

Title	3 項組知識表現< “ 動画知 ” , “ 自然言語知 ” , “ メタデータ ” >とその知識創造支援活動への応用
Author(s)	澤井, 進
Citation	
Issue Date	2006-12
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/3455">http://hdl.handle.net/10119/3455</a>
Rights	
Description	Supervisor: 國藤 進, 知識科学研究科, 博士

3 項組知識表現< “ 動画知 ” , “ 自然言語知 ” , “ メタデータ ” >と  
その知識創造支援活動への応用

北陸先端科学技術大学院大学  
澤 井 進

# 博士論文

## 3項組知識表現<“動画知”, “自然言語知”, “メタデータ”>と その知識創造支援活動への応用

指導教官 國藤 進 教授

北陸先端科学技術大学院大学  
知識科学研究科 知識社会システム学専攻

澤 井 進

2006年12月

## 要 旨

本論文では、独創的な“動画知”、“自然言語知”、“メタデータ”という3項組による新しい知識表現及び、Web ページ検索システムを活用した知識利用の新たな共有手法を提案する。

同時に、特に将来予測が重要になる事前評価(assessment)などで問題になる潜在的問題分析のシステム化研究を行い、新たな知識獲得法である、暗黙知から形式知を表出する一つの手法を提案する。現在まで、潜在的問題分析の日本語でのシステム化研究は皆無で、本研究が初めてである。本論文の研究成果を要約すると、次のようになる。

第1章では、研究の目的と意義、本論文における動画知、自然言語知と潜在問題分析の意味、研究の背景並びに、論文の構成を説明する。

第2章では、<“動画知”、“自然言語知”、“メタデータ”>という3項組の知識表現でデジタルアーカイブ化する手法について述べる。従来の知識創造の研究では、暗黙知とは知っていても言葉にできない経験的で身体的なアナログの知で、形式知は暗黙知を言葉や体系にしたデジタルで共有可能な知と定めている。匠の技は暗黙知の一例である。メタデータはキーワードの集合で、IEEE 準拠の学習オブジェクトメタデータ LOM をメタデータとして採用し、LOM の RDF/RDFS 表現を用い、LOM の構造オントロジーを作成する。

第3章では、上記3項組の知識表現を利用し、KT 法の潜在問題分析システムによる新しい知識獲得法について述べる。複数人により、Web 上の動画知、自然言語知、掲示板やチャット等といった Web テクノロジーを多角的に活用した潜在問題分析システムを評価し、潜在問題分析システムによる新しい知識獲得法を考察する。

第4章では、第2章において提案した3項組の知識表現の知識利用例として、潜在的意味インデキシング (LSI) 法を用いる、自然言語によるWebページ検索について述べる。

第5章では、上記 Web ページ検索システムで“動画知”を利用したサポートベクトルマシン (SVM)による適合性フィードバックを行い、適合性フィードバックあるなしにおけるシステム性能比較を行い、検索精度の向上の度合いを評価する。

第6章では、セマンティック Web 技術を用いた曖昧な検索ができる e-Learning 支援システムを実現するため、LOM の RDF/RDFS 表現を用い、LOM の構造オントロジーを作成し、第4章において説明を行った Web ページ検索システムの検索性能を向上させる手法を検討する。

最後に、第7章において全体の総括を行い、今後の課題について検討する。

本論文では、新たに<“動画知”、“自然言語知”、“メタデータ”>という3項組の知識表現を提案し、デジタルアーカイブ化する手法について述べた。本 Web ページ検索システムで“動画知”を利用し、新たにサポートベクトルマシン (SVM)による適合性フィードバックを行い、検索精度の向上の度合いを評価した。同時に高度な意味処理を行うため、最新技術であるセマンティック Web 技術を用い、曖昧な検索ができる新しい e-Learning 支援システムの試行を行った。

また、知識獲得の一手法として、暗黙知から形式知を表出する一つの手法を提案し、日本で始めて KT 法による潜在問題分析システムを開発した。

## Abstract

In this paper, we propose a new knowledge representation method using a unique triplet knowledge representations, namely "animation knowledge", "natural language knowledge", and "metadata", as well as a new sharing technique utilizing Web page search engines for knowledge use.

At the same time, we conducted systematization research on potential problem analysis that becomes an issue in prior evaluations (assessments), especially those in which future predictions become an especially important element, and propose a new knowledge acquisition method of extracting explicit knowledge from tacit knowledge. Up to date, there has been no precedent of systematization research on potential problem analysis in Japanese, and this study is the first ever to be conducted. The results of our research, which are detailed in this paper, are summarized as follows.

In Chapter 1, we lay out the aim and purpose of our study, explain the significance of motion picture knowledge, natural-language knowledge and potential problem analysis, and account on the background of our research and the structure of this paper.

In Chapter 2, we describe the technique of digital-archiving by way of utilizing the triplet knowledge representations, i.e. "animation knowledge", "natural language knowledge", and a "metadata". Traditional researches on knowledge creation define tacit knowledge as "knowledge of the experiential, physical and analog realm, which cannot be expressed in language even when it is known", and explicit knowledge as "digitally shareable knowledge, created by verbalizing and systemizing tacit knowledge". So called "mastery" is an example of tacit knowledge. Metadata is a set of keywords using IEEE-compliant learning object metadata LOM, and the structure ontology of LOM is consisted of RDF/RDFS expressions of LOM.

In Chapter 3, we introduce a new technique of the knowledge acquisition using a potential problem analysis system by the KT method, applying the aforementioned triplet knowledge representations. Furthermore, we assess the potential problem analysis system multifacetedly using Web technologies such as animation knowledge, natural language knowledge, bulletin boards and chat rooms etc. on the Web by multiple people and consider the usability of the knowledge acquisition technique which uses the potential problem analysis system.

In Chapter 4, we elaborate on Web page searches by natural language in which the LSI method is utilized as an example of knowledge utilization of the triplet knowledge representations explained in Chapter 2.

In Chapter 5, we conduct conformity feedback by the support vector machine (SVM) which uses "animation knowledge" by the aforementioned Web page search engine. Subsequently we compare the performance of the system with and without conducting conformity feedback, and evaluate the improvement of search accuracies.

In Chapter 6, we create the structure ontology of LOM using its RDF/RDFS expressions in order to realize an e-Learning supporting system which allows ambiguous searches using semantic Web technologies, and examine methods for improving the precision of Web page searches explained in Chapter 4.

Finally, we present an overall summary at Chapter 7, and review our future tasks.

In this paper, we have proposed the triplet knowledge representations, i.e. "animation knowledge", "natural language knowledge", and a "metadata", and introduced techniques to digital-archive such.

We utilized "animation knowledge" for this Web page search engine, newly conducted conformity feedback by the support vector machine (SVM), and evaluated the improvement of search accuracies. Furthermore, we tested a new e-Learning supporting system which can perform ambiguous searches using Semantic Web technologies in order to simultaneously conduct more advanced, semantic processings.

Moreover, we have proposed a technique for expressing explicit knowledge from tacit knowledge as a method of knowledge acquisition, and have developed Japan's first potential problem analysis system using the KT method.

# 目次

## 1 序論

1.1 研究の目的と意義	1
1.2 本論文における動画知、自然言語知と潜在問題分析の意味	1
1.3 背景	2
1.4 論文の構成	4

## 2 3項組知識表現<“動画知”, “自然言語知”, “メタデータ”>

2.1 はじめに	7
2.2 3項組知識表現によるデジタルアーカイブ	7
2.3 メタデータ“LOM”	9
2.4 Web ページ	10
2.5 LOM の RDF 表現	10

## 3 潜在的問題分析支援システムによる知識獲得

3.1 はじめに	12
3.2 潜在的問題分析支援システム	15
3.3 eラーニング用Web検索エンジンの潜在的問題分析への適用	19
3.4 潜在的問題分析システムの評価	22
3.5 まとめ	22

## 4 3項組知識表現と LSI 検索を用いた e ラーニング用 Web ページ検索エンジン

4.1 はじめに	24
4.2 Web検索エンジンの特長	24
4.3 自然言語(形式知)をWebページ検索に利用する	26
4.4 潜在的意味インデキシング(LSI)	31
4.5 特異値分解を用いた次元圧縮	31
4.6 LSIを用いるWebページ検索	34

4.7 LSI検索の課題	39
<b>5 “動画知”を利用した SVM による適合性フィードバック</b>	
5.1 はじめに	40
5.2 Web検索エンジン	40
5.3 サポートベクトルマシン (SVM)	41
5.4 暗黙知を利用したSVMによる適合性フィードバック	42
5.5 適合性フィードバックの実行例	44
5.6 具体的な適合性フィードバックの手順	46
5.7 システムの実行	47
5.8 論文を対象としたSVMによる適合性フィードバック	51
5.9 ユーザ・アンケートによる定性評価	57
5.10 システムの検証	59
<b>6 セマンティックWeb技術を用いたe-Learning支援システム</b>	
6.1 はじめに	64
6.2 LOM-RDF変換ツールと相互運用性	64
6.3 RDFの検索及び評価	68
6.5 RDF検索のまとめ	72
<b>7 結論</b>	
7.1 本論文のまとめ	73
7.2 今後の展望	74
<b>謝辞</b>	75
<b>参考文献</b>	76
<b>本研究に関する発表論文</b>	79
<b>【添付資料】</b>	
資料1 花の画像の検索事例	81
資料2 検索データベース素材(1)「埼玉」	87
資料3 検索データベース素材(2)「愛知」	95
資料4 検索システムのユーザ・アンケート用紙	109



# 目 次

1.1 本論文の構成	6
2.1 動画知、自然言語知とメタデータの3項組の知識表現	8
2.2 LOMデータ	9
2.3 LOM データ(対象範囲を特定する項目の例)	10
2.4 Webページとデータベース	10
2.5 Dublin Coreを用いたRDF表現の例	11
3.1 S E C Iモデル	13
3.2 自然言語、動画とメタデータの3項組の知識表現	15
3.3 分析入力フォーム	17
3.4 BBS(掲示板)を含めたシステム画面	18
3.5 投稿されたテーマ	18
3.6 間伐に関する潜在的問題分析処理例	19
3.7 BBS(掲示板)画面のイメージ図	19
3.8 間伐の途中経過画面の例	20
3.9 潜在的問題分析の処理の流れ図	21
4.1 Webページ検索エンジンの概要	25
4.2 LOM 規格概要	26
4.3 Webページ検索システムの特徴	27
4.4 検索システム構成図	28
4.5 本 Web ページ検索システムの事例	34
4.6 検索質問文章入力	35
4.7 400次元での検索結果	36
4.8 水に関するページの1番目	37
4.9 水に関するページの2番目	37
4.10 水に関するページの3番目	38
4.11 水に関するページの4番目	38
4.12 水に関するページの5番目	39

5.1 Web ページ検索エンジン	41
5.2 サポートベクターマシン	42
5.3 暗黙知を利用したSVM による適合性フィードバックStep2	43
5.4 暗黙知を利用したSVM による適合性フィードバックStep3	43
5.5 暗黙知を利用したSVM による適合性フィードバックStep4	44
5.6 暗黙知を利用したSVM による適合性フィードバックStep5	44
5.7 適合性フィードバックの実行例の検索結果	45
5.8 適合性フィードバックの実行例のフィードバック画像	45
5.9 自然言語による文章で入力	47
5.10 本Webページ検索システムの事例	48
5.11 キーワード検索システムNICERの事例	49
5.12 Web ページとデータベース	51
5.13 論文データベースのLOM データ	52
5.14 論文データベースの検索	53
5.15 検索対象Web ページ	54
5.16 フィードバック情報を入力中の検索結果	55
5.17 フィードバック用の論文画像	56
5.18 適合性フィードバック後の検索結果	56
5.19 ユーザの観点から見た検索システムに関する評価結果	57
5.20 ユーザの観点から見た検索システムに関するアンケート結果	58
5.21 ユーザの観点から見た検索システムに関するアンケート結果	58
5.22 ユーザの観点から見た検索システムに関するアンケート結果	59
5.23 ユーザの観点から見た検索システムに関するアンケート結果	59
5.24 適合性フィードバック検証のFB 画像	60
5.25 フィードバック結果のFB 画像	61
5.26 適合率-再現率曲線	62
5.27 訓練データ数の違いによるF 値	63
6.1 サムネイルの項目をLOM-RDFで表現したもの	65
6.2 プロパティをRDFSで定義したもの	65
6.3 RDF/RDFSによる社会教科の階層構造の表現	66
6.4 LOM-RDF変換ツールの入力ファイル	67
6.5 LOM-RDF変換ツールの出力ファイル	67
6.6 LOM-RDF変換ツールのシステム構成図	68
6.7 LOM-RDF変換ツールSeRQLの検索画面	69

# 表 目 次

1.1 暗黙知と形式知の対比	2
3.1 P P Aチャート	16
3.2 P P Aチャートの適用事例	16
3.3 試用調査による評価結果	22
5.1 評価結果（上位20件中の正例数）	49
5.2 適合性フィードバック検証の検索条件	60
5.3 適合性フィードバック検証の検索結果	60
5.4 適合性フィードバック検証のFB情報	61
5.5 適合性フィードバックのフィードバック結果	61
5.6 適合性フィードバックの検証結果	63
6.1 検索結果として得られた教材の例(地理)	70
6.2 検索結果として得られた教材の例(歴史)	71
6.3 検索結果として得られた教材の例(公民)	71

# 第1章

## 序論

### 1.1 本研究の目的

本研究の第一の目的は、知識表現である。動画や静止画に一部含まれる暗黙知は、文章による形式知よりも直感的に人間の記憶に植えつける手段としてその能力を発揮し、自分や他人の状態を理解する手段として最適である。

研究の第二の目的は、知識の利用である。本研究で検索対象とするWebページは、自然言語であるテキストシナリオと動画や静止画のデジタル映像ファイルを持つ。このテキストシナリオは映像ファイルの内容が記されており、これをデータベース化することで検索可能となり、共用を可能とする。更に他のシステムとの相互運用の観点から、オントロジーやRDFといったセマンテックWebの考え方やツールの導入も検討する。具体的には、メタデータであるLOM ( Learning Object Meta data ) をRDFで記述することを試みる。

研究の第三の目的は、知識の獲得である。潜在問題分析 ( P P A ) の研究成果を使い、暗黙知の一部から形式知が表出することを利用し、暗黙知の一部からメタ情報やテキストシナリオを新しい知識として獲得する。コンピュータを使い、< “ 動画知 ” , “ 自然言語知 ” , “ メタデータ ” > という3項組の知識表現で処理できる潜在問題分析ツールを開発することを目指している。これまで問題分析 ( P A ) や決定分析 ( D A ) の研究は行われてきたが、潜在的問題分析 ( Potential Problem Analysis ; P P A ) の研究は皆無である。本論では、特に将来予測が重要になる事前評価などで問題になる、P P A を深く研究する本論文では、動画知と、自然言語知とメタデータの3項組によるコンテキストの表現及び共有手法を提案する。

### 1.2 本論文における動画知、自然言語知と潜在問題分析の意味

本論文では、「創造とは、人が問題を異質な情報群を組み合わせ統合して解決し、社会あるいは個人レベルで、新しい価値を生むこと」で、「創造とは問題の発見と解決だ」という

「創造」の原点に立ち返り、動画知、自然言語知と潜在問題分析について以下のように定義する。

### (1) 動画知

暗黙知(tacit knowledge)とは、言語で明確に表現することができない、もしくはそれが困難な直観知、身体知、体得知、あるいは事実知に対する技能知などをいう(表1.1参照)。従って、技能的なものも含む、広い意味で用いられる。

本研究では、マイケル・ポランニーの暗黙知[1][2]のサブセットとして、言語で明確に表現することができない、もしくはそれが困難だが、一部聴覚を含み、視覚によって得られる直観知、身体知、体得知、あるいは事実知に対する技能知などを、便宜上「動画知」と定義して用いる。

暗黙知のサブセットである「動画知」は、静止画または静止画の連続である動画により映像化・可視化された暗黙知である。

### (2) 自然言語知

形式知とは、哲学の伝統では、言語化された理論知をいう(表1.1参照)。本研究では、形式知のサブセットとして、自然言語で表現した知識を、「自然言語知」と定義して用いる。

表1.1は、一般的にいわれる暗黙知と形式知の対比である。

暗黙知	形式知
主観的な知(個人知)	客観的な知(組織知)
経験知(身体)	理性知(精神)
同時的な知(今ここにある知)	順序的な知(過去の知)
アナログ的な知(実務)	デジタル的な知(理論)

表 1.1 暗黙知と形式知の対比

### (3) 潜在問題分析

潜在問題分析は、今後大いに起こりうる、注意を払うべき問題を明らかにし、それに対応するための体系的なプロセスである。

本研究では、暗黙知から形式知を引き出す知識の表出化の一手法といわれている潜在的な問題分析[28]は、知識獲得の一手法と位置づけている。

## 1.3 背景

本研究は、学習オブジェクトメタデータ(Learning Object Metadata, 以下LOMと称す)と、自然言語を用い類似語検索を行う潜在的意味インデキシング(以下LSI)検索システムを用いたe-learning用Webページ検索エンジン [3] の研究を発展させたものである。

従来、自然言語を用いた検索では、Web ページのテキストシナリオ部分からデータベースを作成し、潜在的意味インデキシング (Latent Semantic Indexing : LSI) の検索手法[3, 4]を使い、入力質問文章と類似の Web ページを検索し、サポートベクターマシン(Support Vector Machines、以下 SVM と称す)による適合性フィードバックシステムを用いた情報検索システムの研究が行われている。

しかし、検索に自然言語 (形式知) を使い、動画知を利用したサポートベクターマシン (Support Vector Machines、以下 SVM と称す) による適合性フィードバックシステムの研究が行われていない[5]。具体的には、サムネイル画像や動画を提示することにより、動画知を利用し、直感的な判断を可能にする情報検索システムが必要になっている。

実際、筆者等が関係する学習ソフトウェア情報研究センターには、「学習情報研究」の論文 100 件と、全国のマルチメディア教材研究会員「GENES (Gakujoken Network Studying group)」が作成した約 4 万点の Web 素材・教材があり、より適確に情報検索する Web ページ検索エンジンが必要となっている[6]。

更に、コンピュータが Web の意味を理解するためのセマンティック Web 技術[7]技術の意味情報を利用してユーザの意図した検索結果や、相互運用性により Web 上にあるあらゆる LOM[8] [9]情報の検索を実現する。そして RDF で表現された LOM の検索について有用性の評価を行い、文科省の NICER (National Information Center for Educational Resources: 教育情報ナショナルセンター) [10]の検索システム[11]では実施していないが、学習指導要領[12] [13]における曖昧さと難易度を組み合わせた検索が望まれている。

知識獲得の研究では、自然言語、動画とメタデータの 3 項組による知識表現を用い、暗黙知から形式知を引き出す知識の表出化の一手法である潜在的問題分析支援システムを新規に提案する。なお、本研究では、マイケル・ポランニーと同様に、暗黙知を言葉では説明できないが、理解して使っている知識と考えている。

K T 法とは社会心理学者ケプナー (C. H. Kepner) と社会学博士トリゴー (B. B. Tregoe) によって開発された思考法であり、「合理的・効率的な意思決定手順の体系化と組織活動への適用」を目的としており、世界で最も使われている意思決定のための創造技法のひとつである。その分析プロセスは、状況分析 (Situation Appraisal ; 以下 S A ) 問題分析 (Problem Analysis ; 以下 P A ) 決定分析 (Decision Analysis ; 以下 D A ) 潜在的問題分析 (Potential Problem Analysis ; 以下 P P A ) という四つに分類される。とりわけ、潜在的問題分析はケプナー・トリゴーのモデルの中でも最も定式化が困難であるといわれている。[14][15]。

潜在的問題分析は、野中の S E C I モデルの暗黙知から形式知を引き出す知識の表出化の一手法といわれている[15]。さらに、共同化から表出化のプロセスにおいて、國藤は「人間のもってる膨大な暗黙知から形式知を取り出す表出化(分節化)のプロセスを支援する発想支援ツールを実現するには発散思考、収束思考、アイデア結晶化のプロセスを支援するツール構築が必要」と述べている[16]。

従来、K T 法の技法の一部を用いて開発したシステムには問題分析 ( P A ) を導入した八

重樫等[17]や中村等が開発したシステム [18] や、決定分析 ( D A ) を導入した佐藤等が開発したシステム [19][20] が存在するが、潜在的問題分析( P P A )を用いた例はまだない。

2006年1月19日発表の国の「IT新改革戦略」[21]の施策と、学習指導要領の目標の一つともなっているように、ネット社会を生きる子どもたちには、情報を正しく選択・判断し活用できる情報活用能力とともに、確固とした倫理感が必要といわれている[22]。

学習目標「自ら学び自ら考える力の育成を図る」が教育目標に登場して久しいが、いまだ学校教育に発想支援・問題解決の手法が十分に活かされているとはいえない。社会教育の分野においてはブレインストーミング法やKJ法が浸透して来ており、コンピュータを使ったツールも出現し始めている。しかし、より深く考えられるリスクを分析する潜在問題分析についてはビジネスの世界ではKT法などで人手により活用されるが、コンピュータを使い問題の発見段階からコンテキスト分析まで一環して処理できるナレッジマネジメント・ツールが現在のところ存在しない。

本論文では、上記の現状認識のもとに、創造とは問題の発見と解決だという立場から、動画知、自然言語知とメタデータの3項組の知識表現を新たに考案し、本知識表現を用いるeラーニング用Web検索エンジン[23]を新規に開発し、自然体験やボランティア活動などで発生する潜在的問題分析を支援する一手法をコンピュータ化する。

## 1.4 論文の構成

本論文は全7章により構成される。図1.1に全章の構成図を示す。

第1章の序論では、本研究の目的、動画知、自然言語知と潜在問題分析の意味と背景を述べ、論文構成について概説する。

第2章では、<“動画知”、“自然言語知”、“メタデータ”>という3項組の知識表現でデジタルアーカイブ化する手法について述べる。これまでの知識創造の研究では、暗黙知は知っていても言葉にできない経験的で身体的なアナログの知で、形式知は暗黙知を言葉や体系にしたデジタルで共有可能な知と定めている。匠の技は暗黙知の一例である。メタデータは、キーワードの集合で、IEEE準拠の学習オブジェクトメタデータLOMをメタデータとして採用し、LOMのRDF/RDFS表現を用い、LOMの構造オントロジーを作成する。

第3章では、上記3項組の知識表現を利用し、KT法の潜在問題分析システムによる新しい知識獲得法について述べる。複数人により、Web上の動画知、自然言語知、掲示板やチャット等といったWebテクノロジーを多角的に活用した潜在問題分析システムを評価し、潜在問題分析システムによる新しい知識獲得法を考察する。

第4章では、第2章において提案した3項組の知識表現の知識利用例として、潜在的意味インデキシング(LSI)法を用いる、自然言語によるWebページ検索について述べる。

第5章では、上記 Web ページ検索システムで“動画知”を利用したサポートベクトルマシン (SVM)による適合性フィードバックを行い、適合性フィードバックあるなしにおけるシステム性能比較を行い、検索精度の向上の度合いを評価する。

第6章では、セマンティック Web 技術を用いた曖昧な検索ができ e-Learning 支援システムを実現するため、LOM の RDF/RDFS 表現を用い、LOM の構造オントロジーを作成し、第4章において説明を行った Web ページ検索システムの検索性能を向上させる手法を検討する。

最後に、第7章において全体の総括を行い、今後の課題について検討する。



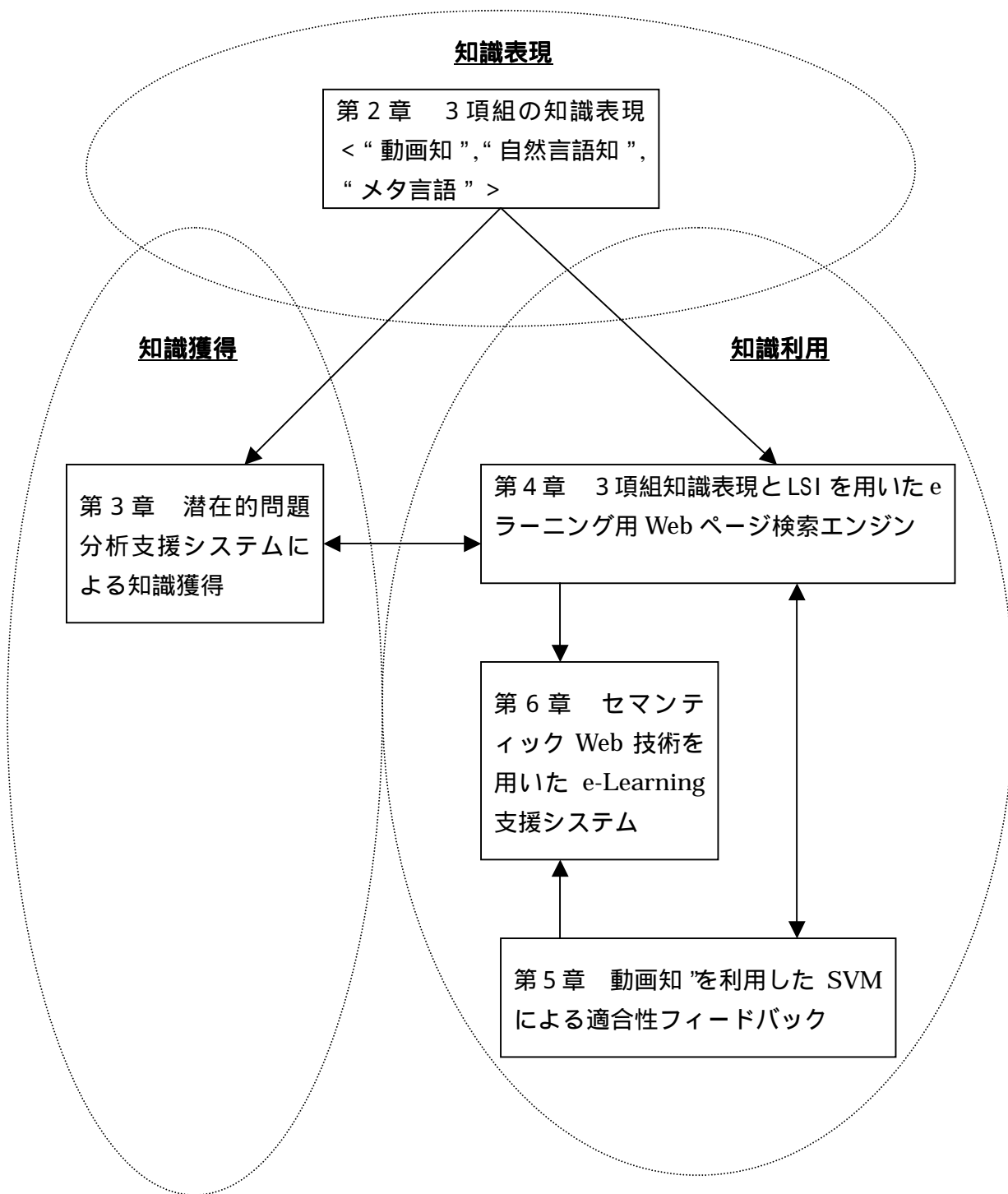


図 1.1 本論文の構成

## 第2章

### 3 項組知識表現

#### < “ 動画知 ” , “ 自然言語知 ” , “ メタデータ ” >

#### 2.1 はじめに

本章では、新たに研究開発した< “ 動画知 ” , “ 自然言語知 ” , “ メタデータ ” > という3項組の知識表現でデジタルアーカイブ化する手法について述べる。

これまでの知識創造の研究では、暗黙知は知っていても言葉にできない経験的で身体的なアナログの知で、形式知は暗黙知を言葉や体系にしたデジタルで共有可能な知と定めている。匠の技は暗黙知の一例です。本章では、匠の技のノウハウを記録した“ 動画知 ”、“ 自然言語知 ”とメタデータの3項組でデジタルアーカイブ化する。メタデータは、キーワードの集合で、本論では IEEE 準拠の学習オブジェクトメタデータ LOM をメタデータとして採用した。現在、研究開発中の LOM の RDF/RDFS 表現と、教材デジタルアーカイブについて言及する。

#### 2.2 3 項組知識表現によるデジタルアーカイブ

「静止画」や静止画の集合体としての「動画」に一部含まれる暗黙知は、文章による形式知よりも直感的に人間の記憶に植えつける手段としてその能力を発揮し、自分や他人のコンテキスト（状態）を理解する手段として最適である。

本論文では「静止画」や静止画の集合体としての「動画」を新規に「動画知」と称し、文章による形式知を新規に「自然言語知」と称す。

図 2.1 は、匠の技のノウハウなどを記録した動画知、素材内容を説明する自然言語知と、IEEE 準拠の学習オブジェクトメタデータ LOM で記述したメタデータという3項組による知識表現である。図 2.1 の動画知、自然言語知とメタデータの3項組による新し

い知識表現を用いる効果は、動画や静止画による暗黙知を、テキスト検索できることにある。

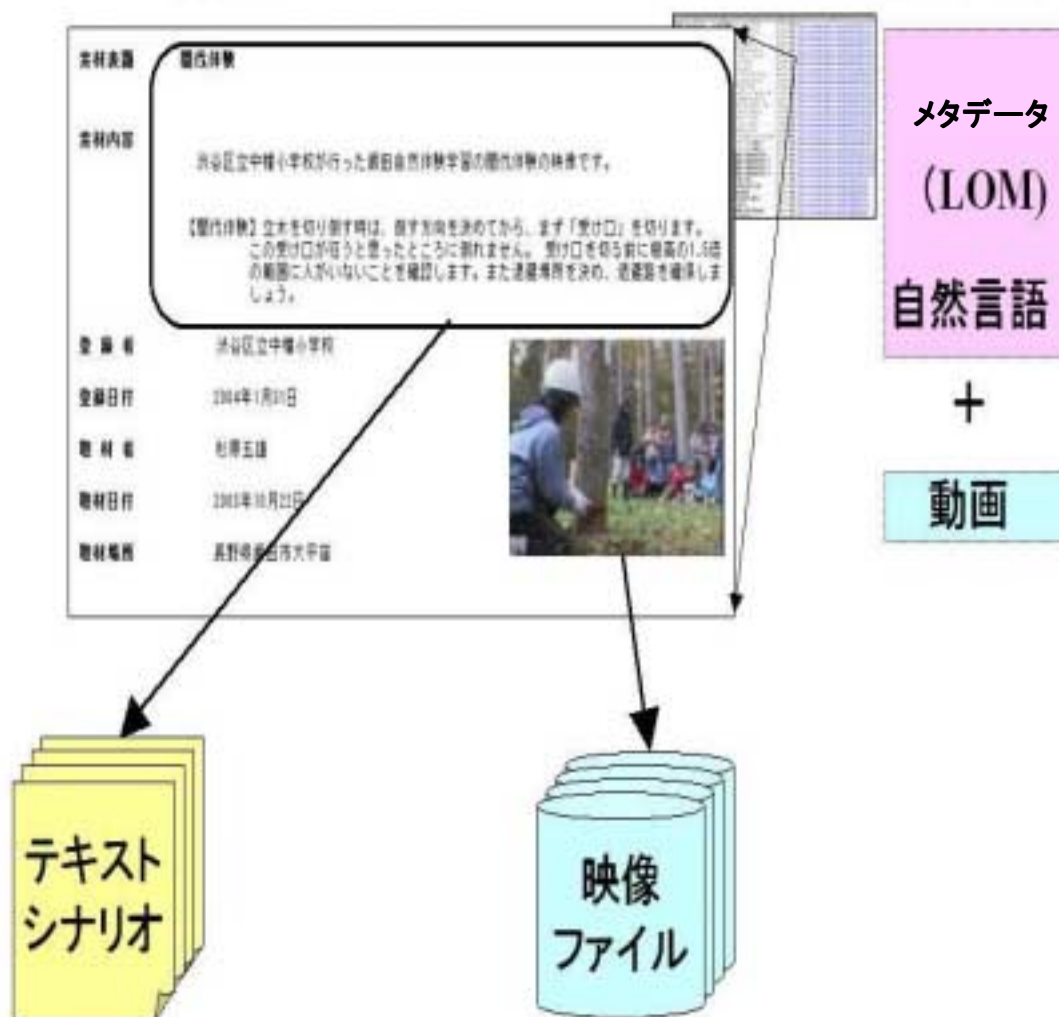


図 2.1 動画知、自然言語知とメタデータの3項組みの知識表現

「デジタルアーカイブ構想」は、人類共通の課題である文化遺産の保存問題解決に向けて、デジタル信号処理やマルチメディア・データベース等の最新技術を積極的に活用することにより、

[ 1 ]有形・無形の文化資産を記録精度が高く、再現性に優れたデジタル情報の形で記録し、

[ 2 ]各所ごとに蓄積、保管して、誰もが自由に閲覧・鑑賞できるようにするとともに、世界規模の情報通信ネットワークを利用して情報の受発信を行うことにより、次の世代に正しく継承することを目的とするものである。

本論の3項組知識表現によるデジタルアーカイブで使用する動画は通常1秒当たり32フレームからなる。つまり1秒間に32枚の静止画像を観客に順次見せることにより、連続した映像が完成する。

## 2.3 メタデータ “ LOM ”

メタデータとは「データに関する情報を記述したデータ」という意味で、データを説明するためのデータである。例えば、本をデータとした場合、その本のタイトルや概要、著者名などが記述されているラベルがメタデータとなる。

LOM(Learning Object Metadata)とは IEEE により 2002 年に認証された学習オブジェクトに関するメタデータの構造と語彙指定の仕組みを定義する規格であり、96 項目が定められている。例えばあるデータについてタイトル、作成者、学習指導要領などの項目を定めることでデータの内容を表現する。LOM の規格概要を図 2.2 に示す。

ID	タイトル	概要	URL
000000000004	自然共生型体験学習	「総合的な学習の時間」などで実践	<a href="http://www.gakujoken.or.jp/lida/index.html">http://www.gakujoken.or.jp/lida/index.html</a>
000000000101	事前準備	「総合的な学習の時間」と自然体験	<a href="http://www.gakujoken.or.jp/lida/video/iidaBlock01.wmv">http://www.gakujoken.or.jp/lida/video/iidaBlock01.wmv</a>
000000000102	出発式・きのご狩り	出発式からきのご狩りまでの映像で	<a href="http://www.gakujoken.or.jp/lida/video/iidaBlock02.wmv">http://www.gakujoken.or.jp/lida/video/iidaBlock02.wmv</a>
000000000103	開村式	大平到着から開村式までの映像です	<a href="http://www.gakujoken.or.jp/lida/video/iidaBlock03.wmv">http://www.gakujoken.or.jp/lida/video/iidaBlock03.wmv</a>
000000000104	夕食準備	夕食準備の映像です。	<a href="http://www.gakujoken.or.jp/lida/video/iidaBlock04.wmv">http://www.gakujoken.or.jp/lida/video/iidaBlock04.wmv</a>

図2.2 LOM データ

LOM とはe-learning 用メタデータであり、学習オブジェクトの構造と語彙指定の仕組みが定義された規格である。データベースとして、ID・タイトル・URLを使用し、検索結果としてID・タイトルを表示する。

LOM 項目のURL から検索対象のe-learning 用Webページを参照することができ、本システムでは、このURL よりWeb ページをダウンロードし、素材内容をテキストシナリオとして抽出し、データベースの作成を行う。また同時に、Web ページのサムネイルを抽出し、SVM による適合性フィードバックの判断の際に一覧の閲覧を可能にし、画像の暗黙知による直感的な判断を可能とする。図2.3 のLOM の項目には、対象範囲を特定する項目が盛り込まれている各項目(対象年齢など)には、適合ならt、不適合ならfがつけられている。

これらの項目を利用することにより、対象範囲を絞り込むカテゴリ検索が可能となる。このような検索を導入することにより、テキストシナリオからでは抽出できないようなLOM 項目からの確定された情報による検索が可能となる。

幼児教育	初等中等教育	高等教育	職業教育	生涯学習	高齢者学習	特別支援教育	教育分野その他	小学校入学前	小学校1年生	小学校2年生	小学校3年生	小学校4年生	小学校5年生	小学校6年生	中学校1年生	中学校2年生	中学校3年生	高校・高专	大学	大学院	20歳代	30歳代	40歳代	50歳代	60歳以上	対象年齢その他	
t	t	f	f	f	f	f	f	t	t	t	t	t	t	t	f	f	f	f	f	f	f	f	f	f	f	f	f
f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	t	t	t	t	t	t	f

図2.3 LOM データ(対象範囲を特定する項目の例)

## 2.4 Web ページ

図2.4に示すように、検索対象となるWeb ページは、自然言語であるテキストシナリオとテキストシナリオの内容を表す画像データをもつ。このテキストシナリオは画像の内容が記されており、このテキストシナリオを抽出しデータベース化することにより検索を可能とする。また、画像ファイルに含まれる暗黙知を利用してe-learning に活用するとともに、フィードバック検索の際に提示することにより画像に含まれる暗黙知をフィードバックに生かすことが可能となる。

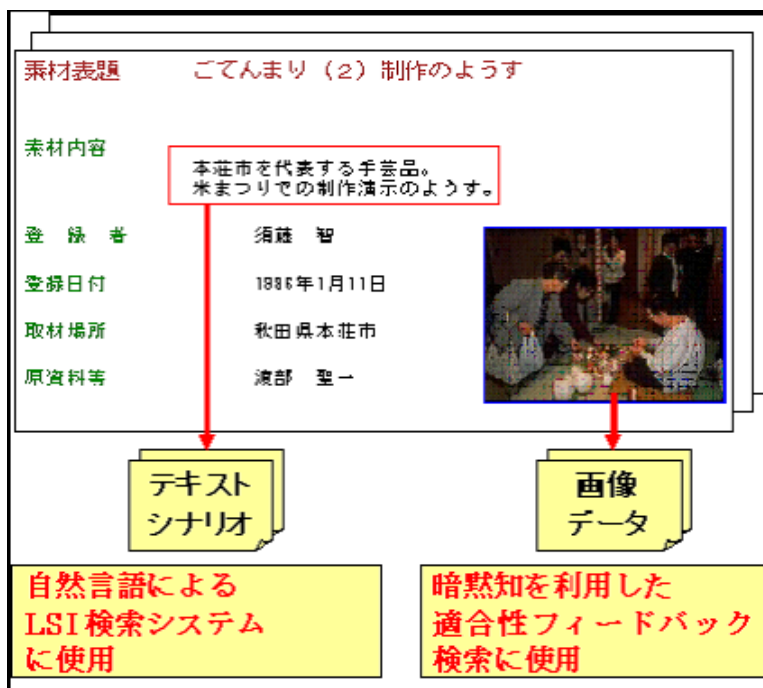


図2.4 Web ページとデータベース

## 2.5 LOM の RDF/RDFS 表現

現在、研究開発中の LOM の RDF/RDFS 表現について言及する。RDF とは、XML ベースの構文を持ちながら意味情報を表現でき、人工知能の意味ネットワークの知識表現

を継承し、主語、プロパティ、目的語の3つ1組で表現する意味モデルである。RDFSはRDFの語彙を定義する際に用いる。LOMをRDFで表現する場合は主語を教材デジタルアーカイブのURL、プロパティを各LOMの各項目、目的語に各項目に対応した内容に応じて記述する。

## 2.5.1 プロパティの記述

プロパティを記述する場合には主に Dublin Core[1]を用いる事例が多い。Dublin Coreとは書誌情報を記述するために定められたメタデータで、15の基本要素と多くの拡張要素からなる。LOMの項目にも類似したのが多くあることから、本研究でもID、タイトル、概要などの複数の項目において、Dublin Coreのメタデータを利用しトリプルで表現している。図2.5にDublin Coreを用いてタイトルの項目を表現したものを示す。

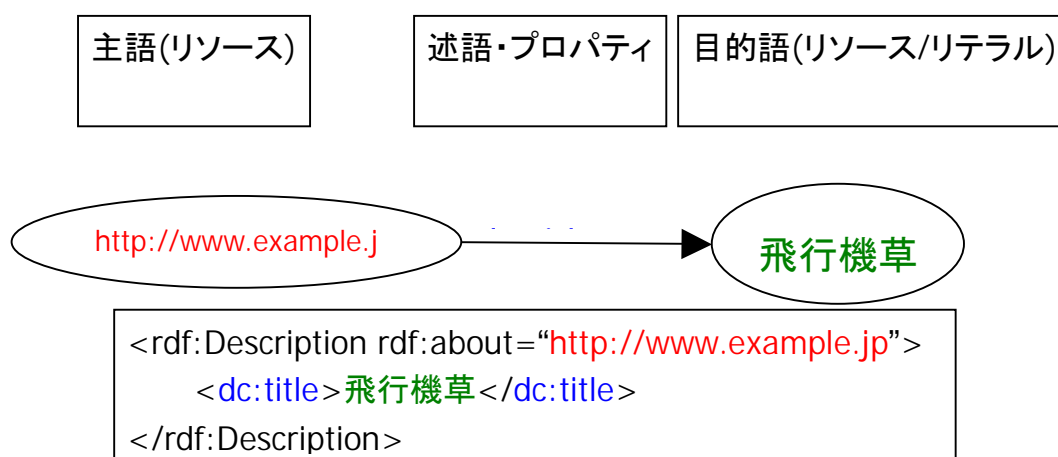


図 2.5 Dublin Core を用いた RDF 表現の例

## 2.5.2 GENES社会とLOM

現在、国の e-Japan 戦略では、世界最先端の IT 国家の実現に向けた「IT新改革戦略」を策定し、教育面で 学校教育の情報化の推進、情報モラル教育の推進、教育用コンテンツの整備と活用、教員評価等の施策を打ち出している。

一方で、e-Japan 戦略の「教育用コンテンツの整備と活用」を推進している NICER (教育情報ナショナルセンター)は、2001年8月に開設し、2005年度各種の教育用コンテンツや教育支援情報数を2万件増加させることを目指している。

他方、(財)学習ソフトウェア情報研究センターのマルチメディア教材研究会 GENES (Gakujoken Network Studying group)が、全国の教師等が提供した教育用コンテンツを11年以上にわたり蓄積し、LOM情報を生成し、活用している。前述のGENES社会とNICERの教育用コンテンツは、メタデータとして共通のLOMを採用している。

## 第3章

# 潜在的問題分析支援システムによる知識獲得

### 3.1 はじめに

本章では、日本で始めて開発されたKT法による潜在的問題分析システムを評価すると共に、本潜在的問題分析システムを利用した新しい知識獲得手法について考察する。

本論文のe-Learning用Webページ検索システムを応用し、複数人により、Web上の動画知、自然言語知、掲示板やチャット等といったWebテクノロジーを多角的に活用し、潜在的問題分析システムによる知識獲得の手法を紹介する。

「暗黙知・形式知を用いたeラーニング用Web検索エンジン」[23]と、ビジネス社会で効力を発揮しているKT法とを援用し、身近な問題を解決したときに発生する、近未来の危険に対して対策を講じるという潜在的問題分析の支援について論じる。

背景には、一方では2006年1月19日発表の国の「IT新改革戦略」の施策があり、他方では学習指導要領の目標の一つともなっているように、ネット社会を生きる子どもたちには、情報を正しく選択・判断し活用できる情報活用能力とともに、確固とした倫理感が必要といわれている[22]。

一方のe-Japan戦略「IT新改革戦略」では、急速な社会のIT化の進展や、インターネット上の違法・有害情報に起因する問題を指摘し、子どもの頃から情報内容を判断できる能力等が必要と述べている。具体的には、教員一人に一台のコンピュータ・ネットワーク環境やサポート体制整備等の学校のIT化、IT指導力の評価など教員のIT活用能力の向上、自ら学ぶITを活用した学習機会の提供、小学校段階からの児童生徒の情報モラルを含む情報活用能力の向上といった施策目標を打ち出している[21]。

他方、小中高等学校の学習指導要領・第1章総則に「学校の教育活動を進めるに当たっては、各学校において、児童に生きる力をはぐくむことを目指し、創意工夫を生かし特色ある教育活動を展開する中で、自ら学び自ら考える力の育成を図るとともに、基礎的・基本的な内容の確実な定着を図り、個性を生かす教育の充実に努めなければならない」とあり、「自然体験やボラ

ンティア活動などの社会体験、観察・実験、見学や調査、発表や討論、ものづくりや生産活動など体験的な学習、問題解決的な学習を積極的に取り入れること」とある。教科でも、例えば高等学校・普通教育・教科「情報B」では、「問題解決においてコンピュータを効果的に活用するための科学的な考え方や方法を習得させる」が目標の一つとなっている[12][13]。

本論文では、上記の現状認識のもとに、創造とは問題の発見と解決だという立場から、自然言語、動画とメタデータの3項組の知識表現を用い e ラーニング用 Web 検索エンジンを用い、自然体験やボランティア活動などで発生する潜在的問題分析を支援する一手法を提案する。

第2章のKT法の潜在的問題分析法では暗黙知[5]から形式知を得る「表出化」[6]の知識獲得手法を説明する。第4章 e ラーニング用 Web 検索エンジンの潜在的問題分析への適用では、本知識表現の創造技法支援での知識の利用の仕方について述べる。

本研究では、自然言語、動画とメタデータの3項組による知識表現を用い、暗黙知から形式知を引き出す知識の表出化の一手法である潜在的問題分析支援システムを新規に提案する。なお、本研究では、マイケル・ポランニーと同様に、暗黙知を言葉では説明できないが、理解して使っている知識と考えている。

KT法とは社会心理学者ケプナー (C. H. Kepner) と社会学博士トリゴー (B. B. Tregoe) によって開発された思考法であり、「合理的・効率的な意思決定手順の体系化と組織活動への適用」を目的としており、世界で最も使われている意思決定のための創造技法のひとつである。その分析プロセスは、状況分析 (Situation Appraisal; 以下 SA)、問題分析 (Problem Analysis; 以下 PA)、決定分析 (Decision Analysis; 以下 DA)、潜在的問題分析 (Potential Problem Analysis; 以下 PPA) という四つに分類される。とりわけ、潜在的問題分析はケプナー・トリゴーのモデルの中でも最も定式化が困難であるといわれている[12][16]。

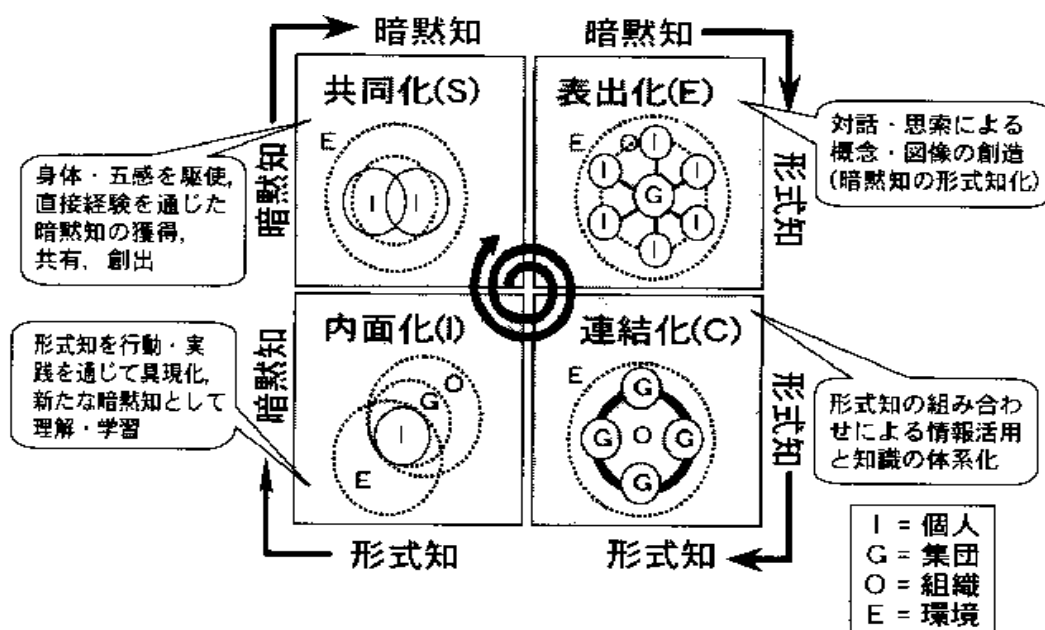


図 3.1 SECIモデル[6]



潜在的問題分析は、図 3.1 の野中の S E C I モデルの暗黙知から形式知を引き出す知識の表出化の一手法といわれている[16]。さらに、図 3.1 の共同化から表出化のプロセスにおいて、國藤は「人間のもっている膨大な暗黙知から形式知を取り出す表出化(分節化)のプロセスを支援する発想支援ツールを実現するには発散思考、収束思考、アイデア結晶化のプロセスを支援するツール構築が必要」と述べている[7]。

従来、K T法の技法の一部を用いて開発したシステムには問題分析(P A)を導入した八重樫等[13]や中村等が開発したシステム [15] や、決定分析(D A)を導入した佐藤等が開発したシステム [14][17] が存在するが、潜在的問題分析(P P A)を用いた例はまだない。

そこで本研究では、P P Aによる意思決定支援システムを Web アプリケーションとして開発して評価する。また、イーラーニング(e-Learning)の研究 [18][20] との関連を考え、学校教育や社会教育への試験的な利用と考察する。

本研究のもたらす効果は、自然言語、動画とメタデータの3項組による知識表現を用いた潜在的問題分析支援システムの開発によって、暗黙知から形式知を引き出す知識の表出化の一手法が明確化し、新しい知識獲得手法を入手することである。

### **3項組知識表現<動画, 自然言語, メタデータ>**

動画や静止画は、文章に追加して、自分や他人のコンテキスト(状態)を理解する手段として有効である。

図 3.2 は、本章で述べる新しい知識獲得手法によって得られた「ボーツと光る星座帽を作ろう」という知識の例である。

新規に得られた知識は、図 3.2 に示すような匠の技のノウハウなどを記録した動画、素材内容を説明する自然言語と、IEEE 準拠の学習オブジェクトメタデータ LOM で記述したメタデータという3項組による知識表現で表記されている。

図 3.2 に示すように、「ボーツと光る星座帽を作ろう」という知識の動画は映像ファイルに記憶され、「ボーツと光る星座帽を作ろう」、「材料」、「星座図」・・・「天球儀を展開した星座図をもとにして、帽子の形をした立体的な星座早見帽を作ります。・・・」や「写真をクリックすると、ムービーを見ることができます」などの文章はテキストシナリオとして記憶される。

本知識表現を用いる効果は、自然言語のみの知識より、視覚・聴覚による暗黙知の一部を活用でき、その結果、テキスト検索結果に対し適合性フィードバックを容易にするという長所がある。

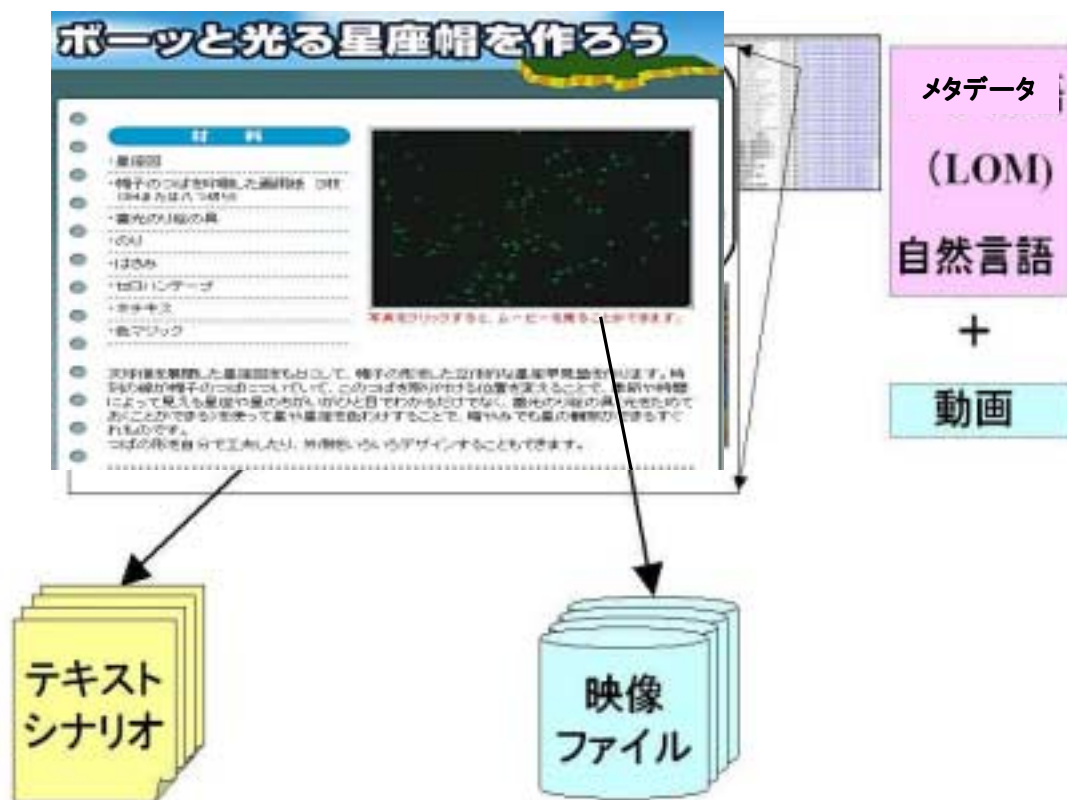


図 3.2 自然言語、動画とメタデータの 3 項組みの知識表現

## 3.2 潜在的問題分析支援システム

潜在的問題分析支援システムでは、KT 法の潜在的問題分析 ( P P A ) を利用して意思決定を支援する。

### 3.2.1 潜在的問題分析の手順

P P A は以下に示す手順で行われる。

- (1) 潜在的問題をリストアップする。
- (2) 潜在的問題が起こる原因は何かを考える。
- (3) 潜在的問題が起こる可能性を考える。
- (4) 未然の予防処置は何かを考える。
- (5) 潜在的問題が発生したらどうするか、  
対緊急処置を考える。

これをチャート化すると、表 3.1 のような型ができる。これに準じた分析入力フォームを設け、複数人の入力した様々な考えが表示されるようなシステムとすれば、他者の考えを見ながら自分の考えを再構築できるブレインストーミング ( Brain Storming ) 効果をもたらす。図 3.1 の野中による知識創造のモデル ( S E C I モデル ) によれば、「表出化作業とは、自己の暗黙知を他人にもわかる形式知へ

転換して表出させる作業」という。

潜在的問題： ...			
原因と考えられるもの	可能性/重大性 (%)	処置	
		予防処置	対緊急処置
1. ...	...	A) ...	A) ...
2. ...	...	B)...	B)...

表 3.1 P P Aチャート

表 3.2 は、「星座を光らせ天球儀で実際に夜空の星座を教えたい」という実行案に対する P P Aチャートの適用事例である。

実際の星空観察映像（動画）から判明した潜在的問題は「実際の星空観察で使えない」ことであった。

ブレインストーミングの結果、原因には 「実際の凹面の星座と凸面の天球儀が合わない」と 「一人で操作できない」の2項目が考えられた。また、対緊急処置として、「事前学習に天球儀として利用する」や 「グループ学習に用いる」が考えられた。最終的な予防処置としては、「凹面の星座天球儀を作る」や 「星座早見帽子を作成し利用する」という新しい知識が獲得された。

潜在的問題：実際の星空観察で使えない			
原因と考えられるもの	可能性/重大性 (%)	処置	
		予防処置	対緊急処置
1. 実際の凹面の星座と凸面の天球儀が合わない	80	A) 凹面の星座天球儀を作る	A) 事前学習に天球儀として利用する
2. 一人で操作できない	20	B) 星座早見帽子を作成し利用する	B) グループ学習に用いる

表 3.2 P P Aチャートの適用事例

図 3.3 に示すように、新たに各人共通の分析入力フォームを Web 上に用意し、ブレインストーミングの結果を穴埋め形式で埋めていき、「星座早見帽子を作成し利用する」等の暗黙知を表出させる。

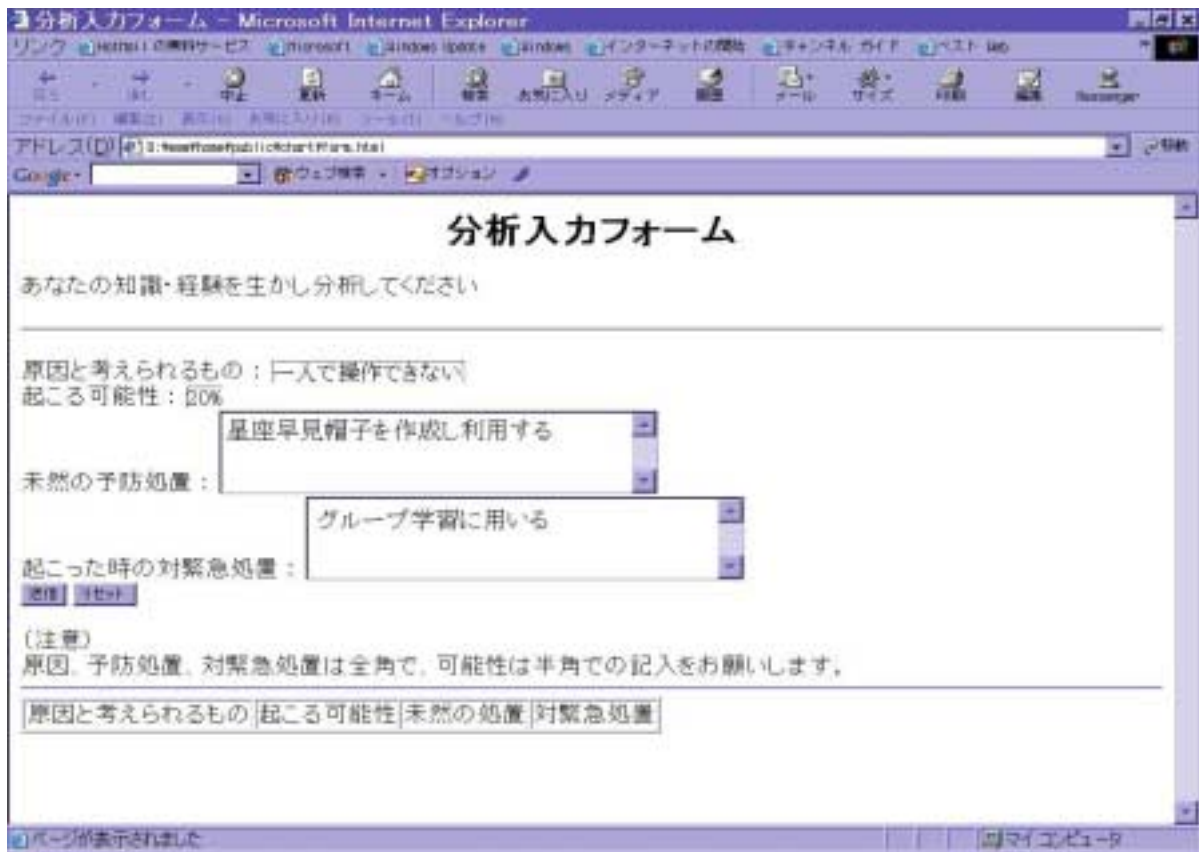


図 3.3 分析入力フォーム

### 3.2.2 意思決定支援の手順

以下に示す手順で意思決定支援を行う。

(1) テーマの投げ掛け

あるユーザが、「もしこの案を実行に移すとしたらどのようなリスクがあるのか」という考えのもと分析入力フォームへの書き込みを行う。

(2) 議論

システムを閲覧したユーザが、問題について考えたことを分析入力フォームに書き込んでいく。数人で書き込む場合、書き込み内容が曖昧だったり矛盾したりするそのような他者との意見の相違を解決するために議論するための BBS(掲示板)[20]を設置する。

(3) リスクの列挙

議論を繰り返すことにより、ユーザに様々なリスクがあることが示され、列挙されたリスクの最大値を使うという P P A の経験則に従った処理がなされていく。

(4) システム画面

図 3.4 は、BBS(掲示板)を含めたシステム画面である。図 3.5 は、本潜在的な問題分析システムに投稿されたテーマ例である。



図 3.4 BBS(掲示板)を含めたシステム画面

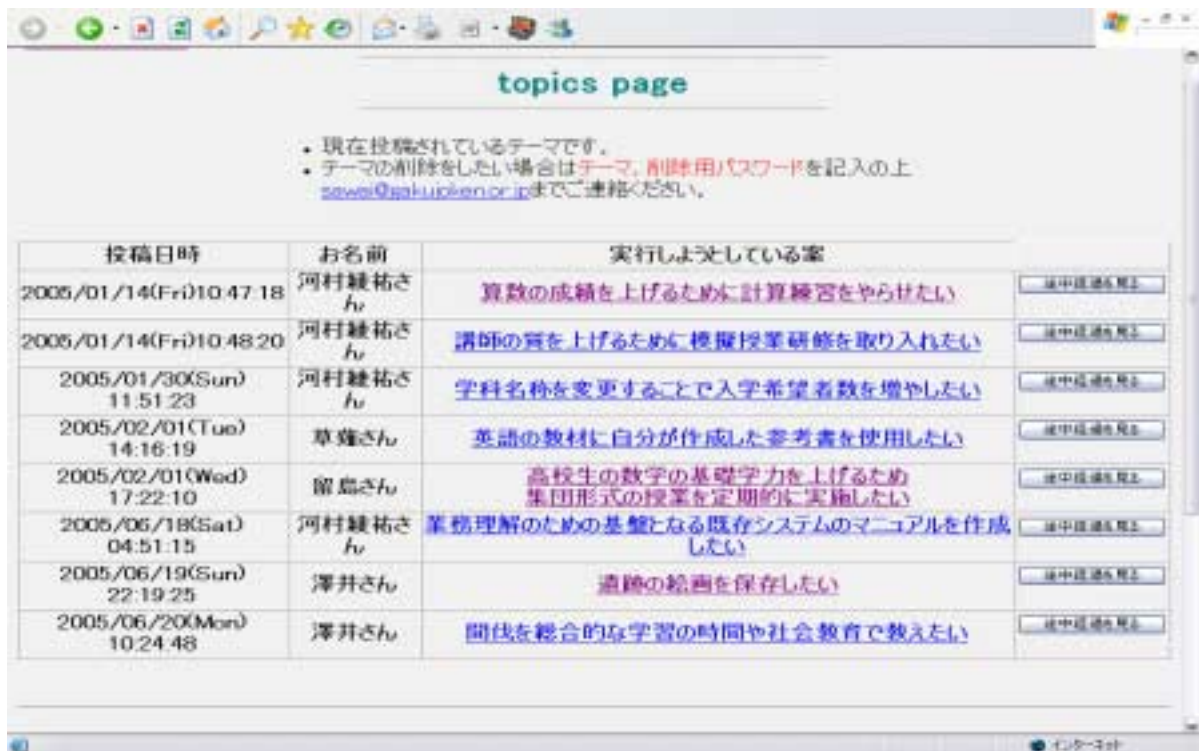


図 3.5 投稿されたテーマ例



### 3.3 e ラーニング用 Web 検索エンジンの潜在的問題分析への適用

図 3.6～図 3.8 は、投稿テーマ「間伐を総合的な学習の時間や社会教育で教えた」を選択し、潜在的問題分析を行い、暗黙知を形式知化する例である。

潜在的問題	原因	可能性	処置	対緊急措置
周囲の器物・家屋を破損する	伐採方向がずれる	80 (9%)	樹木を倒す前には十分時間をかけて安全を確認する	樹木を倒す前に大声で周囲へ知らせる
大ケガをする	身体の一部を倒木に乗られる	30 (9%)	樹木を倒すとき危険な場所を避ける	ヘルメットをはじめ必要な安全具を着用する
間伐がうまくいかない	伐採方向を確認していない	30 (9%)	伐採方向に人や邪魔な枝木がないようにする	重機を用い、伐採方向に、倒す木の高さと比較して1.5倍以上の長さのロープで引っ張る
間伐がうまくいかない	受け口が狂ってしまう	80 (9%)	受け口の深さは倒す木の直径の4分の1以上とします。	大きな木や重心が分からない木の場合は、必ず「くさび」を2本以上使用する

潜在的問題 (全角30文字以内)  
  
 原因 (全角30文字以内)  
  
 可能性 (0～100 半角3文字以内)  
 %  
 処置 (全角100文字以内)  
  
 対緊急措置 (全角100文字以内)

図 3.6 間伐に関する潜在的問題分析処理例



図 3.7 BBS(掲示板)画面のイメージ図

図 3.7 は、暗黙知を複数人で共有し、議論をする BBS(掲示板)の例である。図 3.6 の「議論をしたい!」をクリックすると、図 3.7 の BBS 画面にたどり着く。

図 3.7 の掲示板では、4 章で述べる Web ページ検索システムなどを使い、任意の自然言語で検索し、関連する自然言語、動画とメタデータの 3 項組の知識表現データを参照する。“自然言語知”である説明テキストと、動画を見ることにより、知識の表出化がより容易になった。

理由は、従来のテキストのみ用いる表出化手法と比較して、動画による疑似体験によって対話や思索を促したものと考えられる。

具体的には、潜在的問題分析のフォーム (PPA FORM) は書きやすく、文章化しやすいという特長がある。図 3.8 は、潜在的問題「間伐がうまくいかない」に関する潜在的問題分析の途中経過画面である。

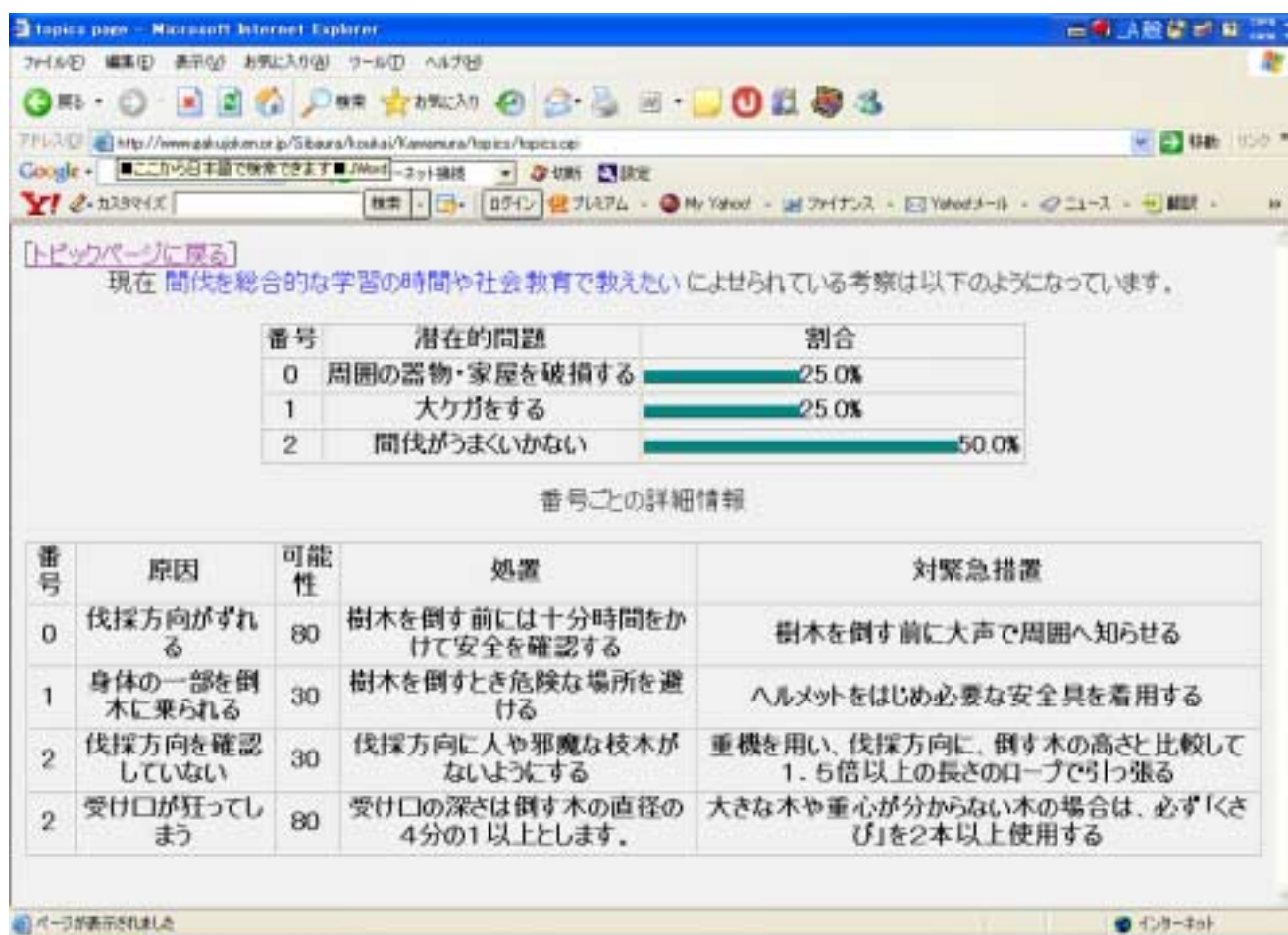


図 3.8 間伐の途中経過画面の例

図 3.9 は、「間伐を総合的な学習の時間や社会教育で教えたい」を選択し、暗黙知を形式知化する潜在的問題分析の処理の流れ図である。

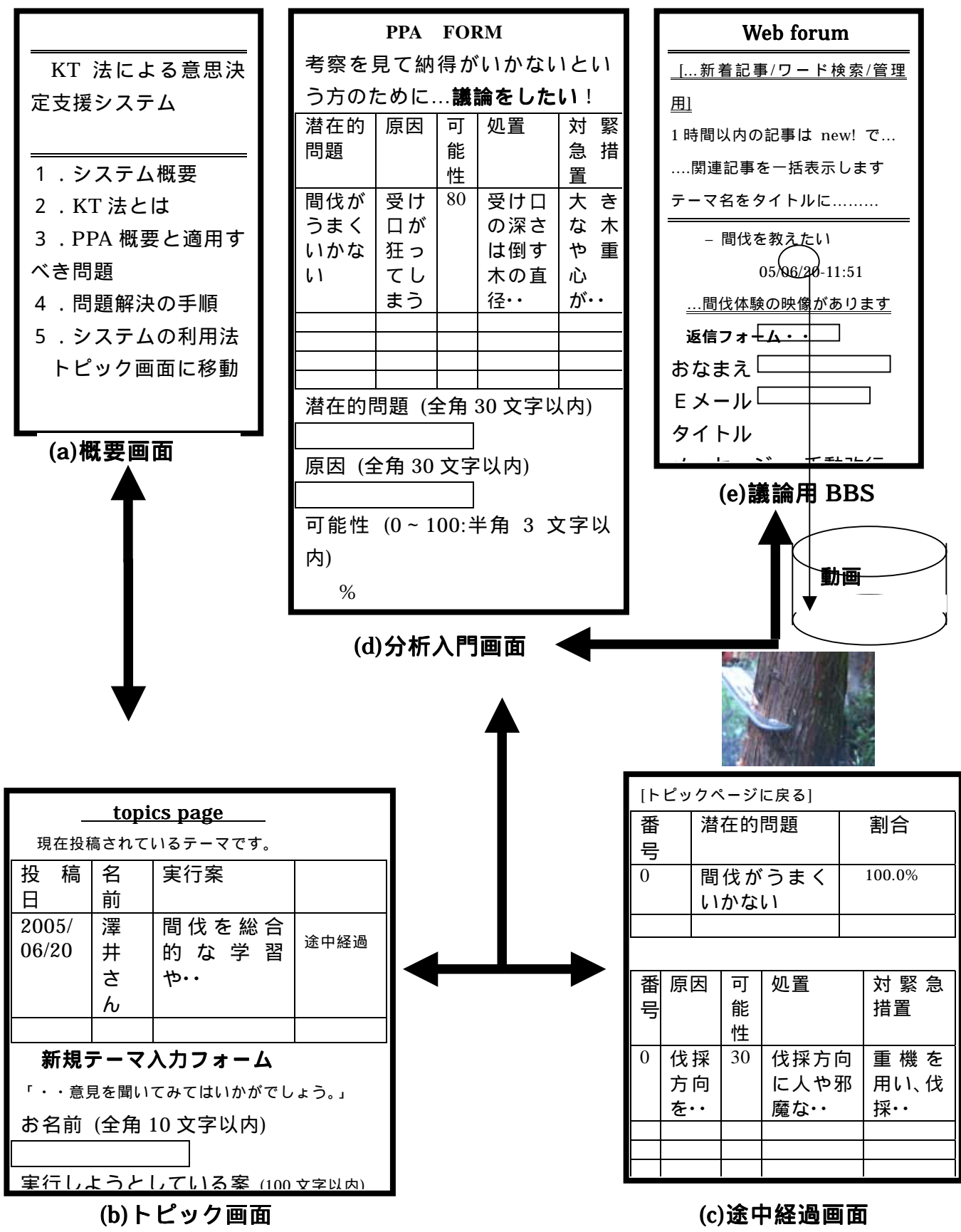


図 3.9 潜在的問題分析の処理の流れ図



### 3.4 潜在的な問題分析システムの評価

本システムの評価は、図 3.5 に示すように、「A 大学 B 学科の入学希望者数を増やしたい」などという 8 事例について各ユーザに P P A の流れに沿って分析を行った。その結果と、事前の紙ベースの P P A チャートを適用し分析した結果とを比較・検討して行う。

本システムはコンピュータが独自に分析を行うわけではなく、公開型 Web アプリケーションという性質上、複数のユーザが個々に主観的に分析を行う。また、本システムの試用調査に参加した各ユーザに対して無記名でアンケートを行い、本システムを評価した。ただし、評価点は表 3.3 の中で平均が示されている。

【評価項目】	【評価点】
「概説・適用事例は有効であったか」	平均6.7点
「他者の分析は有効であったか」	平均7.2点
「議論用 BBS は有効であったか」	平均6.9点
「その他本システムに対する意見」	小さいフォント KT法の色

表 3.3 試用調査による評価結果

評価項目は、

「概説・適用事例は分析を行う上で有効であったか」

「途中経過や他者の分析を見ながら分析をすることは、自分が分析する上で有効であったか」

「議論をするための BBS は有効であったか」

「その他本システムに対する意見」

とし、 、 、 に関しては 10 段階の点数で評価し、 に関しては自由記述とする。表 3.3 は、A 大学の学生 20 名による試用調査による評価結果である。

、 については、6.7点、7.2点と5点以上の評価点を得ることができた。 は当初使い方が分からず、BBSが雑談に近い形で利用されたため、評価点が低かった。Webページ検索などBBSの使い方を議論の例示で教えた後では、評価が高くなり、平均6.9点の評価点を得ることができた。今後、Webページ検索などBBSの使い方を典型的な議論の例で教えることが重要と考える。

のKT法の色や、フォントはソースがHTMLなので改善は比較的容易であった。

### 3.5 まとめ

本論文では、創造とは問題の発見と解決だという立場から、暗黙知・形式知を用いたeラーニング用Web検索エンジンを活用し、この知識表現法の有効性を確認するため、K T法における潜在的な問題分析 ( P P A ) という創造技法支援に応用し、初めて潜在的な問題分析をW e b上で公開した。

無記名のアンケート調査を実施し、P P A 支援インターフェースでの評価実験を行った。そのため、匠の技のノウハウなどを記録した動画、素材内容を説明する自然言語と、IEEE 準拠の学習オブジェクトメタデータ LOM で記述したメタ言語を 3 項組でデジタルアーカイブ化する手法を用いた。

本知識表現の知識利用例として、Web 上の検索エンジン、掲示板やチャットなどといった Web テクノロジーを多角的に活用し、Web ページ検索システムにも利用できることを確認し、他の検索システムとの対照実験を行った。

本研究の成果は、暗黙知から形式知を引き出す「表出化」の一つのツールが得られたことと、本ツールを使って新しい知識が獲得できたことである。

## 第4章

# 3 項組知識表現を用いたeラーニング用 Web検索エンジン

### 4.1 はじめに

本章では、第2章において提案した3項組の新しい知識表現を、知識利用に応用し、独創的なWebページ検索システムを新たに開発した。

本システムは、デジタル化された知識を管理・活用する立場から、自然言語知である「形式知」と、動画知に一部含まれる「暗黙知」と、メタデータの3項組の新しい知識表現をeラーニングに応用したものである。

教師や匠の技のノウハウ等を記録した動画知と、自然言語知と、IEEE準拠の学習オブジェクトメタデータLOMで記述したメタデータの3項組で教材検索を行うeラーニング用Webページ検索システムである。

以下では、eラーニング用Webページ検索エンジンの特長と、自然言語によるWebページ検索について述べる。

### 4.2 Web ページ検索エンジンの特長

図4.1に示すWeb ページ検索エンジンは、以下の特長を持つ。

- ・ 動画知を利用したSVM による適合性フィードバック
- ・ 自然言語によるLSI 検索とLOM カテゴリ検索とのハイブリッドなWeb 検索システム
- ・ データベース構築プログラム(データの追加など)の実装

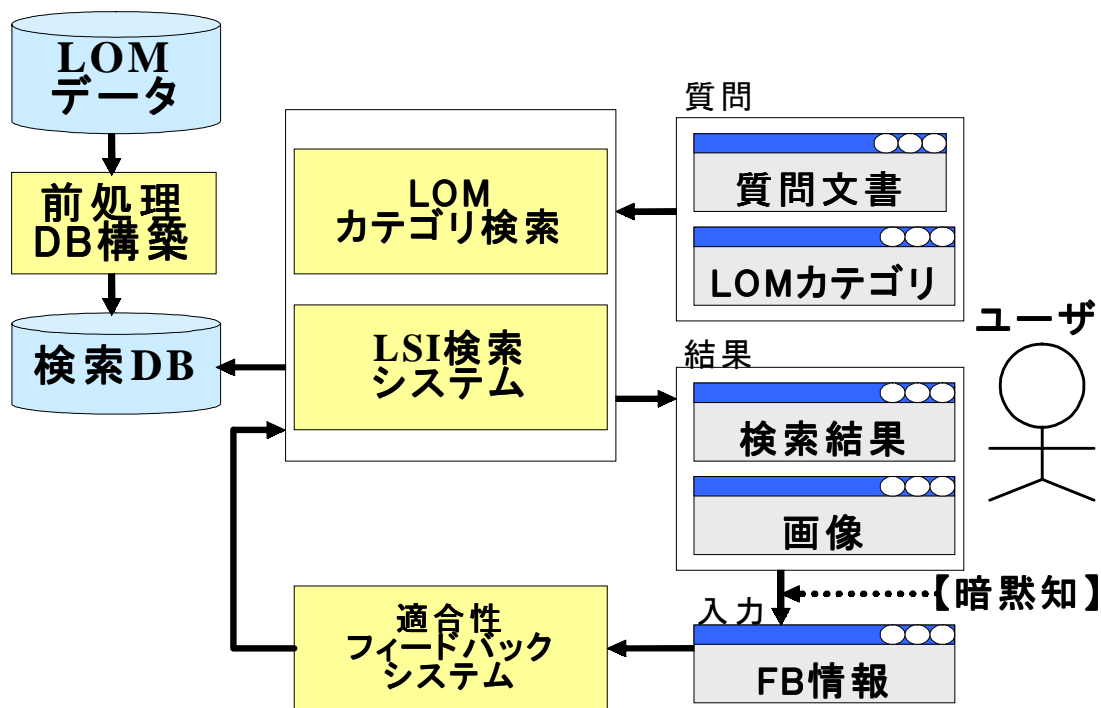


図4.1 Web ページ検索エンジンの概要

#### 4.2.1 暗黙知を利用した適合性フィードバック

直感的に現象を表現している映像は形式知としての文章よりも直感的判断が可能である。本システムでは、検索結果に対してユーザの判断を取り入れる適合性フィードバックを適用し検索結果の改善を行う。この判断を行う際に、画像データを提示することにより、動画知による直感的な判断が可能になる。

#### 4.2.2 自然言語(形式知) によるWeb ページ検索

検索質問に自然言語を用い質問と類似したページを検索することで、一般性をもたせ、よりユーザの意図を反映することができる。また、名詞はWeb ページのテキストシナリオ特徴を表すとして着目し、名詞を索引語とし索引語文書行列を構築することにより、ベクトル空間上で表現が可能となる。ベクトル空間上での類似度を利用することにより、似ているという概念での検索が可能となる。また、検索手法としてLSI 検索を用いることにより、類似語の処理を自動で行うことができ、その際類似語辞書なしで実現できる。

#### 4.2.3 LOMカテゴリ検索

LOM とは、2章で述べたように学習オブジェクトメタデータのことで、メタデータ

として指定する全96項目を規定し、項目ごとに細かく記入制限が設定されている国際規格である(図4.2参照)。任意の文字列を書く項目や複数の選択肢の中から該当するボキャブラリ(語彙)を選択する項目があり、メタデータの項目は階層化されている。LOMデータベースはリレーショナルデータベースであり形式はEXCEL データでなければならない。本システムの検索において、LOM を使用することにより、インターフェースとしてデジタルアーカイブの検索を可能にするまた、これによりNICER(教育情報ナショナルセンター)への検索システムへ利用可能となる。

全96項目

ID	タイトル	...	価格	...	言語
⋮	⋮		⋮		⋮

図4.2 LOM 規格概要

### 4.3 自然言語(形式知)をWebページ検索システムに利用する

本研究で検索対象とするWebページは、形式知の自然言語によるテキストシナリオと、動画や静止画のデジタル映像ファイルを持つ。このテキストシナリオには映像ファイルの内容が記されており、これをデータベース化することで検索可能となり、共用を可能となる。

#### 4.3.1 本システムの特徴

検索は Web ページのテキストシナリオ部分からデータベースを作成し、潜在的意味インデキシング (Latent Semantic Indexing : LSI) の検索手法[3, 6]を使い、入力質問文章と類似の Web ページを検索する。

実際の検索にはインターフェイスとしてデジタルアーカイブの検索を可能にするメタデータ情報 (Learning Object Metadata: LOM) を使用する。これを利用することによって NICER (National Information Center for Educational Resources : 教育情報ナショナルセンター) [4]の検索システムが利用可能となる。

LOM データ

Web ページ

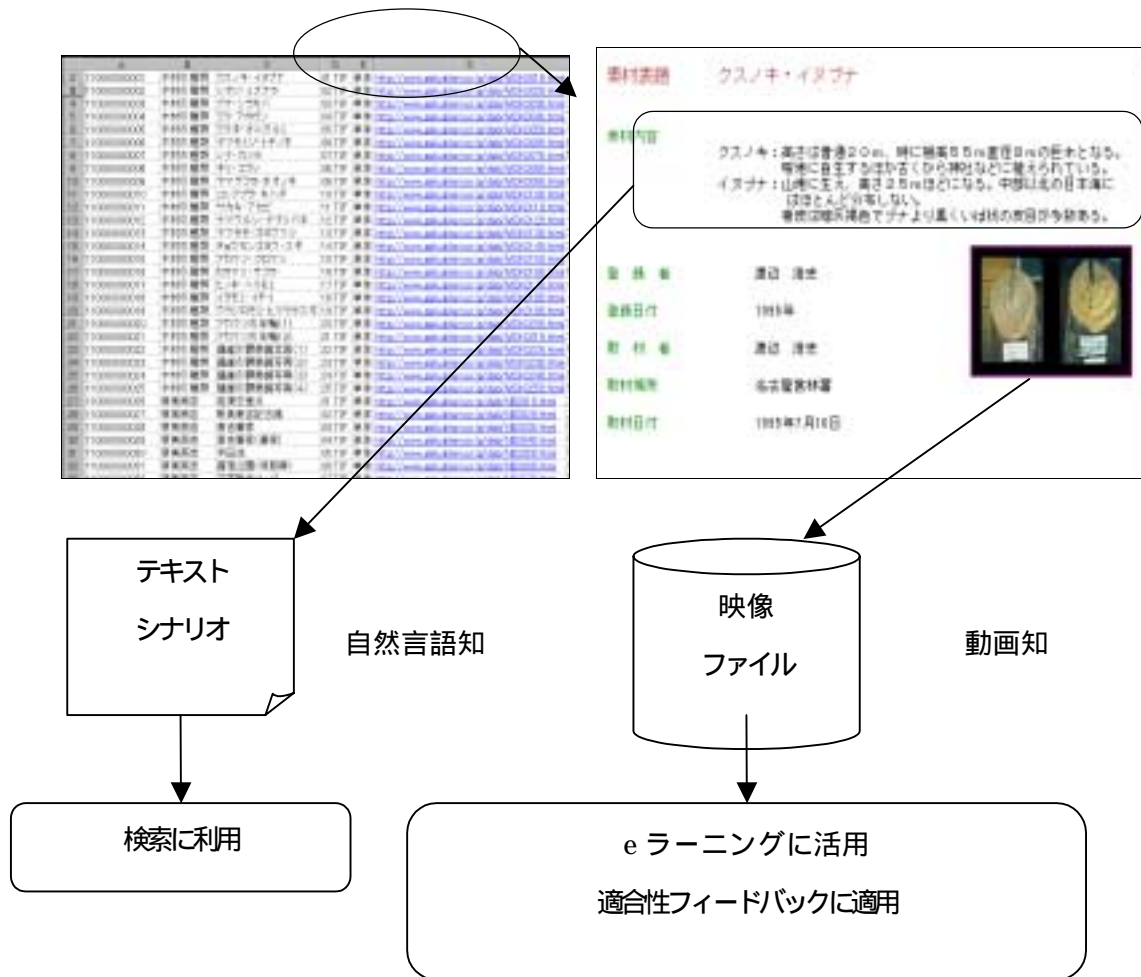


図 4.3 Web ページ検索システムの特徴

また、映像ファイルに含まれる動画知を利用して e ラーニングに活用するとともに、フィードバック検索を行う一助となる。

検索に用いる質問は自然言語による文章入力とする。これにより、一般性が高くなり、多くの利用価値を得られる。

### 4.3.2 本システムの概要

本検索システムは構成図を図 3 に示す。本検索システムでは、検索された結果の精度の良し悪しをユーザが判断し、検索結果から得たフィードバック情報 (FB 情報) を検索システムに与え、適合性フィードバックを行うために、「LOM データ」と「Web ページのダウンロード」が必要なる。

また、本検索システムは、NICER の検索システムに提供したのと全く同一の LOM データ「愛知」を利用しており、NICER の検索システムでも本検索システムでも、検索方式とか検索精度が異なるが、検索可能である。

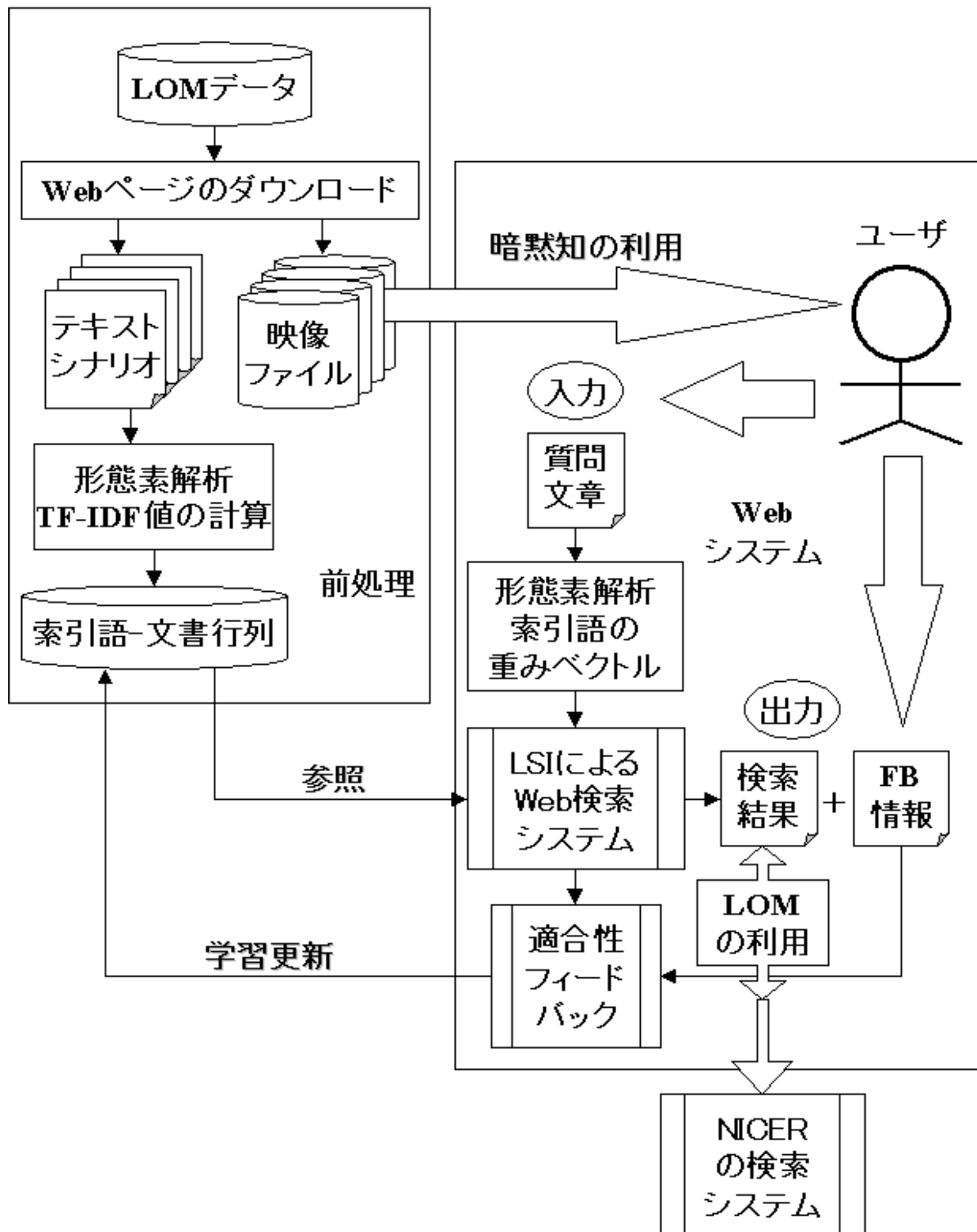


図 4.4 検索システム構成図

### 4.3.3 索引語-文書行列の構築

LOM データから ID ナンバー、タイトル、Web ページの URL などの必要な情報を取り出しファイルに格納するとともに、その URL からテキストシナリオをダウンロードし、形態素解析、TF-IDF 値の計算により索引語を抽出した後、索引語-文書行列を構成する。

#### (1) 形態素解析

本システムでは、Web ページにおいて名詞はそのテキストの特徴を現すものと想定し、Web ページの素材内容の部分をテキストシナリオとして抽出する。その作成されたファイル群に対して、日本語形態素解析ツール茶筌を用いて形態素解析を行い、テキストシナリオを形態素に分解し、その中から名詞のみを抽出し索引語候補としてファイルに保存する。この際、数字や代名詞などの文章を特徴づけないような一般的な名詞は stop list として削除する。取り出された索引語のリストをファイルに格納しておく。

ここで茶筌[7]とは、広く自然言語処理研究に資するため無償のソフトウェアとして開発されたものである。

#### (2) TF-IDF 法

TF-IDF 値は、情報検索における索引語としての単語の重みを表す値で、(i)一つの文書内に多く出現した単語はその文書の内容とより関係がある。(ii)より多くの文書に出現した単語は重要ではない。この2点を考慮しTF-IDFの値は決定され、本研究では下記の式を用いる。

$$tfidf(t, d) = \frac{tf(t, d)}{\sum_{s \in d} tf(s, d)} \left( \log \frac{N}{df(t)} + 1 \right) \quad (1)$$

$t$  は単語、 $d$  は文書、 $N$  は全文書数を表す。 $tf(t, d)$  は文書  $d$  の中での単語  $t$  の出現頻度で、 $df(t)$  は単語  $t$  が出現した文書数を表す。

TF-IDF 法による索引語選択では、 $tfidf(t, d)$  値がある文書  $d$  に対して、索引語候補から所与の閾値以上の単語  $t$  を索引語として選択する。

#### (3) ベクトル空間モデルと索引語-文書行列

本研究ではベクトル空間モデルを採用し、TF-IDF 法により選択された  $m$  個の索引語  $w_1, w_2, \dots, w_m$  から構成される  $m$  次元のベクトルで文書を表現する。Web ページのテキストシナリオである  $n$  個の文書を  $D_1, D_2, \dots, D_n$  とした時、各文書  $D_j$  に対応して文書ベクトル  $\vec{d}_j$  は、次のようなベクトルで表現され、これを文書ベクトル(document vector) と呼ぶ。

$$\vec{d}_j = \begin{bmatrix} d_{1j} \\ d_{2j} \\ \vdots \\ d_{mj} \end{bmatrix} \quad (2)$$

ここで、 $d_{ij}$  は索引語  $w_i$  の文書  $D_j$  における TF-IDF 値、つまり  $tfidf(w_i, D_j)$  である。これらの文書ベクトルの集合を次のような  $m \times n$  行列の  $D$  ( $m$ :索引語数,  $n$ :文書数) で表現することができ、これを索引語 文書行列(term-document matrix) と呼ぶ。



$$\begin{aligned}
D &= \begin{bmatrix} \vec{d}_1 & \vec{d}_2 & \cdots & \vec{d}_n \end{bmatrix} \\
&= \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix}
\end{aligned} \tag{3}$$

検索質問も、文書と同様に、索引語の重みを要素とするベクトルで表現することができ、検索質問ベクトル $\vec{q}$ は次のように表現される。

$$\vec{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{bmatrix} \tag{4}$$

ベクトル空間モデルでの文書検索では、質問ベクトル $\vec{q}$ と各文書ベクトル $\vec{d}_j$ の間の類似度 $sim(\vec{d}_j, \vec{q})$ を計算することにより行う。本研究では、類似度の定義としてコサイン尺度を用いる。

$$\begin{aligned}
sim(\vec{d}_j, \vec{q}) &= \cos(\vec{d}_j, \vec{q}) \\
&= \frac{\vec{d}_j \cdot \vec{q}}{\|\vec{d}_j\| \|\vec{q}\|} \\
&= \frac{\sum_{i=1}^m d_{ij} q_i}{\sqrt{\sum_{i=1}^m d_{ij}^2} \sqrt{\sum_{i=1}^m q_i^2}}
\end{aligned} \tag{5}$$

## 4.4 潜在的意味インデキシング(LSI)

潜在的意味インデキシング(Latent semantic indexing:LSI)は、索引語-文書行列に対して特異値分解を行い、特異値の小さい次元を圧縮して元の索引語-文書行列より低いランクの基底行列を求める。その基底に各ベクトルを射影することにより、文書ベクトルの次元を圧縮する方法である。

潜在的意味インデキシングでは、高次元の空間にある文書ベクトルを低次元の空間へと射影することにより、問題を改善する手法である。

ベクトル空間モデルでは、文書ベクトルの次元数は索引語の総数に等しい。検索対象の文書数が増えるに従い、索引語の総数も増え、文書ベクトルの次元数が増加する傾向にある。次元数が増加すると以下のような問題が出てくる。

- ・ 計算機のメモリによる制限
- ・ 検索時間の増大
- ・ 文書中に含まれる不必要な索引語による検索精度の低下

潜在的意味インデキシングでは、高次元の空間にある文書ベクトルを低次元の空間へと射影することにより、上記の問題を改善する手法である。また、LSI では低次元の空間へと射影する際、索引語間の意味的な関連図化を自動的に行うことができ、類義語処理用の類義語辞典を作成することなく数値的な処理により実現できる。潜在的意味インデキシングでは、特異値分解という技術を用いて高次元ベクトルの次元削減を行う。

## 4.5 特異値分解を用いた次元圧縮

索引語-文書行列を  $m \times n$  行列  $D$  とする。 $D$  の特異値分解は次のように定義することができる。

$$D = U \Sigma V^T \quad (6)$$

$U$  は  $m \times m$  直交行列 ( $UU^T = U^T U = I$ )、 $V$  は  $n \times n$  直交行列 ( $VV^T = V^T V = I$ ) であり、行列  $U, V$  の各列を左特異ベクトル(left singular vector)、右特異ベクトル(right singular vector) という。  $D$  は  $m \times n$  行列であり、 $\text{rank}(D) = r$  とすると、対角線上に  $r$  個の要素  $\sigma_1, \sigma_2, \dots, \sigma_r$  ( $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ ) が並んだ行列である。 $\sigma_i$  のことを特異値(singular value)と呼ぶ。

$$\Sigma = \left[ \begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ \hline & & & 0_{r \times (n-r)} & & \\ \hline & & & & 0_{(m-r) \times r} & \\ & & & & & 0_{(m-r) \times (n-r)} \end{array} \right] \quad (7)$$

索引語-文書行列の特異値分解では、 $U$  の左特異ベクトルが文書を最適に表現する正規直交基底、 $V$  の右特異ベクトルが索引語を最適に表現する正規直交基底となる。 $r$  次元の文書ベクトル  $\vec{d}$  を  $k$  次元の文書ベクトル  $\vec{d}^{(k)}$  で近似するためには、 $U_k$  の張る空間への射影を考えればよい。 $U_k$  は  $U$  の最初の  $k$  個 ( $k < r$ ) の左特異ベクトルのみの  $m \times k$  行列である。

$$\vec{d}^{(k)} = U_k^T \vec{d} \quad (8)$$

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1k} & \cdots & u_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mk} & \cdots & u_{mn} \end{bmatrix} \quad (9)$$

$$U_k^T = \begin{bmatrix} u_{11} & \cdots & u_{m1} \\ \vdots & \ddots & \vdots \\ u_{1k} & \cdots & u_{mk} \end{bmatrix} \quad (10)$$

このようにして、索引語-文書行列を低次元に圧縮することができる。

#### ・類似度の計算

ベクトル空間での文書ベクトル  $\vec{d}$  と質問ベクトル  $\vec{q}$  に対して、余弦  $\cos(\vec{d}, \vec{q})$  を取ることにより、類似度を求める。類似度  $\text{sim}(\vec{d}, \vec{q})$  は、 $k$  次元に圧縮した索引語-文書行列  $D_k$  を再構築することなく特異値分解の結果  $U_k$ 、 $\Sigma_k$ 、 $V_k$  から類似度を計算できる。

具体的には、ベクトル空間での文書ベクトル  $\vec{d}$  と質問ベクトル  $\vec{q}$  との類似度  $\text{sim}(\vec{d}, \vec{q})$  は次のようにして求めることができる。

$$\begin{aligned} \text{sim}(\vec{d}, \vec{q}) &= \cos(U_k^T \vec{d}, U_k^T \vec{q}) \\ &= \frac{(U_k^T \vec{d}) \cdot (U_k^T \vec{q})}{\|U_k^T \vec{d}\| \|U_k^T \vec{q}\|} \end{aligned} \quad (11)$$

行列  $D_k$  は、上で定義した  $U_k$ 、同様に最初の  $k$  個 ( $k < r$ ) の右特異ベクトルのみから構成

される  $n \times k$  行列  $V_k$ 、 $k$  個の特異値のみから構成される  $k \times k$  対角行列  $\Sigma_k$  を用いて以下のように定義する。

$$D_k = U_k \Sigma_k V_k^T \quad (12)$$

行列  $D_k$  は、元の行列  $D$  をランク  $k$  の行列で近似したものとみなすことができる。 $D_k$  が  $D$  にどの程度近似しているかの尺度にフロベニウスノルムを用いる。 $D$  と  $D_k$  に対して、以下の式が成り立つことが知られている。

$$\begin{aligned} \|D - D_k\|_F &= \min_{\text{rank}(D') \leq k} \|D - D'\|_F \\ &= \sqrt{\sigma_{k+1}^2 + \dots + \sigma_r^2} \end{aligned} \quad (13)$$

$D_k$  が元の行列  $D$  と比べ、どの程度、異なっているかは以下の式により計算することができる。

$$\frac{\|D - D_k\|_F}{\|D\|_F} \quad (14)$$

近似行列  $D_k$  を用いた検索について考えてみる。ベクトル  $\vec{e}_j$  を  $n$  次元の単位ベクトルとすると、 $D_k$  の  $j$  番目文書ベクトルは  $D_k \vec{e}_j$  で表すことができる。文書ベクトル  $D_k \vec{e}_j$  と検索質問ベクトル  $\vec{q}$  との類似度は以下のようになる。

$$\begin{aligned} \cos(D_k \vec{e}_j, \vec{q}) &= \frac{(D_k \vec{e}_j) \cdot \vec{q}}{\|D_k \vec{e}_j\| \|\vec{q}\|} = \frac{(D_k \vec{e}_j)^T \vec{q}}{\|D_k \vec{e}_j\| \|\vec{q}\|} \\ &= \frac{(U_k \Sigma_k V_k^T \vec{e}_j)^T \vec{q}}{\|U_k \Sigma_k V_k^T \vec{e}_j\| \|\vec{q}\|} = \frac{\vec{e}_j^T V_k \Sigma_k U_k^T \vec{q}}{\|\Sigma_k V_k^T \vec{e}_j\| \|\vec{q}\|} \\ &= \frac{(\Sigma_k V_k^T \vec{e}_j)^T (U_k^T \vec{q})}{\|\Sigma_k V_k^T \vec{e}_j\| \|\vec{q}\|} \end{aligned} \quad (15)$$

式(15) は、 $k$  次元に圧縮した索引語-文書行列  $D_k$  を再構築することなく特異値分解の結果  $U_k$ 、 $\Sigma_k$ 、 $V_k$  から類似度を計算できることを示している。

潜在的意味インデキシングにおける類似度の計算式は、式(11) と式(15) の2つがでてきた。2つの式は考え方が若干異なっているが、どちらを使うかは好みの問題であるが、多くの文献では式(15) の方を使っている。図3.5は本Webページ検索システムの一例である。

## 質問文章

織田信長が桶狭間の戦いに出陣した歴史について知りたい



## 検索結果

ID	タイトル	類似度
110000000055	熱田神宮	0.478389
110000000059	今川義元が討たれたといわれる場所	0.411242
110000000060	熱田神宮にある信長塀	0.360515
110000000058	今川義元が本陣を置いた場所	0.360025
110000000053	現在の清洲城の天守閣	0.336507
110000000056	桶狭間の合戦の軍団進路	0.290821
110000000144	砦前・鷲津砦跡の様子	0.245510
110000000057	桶狭間の合戦古戦場	0.204634
110000000140	砦前・大高城跡の様子	0.201269
110000000065	長島一向一揆攻略図	0.183833

図 4.5 本 Web ページ検索システムの事例

図 4.5 に示す具体例で、ベクトル空間モデルによる検索並びに類似度計算を行ってみる。索引語として「信長」、「秀吉」、「桶狭間」、「義元」など 8 語を、文書  $D$  として「今川義元が討たれたという場所：桶狭間の合戦時、義元が信長の家来毛利新助に討たれたと伝えられる場所」など 6 文書を、質問文として「織田信長と桶狭間」を選ぶ。

文書ベクトル  $\vec{d}$  は、 $8 \times 6$  行列となり、質問ベクトル  $\vec{q}$  は、1 行目と 3 行目の 2 が「1」で残りが零という  $8 \times 1$  行列となる。類似度計算は、文書ベクトル  $\vec{d}$  と質問ベクトル  $\vec{q}$  の余弦 (Cosine) をとることにより、類似度  $\text{sim}(\vec{d}, \vec{q}) = \cos(\vec{d}, \vec{q}) = 0.411242$  から求まる。

Web 検索結果は、類似度  $\text{sim}(\vec{d}, \vec{q})$  の値の大きいものから順に並べたものである。

## 4.6 LSI法を用いるWebページ検索

自然言語で表現された質問文書に対して索引語の重みベクトルを生成し、LSI 法により索引語-文書行列を参照して類似度の高い Web ページを検索・出力する。

LSI 検索では、学情研より提供いただいた教育コンテンツ(デジタルアーカイブ)の LOM

登録データを使用し、自然言語による質問文書からの類似のWebページを検索するシステムを実装している。検索手法として潜在的意味インデキシングを用いている。また、LSI 検索と平行してLOM カテゴリ検索を実装している。これにより、LOM の項目による絞込みが可能となっている。

#### 実行手順

図4.6に示すようにトップページの検索質問入力フォームに、自然言語にて入力を行う。入りに自然言語を用いることが本システムの大きな特徴の一つである。

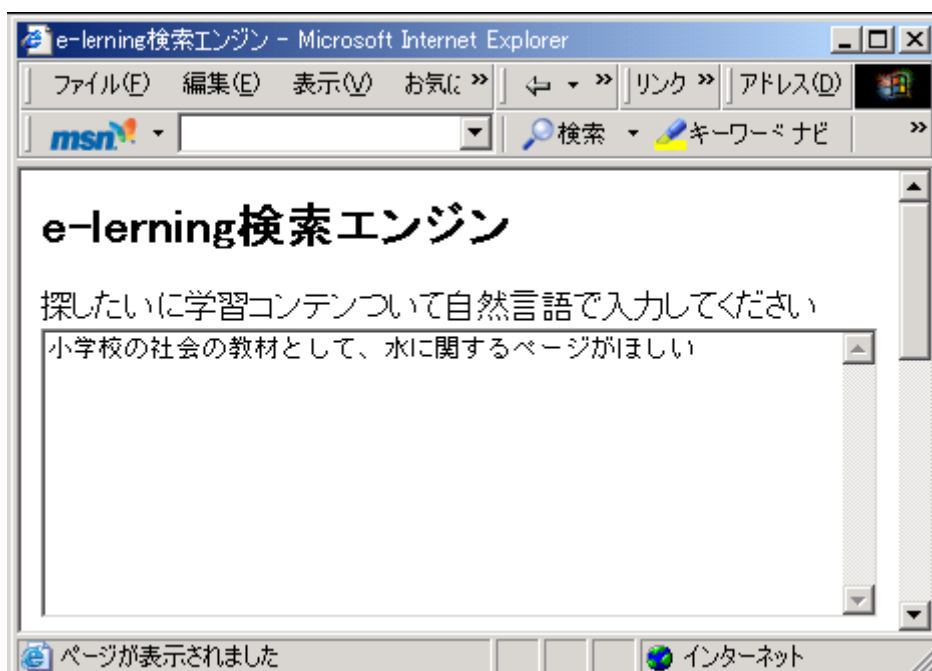


図4.6 検索質問文章入力

次に、検索条件を入力していく。図4.6に示すように、まずLSI における次元数を入力する。入力方法は、圧縮率と直接入力のとおり用意している。圧縮率は、0 で次元圧縮なし、50 で半分、100 で1 となるように設定されている。直接入力は、入力次元数が検索に使用する次元数となる。

入力する次元数は、検索の際の索引語数に当たり、大きな値にすると類義語の処理があまり行われずあいまい性が考慮されなくなり、少ない値にするとページを特徴づける索引語が減っていきデータが壊れていく。これらのことを考慮し、次元数を適度な数にするとよい検索精度が得られる。本システムで登録されているデータでは、300～500 程度にするとよい検索精度が得られる。図4.6では、次元数を400にする。

出力文書数は、検索結果の出力数で、LSI 検索後に行う適合性フィードバックの際の訓練データ数になる。図4.6では出力文書数は5とする。

検索データベースは、地域によりデータベースを分割し選択する方式をとっている。図4.6では、埼玉県の記事769件を対象とする。これはLSI 検索では、データ数が1000件を超えると格段に処理速度が遅くなるので、回避するためである。また、データベース作成時の計算量の問題を回避するために、地域で分割している。

図4.7は、400次元での検索結果である。図4.8～図4.12は1番から5番の検索データの内容である。

FB用画像

### 400次元での検索

該当数: 769

num	LOMID	TITLE	SIM	FB
1	050000000426	水の浄化と川	0.282743	<input type="radio"/>
2	050000000430	水の浄化と川	0.244670	<input type="radio"/>
3	050000000429	水の浄化と川	0.238654	<input type="radio"/>
4	050000000428	水の浄化と川	0.232757	<input type="radio"/>
5	050000000427	水の浄化と川	0.213549	<input type="radio"/>

出力文書数 5

FBの実行

図4.7 400次元での検索結果

<b>素材表題</b>	<b>着水池</b>	
<b>素材内容</b>	井戸からの水を取り入れ、砂を取り除く。また、消毒剤を入れる。	
<b>登録者</b>	越谷市教育工学研究会	
<b>取材者</b>	黒崎政弘	
<b>取材場所</b>	埼玉県越谷市	
<b>取材日付</b>	1995. 8. 29	



図4.8 水に関するページの1番目

<b>素材表題</b>	<b>室内にあるろ過池</b>	
<b>素材内容</b>	水をこしてきれいにする。細かいゴミや鉄やマンガンなども取り除く。	
<b>登録者</b>	越谷市教育工学研究会	
<b>取材者</b>	黒崎政弘	
<b>取材場所</b>	埼玉県越谷市	
<b>取材日付</b>	1995. 8. 29	



図4.9 水に関するページの2番目




<b>素材表題</b>	<b>沈澱池</b>	
<b>素材内容</b>	大きなプールのような形で静かに水が貯められる。水に含まれている砂を沈澱させて取り除く。	
<b>登録者</b>	越谷市教育工学研究会	
<b>取材者</b>	黒崎政弘	
<b>取材場所</b>	埼玉県越谷市	
<b>取材日付</b>	1995. 8. 29	

図4.10 水に関するページの3番目

<b>素材表題</b>	<b>混和池の消毒剤と水を混ぜる装置</b>	
<b>素材内容</b>	水の中にあり、流れてくる水と消毒剤がよく混ざりあうようにしている。	
<b>登録者</b>	越谷市教育工学研究会	
<b>取材者</b>	黒崎政弘	
<b>取材場所</b>	埼玉県越谷市	
<b>取材日付</b>	1995. 8. 29	

図4.11 水に関するページの4番目

素材表題	混和池	
素材内容	いくつかに分けされたり段差をつけたりして水が消毒剤とよく混ざるようになっている。	
登録者	越谷市教育工学研究会	
取材者	黒崎政弘	
取材場所	埼玉県越谷市	
取材日付	1995. 8. 29	

図4.12 水に関するページの事例の5番目

## 4.7 LSI検索の課題

現在、本システムの動作環境は、ペンティアム 2GPS、512MB メモリ搭載の Linux サーバと、ペンティアム 1GPS、256MB メモリ搭載の Windows XP クライアント PC であるが、LSI 検索において検索対象データが 1000 件を越えると、検索スピードが落ちると言われている。

この点について、県別データで検索対象を絞り込むといった処理によりデータ数を 1000 件以内におさめることにより解決を図っている。

## 第5章

# “ 動画知 ” を利用した適合性フィードバック

### 5.1 はじめに

本章では、Web ページ検索システムで “ 動画知 ” を利用し、新たにサポートベクトルマシン (SVM)による適合性フィードバックを行い、適合性フィードバックあるなしにおけるシステム性能比較を行い、検索精度の向上の度合いを評価する。

本研究は、学習オブジェクトメタデータ(Learning Object Metadata、以下LOMと称す)と、自然言語を用い類似語検索を行う潜在的意味インデキシング(以下LSI)検索システムを用いたe-learning用Webページ検索エンジン [1] の研究を発展させたものである。

本研究では、従来の自然言語(形式知)を用いた検索に、新規に画像に含まれる暗黙知を利用したサポートベクターマシン(Support Vector Machines、以下SVMと称す) [2]による適合性フィードバックシステムを加えることにより検索精度改善を行った[3]。具体的には、サムネイル画像を提示することにより、画像に含まれる暗黙知を利用し、直感的な判断を可能にしている[4]。

本システムは、教育とネットワークサービスに関する「社会システム」共同研究プロジェクトとして開発したWeb ページ検索エンジンである。[5]本エンジンは、「学習情報研究」の論文100件と、全国のマルチメディア教材研究会員が作成した約4万点のWeb素材・教材を、より適確に検索すると判明した。

### 5.2 Web ページ検索エンジン

第4章で示したように、図5.1の新しいWeb ページ検索エンジンは、1) 画像における暗黙知を利用したSVM による適合性フィードバック、2) 自然言語によるLSI 検索とLOM カテゴリ検索とのハイブリッドなWeb 検索システムと、3) データベース構築プログラム

(データの追加等)の実装という3つの特長を持つ:

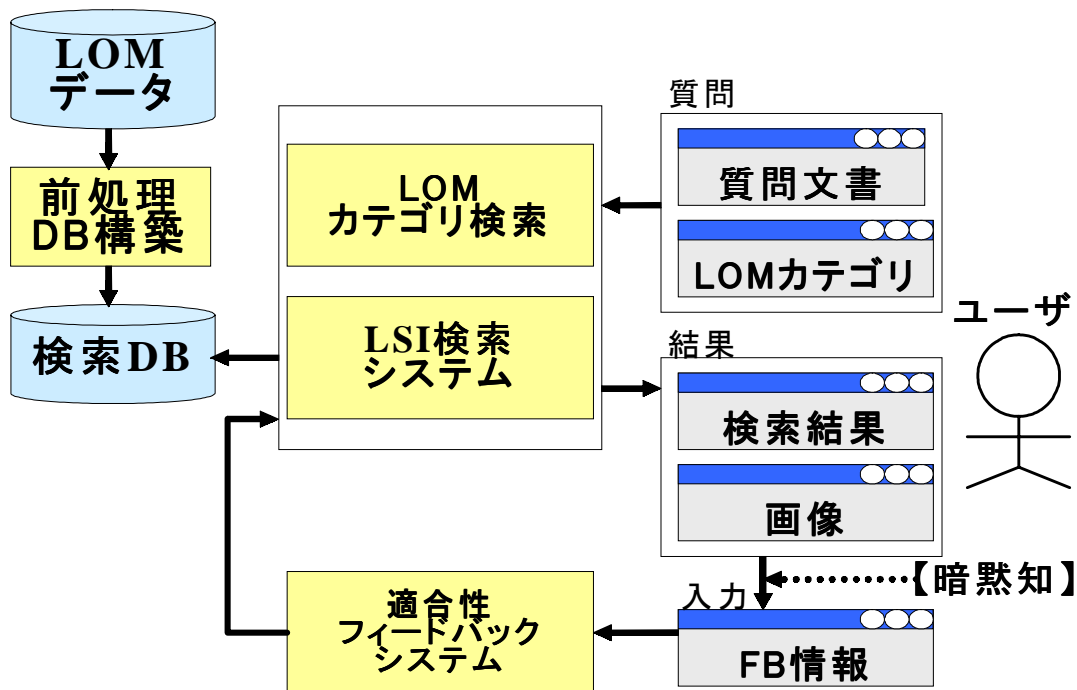


図 5.1 Web ページ検索エンジン

### 5.3 サポートベクターマシン(SVM)

図5.2に示すように、サポートベクターマシン(Support Vector Machine:SVM) では、ベクトル空間に正例と負例が既知の訓練データをプロットする。そして、重みベクトル $w$  に対して訓練データ中の各オブジェクトを $w$  に射影したときに、正例と負例の中で最も近いオブジェクトの距離(margin: マージン) が最大となるようなベクトル(サポートベクトル) を求め、空間を2分割するような超平面を学習する。

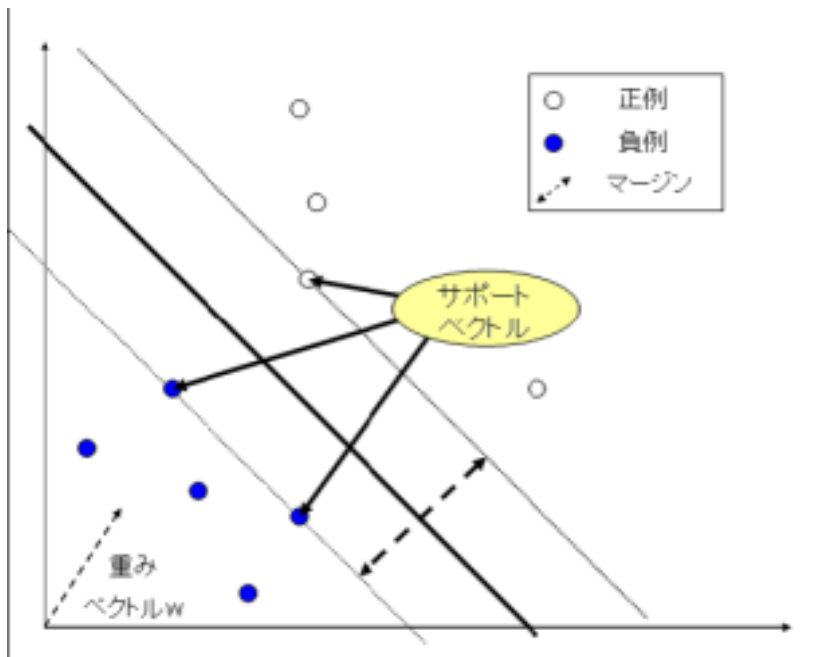


図5.2 サポートベクターマシン

## 5.4 暗黙知を利用した SVM による適合性フィードバック

分割するデータを文書集合とした場合、SVM を用いてある基準に対して文書集合を適合、不適合に分割可能である。このことを用いて、利用者が適合・不適合の判別を行いその判別結果をシステムにフィードバックすることによりさらに適合性の高い文書を検索する。

以下に処理の流れを記す。

### Step1 初期検索

ベクトル空間モデルを用い、質問に対し検索を行い、類似度の高い上位N 文書とそのフィードバック用画像を利用者に提示する。

### Step2 利用者による判断

Step1 により提示されたN 文書(訓練データ) に対して、画像の暗黙知を利用し判断を行う。適合ならラベル+1、不適合ならラベル-1 をつける(図5.3参照)。

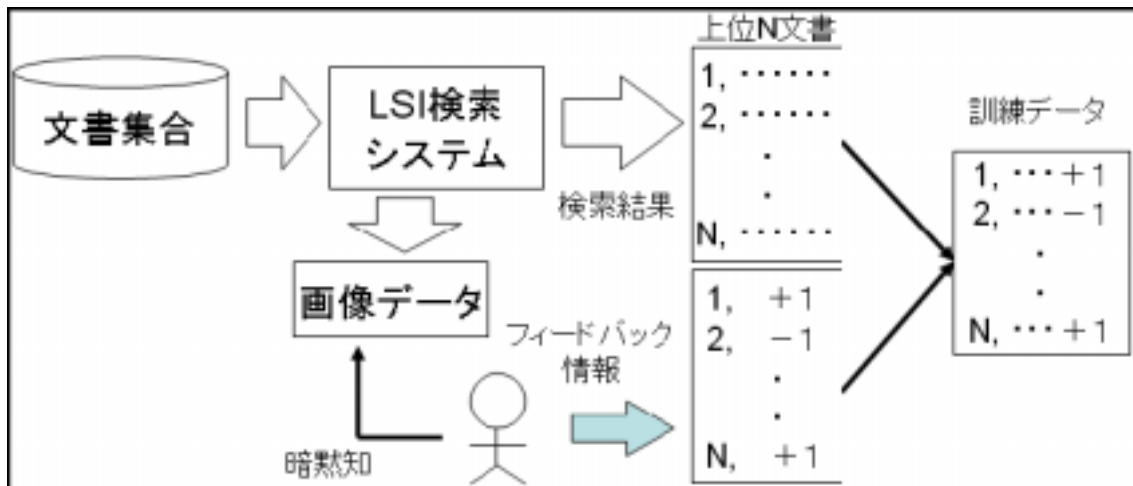


図5.3 暗黙知を利用したSVM による適合性フィードバックStep2

#### Step3 分離平面の決定(SVM 学習)

利用者が判別したN文書(訓練データ)を用いて、SVM学習を行い、訓練データに対して、適合・不適合に分類する最適判別平面を決定する(図5.4参照)。

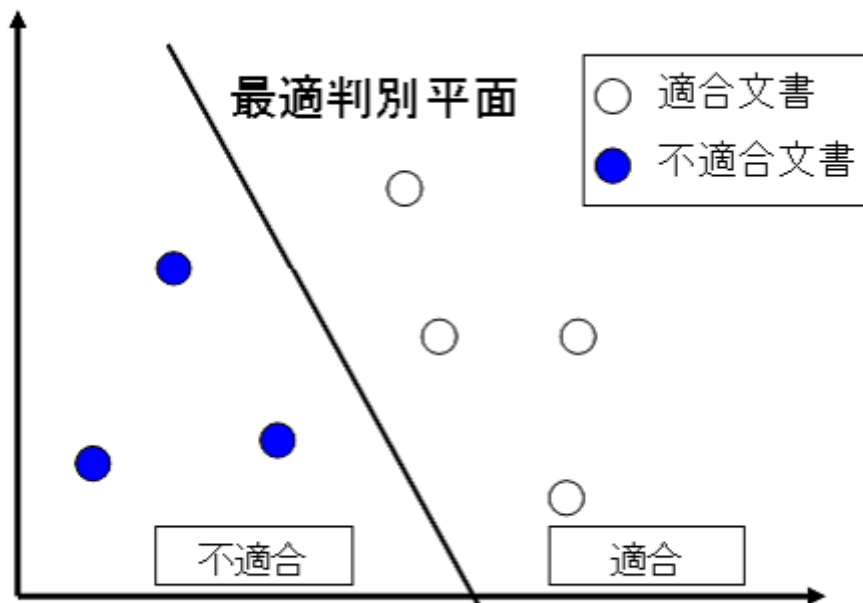


図5.4 暗黙知を利用したSVM による適合性フィードバックStep3

#### Step4 適合・不適合の分類(SVM 分類)

検索対象文書全体に対して決定された分類平面を用いて、適合・不適合の分類を行う(SVM 分類)(図5.5参照)。

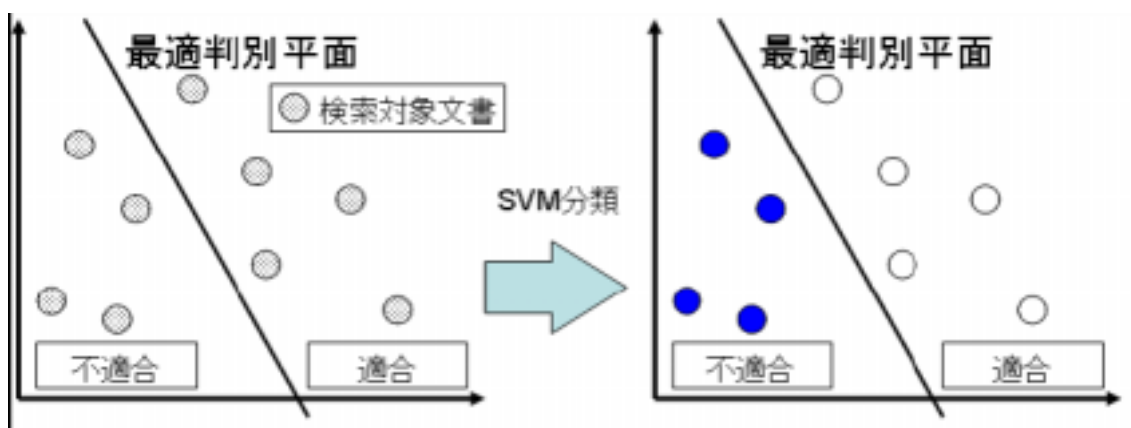


図5.5 暗黙知を利用したSVM による適合性フィードバックStep4

#### Step5 検索結果出力

検索対象文書全体に対して、決定された分離平面により適合と分類された文書集合に対し、再度ベクトル空間モデルを用いた検索を行いStep1の質問に対し、類似度の高いM文書をシステムからの検索結果として出力する(図5.6参照)。

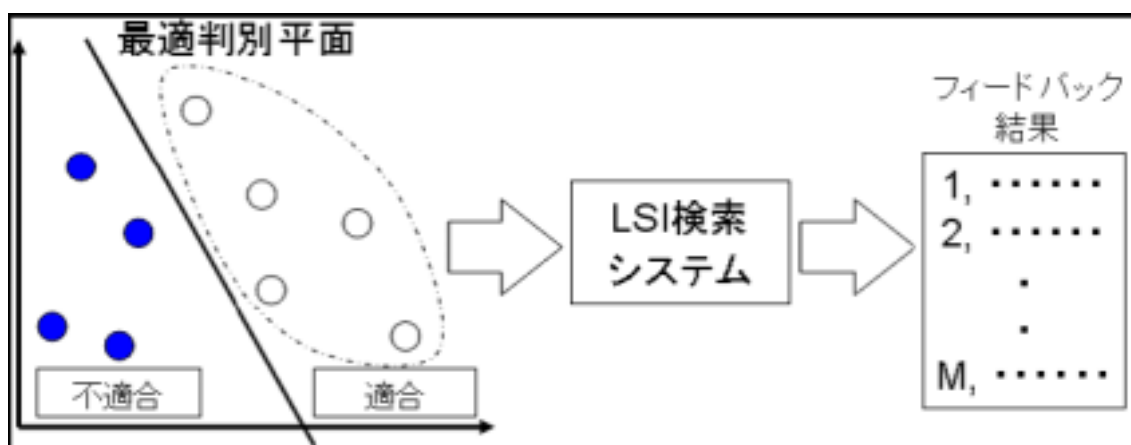


図5.6 暗黙知を利用したSVM による適合性フィードバックStep5

## 5.5 適合性フィードバックの実行例

LSI 検索結果に対し、適合・不適合の情報を入力することにより適合の傾向にあるWeb ページを結果として出力するシステムとなる。適合・不適合の判断の際に、各Web ページを表しているサムネイル画像を提示することにより、暗黙知による直感的な判断が可能となっている。

#### 実行例

実行例では、図5.7に示す検索結果の左上のFB 画像をクリックし、図5.8に示すフィードバ

ック用画像を表示する。

FB用画像

### 300次元での検索

該当数: 769

num	LOMID	TITLE	SIM	FB
1	<a href="#">050000000543</a>	校庭の草花	0.566464	<input type="radio"/> ▾
2	<a href="#">050000000189</a>	四季の草花牛島小	0.542732	<input type="radio"/> ▾
3	<a href="#">050000000538</a>	校庭の草花	0.506694	<input type="radio"/> ▾
4	<a href="#">050000000539</a>	校庭の草花	0.443576	<input type="radio"/> ▾
5	<a href="#">050000000181</a>	四季の草花牛島小	0.443520	<input type="radio"/> ▾

出力文書数

入力された文章

---

理科の教材として花の画像がほしい

---

[戻る](#)

図5.7 適合性フィードバックの実行例の検索結果

ID & TITLE	THUMBNAIL
050000000543 校庭の草花	
050000000189 四季の草花牛島小	

図5.8 適合性フィードバックの実行例のフィードバック画像



各画像・タイトルなどを参考に、検索結果の一番右側のFB 情報（フィードバック情報）を入力する。自分の検索意図したページなら(適合) なら ・関係のないページ(不適合) なら x を入力する。適合性フィードバックでは、適合数と不適合数があまりにも偏っている場合、正しく学習できない場合がある。そのため、本システムでは、多い方が少ないほうの4倍より多い場合、エラーを表示するように作成している。その後、フィードバック後のWeb ページの出力数を入力し、FBの実行をクリックする。検索結果として、LSI 検索と同様のフォーマットで出力する。

## 5.6 具体的な適合性フィードバック手順

適合性フィードバックとは、具体的には検索結果に対し、ユーザが適合( )、不適合(x)の判断を行い、その情報をシステムにフィードバックする手法である。

具体的には、本稿ではサポートベクターマシン (SVM) による適合性フィードバック手法を用いている。SVM は、学習データより決定される分離平面を用い、データ集合を2クラスに分類する手法であり、高い分離能力があるので広く注目され利用されている。

フィードバックされる文書、つまりユーザが適合( )、不適合(x)の判断を行った文書を、SVM の学習データとして用いれば、検索対象文書全体を適合( )、不適合(x)に分類することが可能である。適合に分類された文書集合に、再度検索を行うことにより、より高精度の情報検索が可能と考えられる。

本稿では、以下に示すステップでフィードバック、検索を行う。[7]

### ・ステップ 1: 初期検索

ベクトル空間モデルを用い、ユーザが要求した質問に対し、検索を行い、類似度の高い上位 N 文書をユーザに示す。

### ・ステップ 2: ユーザによる判断

ステップ 1 で示された文書に対し、ユーザが適合( )、不適合(x)の判断を行う。適合と判断した文書はラベル「1」を、不適合と判断した文書はラベル「-1」をつける。

### ・ステップ 3: 分離平面の決定(SVM の学習)

ユーザが判別した文書を用い SVM の学習を行い、検索対象文書全体を適合、不適合に分類する最適分離平面を決定する。

### ・ステップ 4: 検索結果出力

決定された分離平面により、適合と分類された文書に対し、再度ベクトル空間モデルを用い検索を行い、Step1 で要求された検索質問と類似度の高い M 文書をシステムからの検索結果として出力する(以下では M=10 とした)。

## 5.7 システムの実行

### 5.7.1 検索システムの実現

本システムは現在公開されており、芝浦工業大学・若木研究室のサーバにて稼動中である。公開 URL は「<http://www.ailab.se.shibaura-it.ac.jp/p02078/>」で、質問文章を入力することにより容易に Web ページ検索ができる。

### 5.7.2 検索システムの類似度による定量評価

図 5.9 のように、本 Web ページ検索システムの定量評価は、検索したい内容についての質問を自然言語による文章で入力し、次元数、結果の出力数を入力することから開始される。サーバは入力された情報を受け取り、類似度を計算し、検索結果を表示する。

**e-learning検索エンジン**

探したいに学習コンテンツについて自然言語で入力してください

織田信長が桶狭間の戦いに出陣した歴史について知りたい

LSI次元  圧縮率  直接入力 出力文書数   
 検索データベース

**LOM検索**

使用しない  and検索  or検索

[help](#)

図 5.9 自然言語による文章で入力

### (1)検索結果

検索結果は、図 5.10 のようになり、入力された情報に対する検索結果を類似度の高い順に 10 文書を表示する。

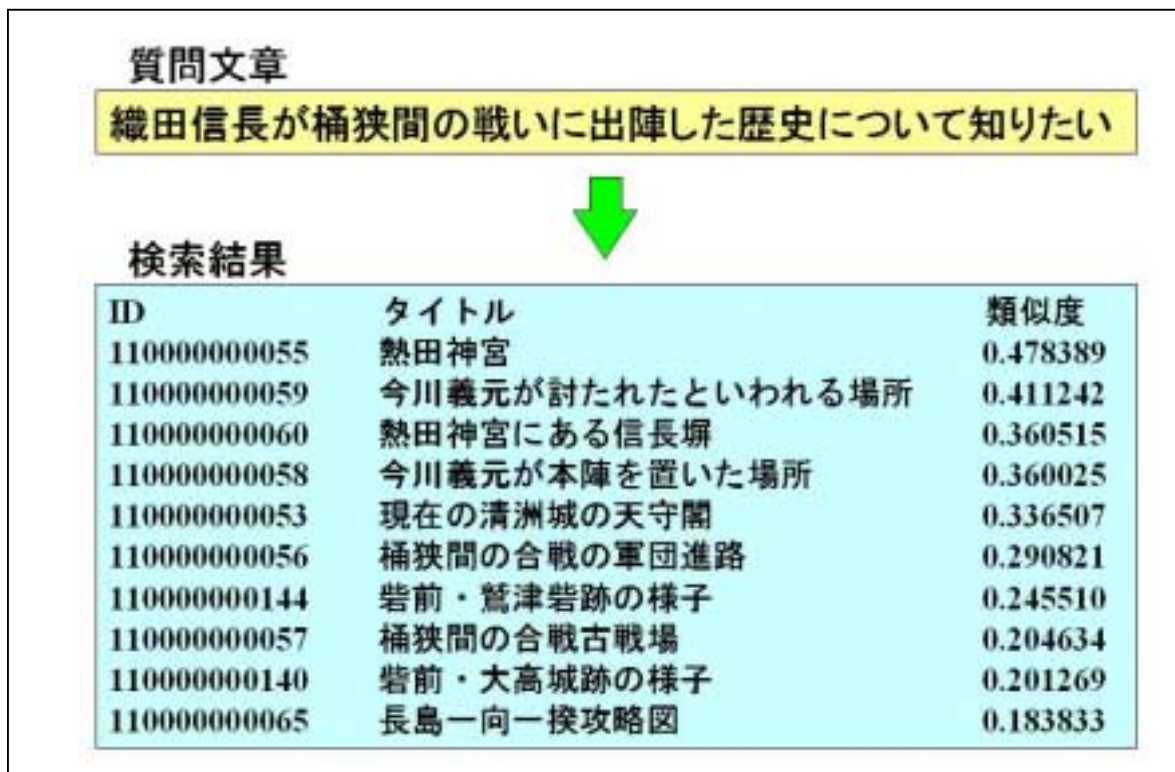


図 5.10 本 Web ページ検索システムの事例

比較のため、ほぼ同じ評価対象を持つキーワード検索システム NICER で検索を実施する。

図 5.11 のように、「織田信長」と「桶狭間の戦い」のキーワードでは、例えば、「桶狭間の合戦、今川義元、織田信長」というキーワードを持つ「今川義元が討たれたといわれる場所」といった教育用コンテンツ (Web ページ) が入手できないと分かる。

質問文章	キーワード
織田信長が	織田信長 桶狭間の戦い
	いて知りたい

✖

**素材表題** 今川義元が討たれたといわれる場所

**素材内容** 桶狭間の合戦の時、義元が信長の家来、毛利新助に討たれたと伝えられる場所。

**登録者** 藤原貴久

**登録日付** 1995年

**取材者** 藤原貴久

**取材場所** 名古屋市緑区桶狭間宇広坪

**取材日付** 1995年7月25日



図 5.11 キーワード検索システム NICER の事例

この例のように、NICER の検索システムではキーワードの付け方によって検索できないコンテンツがでてくるが本システムを用いることによってより広範囲な検索を行うことができる。

表 5.1 評価結果（上位 20 件中の正例数）

正例数	136次元	100次元	50次元
質問 1	19	18	17
質問 2	14	14	12
質問 3	14	15	14
質問 4	12	12	12
質問 5	17	18	18
平均	15.2	15.6	14.6

表 5.1 は、検索の正確性を評価した結果である。類似度を計算し、検索した上位 20 件中の正例数である。表 1 の検索結果から、文書数が増える（次元数が増える）程、正例数が増え

る傾向があると分かる。しかし、現在は文書数が136文書と少ないので、今後は文書数を増やし、検索精度の評価を検討する必要があると分かった。[8]

## (2)検索結果のまとめ

本論では、具体的に、潜在的意味インデキシング (LSI) 法で、文書ベクトルと質問ベクトルとの類似度を計算し、Web ページ検索システムを実現した。

Web 検索結果から、

任意の自然言語による検索で教育用コンテンツ (Web ページ) が類似度順に得られる、  
文書数が増える (次元数が増える) 程、正例数が増える傾向がある、  
今後はデータ量を増やし、検索精度の評価を検討する必要がある  
とわかった。

## 5.8 論文を対象とした SVM による適合性フィードバック

### 5.8.1 論文を対象とした検索データベース

#### (1) Web ページ

検索対象となる Web ページは、自然言語であるテキストシナリオとテキストシナリオの内容を表す画像データをもつ(図 5.12)。

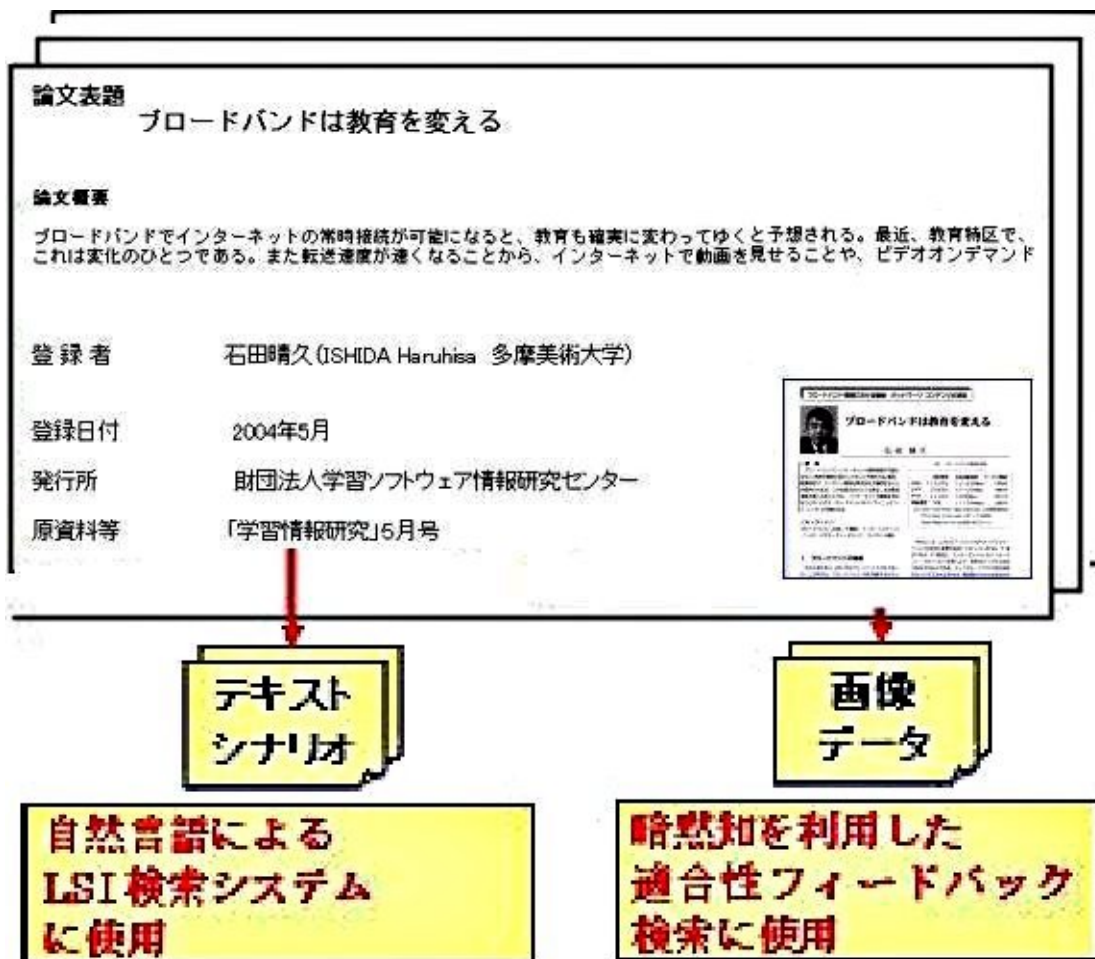


図 5.12 Web ページとデータベース

このテキストシナリオは画像の内容が記されており、このテキストシナリオを抽出しデータベース化することにより検索を可能とする。

本研究では新規に、画像ファイルに含まれる暗黙知を、IT リテラシー向上を目指すという教育プロセス的な側面から e-learning に活用する。具体的には、論文画像を、フィードバック検索の際に提示し、画像の暗黙知をフィードバックに生かしている。

## (2) LOM データ

LOM とは e-learning 用メタデータであり、学習オブジェクトの構造と語彙指定の仕組みが定義された規格である。データベースとして、ID・タイトル・URL を使用し、検索結果として ID・タイトルを表示する。

図 5.13 は、LOM 項目の URL から検索対象の e-learning 用 Web ページを参照することを示す。本システムでは、この URL より Web ページをダウンロードし、素材内容をテキストシナリオとして抽出し、データベースの作成を行う。また同時に、Web ページのサムネイルを抽出し、SVM による適合性フィードバックの判断の際に一覧の閲覧を可能にし、画像の暗黙知による直感的な判断を可能とする。また、LOM の項目には、対象範囲を特定する項目が盛り込まれている(対象年齢など)。その項目では、適合なら T、不適合なら F がつけられている。

ID	タイトル	概要	URL
000020040501	ブロードバンドは教育を充実	ブロードバンドでインターネ	<a href="http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040501.pdf">http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040501.pdf</a>
000020040502	ブロードバンドの教育利用	教育において、eラーニング	<a href="http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040502.pdf">http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040502.pdf</a>
000020040503	授業で使えるコンテンツを採	教育情報ナショナルセンタ	<a href="http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040503.pdf">http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040503.pdf</a>
000020040504	デジタル情報の自己管理	今まで文書が中心だったデ	<a href="http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040504.pdf">http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040504.pdf</a>
000020040505	ネットの未来とく友だちの輪	ウェブは、ばらばらの情報	<a href="http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040505.pdf">http://www.gakujoken.jp/gakuiyousi/gakuiyousi178/photo040505.pdf</a>

図 5.13 論文データベースの LOM データ

## 5.8.2 論文を対象とした Web 検索エンジン

### LSI 検索

LSI 検索では、専門誌「学習情報研究」の論文コンテンツ(デジタルアーカイブ)の LOM 登録データを使用し、自然言語による質問文書からの類似の Web ページを検索するシステムを実装している。[9]検索手法として潜在的意味インデキシングを用いている。また、LSI 検索と平行して LOM カテゴリ検索を実装している。これにより、LOM の項目による論文の絞り込みが可能となっている。

### 実行手順

図 5.14 に示すように、トップページの検索質問入力フォームに、自然言語にて入力を行う。入力に自然言語を用いることが本システムの大きな特徴の一つで、検索条件を入力していく。まず LSI における次元数を入力する。入力方法は、圧縮率と直接入力の二通り用意している。圧縮率は、0 で次元圧縮なし、50 で半分、100 で 1 となるように設定されている。直接入力では、入力次元数が検索に使用する次元数となる。本事例では、500 にする。

出力文書数の 10 件は、検索結果の出力数で、LSI 検索後に行う適合性フィードバックの際の訓練データ数になる。検索データベースは、論文を選ぶ。



## e-learning検索エンジン

探したいに学習コンテンツについて自然言語で入力してください

小学校の社会科の教材として、ブロードバンドに関する資料がほしい

LSI次元	<input checked="" type="radio"/> 圧縮率 <input type="radio"/> 直接入力	出力文書数	<input type="text" value="10"/>
<input type="text" value="500"/>		検索データベース	<input type="text" value="論文"/>

<b>LOM検索</b>	
<input checked="" type="radio"/> 使用しない <input type="radio"/> and検索 <input type="radio"/> or検索	
<input type="text" value="初等中等教育"/> <input type="text" value="高等教育"/>	
<input type="button" value="実行"/>	<input type="button" value="キャンセル"/> <a href="#">help</a>

図5.14 論文データベースの検索

また、LOMカテゴリ検索を並行して実行する場合は、and・or検索から選択し、その後絞り込みたい項目を選択していく。LOMカテゴリ検索では、絞り込む項目は複数選択可能となっている。すべての入力終了した後、実行をクリックする。

検索結果として、順位・LOMID・タイトル・LSI検索の類似度が表示され、図5の結果ページにはそのほかに、入力質問文章・次元数・該当数が表示される。



## 論文表題

ブロードバンドは教育を変える

## 論文概要

ブロードバンドでインターネットの常時接続が可能になると、教育も確実に変わってゆくと予想される。最近、教育特区で、オンラインがこれら変化のひとつである。また転送速度が速くなることから、インターネットで動画を見せることや、ビデオオンデマンドやイーター：

登録者 石田晴久 (ISHIDA Haruhisa 多摩美術大学)

登録日付 2004年5月

発行所 財団法人学習ソフトウェア情報研究センター

原資料等 「学習情報研究」5月号



図5.15 検索対象Web ページ

検索結果のLOMID をクリックすることにより、図5.15の検索対象のWebページにアクセスすることができる。

### 適合性フィードバック手順

LSI 検索結果に対し、適合・不適合の情報を入力することにより適合の傾向にあるWeb ページを結果として出力するシステムとなる。適合・不適合の判断の際に、各Web ページを表示しているサムネイル画像を提示することにより、暗黙知による直感的な判断が可能となっている。

### 実行手順

検索結果の左上のFB 画像をクリックし、フィードバック用論文画像を表示する。

論文画像の暗黙知と、タイトルなどを利用して、検索結果の一番右側のFB 情報(フィードバック情報)を入力する。図5.16に示すように、自分の検索意図したページなら(適合)なら ・ 関係のないページ(不適合)なら × を入力する。

## 300次元での検索

該当数:119

num	LOMID	TITLE	SIM	FB
1	000020040507	高画質な動画資料の活用②小学校社会科	0.249999	<input type="radio"/>
2	000020040511	高画質な動画資料の活用⑥高等学校理科新しいスタイルの動画素材の活用	0.208514	<input checked="" type="checkbox"/>
3	000020050703	高画質な動画資料の活用⑥高等学校理科新しいスタイルの動画素材の活用	0.204124	<input checked="" type="checkbox"/>
4	000020050705	高画質な動画資料の活用⑥高等学校理科新しいスタイルの動画素材の活用	0.114708	<input checked="" type="checkbox"/>
5	000020040510	高画質な動画資料の活用⑤中学校社会科スチル動画コンテンツによる中学校社会科授業	0.114708	<input checked="" type="checkbox"/>
6	000020050109	高画質な動画資料の活用⑤中学校社会科スチル動画コンテンツによる中学校社会科授業	0.104257	<input checked="" type="checkbox"/>
7	000020040711	高画質な動画資料の活用⑤中学校社会科スチル動画コンテンツによる中学校社会科授業	0.104257	<input checked="" type="checkbox"/>
8	000020040501	ブロードバンドは教育を変える	0.104257	<input type="radio"/>
9	000020050104	ブロードバンドは教育を変える	0.102062	<input checked="" type="checkbox"/>

図5.16 フィードバック情報を入力中の検索結果

図5.17の論文画像の暗黙知を用いたFB 情報を利用すると、意図に反する論文を排除できる特長がある。例えば、外観的特徴から、要約、キーワードや本文が、意図に合わないものや悪文である場合、論文画像を目で確認し、不適合としてxを入力し排除することができる。

適合性フィードバックでは、適合数と不適合数があまりにも偏っている場合、正しく学習できない場合がある。

なお、TinySVM[10]を用いた本システムでは、多い方が少ないほうの4倍より多い場合、エラーを表示するように作成している。その後、フィードバック後のWeb ページの出力数を入力し、FBの実行をクリックする。



図5.17 フィードバック用の論文画像

検索結果として、図5.18に示す LSI 検索と同様のフォーマットで出力する。



図5.18 適合性フィードバック後の検索結果

論文検索で、論文画像の暗黙知とLOM 特有の項目を利用した検索を導入したことにより、より適切な論文検索ができ利便性が増したと考えられる。

## 5.9 検索システムのユーザ・アンケートによる定性評価

### (1) ユーザ・アンケートの実施

検索システムのユーザに対するアンケートは、検索システムのアンケート法の説明、使いやすさ、見易さ、適切性と有用性に関する一対比較による「Web ページ検索システム (WPS) に関するアンケート」、動画検索の使い易さ等を 5 段階評価する「Web ページ検索システム (WPS) に関するアンケート」の 3 種類のアンケート・シートを用いて実施した。

### (2) AHP によるアンケート結果のまとめ

ユーザの観点から見た検索システムの「評価基準」として、「使いやすさ」、「見易さ(わかりやすさ)」、「適切性」、「有用性」という 4 つを選ぶ。適切性(pertinence)とは、「検索質問に対して、ユーザが目的とする文書を検索したか?」である。有用性(usefulness)とは、「ユーザにとって役立つ文書を検索したか?」である。

AHP の手法を用い、まず、「使いやすさ」、「見易さ(わかりやすさ)」、「適切性」、「有用性」という 4 つの評価基準の項目を次のような一対比較をする。次に各評価基準について(1)Web ページ検索システム (WPS)、(2)NICER、(3)YAHOO を一対比較する。図 5.19 は、学情研ネットワーク研究会員 12 名から得た、ユーザの観点から見た検索システムに関する評価結果である。

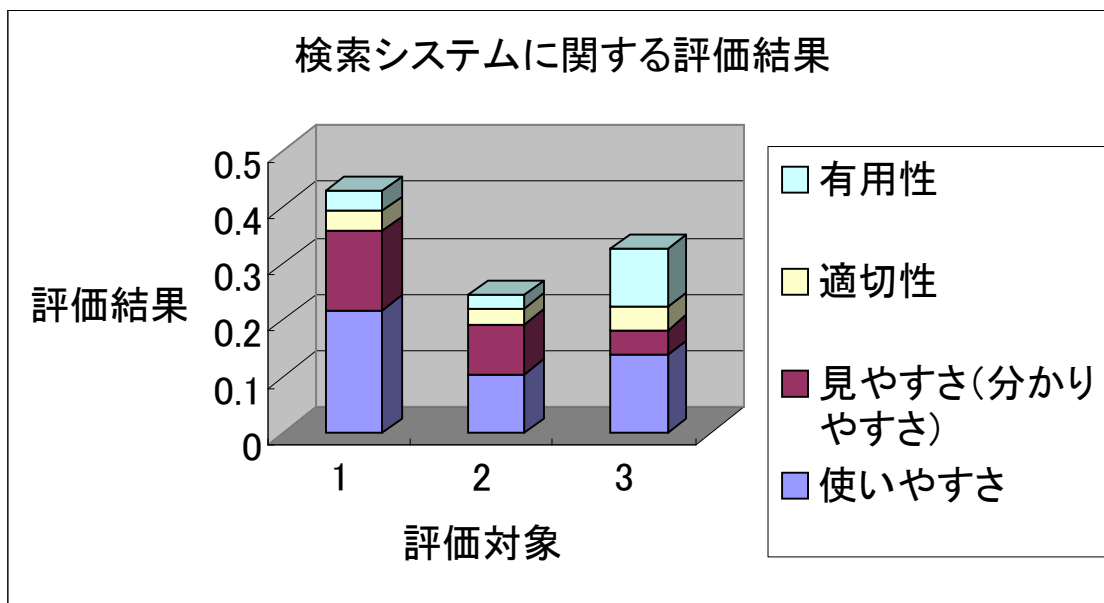


図 5.19 ユーザの観点から見た検索システムに関する評価結果

ユーザの観点を重視した評価結果、特にユーザの主観的な評価基準と代替案による評価結果から、Web ページ検索システム (WPS) が総合的に一番優れていると判断できる。

Web 上の動画検索システムとして、WPS が既存の NICER や YAHOO より使いやすく、見やすく

分かりやすいことが分かる。只、「ユーザにとって役立つ文書を検索したか？」という有用性の観点からは、汎用 Web 検索システムである YAHOO が優れていることが分かる。

図 5.20～図 5.23 は、図 5.19 の評価結果の元データで、ユーザの観点から見た検索システムに関するアンケート結果である。

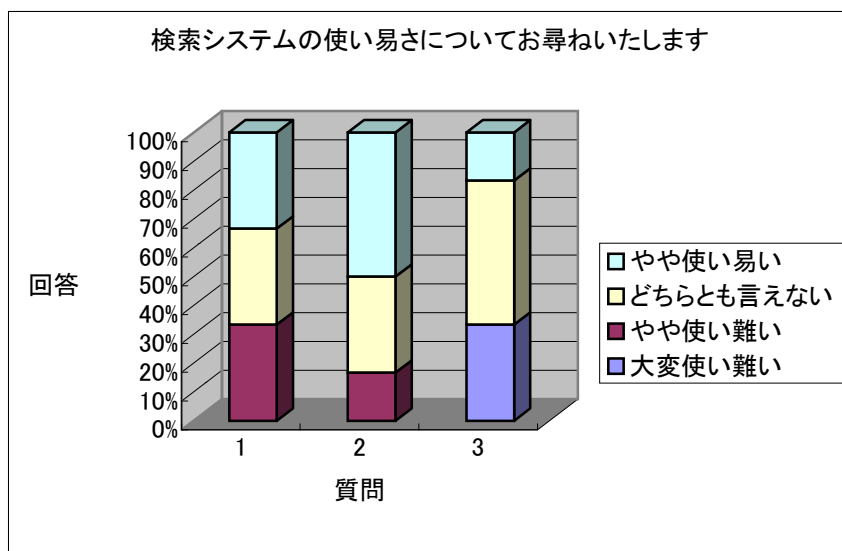


図 5.20 ユーザの観点から見た検索システムに関するアンケート結果

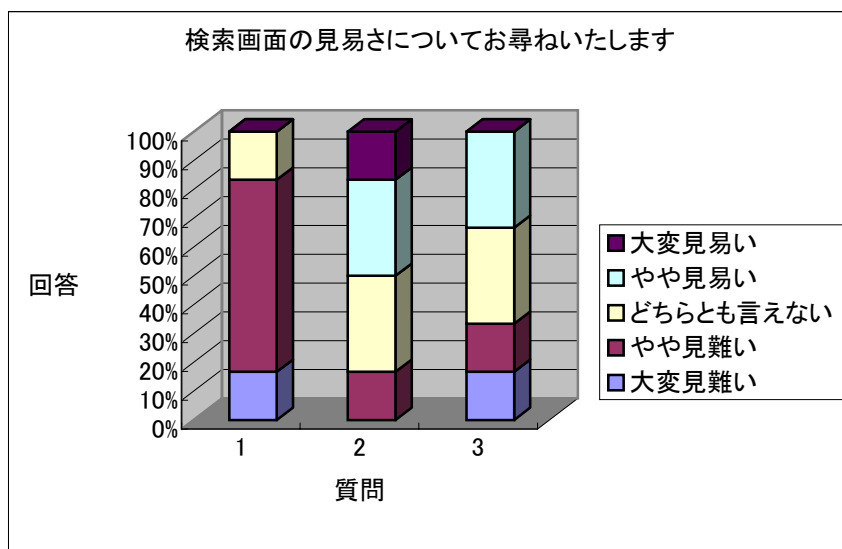


図 5.21 ユーザの観点から見た検索システムに関するアンケート結果

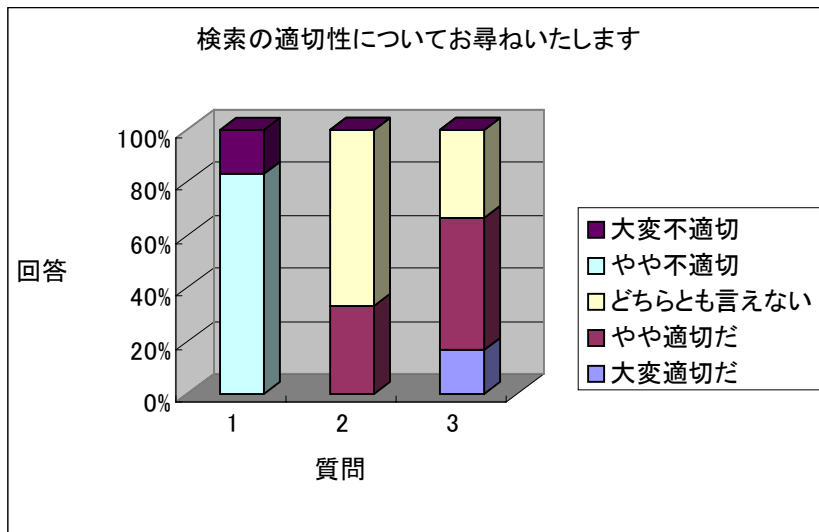


図 5.22 ユーザの観点から見た検索システムに関するアンケート結果

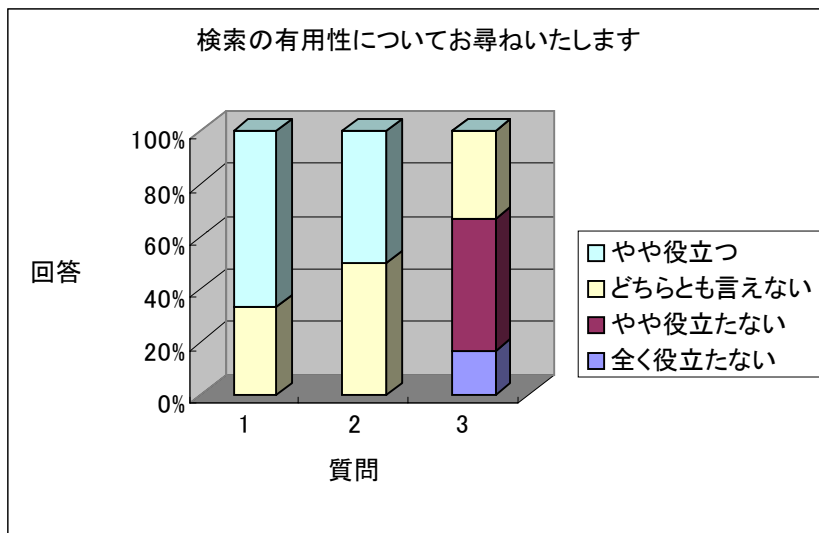


図 5.23 ユーザの観点から見た検索システムに関するアンケート結果

## 5.10 システムの検証

### 5.10.1 画像の暗黙知を用いたSVM による適合性フィードバックの検証

本研究の特徴である画像の暗黙知を用いたSVM による適合性フィードバックについて、検索精度の検討を行う。検索結果に対してユーザが意図するものとそうでないものを判断しそれにより検索精度について考察する。この判断の際に、フィードバック画像を利用する。



表5.2 適合性フィードバック検証の検索条件

背景	: ユーザは理科の教材として植物の花の画像を探している。
検索DB	: 教育コンテンツ埼玉県LOM 登録データ(データ数769)
検索条件:	次元数300、出力数10
質問文章:	理科の教材として、花の画像がほしい。

表5.3 適合性フィードバック検証の検索結果

ID	タイトル	類似度
50000000545	校庭の草花	0.621
50000000196	四季の草花牛島小	0.600
50000000132	四季の草花牛島小	0.516
50000000498	県木1	0.504
50000000207	四季の草花牛島小	0.503
50000000542	校庭の草花	0.498
50000000451	植物と環境に合わせた形態	0.496
50000000194	四季の草花牛島小	0.463
50000000190	四季の草花牛島小	0.446
50000000200	四季の草花牛島小	0.440

0

検索結果に対して、図2.24の像を参照し、FB 情報を付与していく。背景として、植物の画像を探しているという点から、FB 画像に花の映像が入っているものを適合とみなし、入っていないものを不適合とみなす。なお、判断しにくいものも不適合とみなしている。

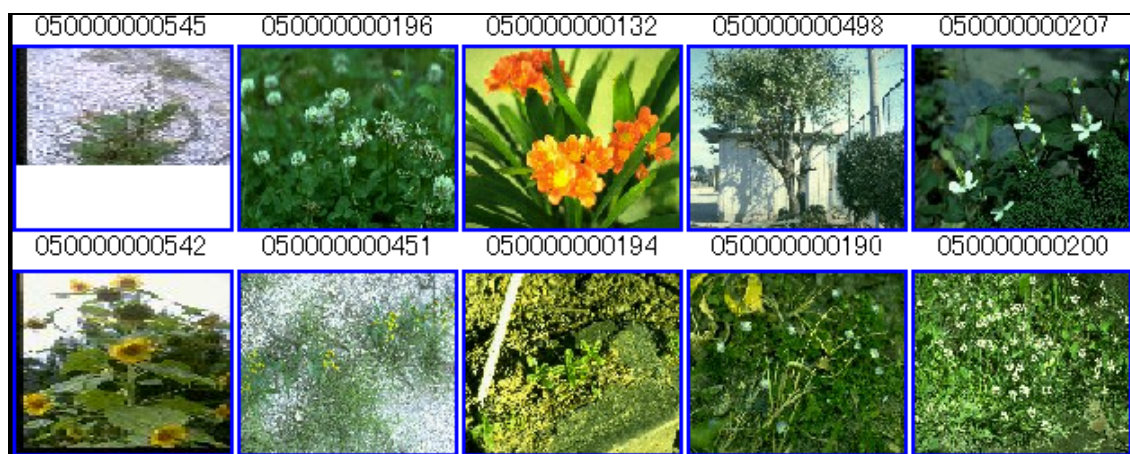


図5.24 適合性フィードバック検証のFB 画像

表5.4 適合性フィードバック検証のFB情報

ID	FB 情報	ID	FB 情報
50000000545	×	50000000542	
50000000196		50000000451	×
50000000132		50000000194	×
50000000498	×	50000000190	
50000000207		50000000200	

このFB 情報をもとにフィードバックを実行する。

表5.5 適合性フィードバックのフィードバック結果

ID	タイトル	類似度
50000000196	四季の草花牛島小	0.600017
50000000132	四季の草花牛島小	0.516936
50000000207	四季の草花牛島小	0.503596
50000000542	校庭の草花	0.498109
50000000190	四季の草花牛島小	0.446735
50000000200	四季の草花牛島小	0.440467
50000000543	校庭の草花	0.415731
50000000189	四季の草花牛島小	0.411954
50000000179	四季の草花牛島小	0.405329
50000000187	四季の草花牛島小	0.334341

FB 画像を表示し、画像を確認して花の画像のページであるかどうかを確認する。

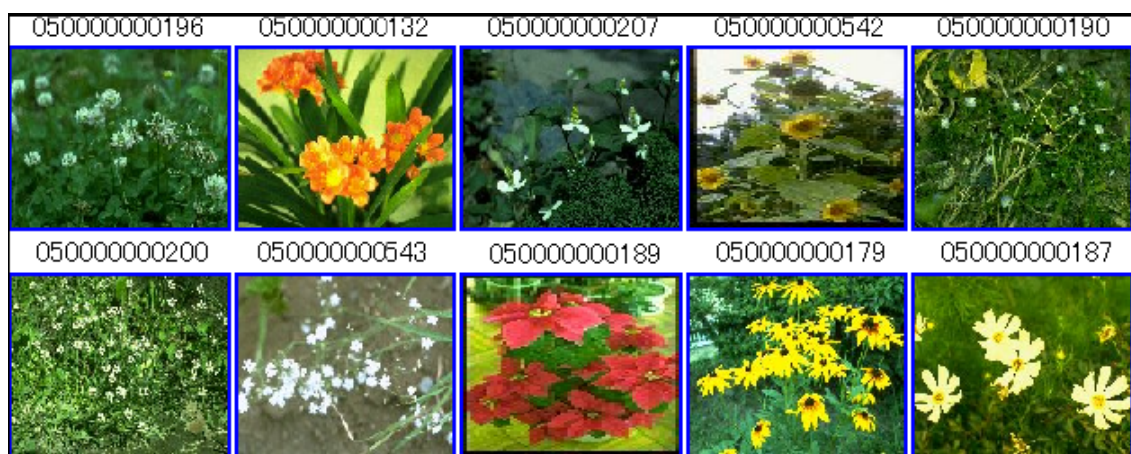


図5.25 フィードバック結果のFB 画像

図5.25により、すべて花の画像であり適合文書が検索されたことが確認できる。このように画像の暗黙知を利用したSVM による適合性フィードバックを使用することにより、ユーザ



の判断をシステムに学習させることが可能となり、 検索精度の改善が可能となる。

### 5.10.2 適合性フィードバックありなしにおけるシステム性能比較

次に、 適合性フィードバックのありなしにおけるシステムの性能比較を行う。 検索システムの精度評価には、 一般的には適合率・再現率を個別に用いて評価を行うが本稿では、 一般にランク付け検索システムの評価に用いられる再現率-適合率曲線を用いシステムの評価を行う。教育コンテンツ埼玉県LOM 登録データ(データ数769) に対し、 FB 画像を確認し、 花の画像であれば適合、 違えば不適合の判断をあらかじめ付加しておく。 確認した結果適合数は141 である。 それによりLSI 検索とLSI 検索+適合性フィードバックを行った実行結果に対して適合数・不適合数を数え適合率-再現率曲線を作成する。 検索条件はさきほどと同様とし、 出力数を500 とする。

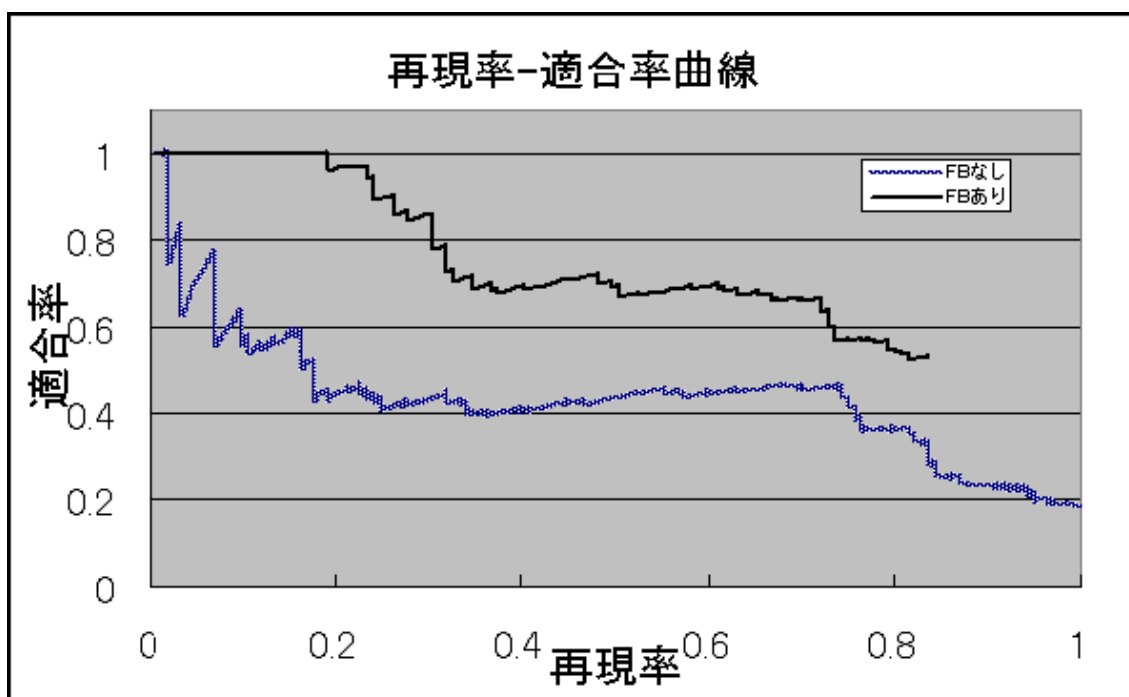


図5.26 適合率-再現率曲線

適合率-再現率曲線では、 2つの曲線が交わらない場合、 右上にある曲線のシステム性能が良いという性質がある。 また、 適合率が1 の際、 重なっているが1 が最大値であるため考慮しない。 図5.26 より、 フィードバックありのグラフが右上に来ているためフィードバックを行うことにより、 システムとしての精度向上が確認できた。

### 5.10.3 訓練データ数における適合性フィードバックの検証

次にさきほどと同じデータベース情報を使用し、 訓練データ数における検索精度の評価を行う。 評価にはF 値を用いる。 F 値はしばしばシステム評価に用いられる値であり、 適合率・

再現率の両方を考慮した値で、お互いに高い場合に高い値になる数値である。

表5.6 適合性フィードバックの検証結果

訓練データ数	適合率	再現率	F 値	該当数
なし	0.332	0.721	0.391	769
10	0.874	0.539	0.606	54
20	0.380	0.764	0.430	632
30	0.450	0.768	0.483	475
40	0.458	0.739	0.481	502
50	0.718	0.613	0.585	225

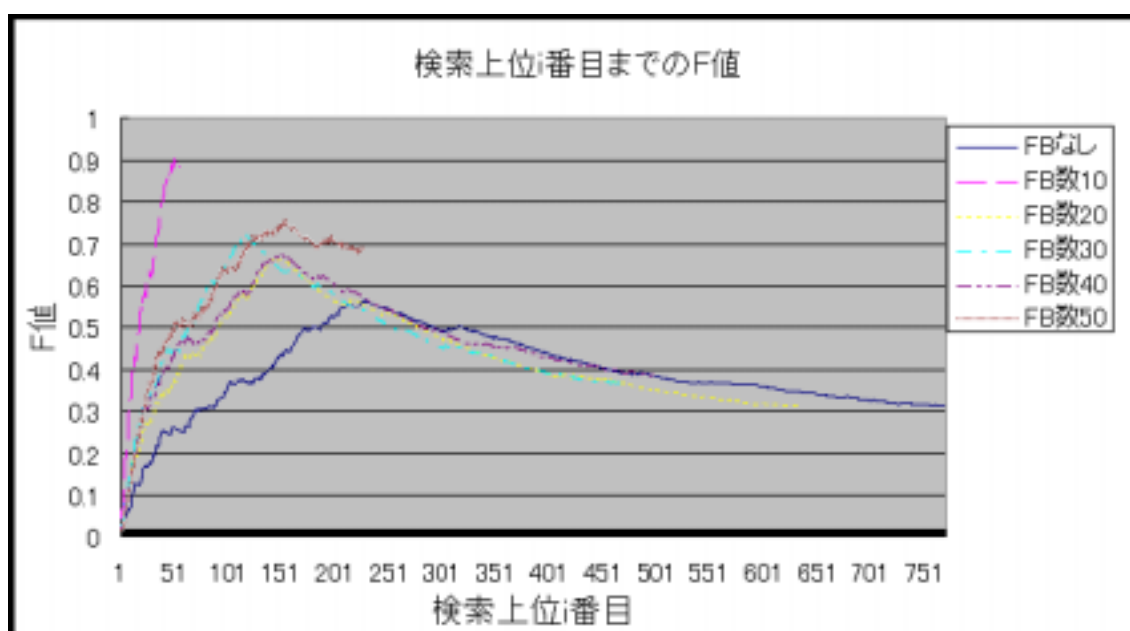


図5.27 訓練データ数の違いによるF 値

図5.27 が得られる。訓練データ数10を除き、訓練データ数の増加によるF 値の増加が確認できた。訓練データ数10の場合、とても高い値が検出されている。この原因としてはSVMにおける学習時の精度が考えられる。すべてにおいてF 値は、最大値より値が下降している。これは下位部分に類似度の低いゴミデータがたまっているためである。訓練データの増加により検索数が絞り込まれ類似後下位部分にあるゴミデータの削減も確認できる。これらの結果から、訓練データ数を増加にすることにより検索精度が向上される。

## 第6章

# セマンティックWeb技術を用いたe-Learning 支援システム

### 6.1 はじめに

本章では、意味処理を行うため、最新技術の一つであるセマンティック Web 技術を用い、曖昧な検索ができる新しい e-Learning 支援システムについて述べる。

曖昧な検索ができる新しい e-Learning 支援システムを実現するため、LOM の RDF/RDFS 表現を用い、LOMの構造オントロジーを作成し、第4章において説明を行った Web ページ検索システムの適合率を向上させる手法を検討する。

本研究ではコンピュータがWebの意味を理解するためのセマンティックWeb技術[2]技術の意味情報を利用してユーザの意図した検索結果や、相互運用性によりWeb上にあるあらゆるLOM[3]の検索を実現することを目的とする。これを実現するためにメタデータ記述に適したRDFを用いて、内容の類似や包含など意味的關係を考慮しつつLOMのRDF表現の仕様を決定すると共に大量のLOMを容易にRDFへ変換するためのLOM-RDF変換ツールの設計と実装を行った。そしてRDFで表現されたLOMの検索について有用性の評価を行い、文科省のNICER(National Information Center for Educational Resources: 教育情報ナショナルセンター)検索システム[5]では実施していない学習指導要領における曖昧さと難易度を組み合わせた検索により、比較的容易に意図した検索結果を得ることができた。

### 6.2 LOM-RDF 変換ツールと相互運用性

#### 6.2.1 RDFSを使用して独自に語彙を定義

LOM で既知の語彙を利用することができない項目では、新規にプロパティやクラスを定義する。次の図 6.1 はサムネイルの項目を表現したもので、プロパティを図 6.2 のように

RDFS で定義している。サムネイルは画像であることから、プロパティの値域に `dct:Image`(Dublin Core のタイプ要素、文字以外の視覚的な象徴表現を示すクラス)を流用して定義し、目的語に画像が来ることを明示する。

```
<rdf:Description rdf:about="URL">
  <nicer-lom:thumbnail>
    <rdf:Description rdf:about="01.jpg">
      <rdf:type rdf:resource="¥&dct;Image"/>
    </rdf:Description>
  </nicer-lom:thumbnail>
</rdf:Description>
```

図 6.1 サムネイルの項目をLOM-RDFで表現したもの

```
<rdf:Property rdf:ID="thumbnail">
  <rdfs:range rdf:resource="¥&dct;Image"/>
</rdf:Property>
```

図 6.2 プロパティをRDFSで定義したもの

## 6.2.2 学習指導要領

図6.3にLOMの1項目である学習指導要領の社会教科をRDF/RDFSで定義、表現したものを示す。学習指導要領とは文部科学省が告示する教育課程の基準であり、小学校から高等学校までの各学校で教えられる内容とその詳細が定められている。NICERのLOMにおいてはそれを学校名/教科/項目のようにツリー上で表現している。これをRDF/RDFSで表現する場合には学校間や分野が近い教科のつながりに着目し、高等学校など上位校種で細かく分けられている科目を中学校、小学校の下位校種にある対応した科目などでまとめてクラスを定義し上下関係を構成する。分野別の例としては日本史と世界史の上位概念に歴史を定義したり、校種別の例としては中学の公民的分野と高等学校の公民を科目の種類としての公民の下位クラスにまとめることなどによりクラス階層を構成する。日本の地理、歴史、及び公民においては最終的に社会科としてまとめ、社会を頂点とした階層となる。このように階層を構成することで上位にある曖昧な概念を使用した検索が可能になる。

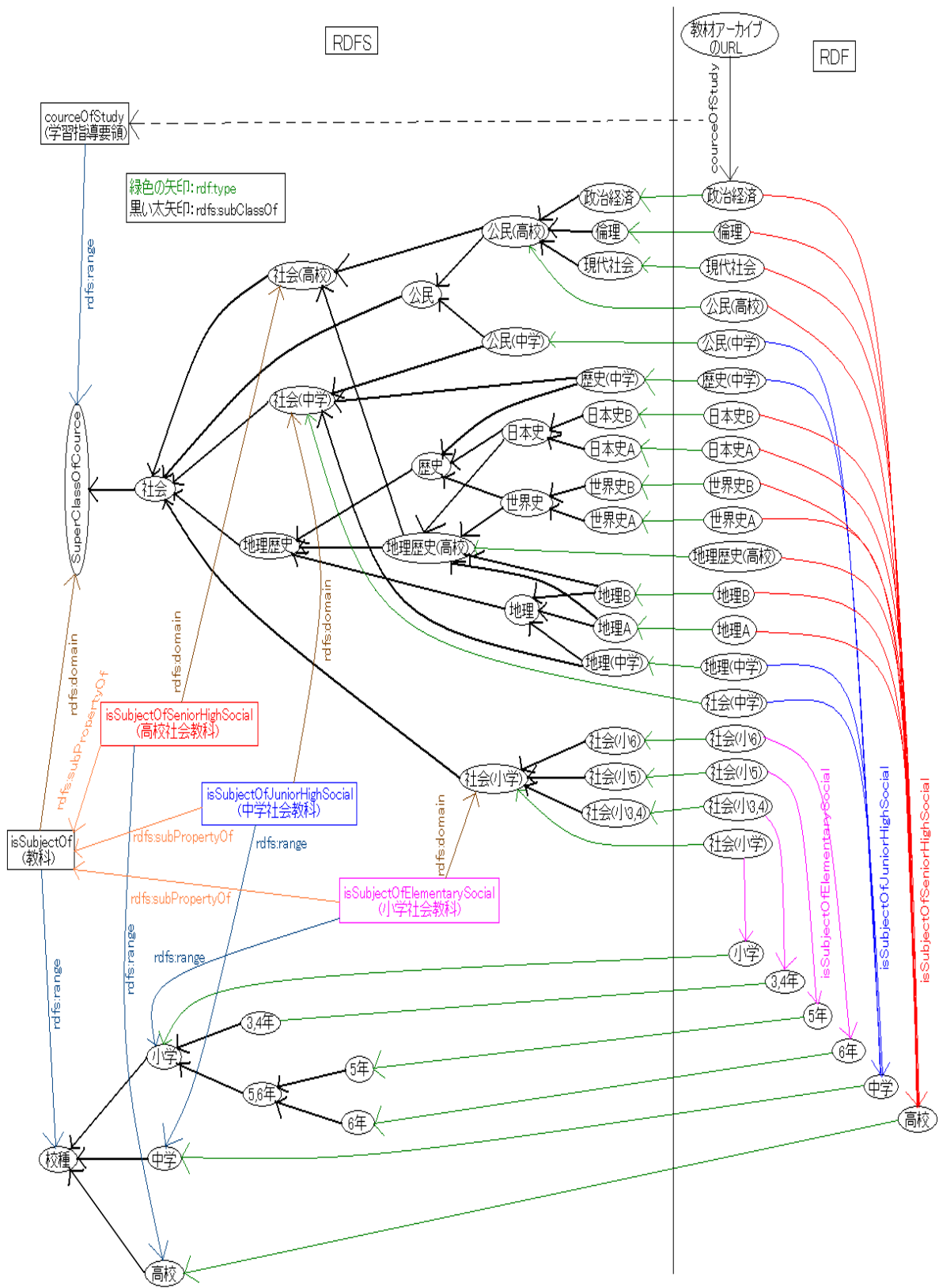


図6.3: RDF/RDFSによる社会教科の階層構造の表現

### 6.2.3 LOM-RDF変換ツールと相互運用性

LOMのRDF表現への変換方法には大きく分けて2種類のタイプがある。一方はタイトルや概要など内容を自由に入力できる項目はあらかじめ定義したRDFのテンプレートに内容を当てはめる。他方では学習指導要領や言語など選択項目は項目の内容に対応したRDF表現に変換する。出力は複数件のLOMを1ファイルにまとめて出力する。図6.4に入力ファイル、図6.5にLOMをRDF表現に変換した出力ファイルの一部、図6.6にシステム構成図を示す。

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	93000001,	三英傑	(信長・秀吉・家康)	:	万松寺の様子,	万松寺の様子,	,,	単体,	,,,								
2	93000002,	三英傑	(信長・秀吉・家康)	:	信長の父、信秀の墓,	信長の父、信秀の											
3	93000003,	三英傑	(信長・秀吉・家康)	:	政秀寺の様子,	政秀寺の様子,	,,	単体,	,,,								
4	93000004,	三英傑	(信長・秀吉・家康)	:	信長の頃の清洲城跡,	信長の頃の清洲城											
5	93000005,	三英傑	(信長・秀吉・家康)	:	現在の清洲城の天守閣,	現在の清洲城の											
6	93000006,	三英傑	(信長・秀吉・家康)	:	清洲城天守から見た城下の様子,	清洲城											
7	93000007,	三英傑	(信長・秀吉・家康)	:	熱田神宮,	熱田神宮,	,,	単体,	,,,,,,								
8	93000008,	三英傑	(信長・秀吉・家康)	:	桶狭間の合戦の軍団進路,	桶狭間の合戦											
9	93000009,	三英傑	(信長・秀吉・家康)	:	桶狭間の合戦の軍団進路,	桶狭間の合戦											

図6.4 LOM-RDF変換ツールの入力ファイル

```

^ xmlns:nicer-course="&nicer-course;"
^ xml:base="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
<rdf:Description rdf:about="http://www.gakujoken.or.jp/dab/SNE0018.html">
^   <dc:identifier>93000001</dc:identifier>
^   <dc:title xml:lang="ja">三英傑 (信長・秀吉・家康) : 万松寺の様子</dc:title>
^   <dc:description xml:lang="ja">万松寺の様子</dc:description>
^   <nicer-lom:aggregationLevel rdf:resource="&nicer-lom;Unit"/>
^   <dc:subject xml:lang="ja">織田信秀</dc:subject>
^   <dc:subject xml:lang="ja">織田信長</dc:subject>
^   <dc:subject xml:lang="ja">万松寺</dc:subject>
^   <dc:subject xml:lang="ja">名古屋市</dc:subject>
^   <nicer-lom:context rdf:resource="&nicer-lom;SchoolEducation"/>
^   <nicer-lom:ageRange rdf:resource="&nicer-lom;HighSchool"/>

```

図6.5 LOM-RDF変換ツールの出力ファイル

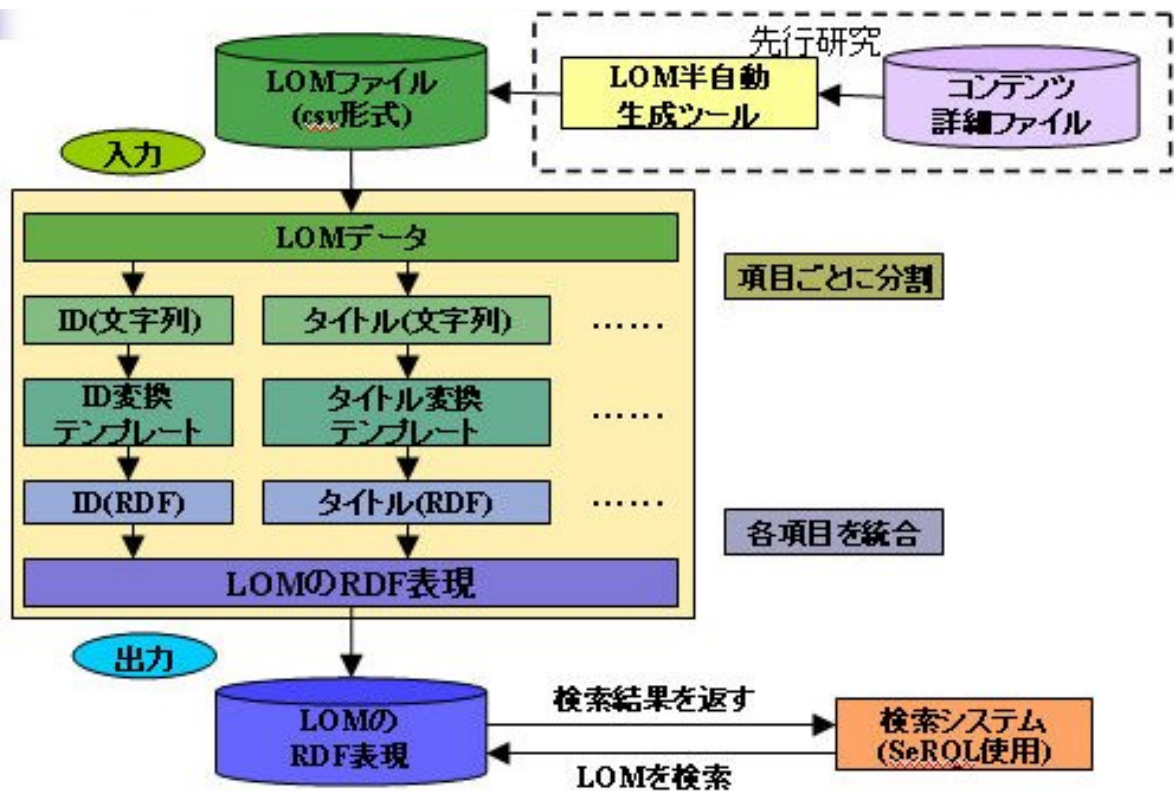


図6.6 LOM-RDF変換ツールのシステム構成図

## 6.3 RDFの検索及び評価

### 6.3.1 検索の概要

RDFで表現されたLOMによる教材アーカイブの検索の有用性の評価を行うために、RDFをデータベースとして扱い、SeRQLと呼ばれるSQLライクな言語で検索を行うツールであるSesame[4]を用いた。評価に使用したLOMについて学習指導要領と対象年齢に条件を課した時の問い合わせ文と検索結果の一部を図6.7に示す。

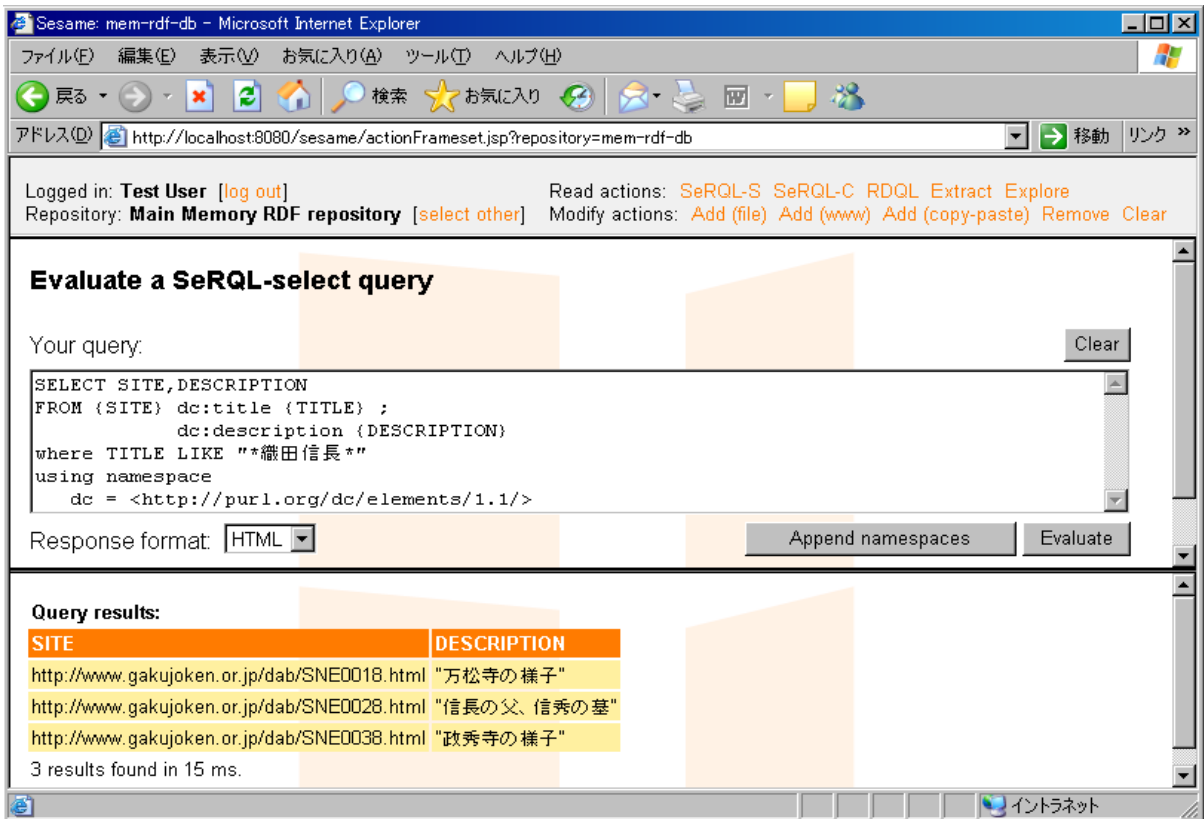


図6.7 LOM-RDF変換ツールSeRQLの検索画面

### 6.3.2 評価

小学校から高等学校までの社会系に属する教材アーカイブ400件のLOMを用いて、学習指導要領及び対象年齢の項目を利用した検索での妥当性の評価を行った。表6.1、6.2、6.3は小学校レベルの地理、歴史、公民系のアーカイブを異なった条件で検索し結果を比較したものである。それぞれ条件は

条件A 対象年齢:[小学3or4or5年生]

条件B 学習指導要領:[小学校/社会/第3学年及び第4学年]or[小学校/社会/第5学年]

条件C 学習指導要領:[地理系科目]and対象年齢:[小学3or4or5年生]

条件D 学習指導要領:[小学校/社会/第6学年]

条件E 学習指導要領:[歴史系科目]and対象年齢:[小学6年生]

条件F 学習指導要領:[公民系科目]and対象年齢:[小学6年生]

である。この中で条件C、D、Eの学習指導要領の条件に上位概念を使った検索を行っている。



表6.1 検索結果として得られた教材の例(地理)

教材アーカイブのタイトル	検索条件		
	A	B	C
長崎紹介/長崎の市電	○	○	○
名古屋周辺の地図	○	○	○
電車の乗り方	○	○	
市民の生活に必要な交通機関	○	○	
ドイツの民族の構成と衣装	○		○
トルコの民族の構成と衣装	○		○
アメリカ・ニュージャージー州	○		○
岐阜の農業	○		
事件や事故がおきたら		○	
検索条件の該当数	163	155	74

#### 地理について

地域や産業に関する分野などいかにも地理的なアーカイブについてはだいたい検索されているが、「電車の乗り方」に代表されるように地理というよりも一般常識的なことを学ぶためのアーカイブは検索されない。このことから初期段階の社会科目に関するアーカイブを検索する場合において、地域や産業に関する分野を検索する場合のみ学習指導要領を地理に指定し、その他のアーカイブについては地理と限定せずに検索したほうが良いと考えられる。

表6.2 検索結果として得られた教材の例(歴史)

教材アーカイブのタイトル	検索条件	
	D	E
豊国神社・庭園	○	○
岐阜城	○	○
トンネル窯の焼成	○	○
名古屋城・天守（愛知県）	○	
大名行列などの時代祭り	○	
最高裁判所バーチャルツアー	○	
デフレとは	○	
検索条件の該当数	170	134

歴史について

検索結果を比較すると、条件Dに比べて条件Eでは該当数が少なく、本来は検索結果として出力してほしいものから多くの漏れがある。ただし、小学6年次には歴史と公民にあたるものを同じ学年で扱うため条件Dでは公民に関するアーカイブも検索してしまうが、条件Eでは明確に歴史と指定しているため公民的分野のアーカイブは検索されない。このことから、公民的分野のアーカイブが多く混ざって検索されてしまう時に有効であると考えられる。

表6.3 検索結果として得られた教材の例(公民)

教材アーカイブのタイトル	検索条件	
	D	F
豊国神社・庭園	○	
名古屋城・天守（愛知県）	○	
大名行列などの時代祭り	○	
岐阜城	○	
最高裁判所バーチャルツアー	○	○
日本銀行バーチャル見学ツアー		○
商店街再生の方向		○
検索条件の該当数	170	27

## 公民について

検索条件の該当数を見ると条件Dの方が多く検索されているが、これは小学6年次には歴史と公民にあたるものを同じ学年で扱うために「岐阜城」など歴史に関するアーカイブが非常に多く検索されておりユーザの意図した検索を行うことができていない。これに対し条件Fにおいて学習指導要領は公民の分野として抽象的に指定しておき難易度を対象年齢で絞ることで、小学校レベルの公民系分野の教材アーカイブを得ることを示している。

## 共通

これらに共通していえることは、それぞれのアーカイブが限定された校種のみで使われるものではなく、複数の校種にまたがって使用されるアーカイブが数多く存在することで、有効的に検索することができることである。

例えば、ある歴史のアーカイブでは学習指導要領に[小学校/社会/第6学年]と[中学校/社会/歴史的分野]の両方がLOMに登録されており、教科については中学校の方の教科及び上位概念を参照し、実際の学習レベルは他の項目を参照することにより検索を実現している。特に、学習指導要領をRFS/RDFSで定義した上位概念で曖昧に指定することでより広範囲のアーカイブから検索することができる。このことからNICERでは実施していない、または検索に複雑な条件を指定する必要があるものを、RDFの上下関係を利用した曖昧な分野での検索と難易度の指定を組み合わせることで、検索したい分野とは違う分野を取り除きつつ適切な難易度の教材アーカイブを検索結果として得ることができた。

## 6.4 RDF検索のまとめ

RDFを利用した検索において、ある程度妥当な検索結果を得ることができた。しかし、現在はLOMの件数が少ないことや検索対象の教科が限られていることからLOMの件数や検索対象の教科をさらに増やして評価を行う。また、教材アーカイブの内容の詳細にまではあまり踏み込まずに評価を行っているため、詳細な内容を加味した評価、検討を行う。

## 第7章

# 結論と今後の展望

### 7.1 結論

本論文では、新たに<“動画知”, “自然言語知”, “メタデータ”>という3項組の知識表現を提案し、デジタルアーカイブ化する手法について述べた。以下に本研究の成果についてまとめる。

第1章では、本研究の目的と背景について述べるとともに、論文の構成で本研究の位置づけを示した。

第2章では、知識利用に、新しい知識表現を応用し、LSI法やSVMを導入し、独創的なWebページ検索システムを新たに開発した。

第3章では、知識獲得の一手法として、日本で始めてKT法による潜在問題分析システムを開発し評価した。

第4章では、本Webページ検索システムは、新規拡張機能として、潜在的意味インデキシング(LSI)法を用い、新たにサポートベクトルマシン(SVM)による適合性フィードバックを導入した。LSIは、索引語-文書行列に対して特異値分解を行い、類義語処理用の類義語辞典を作成することなく数値的な処理により実現できるという特長を持つからである。

第5章では、適合性フィードバックあるなしにおけるシステム性能比較を行い、検索精度の向上の度合いを評価した。

第6章では、より高度な意味処理を行うため、最新技術であるセマンティックWeb技術を用い、曖昧な検索ができる新しいe-Learning支援システムの試行を行った。

本論文の成果は、以下の6点である。

- (1) 動画知、自然言語知とメタデータの3項組の新しい知識表現を用いる効果は、暗黙知を一部含む動画や静止画をテキスト検索で抽出できたこと。
- (2) 動画知の検索に、LOMを使用することにより、インターフェースとしてデジタルアーカイブの検索を可能にし、これによりNICER(教育情報ナショナルセンター)の検索システムが利用可能となること。

- ( 3 ) 暗黙知から形式知を引き出す「表出化」の一つのツールがコンピュータ化でき、Web ページ検索結果においても当初期待した結果を得ることができた。
- ( 4 ) 自然言語で表現された質問文書に対して索引語の重みベクトルを生成し、LSI 法により索引語-文書行列を参照して類似度の高い Web ページを検索・出力する。本システムで登録されているデータでは、入力する次元数を 300 ~ 500 程度にするとよい検索精度が得られると分かったこと。
- ( 5 ) NICER の検索システムではキーワードの付け方によって検索できないコンテンツがでてくるが本システムを用いることによってより広範囲な検索を行うことができる。適合性フィードバックを行うことにより、システムとしての精度向上が確認できたこと。
- ( 6 ) 曖昧な検索ができる新しい e-Learning 支援システムを実現するため、LOM の RDF/RDFS 表現を用い、LOM の構造オントロジーを作成し、第 4 章の Web ページ検索システムと同様に情報検索ができたこと。

## 7.2 今後の展望

今後の展望として、以下の 3 点がある。

- ( 1 ) K T 法による潜在的問題分析システムで表出化された新しい知識を、Web ページ検索システムで利用できるようにしたい。
- ( 2 ) 暗黙知の一部を含む映像ファイルのメタデータのつけ方に工夫をしたい。例えば、どんな場面や音ならどんなメタデータを付けるかシソーラスの調査が必要である。
- ( 3 ) 第 4 章で論じた Web ページ検索システムと、第 6 章で論じたより高度な意味処理を行う新しい e-Learning 支援システムとの比較・評価が、今後に残された課題である。1,000 件程度の検索で良好な性能がでるなら、最新技術であるセマンティック Web 技術を用い、曖昧な検索ができる新しい e-Learning 支援システムに切り替えたい。良好な性能がでなくても、より高度な意味処理を必要とする検索分野で活用したい。

# 謝辞

本研究を行うにあたり、お世話になった方々へ感謝の意を述べさせていただきます。

主指導教官である北陸先端科学技術大学院大学 知識科学研究科 國藤 進 教授には、博士後期課程の3年間において、温かくご指導及び御鞭撻を賜り、深謝いたします。

本論文をまとめるにあたり、ご尽力頂きました静岡大学 弓野 憲一 教授、北陸先端科学技術大学院大学 知識科学教育研究センター 宮田 一乗 教授、北陸先端科学技術大学院大学 知識科学研究科 藤波 努 助教授、北陸先端科学技術大学院大学 知識科学教育研究センター 金井 秀明 助教授に感謝いたします。

博士後期課程の副テーマを快くご指導いただきました慶応義塾大学 常務理事の斉藤信男教授、北陸先端科学技術大学院大学 知識科学研究科 池田 満 教授に感謝いたします。

本研究の共同研究者の芝浦工業大学の若木利子教授に感謝いたします。また、潜在的問題分析システムの開発並びにアンケート調査にご尽力されました芝浦工業大学の河村綾祐様はじめ、若木研の研究室の方々に感謝申し上げます。

日頃から有益なご助言を頂き、励ましていただいた北陸先端科学技術大学院大学 知識科学研究科 助手 三浦 元喜 博士、羽山 徹彩 博士と、お世話になった國藤研の研究室の方々に感謝いたします。

## 参考文献

- [1] Michael Polanyi : THE TACIT DIMENSION , Routledge & Kegan Paul Ltd. ,London ,1966( マイケル・ポランニー: 暗黙知の次元: 言語から非言語へ, 佐藤敬三訳, 伊東俊太郎序, 紀伊国屋書店, 1980)
- [2] 野中郁次郎, 紺野登: 知識創造の方法論【ナレッジワーカーの作法】, 東洋経済新報社, 2003
- [3] 督永涼平, 樋口茂樹, 若木利子, 新田克己: 判例データベースからの類似文書検索システムの開発と評価, 平成 15 年度電子情報通信学会東京支部学生会研究発表会 D-8 p108, 2004
- [4] 北 研二, 津田 和彦, 獅々堀正幹: 情報検索アルゴリズム, 共立出版, pp.51-89, 2002
- [5] 拓殖 覚, 獅々堀正幹, 北 研二: サポートベクターマシンによる適合性フィードバックを用いた情報検索, 自然言語処理 141-14, 2001, pp.83-88
- [6] S. Sawai, T. Wakaki, M. Ikeda, S. Kunifuji: Ubiquitous Networking for GENES Society: e-learning Tools and Digital Archives for Education with Significant Use of Cultural Heritage Contents, proc. of the 1st Int. Workshop on Interoperability of Web-based Educational Systems, WWW2005, pp. 71-74, 2005]
- [7] G.Antoniou, F.v.Harmelen: A Semantic Web Primer, The MIT Press, pp. 61-108, 2004
- [8] IEEE1484.12.1-2002 LOM Draft Standard:  
<http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html>
- [9] Dublin Core Metadata Initiative: <http://dublincore.org/>
- [10] 教育情報ナショナルセンター(NICER) : <http://www.nicer.go.jp/>
- [11] 仲林 清, 清水 康敬, 山田 恒夫: e-Learning 標準化技術の開発と実践の新しい展開 -SCORMとLOMを中心に-, 人工知能学会誌21巻1号, pp. 92-98, 2006

- [12]小学校学習指導要領,文部科学省,平成 10 年 12 月告示,平成 15 年 12 月一部修正  
中学校学習指導要領,文部科学省,平成 10 年 12 月告示,平成 15 年 12 月一部修正  
高等学校学習指導要領,文部科学省,平成 11 年 3 月告示,平成 15 年 12 月一部修正
- [13] 文部科学省, 学習指導要領: [http://www.mext.go.jp/b¥\\_menu/shuppan/sonota/990301.htm](http://www.mext.go.jp/b¥_menu/shuppan/sonota/990301.htm)
- [14]C. H. Kepner,B. B. Tregoe,上野一郎訳: 新・管理者の判断力,産業能率大学出版部,1999
- [15]高多清在,問題解決と意思決定の世界標準・KT 法,実業之日本社,1999
- [16]國藤 進: 知的グループウェアによるナレッジマネジメント,日科技連出版社,2001
- [17]八重樫理人,中村恵一,鈴木智,橋浦弘明,小宮誠一: KT 法を改良したグループ演習における個人の成績評価方法の提案,情報教育シンポジウム論文集,情報処理学会,pp167-172,2005
- [18]中村恵一,鈴木智,橋浦弘明,八重樫理人,小宮誠一: KT 法の導入によるグループ問題解決支援システム - 適用実験によるコーディネータ支援機能の有効性確認,情報処理学会論文誌,Vol.46 No.1,pp127-137, 2005
- [19]佐藤雅彦,國藤 進: KT 法支援システムの実現に向けて,北陸先端科学技術大学院大学 國藤研究室主催発想支援ツール・シンポジウム, 1992
- [20]佐藤雅之: 思考支援システムのアプローチによる意思決定の試み - KT 法に基づいた意思決定支援システム -,北陸先端科学技術大学院大学修士学位論文,1994
- [21] IT 戦略本部: IT 新改革戦略, 2006
- [22]井上史子,林徳治: 情報化の影に対応するための力を育成する授業実践,学習情報研究, Vol.184,pp35-38, 2005
- [23]澤井 進,若木 利子,池田 満,國藤 進(2006),暗黙知・形式知を用いたeラーニング用Web検索エンジン,日本教育情報学会,Vol.21 No.2,pp13-18,2006
- [24] 茶筌: <http://chasen.aist-nara.ac.jp/>
- [25]TinySVM, <http://chasen.org/~taku/software/TinySVM/>



[26]検索エンジン: <http://www.ailab.se.shibaura-it.ac.jp/p02078a/>

[27] Sesame: <http://www.openrdf.org/>

# 本研究に関する発表論文

## 論文誌

- [1] 神田泰典、杉本正勝、澤井 進：エンドユーザのための日本語によるプログラミング、情報処理学会誌、Vol.21, No.3, pp.197-199,1980年3月
- [2] 澤井 進: デジタルコンテンツ、電子情報通信学会, Vol.80, No.2, pp.197-199, 1997年2月
- [3] 澤井 進、若木 利子、池田 満、國藤 進：暗黙知・形式知を用いたeラーニング用Web検索エンジン、日本教育情報学会誌第21巻第2号、pp.13-20、2006年
- [4] 澤井 進、若木 利子、國藤 進：潜在的問題分析システムの教育への一応用、日本教育情報学会誌、2006年（条件付き採録）

## 査読付き国際会議

- [5] Susumu Sawai, Masanori Fukusimai, Masakatsu Sugimoto, Naoya Ukai: Knowledge Representation and Machine Translation, COLING82 (9th International Conference on Computational Linguistics), July 5-10, Prague, 1988
- [6] Susumu Sawai, Terano Takao: "YUAI, A Knowledge acquisition support system for classification-type problem", TIMS89, Osaka, 1989
- [7] Susumu Sawai, Tosiko Wakaki, Mitsuru Ikeda, Susumu Kunifuji: "UBIQUITOUS NETWORKING FOR GENES SOCIETY", proc. of the 1st Int. Workshop on Interoperability of Web-based Educational Systems, WWW2005, pp.71-74, 2005
- [8] Susumu Sawai, Tosiko Wakaki, Masakazu Tachikawa, Susumu Kunifuji: "The Application of Semantic Web Technology to e-Learning Support System", proc. of the 1st Int. Workshop on Semantic Web Annotations for Multimedia 2006, WWW2005, 2006

## 国際会議(招待)

[9] Susumu Sawai: GENES Project - A Japanese Initiative for Digital Archives for Education with Significant Use of Cultural Heritage Content, 12th World Wide Web Conference, Budapest, 2003

[10] Susumu Sawai, Tosiko Wakaki, Tomio Saku, Junnichi Yamanisi, Susumu Kunifuji: Mobile and Ubiquitous Networking for GENES Society, Workshop on Ubiquitous and Mobile Computing for Educational Communities, Amsterdam, 2003

[11] Susumu Sawai, Tosiko Wakaki, Tomio Saku, Junnichi Yamanisi, Susumu Kunifuji: "Mobile and Ubiquitous Networking for GENES Society", 13th World Wide Web Conference, New York, May 2004

[12] Susumu Sawai, Tosiko Wakaki, Susumu Kunifuji: "Long-term preservation of digital archives by using potential problem analysis and natural language searching", 14th World Wide Web Conference, Makuhari, May 2005 ( Online Journal )

## 査読なし国内会議

[13] 沢井進ほか：知識表現 F K R - 0 とその機械翻訳への応用、情報処理学会、人工知能と対話技法研究会資料19、1980

[14] 澤井 進、森 勇喜、若木 利子、板倉 弘幸: 自然言語による e-learning 用 Web ページ検索システム、情報処理学会第 67 回全国大会、2005 年

[15] 澤井 進、若木 利子、國藤 進：暗黙知・形式知一対の知識表現及び創造支援への応用、第 2 回知識創造支援システム・シンポジウム、日本創造学会、2005 年

## 花の画像の検索事例

### e-learning検索エンジン

探したいに学習コンテンツについて自然言語で入力してください  
理科の教材として花の画像がほしい

<input type="radio"/> 圧縮率 <input checked="" type="radio"/> 直接入力	出力文書数 <input type="text" value="5"/>
LSI次元 <input type="text" value="400"/>	検索データベース <input type="text" value="埼玉"/>

**LOM検索**

使用しない  and検索  or検索

初等中等教育  
 高等教育

実行  [help](#)

## 400次元での検索

該当数: 769

num	LOMID	TITLE	SIM	FB
1	<a href="#">050000000543</a>	校庭の草花	0.541627	<input type="radio"/> <input type="checkbox"/>
2	<a href="#">050000000189</a>	四季の草花牛島小	0.500281	<input type="radio"/> <input type="checkbox"/>
3	<a href="#">050000000538</a>	校庭の草花	0.469718	<input type="radio"/> <input type="checkbox"/>
4	<a href="#">050000000545</a>	校庭の草花	0.444969	<input type="radio"/> <input type="checkbox"/>
5	<a href="#">050000000539</a>	校庭の草花	0.436347	<input type="radio"/> <input type="checkbox"/>

出力文書数<sup>5</sup>

FBの実行

入力された文章

---

理科の教材として花の画像がほしい

---

戻る

ID & TITLE	THUMBNAIL
050000000543	
校庭の草花	
050000000189	
四季の草花牛島小	
050000000538	
校庭の草花	
050000000545	
校庭の草花	
050000000539	
校庭の草花	

素材表題      こぼれ種から成長したカスミソウの仲間

素材内容

こぼれた種から成長して白っぽい花を咲かせている。  
花は小さいが数が多い。

登 録 者                      越谷市教育工学研究会  
登 録 者                      1995. 12. 15  
取 材 者                      黒崎政弘  
取材場所                      埼玉県越谷市  
取材日付                      1995. 8. 27  
著作権等                      越谷市教育工学研究会 黒崎政弘



素材表題      ポインセチア

素材内容

ポインセチア  
中南米高地原産。寒さにも強い。  
クリスマスを彩る花として好んで作られている。

登 録 者                      越谷市教育工学研究会  
登 録 者                      1995. 12. 15  
取 材 者                      辺見 愉偉  
取材場所                      埼玉県春日部市牛島地区  
取材日付                      1990. 4. 10-1995. 10. 30



素材表題 成長したハルノノゲシ

素材内容

茎を長く成長させ黄色い花を咲かせている。  
ぎざぎざした葉が上部に近くなるほど小さく細くなっている。

登録者 越谷市教育工学研究会  
登録者 1995. 12. 15  
取材者 黒崎政弘  
取材場所 埼玉県越谷市  
取材日付 1995. 8. 27  
著作権等 越谷市教育工学研究会 黒崎政弘



素材表題 オニタビラコの花

素材内容

キク科の植物で、黄色い花が咲く。  
似た仲間にコオニタビラコがある。

登録者 越谷市教育工学研究会  
登録者 1995. 12. 15  
取材者 黒崎政弘  
取材場所 埼玉県越谷市  
取材日付 1995. 8. 27  
著作権等 越谷市教育工学研究会 黒崎政弘





素材表題 大きく成長したハルノノゲシ

素材内容

80cmほどにも成長している。  
花をつけるくきが分かれ、それぞれに黄色い花をつけている。  
大きくよく目立つ。

登録者 越谷市教育工学会  
登録者 1995. 12. 15  
取材者 黒崎政弘  
取材場所 埼玉県越谷市  
取材日付 1995. 8. 27  
著作権等 越谷市教育工学会 黒崎政弘



## 検索データベース素材(1)「埼玉」

ネットワークパスワードの入力	
認証情報を入力してください。	
<input type="button" value="OK"/>	
<input type="button" value="キャンセル"/>	
リソース:	教育コンテンツ(デジタルアーカイブ)@学情研
ユーザ名(U):	<input type="text" value="▲▲▲▲▲"/>
パスワード(P):	<input type="text"/>
<input checked="" type="checkbox"/>	このパスワードを保存する(S)

## 教育コンテンツ(デジタルアーカイブ) @学情研



Copyright (c) 1997-2001 Gakujoken All Rights reserved.

財団法人 学習ソフトウェア情報研究センター

(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研
(財)学情研	(財)学情研	(財)学情研	社)教科書協会	電子開発学園	東京書籍(株)	(株)日本IBM
日本教育新聞	日本電気(株)	(株)日立製作所	富士通(株)	(株)文溪堂		教育出版(株)
磁光電子(株)	シャープSP(株) 高精細国宝仏像	帝国書院(株)	財団法人日本児童教育振興財団	松下電器産業株	光村図書出版株	
	(株)PFU	マイクロソフト(株)	日本ビクター株			
				(財)学情研	(財)学情研	(財)学情研
(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研
	JAPET					

[\[操作ガイド\]](#)

[\[日本の伝統的工芸、工業\]](#)

[\[地区版\]](#)

[\[海外\(国際理解\)版\]](#)

埼玉

ページ移動

※本映像には電子透かしが入っています。学校の授業、教師の教授法研究以外で本映像の複写・改変は著作権法で禁じられていますので、ご注意ください。

※現在、「日本の伝統的工芸、工業」では北海道・秋田・岩手を、「地区版」では、北海道・秋田・岩手・群馬・埼玉・千葉・東京・神奈川・富山・石川・愛知・岐阜・奈良・兵庫・岡山・鳥取・愛媛・高知・鹿児島・沖縄、「海外」を公開中です。

今後、随時情報を追加予定です。

[学情研ホームページTOPへ](#)

# 埼玉 地区版

## 目次

シリーズ名	シリーズ名	シリーズ名
<a href="#">渡良瀬川(源流から利根川合流まで)</a>	<a href="#">樹木の花(春の花)</a>	<a href="#">植物と環境に合わせた形態</a>
<a href="#">水のゆくえ</a>	<a href="#">樹木の花(夏、秋の花)</a>	<a href="#">県木1</a>
<a href="#">水の浄化と川</a>	<a href="#">草花の花(春の花)</a>	<a href="#">高山植物(大雪山の植物)</a>
<a href="#">ゴミのゆくえ(1)</a>	<a href="#">草花の花(夏、秋、冬の花)</a>	<a href="#">電車・バスの乗り方</a>
<a href="#">ゴミのゆくえ(2)</a>	<a href="#">学校の周りの雑草(春、夏)</a>	<a href="#">椅子</a>
<a href="#">埼玉県越谷市の伝統工芸 だるま</a>	<a href="#">学校の周りの木(常緑樹・針葉樹)1</a>	<a href="#">商店街のお店</a>
<a href="#">伝統工芸 岩槻人形(ひな人形づくり)</a>	<a href="#">学校の周りの木(落葉樹・針葉樹)2</a>	<a href="#">たまごの料理</a>
<a href="#">大名行列などの時代祭り</a>	<a href="#">学校の周りの木(常緑樹・広葉樹)3</a>	<a href="#">算数・数学の図形</a>
<a href="#">お線香づくり 1 粉までの過程</a>	<a href="#">学校の周りの木(広葉樹・落葉樹)4</a>	<a href="#">図形</a>
<a href="#">お線香づくり 2 杉の粉を製品にする過程</a>	<a href="#">学校の周りの木(広葉樹・落葉樹)5</a>	<a href="#">書写の基本</a>
<a href="#">校庭の草花</a>	<a href="#">学校の周りの木(常緑竹・常緑笹)6</a>	<a href="#">高麗川 巾着田・曼珠沙華</a>

## ～関連リンク集～

[「埼玉県ホームページ」](#)

[「全国旅行・観光情報\(埼玉県\)」](#)

[表紙にもどる](#)

シリーズ表題 水の浄化と川

校 種	小学校
学 年	4年, 5年, 6年
教 科 名	社会, 理科, 環境教育
単 元 名	水との戦い
内 容	昔から農業用水の確保、水の確保は大切なことであった。 4年の社会の学習の発展、環境教育の視点から資料とする。 現在の飲み水の浄化と配水などの様子である。
登 録 者	越谷市教育工学研究会
著作権等	越谷市教育工学研究会 黒崎政弘



[シリーズを見る](#)

[「埼玉」の目次へもどる](#)



# 水の浄化と川

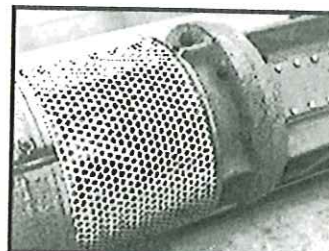
深井戸



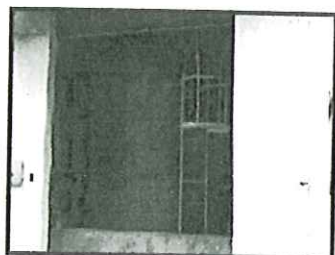
深い井戸で使われるくだ



深井戸で使われる管のアップ



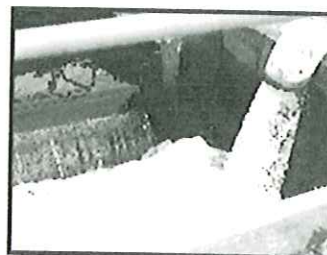
消毒用の次亜塩素酸ソーダのタンク



着水池



混和池



混和池の消毒剤と水を混ぜる装置



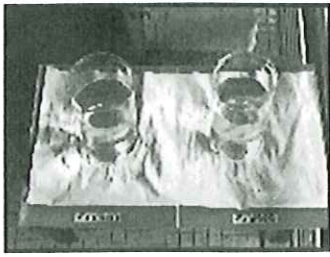
沈澱池



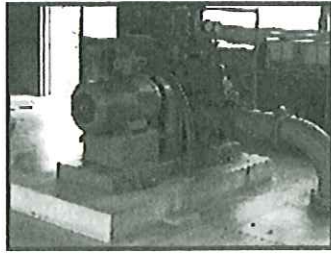
室内にあるろ過池



ろ過前とろ過後の水



水道水を送るポンプ



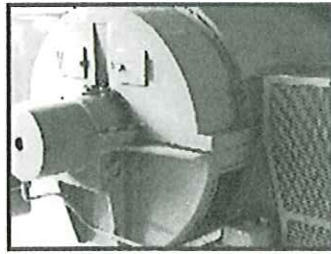
水道水を送るポンプが並んだ様子



ポンプの電気を調節する配電盤



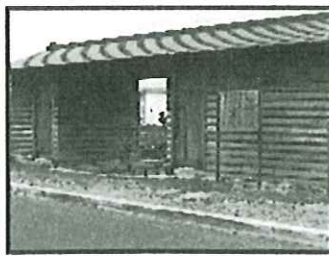
ポンプの予備電源となる発電機



水利用のために川の流れを変える



綾瀬川と出羽堀の合流点にある河岸後 蒲生藤助河岸跡の説明



藤助河岸跡の前にある藤助河岸で舟運してい



藤助河岸跡と川



川を広げる工事



かつての川の跡



もどる





## 検索データベース素材(2)「愛知」

### 教育コンテンツ(デジタルアーカイブ) @学情研



Copyright (c) 1997-2001 Gakujoken All Rights reserved.

財団法人 学習ソフトウェア情報研究センター

(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研
(財)学情研	(財)学情研	(財)学情研	社)教科書協会	電子開発学園	東京書籍(株)	(株)日本IBM
日本教育新聞	日本電気(株)	(株)日立製作所	富士通(株)	(株)文溪堂		教育出版(株)
磁光電子(株)	シャープSP(株) 高精細国宝仏像	帝国書院(株)	財団法人日本児童教育振興財団	松下電器産業株	光村図書出版株	
	(株)PFU	マイクソフ(株)	日本ビクター株			
				(財)学情研	(財)学情研	(財)学情研
(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研	(財)学情研
	JAPET					

[\[操作ガイド\]](#)

[\[日本の伝統的工芸、工業\]](#)

[\[地区版\]](#)

[\[海外\(国際理解\)版\]](#)

愛知

ページ移動

※本映像には電子透かしが入っています。学校の授業、教師の教授法研究以外で本映像の複写・改変は著作権法で禁じられていますので、ご注意ください。

※現在、「日本の伝統的工芸、工業」では北海道・秋田・岩手を、「地区版」では、北海道・秋田・岩手・群馬・埼玉・千葉・東京・神奈川・富山・石川・愛知・岐阜・奈良・兵庫・岡山・鳥取・愛媛・高知・鹿児島・沖縄、「海外」を公開中です。

今後、随時情報を追加予定です。

[学情研ホームページTOPへ](#)

# 愛知 地区版

## 目次

シリーズ名	シリーズ名
<a href="#">三英傑(織田信長 豊臣秀吉 徳川家康)</a>	<a href="#">地名からさぐる地域の歴史</a>
<a href="#">低地の人々のくらし</a>	<a href="#">木材の種類</a>
<a href="#">新美南吉</a>	

## ～関連リンク集～

[「愛知県ホームページ」](#)

[「全国旅行・観光情報\(愛知県\)」](#)

[表紙にもどる](#)

シリーズ表題 三英傑（織田信長 豊臣秀吉 徳川家康）

校 種	小学校
学 年	小学校 6 年
教 科 名	社会科
単 元 名	日本の歴史（安土、桃山時代）
内 容	織田信長に関する史跡やその地域の様子。 豊臣秀吉に関する史跡やその地域の様子。 徳川家康に関する史跡やその地域の様子。 愛知県内及び、近郊にある歴史的資料の紹介。
登 録 者	榊原貴久
登録日付	1995年



[シリーズを見る](#)

[「愛知」の目次へもどる](#)



三英傑（織田信長 豊臣秀吉 徳川家康）

万松寺の様子



信長の父、信秀の墓



政秀寺の様子



信長の頃の清洲城跡



現在の清洲城の天守閣



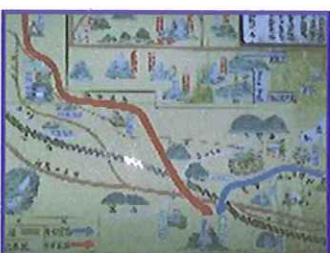
清洲城天守から見た城下の様子



熱田神宮



桶狭間の合戦の軍団進路



桶狭間の合戦古戦場



今川義元が本陣を置いた場所

今川義元が討たれたといわれる場所

熱田神宮にある信長塚



岐阜城天守閣

岐阜城天守閣から見た岐阜城下（西方）

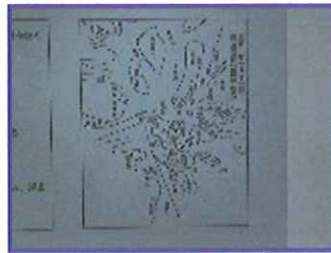
岐阜城天守閣から見た岐阜城下



信長が住んだ御殿跡

長島一向一揆攻略図

長島の田園地帯の様子



長島城の模型

総見寺

秀吉が生まれたと伝えられる場所



秀吉が使ったと伝えられる場所

秀吉が植えたと伝えられる柵（ひいらぎ）

秀吉が初めて瓢箪（ひょうたん）



墨俣の一夜城

犬山城全景

犬山城天守閣



犬山城天守閣から見た犬山城下

犬山城天守閣から小牧山方面を望む

大阪城全図





大阪城天守閣



豊国神社正面



小牧山の様子



小牧城（歴史資料館）



小牧城から犬山方面を望む



関ヶ原合戦陣形図



家康が最初に陣を置いた桃配山

桃配山から見た関ヶ原

家康が腰掛けて采配をふるったといわれる





石田三成が陣を置いた笹尾山入り口

石田三成の陣跡

石田三成の陣から見た家康の（桃配山）方



小早川秀秋が陣を置いた松尾山

関ヶ原全景

関ヶ原の様子



関ヶ原の合戦開戦地

関ヶ原の合戦決戦地

家康が最後に陣を置いた場所



家康が戦勝を祝った場所

関ヶ原の西首塚

名古屋城天守閣



もどる



## シリーズ表題

## 地名からさぐる地域の歴史

校 種	小学校
学 年	小学校6年
教 科 名	社会科
単 元 名	日本の歴史 ー地域の歴史ー
内 容	身近な地域の地名や街道名（通りの名前）について調べていくことで、それぞれの時代の歴史上の出来事や代表的な人物とのつながり、ある時代の人々の生活をさぐることができる。
登 録 者	余合久恭, 立花賢修
登録日付	1995年



[シリーズを見る](#)

[「愛知」の目次へもどる](#)



## 地名からさぐる地域の歴史

太閤通の現在の様子



太閤通 豊臣秀吉ゆかりの豊国神社



砦前 案内板



砦前 散策路図



砦前 大高城跡の地形



砦前 大高城跡石碑



砦前 大高城跡の様子



砦前 丸根砦跡の看板



砦前 丸根砦跡の様子



砦前 鷺津砦公園

砦前 鷺津砦跡の様子

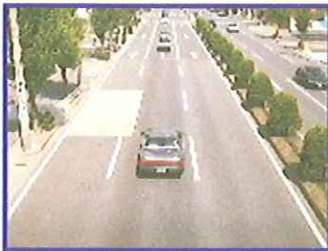
妙音通(ミヨウノド' 利)



妙音通の様子

妙音通 藤原師長ゆかりの師長橋

妙音通 藤原師長 居跡



鎌倉街道

塩付通(ソツトノ木)散策路図

塩付通の様子



門前町(モンゼンチョウ)

門前町 西本願寺別院



もどる

## 【添付資料4】

### 検索システムのユーザ・アンケート用紙

#### □ 検索システムのアンケートについてご説明いたします。

1. 「問題」の設定：「検索システムの選定」を「問題」とする。
2. 「検索システムの選定」のブレイクダウン：「評価基準」として、「使いやすさ」、「見易さ（わかりやすさ）」、「適切性」、「有用性」という4つを選ぶ。

適切性(pertinence)：検索質問に対して、ユーザが目的とする文書を検索したか？

有用性(usefulness)：ユーザにとって役立つ文書を検索したか？

3. 「代替案」の選定：「代替案」として、「Web ページ検索システム (WPS)」、「NICER」、「YAHOO」3つを選ぶ。
4. 今決めた3つの事を階層図にする。

#### □ 一対比較をする

4つの評価基準の項目を次のような一対比較をします（一対比較表）。

対	使いやすさ	見易さ	適切性	有用性
使いやすさ				
見易さ				
適切性				
有用性				

次に一対比較値の数値について説明します。

一対比較値	意 味
1	両方の項目が同じくらい重要
3	前の項目の方が後の方より若干重要
5	前の項目の方が後よりも重要
7	前の項目の方が後の方よりもかなり重要
9	前の項目よりも後の方が絶対的に重要

例) 「使いやすさ」より「見易さ」のほうが重要ならば表に記入すべき数値は5であり、反対の「見易さ」対「使いやすさ」は必然的1/5という事になります。



例) 「使いやすさ」より「見易さ」のほうが重要ならば表に記入すべき数値は5であり、反対の「見易さ」対「使いやすさ」は必然的1/5という事になります。

対	使いやすさ	見易さ	適切性	有用性
使いやすさ		1	5	
見易さ	1/5		1	
適切性				1
有用性				1

#### □ ウエイトの総合化

次に各評価基準について Web ページ検索システム (WPS), NICER, YAHOO を一対比較していきます。

(例) WPS と NICER どちらがどれだけ使いやすいかについて、WPS のほうが NICER よりはるかに使いやすければ9ないしは8という具合に一対比較表に書き込んでください。

使いやすさについて

対	WPS	NICER	YAHOO
WPS	1	2	3
NICER	1/2	1	2
YAHOO	1/3	1/2	1

見易さ (わかりやすさ) について

対	WPS	NICER	YAHOO
WPS	1	1/5	1/2
NICER	5	1	7
YAHOO	2	1/7	1

適切性について

対	WPS	NICER	YAHOO
WPS	1	3	2
NICER	1/3	1	1/2
YAHOO	1/2	2	1

有用性について

対	WPS	B車	YAHOO
WPS	1	1/2	1/2
NICER	2	1	1
YAHOO	2	1	1

(2) Web ページ検索システム (WPS) に関するアンケート□

□ 検索システムの4つの評価基準について一対比較し、ウエイトを算出して下さい。

対	使いやすさ	見易さ	適切性	有用性
使いやすさ				
見易さ				
適切性				
有用性				

使いやすさについて

対	WPS	NICER	YAHOO
WPS			
NICER			
YAHOO			

見易さ (わかりやすさ) について

対	WPS	NICER	YAHOO
WPS			
NICER			
YAHOO			

適切性について

対	WPS	NICER	YAHOO
WPS			
NICER			
YAHOO			

有用性について

対	WPS	NICER	YAHOO
WPS			
NICER			
YAHOO			

(3) Web ページ検索システム (WPS) に関するアンケート□

(1)比較対象：Web ページ検索システム (WPS), NICER, YAHOO

(2)比較条件：

学情研ネットワーク研究グループ (GENES) 愛知地区の Web ページを検索対象とする。

(理由：3つの比較対象の検索システムで同じデータが検索できるので。)

検索速度は、データ量が3つの比較対象の検索システムで異なるので問わない。

動画には、動画の一形態の静止画も含まれます。

(3)評価基準 (比較カテゴリー)：

- 「動画検索」画面についてお尋ねいたします。使い易さ、見易さ (分かり易さ),
- ユーザから見た検索結果の有効性についてお尋ねします。適切性, 有用性

□ 検索システムの使い易さについてお尋ねいたします

(たとえば、検索システムで同じデータが参照できる「愛知」で「織田信長」と「桶狭間」の戦いを検索キーワードとして「今川義元が打たれた石碑」の写真を検索結果として得る場合、自然言語の質問文は「織田信長が桶狭間に陣出した歴史について知りたい」となる)

【質問1】 Web ページ検索システム (WPS) の「動画検索」画面で「探したい動画を自然言語で入力する」を用いて検索し易いですか？

- 大変使い易い, □やや使い易い, □どちらとも言えない, □やや使い難い, □大変使い難い

【質問2】 YAHOO で探したい動画を検索し易いですか？

- 大変使い易い, □やや使い易い, □どちらとも言えない, □やや使い難い, □大変使い難い

【質問3】 NICER の「検索」画面で探したい動画を検索し易いですか？

- 大変使い易い, □やや使い易い, □どちらとも言えない, □やや使い難い, □大変使い難い

□ 検索画面の見易さについてお尋ねいたします

【質問4】 Web ページ検索システム (WPS) の「動画検索」画面は役立つですか？

- 大変見易い, □やや見易い, □どちらとも言えない, □やや見にくい, □大変見にくい

【質問5】 YAHOO の画面は役立つですか？

- 大変見易い, □やや見易い, □どちらとも言えない, □やや見にくい, □大変見にくい

【質問6】 NICER の「検索」画面は役立つですか？

- 大変見易い, □やや見易い, □どちらとも言えない, □やや見にくい, □大変見にくい

□ 検索の適切性についてお尋ねいたします

【質問 7】 Web ページ検索システム (WPS) は目的とする文書を検索したか？

大変適切だ,  やや適切だ,  どちらとも言えない,  やや不適切,  大変不適切

【質問 8】 YAHOO は目的とする文書を検索したか？

大変適切だ,  やや適切だ,  どちらとも言えない,  やや不適切,  大変不適切

【質問 9】 NICER は目的とする文書を検索したか？

大変適切だ,  やや適切だ,  どちらとも言えない,  やや不適切,  大変不適切

四 検索の有用性についてお尋ねいたします

【質問 10】 Web ページ検索システム (WPS) は役立つ文書を検索したか？

大変役立つ,  やや役立つ,  どちらとも言えない,  やや役立たない,  全く役立たない

【質問 11】 YAHOO は役立つ文書を検索したか？

大変役立つ,  やや役立つ,  どちらとも言えない,  やや役立たない,  全く役立たない

【質問 12】 NICER は役立つ文書を検索したか？

大変役立つ,  やや役立つ,  どちらとも言えない,  やや役立たない,  全く役立たない