

## 精巧なスライドに注目したスライド検索システム

田中 清太郎<sup>I</sup> 手塚 太郎<sup>II</sup> 青山 敦<sup>III</sup> 田尾 啓一<sup>III</sup> 木村 文則<sup>II</sup> 前田 亮<sup>II</sup>

I II 立命館大学情報理工学部 〒525-8577 滋賀県草津市野路東 1-1-1

III 立命館大学 MOT 大学院 テクノロジー・マネジメント研究科

〒525-8577 滋賀県草津市野路東 1-1-1

E-mail: I cm007073@de.ritsumei.ac.jp II {tezuka, amaeda}@media.ritsumei.ac.jp

III {aoyama, taok}@mot.ritsumei.ac.jp II fkimura@is.ritsumei.ac.jp

**あらまし** 近年、企業などでは過去の PowerPoint 文書を再利用し資料などを作成する機会が多く、効率よく文書データを探し出せる検索システムの需要が高まってきている。本研究では、企業などに埋蔵されている資料の再利用を目的として新たに提案した手法を元にスライド検索システムの構築を行った。本手法では、「凝った又は、手間のかかった（以降、精巧と呼ぶ）スライド程、検索して見つけたい重要なスライドである」と仮定しスライドごとに図形・画像数や文字情報などによって重みを計算し、重要度を付与した。この重要度およびキーワードによるスライド検索を実施し、スライド並びにその集合体である PowerPoint 文書をランキングするシステムの構築を行った。

**キーワード** Windows, PowerPoint, OpenXML, スライド, 検索

## A Slide Search System for Finding Elaborate Slides

Seitaro TANAKA<sup>I</sup> Taro TEZUKA<sup>II</sup> Atsushi AOYAMA<sup>III</sup>

Keiichi TAO<sup>III</sup> Fuminori KIMURA<sup>II</sup> Akira MAEDA<sup>II</sup>

I II College of Information Science and Engineering, Ritsumeikan University,

1-1-1 Noji-higasi, Kusatu, Shiga, 525-8577 Japan

III Graduate School of Technology Management, Ritsumeikan University,

1-1-1 Noji-higasi, Kusatu, Shiga, 525-8577 Japan

E-mail: I cm007073@de.ritsumei.ac.jp II {tezuka, amaeda}@media.ritsumei.ac.jp

III {aoyama, taok}@mot.ritsumei.ac.jp II fkimura@is.ritsumei.ac.jp

### 1. はじめに

IT 社会と呼ばれる現在、内閣府調査によるとパーソナルコンピュータの普及率は 1990 年代前半では 10% 近くであったが年々増加し 2010 年には、74.6% まで上昇し過去最高となった[1]。それに伴い世界中に存在するデータ量も急激な増加が進み、2009 年には 2008 年から 62% の成長を見せデータ容量の総数は 8,000 億ギガバイトに達した。また、2020 年には年間のデータ量が 2009 年の 44 倍にも達するとも言われている[2]。これら膨大なデータ量を資源として効率的に管理し利用するには各分野に対応する検索システムの導入が必須

である。本研究では、電子文書の中の 1 つである PowerPoint 文書を対象にスライド単位での検索を行い、過去に蓄積した PowerPoint 文書の中から有益な情報を利用者に提供するシステムの提案・構築を行う。

近年では、様々な紙媒体が電子化され、文書データは増加し続けている。それに伴って従来の文書だけでなく、PowerPoint 文書で作成されたプレゼンテーション資料も文書データとして蓄積されるようになった。また、企業などでは資料を PowerPoint 文書により作成する機会が増え続けている。企業になると前任者などの他の社員が作成した資料なども利用する機会があり、

過去に作成した資料の保存場所が解らないことや、場所が解ってもどのデータが求めている資料なのかが解らずに無駄にコストや時間を費やしてしまうことがあり非効率である。しかし、システムで上手く支援を行うことが出来れば、過去の資料を用いることはノウハウの積み重ねにもなり、より効率的に業務を行うことが可能となる。

本研究では、検索キーワードに対応するスライドの検出を行い、重要度を付与しランキングを行った。また、オープンソースの検索システムとして HyperEstraiier [3]を用い、インデックスを利用した高速な検索システムを目指した。これにより過去のスライドの再利用を促進し、プレゼンテーション資料作成の作業効率の向上を図ることを目的としたサポートツールの構築を行った。

## 2. 関連研究

プレゼンテーションスライド検索の研究として、有熊らは企業内スライド検索手法を提案している[4]。有熊らの手法では、「流用性の高いスライドほど他の資料に流用されることが多い」と仮定し、スライド同士の類似度から流用性の高いスライドの集合を探し、検索キーワードと関連のある流用性の高いスライドを検出する。この手法の利点は、流用性の高い定型文のような似た内容を何度も書かなくてもこのシステムで検索したスライドを再利用すれば済むという点であり、将来的にテンプレートとしても用いることができる可能性も考えられる。

しかし、過去に使われた表などを用いたスライドや、あるキーワードに対して詳しく説明されているスライドなど、定型文では無く有熊らの手法では流用性が低いとみなされるスライドであっても再利用できる可能性もあると考えられる。そこで、本研究では、「精巧なスライドほど、有用なスライドであり、再利用に適したスライドである」と仮定し、有熊らの手法とは別の視点からスライド検索を行う手法を提案する。また、有熊らの手法では、流用性の高いスライドを検出するためにはある程度のスライド量が必要となるが、本手法ではデータベースに保存されているスライドが少ない段階でも有効に機能することが出来る。

スライド内の情報抽出に関する関連研究では、羽山らが任意の表示領域に応じて検索要求に関する情報を抽出する手法を提案している[5]。この研究では本研究がキーワードに関連する精巧なスライドを提示するのに対して検索要求に関連する部分の情報のみを抽出し提示することで有用なスライド情報の抽出を行っている。また、スライド情報を対象とした構造化手法を提案した研究がある[6]。本手法では、XML文書のタグ情報から使用されているオブジェクトなどの情報など

の要素を抽出しているが、この[6]ではスライドに含まれるオブジェクトやレイアウト情報などの機能的な纏まりを組織化し、構造化し計算機でも扱える形に整理するといった研究を行っている。別のスライドを再利用する研究としては、本研究の様な単一なデータベースを利用するのではなくイントラネットによるスライドの共有を利用した手法がある[7]。

## 3. 提案システム

### 3.1. 概要

本システムでは、精巧なスライドの検索システムの提案を行う。本手法においてスライドの精巧さを測る指標は各スライドに対して、画像が多い、図形が多い、キーワードの含まれている数が多い、文字数が多いこととした。これらの指標を基に、各スライドの重要度を算出する。検索時には、この重要度を利用して適正文書（スライド）をランキングし利用者に提示する。

システム内の処理の流れを図1に示す。図中の「データ格納モジュールの流れ」では、検索対象のPowerPoint文書をデータベースに登録するまでの流れを、「検索時の流れ」では、利用者が検索を行い、結果が返ってくるまでの流れを示している。

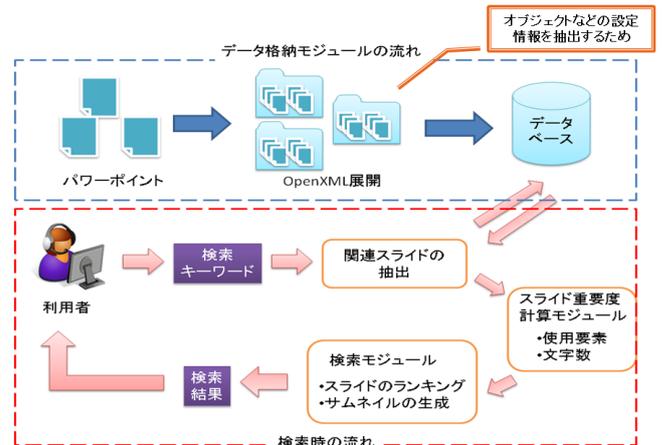


図1: システムの概要図

#### データ格納モジュールの流れ

管理対象となるPowerPoint文書群をOpenXMLに展開することでXML文書群を含むディレクトリ構造に変換し、データベースとして指定された場所に格納する。

#### 検索時の流れ

入力された検索キーワードに関連する各スライドをデータベースから抽出し、検索キーワードとの類似性およびスライドの重要度から各スライドの適合度の計算を行う。計算した適合度を基に各スライドをランキングし、検索結果として表示する。

### 3.2. システム画面

本システムは4つの画面からなっており、各画面で行う主な機能は以下のとおりである。



図 2：設定画面

**設定画面**

データベースを作成するディレクトリの指定や評価バランスの設定を行う(図 2)。評価バランスの設定とは、表 1 の重み係数の数値を利用者がある程度変更出来るようにしたものである。

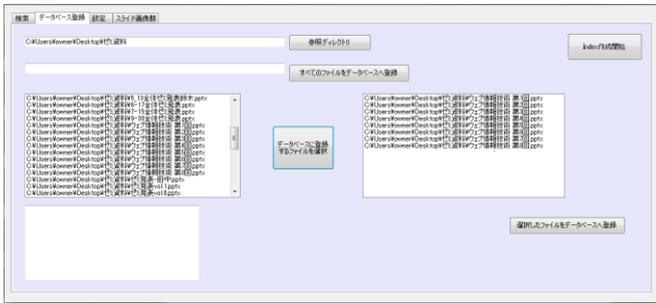


図 3：データベース登録画面

**データベース登録画面**

検索対象となる pptx (PowerPoint) 文書の登録とインデックスの作成を行う(図 3)。



図 4：検索画面

※①, ②の一部拡大したものが図 9, 10 である。

**検索画面**

入力されたキーワードに関連する PowerPoint 文書をランキングやスライドのランキングを表示する。また、対象となっているスライド本文とサムネイルの表示を行う(図 4)。



図 5：スライド画像群画面

**スライド画像群画面**

各スライドの一覧を表示する(図 5)。

**3.3. データ格納モジュール**

**3.3.1. データベース登録**

Microsoft Office 2007 から採用された文書フォーマットである OpenXML は従来のバイナリ形式のデータと異なり、XML 文書として記述されたテキスト部分と画像などのバイナリデータを ZIP 形式で圧縮した形式となっている。XML 文書はタグで構成されたテキストデータであり、要素ごとに整理されているため、各要素の情報を取得しやすい。そのため、PowerPoint 文書の細かな情報を要素ごとに管理することができる為、OpenXML を採用した。

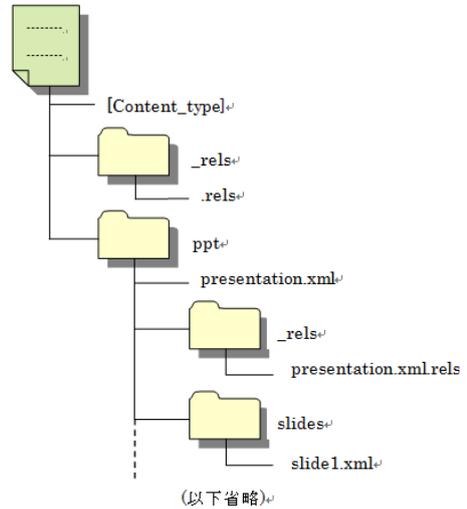


図 6：pptx 文書のパッケージ構造 [8]

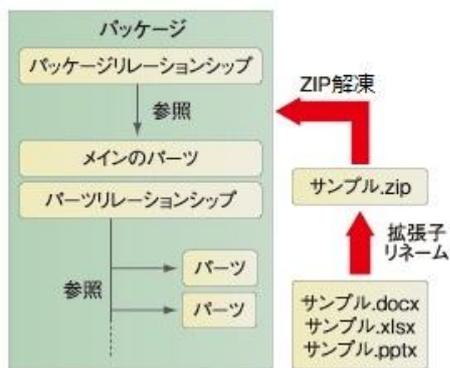


図 7：OpenXML のディレクトリ構造

OpenXML では、図 6 のように各情報が分けられている [8] ため、細かな設定が不要であることから OpenXML の ppt 対応拡張子 pptx の PowerPoint 文書を検索元として扱っている。pptx 文書を zip 拡張子にリネームし、展開することで、図 6 のようなパッケージ構造をもつディレクトリに変換できる。そのため、図 3 のデータベース登録画面内で検索元となる pptx 文書を登録するために、指定のディレクトリに pptx 文書のコピーが作成され、図 7 の様に展開したディレクトリ（パッケージ）群をまとめ、各 PowerPoint 文書のデータベースとして使用している。

また、各スライド画面の画像を JPEG 画像として保存し、データベース内に保存している。これは、3.6 節で述べるサムネイルに使用される。

### 3.3.2. テキストデータの抽出

データベースの登録と同時に各 PowerPoint 文書のスライド内に含まれるテキストデータを抽出し、txt 文書として保存する。この txt 文書は後に述べる 3.5.1 節で Hyper Estraier を利用したインデックスの作成を行うときにインデックスの元となるデータとして使用する。また検索結果として表示されたリスト内から選択対象となっているスライドに対して、スライド内のテキストデータも表示している。これは文章によってスライドの内容が判断できるようにするためである。この表示されるテキストデータは、対象となるスライドに合致する txt 文書から表示されている。

## 3.4. 検索

### 3.4.1. スライド重要度計算モジュール

スライドの要素は、図 6 に示されている slides ディレクトリの中にある各スライドの XML 文書中のタグ情報から抽出する。

スライドの重み付けは、表 1 に示した要素によってそれぞれ加算される。さらに、「凝ったスライドとは、画像・図形が多いスライドである」と考えられるため、対象のスライド内の画像・図形の個数も表 2 の要素と

して加えている。スライドのランキングを行うスコアはこれらの要素から加算される合計値である。また“タイトルに含まれている”、“内容に含まれている”の二つ以外の要素については、複数ある場合はその個数分の重みを掛け合わせる。

スライドのスコアの計算式は以下のとおりである。

$$\text{スコア}(S) = \sum_{n} \{b_n + s_n + u_n + t + c\} + (k \times 3) + (i \times 3) + (f \times 2)$$

※n はスライド内に含まれるキーワードの保有数

表 1：キーワードに対する重み付け要素

要素	重み(係数)
キーワード自体の重み(k)	3
タイトルに含まれている(t)	20
内容に含まれている(c)	1
キーワードが(デフォルトより)大きい文字サイズの時の数(s)	5
キーワードが太文字の数(b)	5
キーワードに下線が付く数(u)	5

表 2：スライドに対する重み付け要素

要素	重み(係数)
画像数(i)	3
図形数(f)	2

上の表 1,2 の要素のほかに、凝ったスライドとは、文字数の多いスライドであるとも考えられるため、スライドの XML 文書のタグを含んだ文字数もスコアの要素とし、下の表 3 の係数をスコアに乗算している。

表 3：XML 文書の文字数に対する係数

文字数	係数
0 ~ 100	1.0
101 ~ 200	1.1
201 ~ 300	1.2
301 ~ 400	1.3
401 ~	1.4

### 3.4.2. 検索モジュール

スライドの検索は、キーワードによる全文検索を行う。3.4.1 節で行った重み付けの後、各スライドの XML 文書から全文検索を行い、キーワードが含まれるスライドを抽出する。

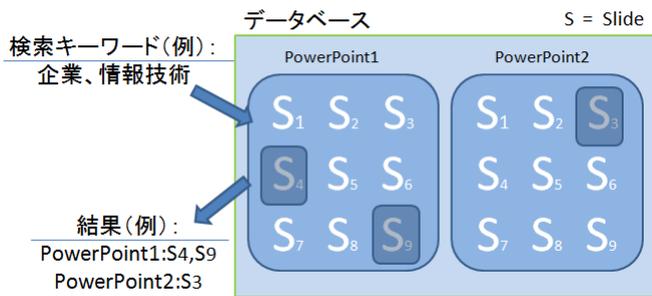


図 8 : 検索の流れ

図 8 に示すように、検索キーワードからデータベース登録されている各 PowerPoint 文書のスライドに対してインデックスを用いて検索を行い、キーワードが含まれるスライドを抽出し、結果として出力されたスライドに 3.4.1 節の重み付けを行い、各スライドのスコアを求める。PowerPoint 文書全体のスコアは、重みを付けたスライドのスコアの合計値であり、図 8 の例であれば、PowerPoint 文書 1 のスコアは S4 と S9 のスコアの合計となる。それぞれスコアを計算した PowerPoint 文書に対してランキングを行う。検索結果表示画面上でランキングされた PowerPoint 文書のリスト (図 9) から一つを選択すると、その PowerPoint 文書内で抽出されたスライド (図 8 の例であれば、S4, S9 など) を新たなリストとして表示し、図 10 のようにスライド単位でのランキングを行う。検索キーワードが複数存在する検索を行った場合は、1 番目のキーワードから最後のキーワードまで図 8 の流れを順々に繰り返す。この時、現在検索しているキーワードで抽出された PowerPoint 文書が前のキーワードでも抽出されていた場合は、各キーワードで得られたスコアをスライドごとに加算し、PowerPoint 文書のスコアを求める。

スコア	FailName
297	ウェブ情報技術 第4回.pptx-OpenXml
237	ウェブ情報技術 第1回.pptx-OpenXml
122	ウェブ情報技術 第7回.pptx-OpenXml
116	ウェブ情報技術 第5回.pptx-OpenXml
55	ウェブ情報技術 第6回.pptx-OpenXml

図 9 : PowerPoint のランキングリストの一部

スコア	タイトル
65	研究概要 - 多言語Webページ
59	システムの構成
59	System overview
59	System overview
44	Webプロキシサーバの動作
40	第1回「World Wide Webの概要」Web...
40	Webアプリケーションの事例3Webグルー...
26	Webページの例
23	通常のWebアプリケーション同期通信サ...

図 10 : スライドのランキングリストの一部

### 3.5. Hyper Estraier

本システムでは、検索速度向上を目的に LGPL(GNU Lesser General Public License)で提供されているインデックス型の全文検索エンジンライブラリーである Hyper Estraier をシステム内に導入した。

Hyper Estraier は、世界各国の言語に対応しており、形態素解析と N-gram の二つの方式から検索が行えるという特徴があり、日本語や英語の入り混じるスライド検索には最適な検索エンジンである。また、属性検索や類語検索など多くのサポート検索が行える。さらに大量の文書データから短時間でインデックスの作成が行え、このインデックスを利用した高速な検索が行える。これにより Hyper Estraier 導入前では、検索キーワードを各スライドから一つ一つ順次検索で行い検索キーワードの含まれたスライドを抽出していたが、インデックスの作成を Hyper Estraier で自動的に行うことにより、転置索引による検索で効率よく適切なスライドを抽出することが可能となった。

#### 3.5.1. N-gram 方式

この方式は、検索対象となる文章を形態素解析のように単語単位ではなく、文字単位で分割し、検索を行う全文検索の一種である。具体的には、N 個の文字を文字列として分割を行っている。この方式の特徴は、検索漏れが少なく文法解析を行わずに多言語が扱える点である。逆に問題としては、インデックスのサイズが大きくなることがあげられる。形態素解析と N-gram 方法の比較 [9] を表 4 に示す。

表 4 : 形態素解析と N-gram 方法の比較 [9]

	形態素解析	N-gram
インデクシング速度	遅い	速い
インデックスサイズ	小さい	大きい
検索ノイズ	少ない	多い
検索漏れ	多い	少ない
検索速度	速い	遅い
言語依存	辞書が必要	辞書が不要

本システムでは膨大なデータ量から検索を行うため、情報漏れが少なく今後も pptx 文書が追加されていくため辞書が不要で言語による縛りが無い N-gram 方式を採用している。また、Hyper Estraier は N-gram 方式で検索を行っているが形態素解析とのハイブリッドによる検索精度の向上が行えるため、検索ノイズなどの問題にも対応することが出来る。

### 3.6. サムネイル

3.4.2 節の最後でリスト化された図 9 の PowerPoint のうちの一つを選択するとその PowerPoint の一枚目のスライドがサムネイル表示される。また、その中でランキングされた図 10 のスライドのうち 1 つを選択すると、3.3.1 節の最後で述べた JPEG 画像をそのスライドの全体画像としてサムネイル表示するようになっており、視覚的にスライドをとらえることができる。

### 3.7. スライド画像群

スライド画像群の画面では、検索画面でランキングされた中から選択された PowerPoint 文書に含まれているすべてのスライド画像をサムネイル表示している。これにより、図 5 のようにランキングされたスライドが含まれる PowerPoint 文書の各スライドを一覧にして表示し、利用者が視覚的にどのような内容の PowerPoint 文書に含まれるスライドであるかを判断できるようにしている。

## 4. 評価実験

現在、システム上に 76 件の PowerPoint 文書がデータベースに登録され総数 1064 件のスライドが存在している。本章では、「検索キーワードが含まれるスライドが実際に出てきているか」という、精度を評価する方法と、「精巧なスライドが上位に来ているか」を評価する 2 つの方法によりシステムの評価を行った。

### ・精度に対する評価方法

検索する PowerPoint 文書の中からあらかじめランダムで選択した 10 件のキーワード名と検索結果の中に含まれるであろう PowerPoint 文書名などのスライド情報の詳細を表 5 に示す。

表 5：キーワードが記入されているスライド群

キーワード	PowerPoint 文書名	スライドページ	備考
植物	9-30 全体ゼミ発表表.pptx	3 ページ目	なし
版元	2010.10.21(村岡).pptx	8 ページ目	テーブル内の文章
九天社	2010617.pptx	5 ページ目	最後のスライド
定正氏	矢印クリップアート.pptx	2 ページ目	テーブル内の文章
沿革	FSI プロローシャー_FY10_v1.pptx	5 ページ目	タイトル
統括	FSI プロローシャー_FY10_v1.pptx	1 ページ目	先頭スライド
『吾妻鏡』	古典史料における人名・地名・・・	4 ページ目	特殊記号 + 文章
θ	小西_2010_11_11.pptx	5 ページ目	特殊記号
PMO	FSI プロローシャー_FY10_v1.pptx	15 ページ目	英数
Title	ウェブ情報技術第 8 回.pptx	9 ページ目	図形内に記入

### ・精巧なスライドに対する評価方法

実際に精巧なスライドがランキングの上位に表示されているかを評価する方法として、図 11~22 の 12 件のスライドを精巧なスライドと仮定し、それぞれのスライドに含まれる文章の中からキーワードを任意に選択し、システムで検索を行い検索結果として表示されるリストの中から全体件数の中から、上位に目的のスライドが表示されるか調査を行った。任意に選択したキーワードと検索によってヒットしたスライド番号、その順位と全体の件数を表 6 に示す。また、表 6 のスライド番号は、図 11~22 の 12 件のスライドと対応している。



図 11：スライド 1

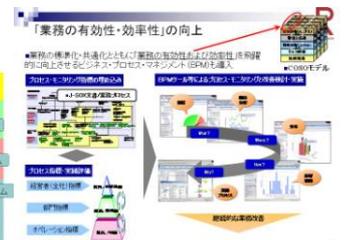


図 12：スライド 2

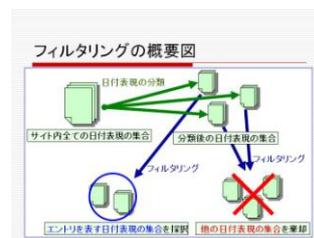


図 13：スライド 3

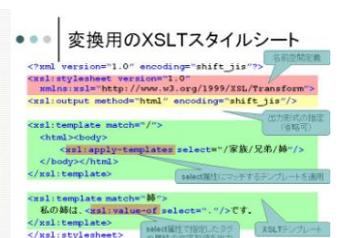


図 14：スライド 4

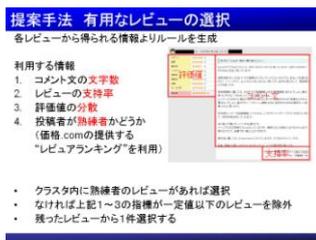


図 15: スライド 5

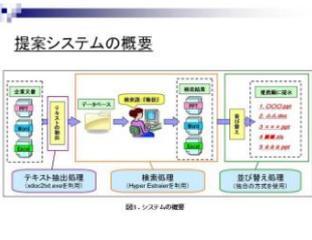


図 16: スライド 6



図 17: スライド 7



図 18: スライド 8

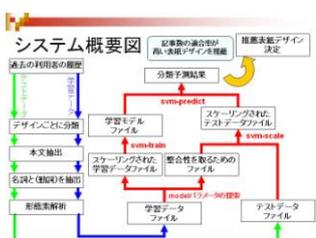


図 19: スライド 9

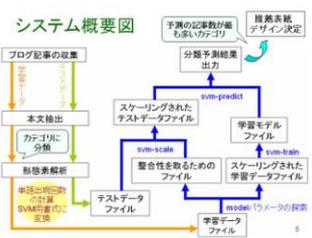


図 20: スライド 10

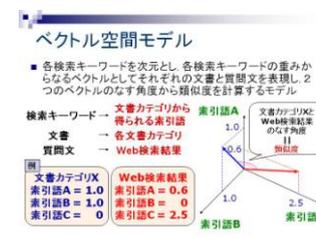


図 21: スライド 11

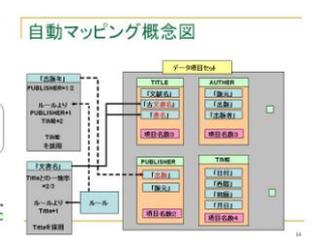


図 22: スライド 12

表 6: 実用性の調査結果

キーワード	スライド番号	ランキング順位	件数
ビジネス	2, 7	7, 1	16 件
サービス	8	2	14 件
概要	3, 6, 9, 10	9, 8, 17, 5	46 件
評価	2, 5, 7	2, 19, 1	78 件
キーワード	11	2	19 件
モデル	2, 7, 10, 11	3, 10, 1, 18, 9	60 件
タグ	1, 4	6, 3	27 件
自動	12	10	36 件

また、これら小規模な実験と並行してシステム全体で同様に精巧と呼ぶスライドが上位を占めているかどうかを調べる為に簡易的ではあるが全スライドを対象に実験を行った。評価実験の準備として本研究で使用した 1064 件のスライドを製作者の任意によって精巧な

スライドとして判断された「グループ 1」とそうでないスライドの「グループ 2」という二つのグループに分けた。キーワードによる検索を行い結果として表示されるランキング最上位から 10 位までのスライドに以下の式に従って数値を割り当てる。

$$x_n = \frac{1}{n}$$

$n$ にはランキングの順位が入力され、 $x_n$ に各スライド順位に割り当てられた評価数値が入る。

リスト内で表示されている各スライドを 1 つずつグループ 1, 2 のどちらかに所属しているかを調べ、グループ 1 に属しているスライドである時に限り、ランキングの順位に合わせた評価数値を加算し、リスト内でその合計数値を導き出す。この合計値が高いほど精巧と判断したスライドがランキング上位に表示されているとして各キーワードに対するそれぞれの合計値を求めた。また、この実験で検索するキーワードは小規模な実験で利用したキーワードと同様である。表 7 には各キーワードとこの実験で計算した値を示している。また、評価点数は小数点以下 4 桁で四捨五入を行った。

表 7: キーワードに対する評価点数

キーワード	評価点数
ビジネス	2.9290
サービス	2.6623
概要	2.5956
評価	2.4678
キーワード	2.4123
モデル	2.9290
タグ	2.6611
自動	2.5373

※グループ 1 に所属するスライド数: 452

グループ 2 に所属するスライド数: 612

## 5. 考察

1 つ目の評価方法では、実際にシステム内で表 5 のキーワードをそれぞれ検索し、スライドが検索結果として表示されるかを調査した結果、特殊記号が含まれているものや、日本語、英語で書かれたキーワードであっても 100% の精度でキーワードが含まれるスライドが検出された。これは、PowerPoint 文書に直接書かれた文章を対象にした場合であるからともいえる。現段階では、画像や、写真に含まれた PowerPoint 文書に直接的に書かれていない文章からは情報抽出が出来ないためである。これは、画像内に含まれている文章は、文書データではなく画像という属性でシステムがとらえている為である。だが直接 PowerPoint 文書内に記入されたテキストデータに対しては上記のとおり高い精度となっておりシステムの精度評価としては満足のか

く結果となった。今後は、スライドの中に含まれる画像などからも情報を抽出し、データの充実度を高めていく予定である。

2 つ目の評価方法では、いくつかのスライドに含まれている検索キーワードで検索結果を確認した。検索キーワードには他の多くのスライドにも含まれている一般的な単語を選択している。これは、一般的な単語であっても精巧なスライドを提供できるかを実験するためと検索対象となるスライドの特色が大きく含まれた単語を検索キーワードに設定した場合は、結果として出てくるスライドが特定し易く、スライド数も少数となるためランキングを行う必要性が低くなるからである。結果、開発者自身が「精巧なスライドであろう」と任意で選んだサンプルスライド群ではあるが、全てのサンプルスライドが検索結果の上位にランキングされていることが確認出来た。また、全スライドを対象に行った実験でも各キーワードに対する評価点数の平均が 2.6493 となった。これは評価点数の最大値が 2.9290 なので十分高い数値を得たといえる。この数値から各キーワードのランキング 1 位には全て精巧なスライドに所属するグループ 1 の中から表示され、その他にも多くの精巧スライドがランキング上位にきていることが解る。これらの結果から提案通りシステムが機能していると考えることが出来る。

本研究では精巧なスライドを検索する上でスライドに含まれる画像や図形の数が大きく関連している。その数が多ければ多いほどスライドのスコアが上昇し、ランキングの上位に向かう。これは、画像や図形に含まれる情報が精巧なスライドであるかを判断する上で重要な要素となると考えたからである。しかし画像の場合、1 枚の画像であっても極めて重要な情報が含まれている可能性があり現状の個数だけでは情報が重要な画像であるのかどうかを判断することが出来ない。この問題を解決するためには、何らかの形で画像に含まれた情報を抽出する必要があるため今後の発展として考慮していく。

## 6. おわりに

本論文では、スライドの再利用を目的とし、その一つの手法として精巧性に着目したスライド検索システムを提案した。なぜなら作り込まれたスライド程参考として、もしくは資料として再利用できるスライドだと考えているためである。評価実験の結果から、提案通りのスライド検索システム構築が成功したと判断する。今回の研究では検索対象となるスライドに対して条件を設けていなかった。しかし、研究の過程でスライドには配布資料としての側面を持った一般的に企業内で作成されるような細かな情報も記載されているスライドと、プレゼンテーションの要約のみを簡潔に列

挙されているスライドの 2 種類の形が存在することが解った。

今後は機能評価だけでなくユーザ評価を行い、本手法の有用性を示していく予定である。さらに、対象となるスライドの形を絞って、どのような場面で利用されるシステムなのかを検討し実験を進めていく。また、重み付けを行う各要素の重みは、使用者側がどのようなスライドを要求したいかによって、システム内で利用者側からある程度数値を変更出来るようになっていくが、要素自体に学習機能を設け、利用者に対して各要素が適切な数値(重み)になるように構築していきたい。その他にも細かな要素の追加や要素の追加とキーワードに対する同義語や類義語の対応なども行いより精度の高いスライド検索システムの研究を行っていく予定である。

## 参 考 文 献

- [1] ESRI,消費動向調査 22 年 3 月  
<http://www.esri.cao.go.jp/jp/stat/shouhi/2010/1003shouhi.html>
- [2] asahi.com (朝日新聞社) : 2020 年には年間データ量が 45 倍近くまで成長する見通し - 企業リリース - ビジネス・経済  
<http://www.asahi.com/business/pressrelease/CNT201006290046.html>
- [3] 全文検索システム Hyper Estraier ,  
<http://fallabs.com/hyperestraier/>
- [4] 有熊威, 白石展久 “スライドの流用性に着目した企業内スライド検索手法の提案”, 情報処理学会第 71 回全国大会論文集, pp 489-490, 2009.
- [5] 羽山徹彩, 国藤進 “プレゼンテーションスライド情報検索のためのスライドページからの要求関連乗法の抽出” 情報処理学会研究報告 2010-DD-76(2),1-7, 2010-07-15
- [6] 羽山徹彩, 難波英嗣, 国藤進 “プレゼンテーションスライド情報の構造化” 電子情報通信学会技術研究報告.OIS, オフィスインプォメーションシステム 108(156),45-50, 2008-07-17
- [7] 木村聡宏, 梶原清彦 “プレゼンテーション資料共有システムの一提案” 情報処理学会研究報告.DD[デジタル・ドキュメント]98(59),31-37,1998-07-17
- [8] Grier 陽子,高山佳文,森本孝司 “入門 Office OpenXML”,ソフトバンク クリエイティブ株式会社,2007/9/7 ,p130
- [9] 全文検索-Wikipedia ,  
<http://ja.wikipedia.org/wiki/%E5%85%A8%E6%96%87%E6%A4%9C%E7%B4%A2>