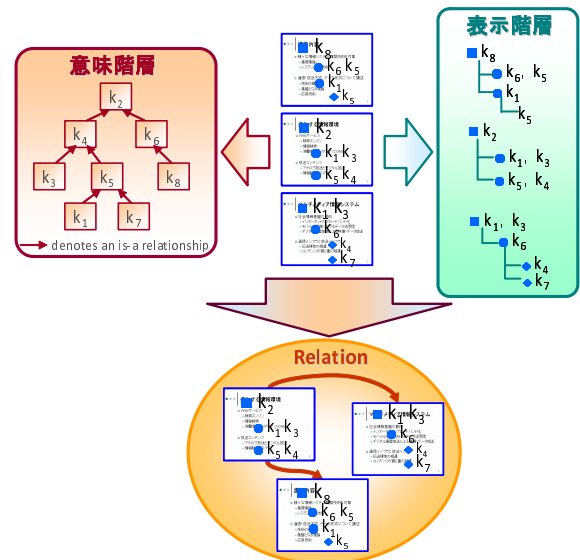


図1 意味階層と表示階層に基づくスライド間の関係判定の概念図

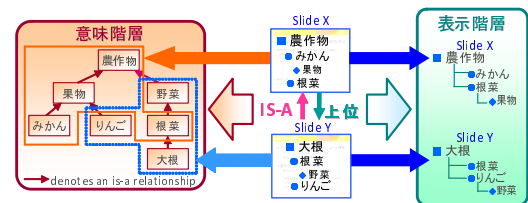


図2 スライド間の関係の例

表1 関連研究との比較

		要約手法		補足手法	提案手法
		小林ら [4]	Le ら [5]	山崎ら [6]	
1	全体要約	○	○	×	○
2	部分要約	×	○	×	△
3	トピック検索	○	×	○	○
4	文脈判定	×	×	×	○
5	用語説明	×	×	○	×

係を用いてスライド間の関係を判定する手法を提案する。本手法によりユーザはコンテンツを入力することで、スライドの内容を表すキーワードの意味と、インデントによる意味を考慮したスライド間の関係を得ることが可能となる。

2.2 関連研究

表1に本手法と関連研究との比較を示す。要約手法では、小林ら [4] の提案する UPRISE のようなメタデータを用いて資料や映像の疎結合を行い、コンテンツを蓄積し、提供する枠組みを提案し、コンテンツ全体から重要なシーンを判定し、要約を生成する手法がある。この研究では重要なシーンを抽出することでコンテンツの全体を要約することができる。Le ら [5] は、動画とスライドからコンテンツを作成するシステムである MPMeister [2] で収録された講義講演の情報から、重要シーンを抽出しダイジェストを自動的に作成する手法を提案している。各スライドのダイジェストを生成することで部分的な要約をすることができる。補足手法として、山崎ら [6] は、音声認識を行うと、各スライドより抽出した用語の単語辞書を作成し、ス

ライドを補足的に説明する手法を提案している．これらの研究では，単なるコンテンツの要約と補足の説明を足しただけではなく，スライド間の関係を抽出することでコンテンツの内容に関する要約，ユーザに合う指定されたスライドの内容を補足説明することもできる．また，特にスライド間に要約のような部分的に概念的な内容を説明している場合で，部分的な要約をすることも可能である．

小山ら [7] は，Web ページ内でキーワードの出現位置を考慮した検索方式を提案している．具体的には，同じキーワードでも，タイトルに現れる場合と，本文中に現れる場合とで，キーワードの意味を区別している．小山らは単一の Web ページの構造を対象として，キーワードの意味的な関係を抽出しているのに対し，本手法はキーワードの意味階層と表示階層を対象として，2 つの階層を組み合わせることでスライド間の関係を抽出するため方向性が異なる．

3. 概念的関係と表層的関係

3.1 キーワードの意味階層

本章では，キーワードの意味階層を用いてスライド間の概念的关系を判定する．キーワード間の意味関係で基礎となるのは is-a 関係 [8] である．「Y は X である (Y is-a X)」は通常，概念 Y は概念 X の特化であり，概念 X は概念 Y の汎化であることを意味する．例として「フルーツ」は「リンゴ」，「オレンジ」などの汎化である．すなわち，リンゴはフルーツである (Apple is-a fruit.) といえる．本稿では，WordNet [3] を利用して，キーワード間の意味的关系を抽出し，意味階層を定義する．

3.2 スライド間の概念的关系

スライド間の概念的关系はキーワードの意味階層における複数のキーワード間に is-a 関係を用いて判定を行う．以下にスライド間の概念的关系の定義を述べる．

IS-A スライド X のキーワードとスライド Y のキーワードの間に is-a 関係となるキーワードペアの個数がスライド X よりスライド Y が多いとき，IS-A と定義し，Y IS-A X と表す．これにはスライド Y がスライド X の内容について特化しているという特徴がある．以下の条件式を満たした場合，スライド間の概念的关系を Y IS-A X と判定する (図 3) ．

$$\sum_{a_i \in X} \sum_{b_j \in Y} hypo(a_i, b_j) > \sum_{b_j \in Y} \sum_{a_i \in X} hypo(b_j, a_i) \quad (1)$$

$$hypo(a_i, b_j) = \begin{cases} 1(b_j \text{ is-a } a_i) \\ 0(b_j \text{ not is-a } a_i) \end{cases} \quad (2)$$

$hypo(a_i, b_j)$ は，スライド Y のキーワード b_j とスライド X のキーワード a_i の間にある is-a 関係を判定する関数である．is-a 関係である場合に 1，そうでない場合に 0 を返す．スライド X より Y から is-a 関係となるキーワードペアの個数が多いかどうかを条件式 (1) により判定する．

PART-OF スライド X のキーワード a_i からなる意味階層 T_X の中で，子が複数個存在するサブツリーの一部にスライド Y のキーワード b_j からなる意味階層 T_Y が，概念的に包含されているとき，PART-OF と定義し，Y PART-OF X と

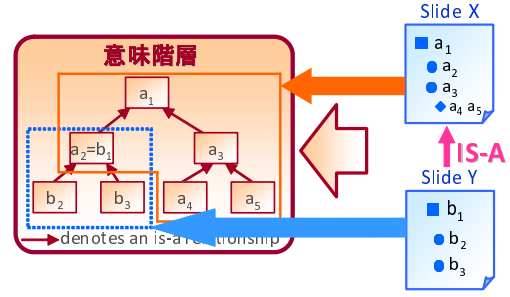


図 3 IS-A 関係の例

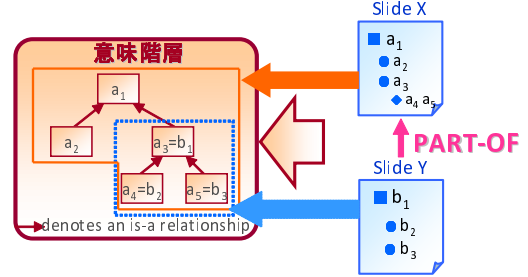


図 4 PART-OF 関係の例

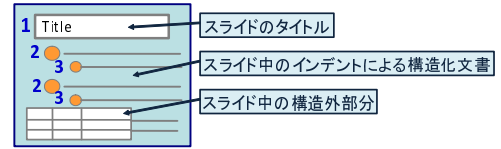


図 5 スライドの階層

表す．これにはスライド Y がスライド X の一部分について説明しているという特徴がある．以下の条件式を満たした場合，スライド間の概念的关系を Y PART-OF X と判定する (図 4) ．

$$T_Y \subset T_X \quad (3)$$

T_X は，スライド X のキーワード a_i からなる意味階層のキーワード集合を表す． T_Y は，スライド Y のキーワード b_j からなる意味階層のキーワード集合を表す．条件式 (3) は T_Y が T_X の中でサブツリーの一部と示す．

一般的にキーワード間の関係として is-a と part-of は並列にない．しかし，スライド X の一部分がスライド Y において特化している場合，IS-A と PART-OF 関係が並列に存在することがある．そのため，スライド間の概念的关系は IS-A のみ，PART-OF のみ，IS-A と PART-OF の並存の 3 つの場合が考えられる．条件式 (1)(2)(3) を満たした場合，スライド間の概念的关系を Y IS-A X，Y PART-OF X と判定する．

3.3 スライドの表層的構造

スライドからはスライドに書かれた文字列とインデントによる構造化情報を得ることができる．図 5 に本稿で扱うスライドの階層を示す．本稿では，スライドのインデントによる階層を以下のように定めた．

- スライドのタイトルに相当する要素を階層の最上位であ

る 1 とする

- インデントの開始要素を 2 として、インデントが深くなるたびに 3, 4 のように 1 階層ずつ深くなるものとする
- 図や表などのインデント外となる要素をスライドの平均階層にあるとして扱う

本研究では、スライド X 内においてキーワード a_x と、is-a 関係が存在するキーワード a_y の位置関係に対するスライド Y 内における位置の変化を表層的関係と定義する。表層的関係はスライド X, Y 間に共通のキーワードごとに存在する。すなわち、スライド間の表層的関係はキーワード間の表層的関係の集合となる。以下に、スライドのインデントによるキーワードの表示階層を定義する。ここでは、スライド間の共通キーワードと is-a 関係を持つキーワードの位置関係における判定について説明する。

上位 スライド X, Y 間に共通のキーワード k_c とスライド X 内のあるキーワード a_i, a_m の関係が k_c is-a a_i, a_m is-a k_c であり、 k_c とスライド Y 内のあるキーワード b_j, b_n の関係が k_c is-a b_j, b_n is-a k_c であるとする。このとき、スライド X で a_i が k_c より上位（同位）階層に出現している、かつ a_m が k_c より下位階層に出現しているという特徴がある。スライド Y で b_j が k_c より上位（同位）階層に出現している、かつ b_n が k_c より下位階層に出現しているという特徴がある。また、スライド X で k_c is-a a_i のキーワードペアの個数と a_m is-a k_c のキーワードペアの個数の割合がスライド Y で k_c is-a b_j のキーワードペアの個数と b_n is-a k_c のキーワードペアの個数の割合より多いという特徴がある。以下の条件式を満たした場合、この表層的関係ではスライド X, Y の共通キーワード k_c の位置づけを上位と判定する（図 6）。

$$\frac{\sum hyper(a_i, k_c) + 1}{\sum hypo(k_c, a_m) + 1} > \frac{\sum hyper(b_j, k_c) + 1}{\sum hypo(k_c, b_n) + 1} \quad (4)$$

$$hyper(a_i, k_c) = \begin{cases} 1(k_c \text{ is-a } a_i) \\ 0(k_c \text{ not is-a } a_i) \end{cases} \quad (5)$$

$hyper(a_i, k_c)$ は、共通キーワード k_c とスライド X のキーワード a_i の間にある is-a 関係を判定する関数である。is-a 関係である場合に 1、そうでない場合に 0 を返す。 $\sum hyper(a_i, k_c) + 1 / \sum hypo(k_c, a_m) + 1$ は、スライド X で is-a 関係となる上位（同位）階層に出現している a_i のキーワードペアの個数と下位階層に出現している a_m のキーワードペアの個数の割合を計算する。 k_c とスライド X 内の is-a 関係となるキーワードペアの個数の割合がスライド Y より多いかどうかを条件式 (4) により判定する。また、条件式 (2)(5) で、1 となるとスライド X で k_c is-a a_i, a_m is-a k_c と判定される。

下位 スライド X, Y 間に共通のキーワード k_c とスライド X 内のあるキーワード a_i, a_m の関係が k_c is-a a_i, a_m is-a k_c である。 k_c とスライド Y 内のあるキーワード b_j, b_n の関係が k_c is-a b_j, b_n is-a k_c である。このとき、スライド X で a_i が k_c より上位（同位）階層に出現している、 a_m が k_c より下位階層に出現しているという特徴がある。スライド Y で b_j が k_c より上位（同位）階層に出現している、 b_n が k_c より下位階層に出現しているという特徴がある。また、スライド Y で k_c is-a b_j のキーワードペアの個数と b_n is-a k_c のキーワードペアの個数の割合がスライド X で k_c is-a a_i のキーワードペアの個数と a_m is-a k_c のキーワードペアの個数の割合に類似しているという特徴がある。以下の条件式を満たした場合、この表層的関係ではスライド X, Y の共通キーワード k_c の位置づけを無変化と判定する（図 8）。

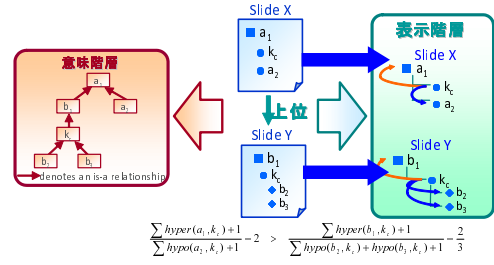


図 6 上位の例

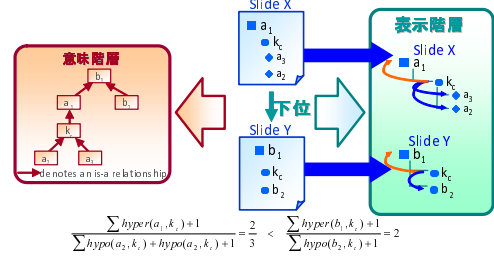


図 7 下位の例

出現しているという特徴がある。また、スライド Y で k_c is-a b_j のキーワードペアの個数と b_n is-a k_c のキーワードペアの個数の割合がスライド X で k_c is-a a_i のキーワードペアの個数と a_m is-a k_c のキーワードペアの個数の割合より多いという特徴がある。以下の条件式を満たした場合、この表層的関係ではスライド X, Y の共通キーワード k_c の位置づけを下位と判定する（図 7）。

$$\frac{\sum hyper(a_i, k_c) + 1}{\sum hypo(k_c, a_m) + 1} < \frac{\sum hyper(b_j, k_c) + 1}{\sum hypo(k_c, b_n) + 1} \quad (6)$$

スライド X, Y で is-a 関係となるキーワード集合を条件式 (2), (5) により判定する。 k_c とスライド Y 内の is-a 関係となるキーワードペアの個数の割合がスライド X より多いかどうかを条件式 (6) により判定する。

無変化 スライド X, Y 間に共通のキーワード k_c とスライド X 内のあるキーワード a_i, a_m の関係が k_c is-a a_i, a_m is-a k_c である。 k_c とスライド Y 内のあるキーワード b_j, b_n の関係が k_c is-a b_j, b_n is-a k_c である。このとき、スライド X で a_i が k_c より上位（同位）階層に出現している、 a_m が k_c より下位階層に出現しているという特徴がある。スライド Y で b_j が k_c より上位（同位）階層に出現している、 b_n が k_c より下位階層に出現しているという特徴がある。また、スライド Y で k_c is-a b_j のキーワードペアの個数と b_n is-a k_c のキーワードペアの個数の割合がスライド X で k_c is-a a_i のキーワードペアの個数と a_m is-a k_c のキーワードペアの個数の割合に類似しているという特徴がある。以下の条件式を満たした場合、この表層的関係ではスライド X, Y の共通キーワード k_c の位置づけを無変化と判定する（図 8）。

$$\left| \frac{\sum hyper(a_i, k_c) + 1}{\sum hypo(k_c, a_m) + 1} - \frac{\sum hyper(b_j, k_c) + 1}{\sum hypo(k_c, b_n) + 1} \right| < \alpha \quad (7)$$

スライド X, Y で is-a 関係となるキーワード集合を条件式 (2), (5) により判定する。 k_c とスライド X 内の is-a 関係となるキー

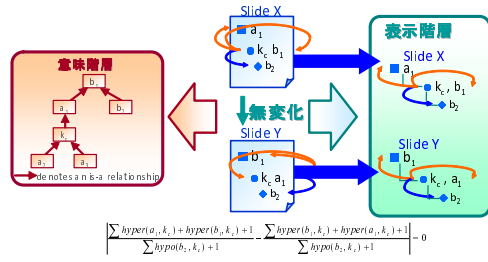


図 8 無変化の例

ワードペアの個数の割合がスライド Y に類似しているかどうかを条件式 (7) により判定する。α は、割合の差異に関する閾値である。

スライド内の位置関係を判定するとき、同じキーワードが複数階層に出現する場合があると考えられる。ここでは、is-a 関係が存在しているキーワードの最下位層同士と最上位層同士を比較する。最下位層のインデントと、最上位層のインデントから、最終的な最上位階層（最下位階層）が異なることによる上位（下位）を判定する。たとえば、あるスライド X 中の共通キーワード k_c とキーワード a_i 間の is-a 関係が存在し（ k_c is-a a_x ）、 a_i が複数階層に出現し、 k_c が階層 2 に出現する場合を挙げる。このとき、 a_i の最上位階層 1 と k_c の最上位階層 2 が異なり、かつ、 a_i の最下位階層 2 と k_c の最下位階層 2 が同じであるとき、最終的に a_i が k_c より上位階層に出現していると判定される。また、複数の共通キーワードが存在する場合もあると考えられる。ここでは、複数の共通キーワードと is-a 関係を持つキーワードの位置関係の集合を使ってスライドの表層の関係の判定を行う。

4. スライドの関係判定

本章では、スライド間の概念的関係と表層的关系を組み合わせてスライド間の関係を判定する手法について述べる。本手法では、スライド間の詳細化/要約関係として以下の 6 種類を定義する。

Generalization あるスライドに対して、概念的内容を説明している他のスライドとの関係である。スライド Y がスライド X より概念的な内容が説明しているとき、スライド X から Y に対する関係として判定する。

Specialization あるスライドに対して、特化した内容を説明している他のスライドとの関係である。スライド Y がスライド X より特化した内容が詳細に説明しているとき、スライド X から Y に対する関係として判定する。

Generic part-of あるスライドに対して、部分内容より一般的に説明している候補スライドとの関係である。スライド Y がスライド X の部分より一般的に説明しているとき、スライド X から Y に対する関係として判定する。

Specific part-of あるスライドに対して、部分内容をより増やして丁寧に説明している候補スライドとの関係である。スライド Y がスライド X の部分より増やして丁寧に説明しているとき、スライド X から Y に対する関係として判定する。

表 2 スライド間の関係判定

スライド間の関係	概念的関係	表層的关系
Generalization	$X IS - A Y$	下位
Specialization	$Y IS - A X$	上位
Generic part-of	$Y PART - OF X$	上位
Specific part-of	$Y PART - OF X$	下位
Partial generalization	$Y PART - OF X, Y IS - A X$	下位（無変化）
Partial specialization	$Y PART - OF X, Y IS - A X$	上位（無変化）

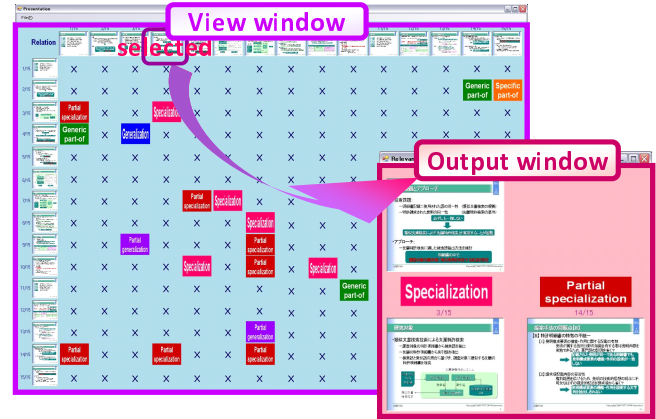


図 9 プロトタイプシステムの画面

Partial generalization あるスライドに対して、概念的な部分内容を説明している他のスライドとの関係である。スライド Y がスライド X の概念的な部分を説明しているとき、スライド X から Y に対する関係として判定する。

Partial specialization あるスライドに対して、特化した部分内容を説明している他のスライドとの関係である。スライド Y がスライド X の特化した部分を説明しているとき、スライド X から Y に対する関係として判定する。

スライド間の関係判定を表 2 に示す。具体的には、スライド間の概念的関係 $IS - A$ 、 $PART - OF$ 、 $PART - OF$ と $IS - A$ の並存と、スライド間の共通キーワードが各スライド内の is-a 関係が存在するキーワードとの表層的关系の変化を組み合わせる関係判定を行う。ここでは、スライド間の共通キーワードがスライド X 内の is-a 関係が存在するキーワードとの位置変化の割合、スライド間の共通キーワードがスライド Y 内の is-a 関係が存在するキーワードとの位置関係の割合を用いて判定を行う。

5. 評価

5.1 プロトタイプシステム

プロトタイプシステムは提案手法に基づき Microsoft Visual Studio 2008 C# によって開発した。プロトタイプシステムは、ユーザの入力部、スライド間の関係判定部と提示部によって構成される。図 9 にプロトタイプシステムの画面イメージを示す。入力部では、ユーザがプレゼンテーションコンテンツを指定することが可能である。システム内部のスライドの関係判定部は、スライド間の概念的関係の判定と表層的关系の判定によって構

表 3 評価実験の結果

セット	関係	人手による判定	提案手法による判定	概念的関係	表層的關係
1	Slide 1 と Slide 2	Partial generalization	Specialization	Slide 2 <i>IS</i> – <i>A</i> Slide 1	上位
2	Slide 1 と Slide 4	Generic part-of	Specialization	Slide 4 <i>IS</i> – <i>A</i> Slide 1	上位
3	Slide 3 と Slide 2	Generalization	Generalization	Slide 3 <i>IS</i> – <i>A</i> Slide 2	下位
4	Slide 4 と Slide 5	Partial specialization	Partial specialization	Slide 5 <i>PART</i> – <i>OF</i> (<i>IS</i> – <i>A</i>) Slide 4	無変化

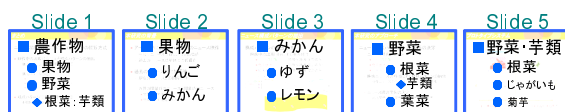


図 10 評価実験に用いたコンテンツ

成される。判定部では、WordNet [3] を使ってキーワードの意味階層を用いてスライド間の概念的関係を判定する。概念的関係が存在する各スライドのキーワードの表示階層によるスライド間の表層的關係を判定し、これを組み合わせてスライドの関係判定を行う。

出力部は、ユーザの入力に基づいて、コンテンツ内のスライド間の関係をユーザに提示する。ユーザが興味のあるスライドを指定した場合、別ウィンドウで関係するスライド群のサムネイルを出力する。スライドからのキーワード抽出は SlothLib [9] [10] より呼び出した MeCab [11] [12] による形態素解析を用いて行う。

5.2 評価実験

提案した関係の判定アルゴリズムを用いて評価実験を行った。評価実験ではある農作物に関するプレゼンテーションコンテンツのスライドを用いた(図 10)。人手により判定した関係と提案手法による出力を用いて 4 セットのスライドの関係を評価した。

その結果を表 3 に示す。人手で判定したセット 1, 2 において、スライド間の関係は Partial generalization, Generic part-of と判定されているため、概念的関係の *PART* – *OF* が存在すると考えられる。しかし、提案手法では *IS* – *A* のみが判定される結果となった。セット 1 の場合、共通キーワードは「果物」だけである。提案手法により *PART* – *OF* を判定する場合、共通キーワードが 1 つではなく、複数の共通キーワードを含む集合が得られるが、また、セット 2 の場合、共通キーワード集合の中の同じ特化概念を持っているキーワードが存在しないため、*PART* – *OF* と判定できない。これにより、*PART* – *OF* を判定する場合には、キーワードの数量、存在する意味などを判定する方式を改良する必要があると考えられる。一方、セット 3, 4 において人手で判定したスライド間の関係は Generalization, Partial specialization であり、提案手法より判定された結果と一致した。そのため、提案した関係のアルゴリズムの中で、スライドのインデントによるキーワードの位置関係の定義は適切であると考えられる。

6. おわりに

本稿では、スライド間の概念的関係と表層的關係を用いて関

係を定義し、スライド間の関係判定方式について提案を行った。その中で、スライド間の関係を抽出するためのキーワード間の意味関係による意味階層と、スライドのインデントによる表示階層を用いてスライド間の概念的関係と表層的關係を定義し、それらの関係を組み合わせた関係について説明した。提案手法の評価実験を行った。今後、評価実験に基づいてアルゴリズムを改良し、プロトタイプシステムを実装し、さらに評価実験を行う予定である。また、本研究では、コンテンツにおける同一意味階層を構成するキーワード間の関係性に基づき判定アルゴリズムを開発した。しかし、コンテンツにおいて、複数の意味階層が存在する場合が一般的であるため、複数の意味階層を用いた関係の判定方式が必要である。

謝 辞

本研究の一部は、平成 21 年度科研費基盤研究 (B)(2)「ユーザの潜在的意図を用いたレス・コンシャス情報検索基盤の構築」(課題番号: 20300039) および平成 21 年度特別研究員奨励費 (21.197) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] Slideshare. <http://www.slideshare.net/>.
- [2] Ricoh. MPMeister II, <http://www.ricoh.co.jp/mpmeister/>.
- [3] Wordnet. <http://wordnet.princeton.edu/>.
- [4] 小林隆志, 村木太一, 直井聡, 横田治夫. 統合プレゼンテーションコンテンツ蓄積検索システムの試作. 電子情報通信学会論文誌, 第 J88-D-I 巻, pp. 715–726, 2005.
- [5] H.-H. Le, T. Lertrusdachakul, T. Watanabe, and H. Yokoda. Automatic digest generation by extracting important scenes from the content of presentations. In *Proc. of the 19th International Conference on Database and Expert Systems Application (DEXA2008)*, pp. 590–594, September 2008.
- [6] 山崎裕紀, 岩野公司, 篠田浩一, 古井貞照, 横田治夫. 講義音声認識における講義スライド情報の利用. 情報処理学会研究報告, pp. 2006–SLP–64–39, 2006.
- [7] 小山聡, 田中克己. 文書構造を利用した web からの話題発見. 電子情報通信学会第 14 回データ工学ワークショップ (DEWS2003), 2-P-06, 2003.
- [8] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to wordnet: An on-line lexical database. In *Journal of Lexicography*, pp. 235–244, 1999.
- [9] SlothLib. <http://www.dl.kuis.kyoto-u.ac.jp/slothlib/>.
- [10] 大島裕明, 中村聡史, 田中克己. Slothlib:web サーチ研究のためのプログラミングライブラリ. 日本データベース学会 Letters, 第 6 巻, pp. 13–116, 2007.
- [11] Mecab. <http://mecab.sourceforge.net/>.
- [12] M. Asahara and Y. Matsumoto. Extended models and tools for high-performance part-of-speech tagger. In *Proc. of the 18th International Conference on Computational Linguistics (COLING 2000)*, pp. 21–27, 2000.