

## 番組メタファーを用いた情報検索結果の提示方式

近藤 宏行<sup>†</sup> 角谷 和俊<sup>††</sup> 田中 克己<sup>††</sup>

本稿では、データベースへの問い合わせの結果表示のための新たな提示方式として、番組メタファー表示を提案する。番組メタファーで表示することで、単なる文字の羅列であるよりも、よりユーザの求めているものを的確にかつ魅力的に表現する提示を実現する。番組メタファーは、既存のテレビ番組にそって作成した疑似テレビ番組スタイルである。検索結果をあたかもテレビで放送しているように見せることで、ユーザにとってより効果的で、親しみやすい表現を行うことができる。適当な番組メタファーを動的に選択することにより、単に検索結果を提示するだけでなく、検索された結果の特徴に応じて提示スタイルを変更することが可能になる。提示スタイルの決定には検索結果のデータ集合の分布状態、結果の内容、注目する属性とメタファーの関係を考慮する。更に、提案した方式をもとに、番組記述言語 TVML を用いて実現したプロトタイプシステム Waffle について述べる。プロトタイプシステムは、入力インタフェース部、メタファー決定部、スクリプト作成部、結果提示部から構成される。

### A TV-program Metaphor Based Presentation System for Retrieval Results

HIROYUKI KONDO<sup>†</sup>, KAZUTOSHI SUMIYA<sup>††</sup> and KATSUMI TANAKA<sup>††</sup>

In this paper, we propose a TV-program Metaphor Based Presentation System for Retrieval Results. With the TV-program metaphor based presentation system we can present information which a user asks for more accurately and attractively. A TV-program metaphor is a pseudo-TV-program. By dynamically selecting a TV-program metaphor we can change presentation styles according to the characteristics of retrieval results. To determine a presentation style we use distribution of retrieval results, contents of retrieval results, and the relationship between attribute and TV-program metaphors. Finally, we present our prototype Waffle which was implemented by TVML. Waffle is composed of four modules, input-interface modules: metaphor-definition module, script-generation module, and answer-presentation module.

#### 1. はじめに

現在、様々な分野でデータベースに莫大な情報が生成、蓄積されるようになってきている。そのため膨大な情報データから迅速かつ的確に必要なものを取り出し提示するということが重要な問題となってきている。ユーザの問い合わせが求める答を的確に表現するために様々な方法が考えられているがその中には画像や、音声、動画などマルチメディアコンテンツを用いたものや、3次元表示を行うものや、キャラクタを用

いるものなどがある。本研究ではユーザに検索データの分布や全体内容の分かり易い提示を行う、親しみやすく飽きさせない新しいインタフェースとして番組メタファー表示を提案する。今回、この番組メタファー表示を実現するために番組記述言語 TVML(TV Program Making Language)<sup>1)2)</sup>を用いた。本研究では素材として、放送のデジタル化に伴いますます多チャンネル化することが予想される、デジタル放送のインデックス情報である電子番組表(EPG)を対象とした、また、これらの考えをもとに試作システムの実装を行っている。

#### 2. 番組メタファー

##### 2.1 番組メタファー

本研究では番組メタファー表示を提案するが、番組メタファーとは、既存のテレビ番組に即した演出を擬似的に行うことである。例えば、ニュース番組メタ

<sup>†</sup> 神戸大学大学院自然科学研究科情報知能工学専攻  
Division of Computer and Systems Engineering,  
Graduate School of Science and Technology,  
Kobe University

<sup>††</sup> 神戸大学大学院自然科学研究科情報メディア科学専攻  
Division of Information and Media Sciences,  
Graduate School of Science and Technology,  
Kobe University

ファーとは、演出としてアナウンサーがいてあたかもニュースを読んでいるかのように振る舞うことである。このような、メタファーを用いることでただ単なる結果の羅列ではない、ユーザにとって分かりやすく親しみやすいインタフェースが提供できると考えられる。図1に番組メタファーの概念を示す。このように、ユーザの問い合わせに対してその答えをテレビ番組風に紹介することにより、ユーザの求めているものをよりの確にかつ魅力的に表現することが可能である。

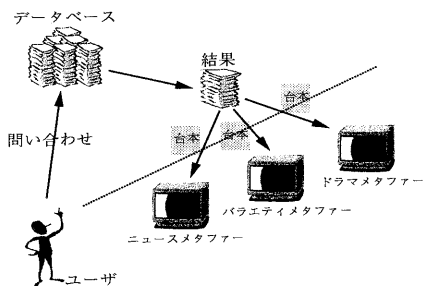


図1 番組メタファーの概念

今回は番組メタファーを番組記述言語 TVML(TV Program Making Language)<sup>1)2)</sup>を用いて作成した。TVMLはキャラクタやスタジオセットを用いて擬似的なテレビ番組を作成するために開発された独自のスクリプト言語であり、そのスクリプトを作成することによって、リアルタイムCG、音声合成、動画像などの技術を最大限に使って、1本のTV番組をすべてデスクトップ上でリアルタイムに生成することが可能である。すなわち、ニュース番組や情報番組を計算機の中だけで作成してしまうことが可能である。

本研究では、スクリプト作成のためのデータをデータベースに保管しておく必要に応じて提示する、動的な番組の生成について述べる。また、検索結果の内容の状態を調べることによりその内容に適した番組メタファーを動的に構成し、よりユーザに分かりやすい検索結果の提示方式を提案する。

## 2.2 種々の番組メタファー

提示を行うための番組メタファーは様々なものが考えられるが、今回は典型的な5種類のメタファーを提案する。これ以外にもさらに様々なメタファーも定義することができる。以下に今回提案する番組メタファーの特徴を説明する。

### A. ニュース

その日に起こったニュースをメインキャスターとサブキャスターが報道する。話題は様々である。

### B. ヘッドラインニュース

ヘッドラインニュースの形式。このメタファーではニュースキャスターがニュースを要約してすばやく伝える。登場するキャスターは一人である。

### C. 討論

多数の登場人物が一つの話について様々に議論する。

### D. ドラマ

ひとつの関連する内容に関してドラマ風に番組を進めるように情報を提供する。

### E. バラエティショー

様々な話題について色々な内容を含むとき、その中から幾つかをピックアップするなどして、漫才風にキャラクタが振る舞う。

## 2.3 番組メタファーの決定

得られた検索結果に関する次の3つの事項を考慮し演出する番組メタファーを決定する。

- 検索結果のデータ集合の分布状態
- 検索結果の内容
- 属性とメタファーとの関係

以後の章でこれらについて詳しく述べる。まず、3章では結果の分布の状態、すなわち結果の相関関係と各メタファーの特徴を考慮しメタファーを決定する手法について述べる。次に、4章では結果の内容、すなわち含まれるキーワードから選ぶメタファーの範囲を決定しそこに結果をマッピングすることで決定する。5章では検索情報の属性と各メタファーの特性から番組メタファーをどのように動的に決定するかについて述べる。最後に、6章ではそれらの方式をもとに作成したプロトタイプシステム Waffle の実装について述べる。

## 3. 結果の分布による検索結果の提示方式

実際に番組メタファーで表示する場合、検索結果と番組メタファーをどのように結び付け、どのメタファーを適用するかを決定することが重要である。本章では検索結果のデータ集合の分布状態から番組メタファーを決定する方法を提案する<sup>4)</sup>。分布による方法とは重み付けによる検索結果の分布からの分類であり、その分布を的確に表現する番組メタファーの定義することにより、演出する番組メタファーを決定する。

本研究で取り扱う電子番組表情報(EPGデータ)は、文書の形で配信されている。よって、検索結果の集合  $D$ (式1)を考える上で、これらの文書群からキーワードを抜き出すことによりインデキシングを行う。抽出

されたキーワードの全体集合を  $K$ (式2)とする。

$$D = \{d_1, d_2, \dots, d_n\} \quad (1)$$

但し,  $n$ は検索結果  $D$ の要素数

$$K = \{k_1, k_2, \dots, k_N\} \quad (2)$$

各番組情報は  $N$ 次元の特徴ベクトルとして表せる(式3)。

$$\mathbf{v}(d_i) = (w_{i1}, w_{i2}, \dots, w_{iN}) \quad (3)$$

但し,  $w_{ij} = tf(d_i, k_j) * idf(k_j)$

これらの各キーワードについては  $tf/idf$ 法<sup>5)6)</sup>によって重みづけすることによって, よりその番組情報の特徴を表すベクトルと考えることができる。

今後, 以上のように番組情報はキーワードの集合, または重みづけされた  $N$ 次元ベクトルとして考える。

### 3.1 検索結果の分類法

検索結果の分布を求めるためにはそれらを分類する必要がある。そのために以下の5個のパラメータを導入する。これらのパラメータの度合いによって分布の状態を決定する。

- (1)  $n$ : 検索結果  $D$ の要素数。
- (2)  $S_i$ : 個々の検索文書  $d_i$ (ベクトル) $\mathbf{v}(d_i)$ と問い合わせ(ベクトル) $\mathbf{q}$ の類似度。  
類似度はコサイン相関値を用いて求める(式4)。

$$S_i = \frac{\mathbf{v}(d_i) \cdot \mathbf{q}}{|\mathbf{v}(d_i)| \cdot |\mathbf{q}|} \quad (4)$$

- (3) 検索結果のクラスタリング。
  - (a)  $m$ : クラスタの数。
  - (b)  $|C_j|$ : 各クラスタ  $C_j$ ( $j \in \{1, \dots, m\}$ )に含まれる検索結果の要素数。
  - (c)  $S_{C_j}$ : 各クラスタ  $C_j$ に対する問い合わせ  $\mathbf{q}$ の類似度。

検索結果  $D$ を個々の類似度によりクラスタリングすることによって, 分布の状態を把握できる。以下に, 検索結果のクラスタリングの手法について述べる。

得られた検索結果は不特定多数のキーワードの集合であると考え。よって, 任意の2つの検索結果の類似度はそれぞれを  $N$ 次元ベクトルと考えたときのコサイン相関値で求まるものとする。

以上のように重み付けを行いクラスタリングしていくが, そのクラスタリングの手順を述べる。

- (1) まず, 問い合わせと個々の文書との類似度を計算し, その順序で並べる。
- (2) 初めに全ての結果は異なるクラスタに属しているとする。
- (3) 次に最初の結果において, これ以外に関して類似度がある閾値以上でさらに一番類似している

ものについて, これを関連集合としそのクラスタの最後に加える。

- (4) もし, 閾値以上で同じ類似度ならば両方の結果を加える。
- (5) 以後順番に同様の操作を行い, 以前にクラスタリングしたものよりもさらに類似したものが見つかった場合, そのクラスタを加えて新しいクラスタを作成する。
- (6) 以上のことを結果全てについて行う。

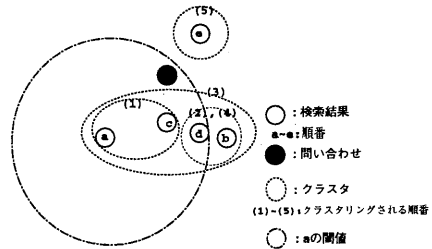


図2 クラスタリングの例

一例を図2に示す。この場合,  $a \rightarrow c, b \rightarrow d, c \rightarrow d, d \rightarrow b, e$ の順にクラスタリングされる。

このようにして, クラスタリングされた検索結果の集合から, クラスタの数  $m$ , この集合に含まれる結果の要素数  $|C_j|$ , このクラスタと問い合わせとの類似度は, このクラスタに含まれる個々の検索結果に対する類似度の平均とし  $S_{C_j}$ (式5)を求める。

$$S_{C_j} = \frac{1}{|C_j|} \sum_{d_i \in C_j} S_i \quad (5)$$

### 3.2 様々な番組メタファ어의定義

3.1節に種々のパラメータを導入したが, そのパラメータの値により演出する番組メタファーが決定される。本節では, 2.2節に示した実際の番組の特徴とこれらのパラメータを関連づけることにより, 動的に番組を選択する。すなわち, 結果の分布の状態がその番組メタファーの特徴に合うならばそのメタファーを選択すると思える。

表1に, 今回提案する5種類の番組メタファーの特徴から定義したパラメータの値を示す。以下に定義した理由を示す。

#### A. ニュース

様々な内容のニュースを扱うと考え, 検索結果数  $n$ が多く, クラスタリングにおいてどれも同じようなクラスタに分割された場合, すなわち  $|C_j| \approx \frac{n}{m}$ の時に

	$n$	$S_i$	$m$	$ C_j $	$S_{C_j}$
A. ニュース	<i>large</i>	-	-	$ C_j  \simeq \frac{n}{m}$	-
B. ヘッドラインニュース	<i>small</i>	<i>high</i>	-	$ C_j  \simeq 1$	-
C. 討論	-	-	<i>small</i>	$ C_j  \simeq n$	<i>high</i>
D. ドラマ	-	-	$m = 1$	$ C_j  = n$	-
E. バラエティショー	-	-	<i>large</i>	<i>small</i>	-

表1 番組メタファー決定のためにパラメータがとる値

このメタファーを適用する。

#### B. ヘッドラインニュース

様々な内容のものを扱うが結果の数 $n$ は少なく、それらひとつひとつがばらばらであると考え、検索結果のクラスタリングにおいてどのクラスタにも一つか二つぐらいの結果しか含まれず ( $|C_j| \simeq 1$ ) かつ、問い合わせへの類似度 ( $S_i$ ) が高い場合にこのメタファーを適用する。

#### C. 討論

あることについて議論すると考えるから、検索結果が大きなクラスタに属し ( $|C_j| \simeq n$ ) かつ、そのクラスタが問い合わせに類似する場合、すなわち  $S_{C_j}$  が大きい場合に適用する。

#### D. ドラマ

一連の内容について連続的に紹介するので、検索結果が大きな一つのクラスタに属する場合 ( $|C_j| = n$ ) で、問い合わせに類似しない場合、すなわち  $S_{C_j}$  が小さい場合でもよい。

#### E. バラエティショー

様々ジャンルのことを幅広く伝えるために、結果をたくさん含むクラスタが複数出現した場合 ( $m$  が大きく、 $|C_j|$  が小さい) に適用する。その中からの確性の高いものを抜き出すなどを行う。

これ以外にもさらに細かいルールを適用することで、他の様々なメタファーも定義することが可能である。

更に、その定義より幾つか番組メタファーの候補があがるが、ユーザはその中から任意に選択することができるとする。

### 4. 検索結果の内容からの番組メタファー決定

前章では、検索結果情報の相互関係による分布状態と番組メタファーの特徴を考慮して番組メタファーを決定する手法を提案したが、本章では検索結果群の内容から番組メタファーを決定する手法を提案する。すなわち、得られた結果の集合がどのメタファーの内容に近いかということから、演出する番組メタファーを決定することを考える。

#### 4.1 メタファーマッピング

番組情報は文書の形で提供される。その中から幾つかのキーワードを抜き出すことにより、それらキーワードの集合であると考えることができ、更に  $N$  次元ベクトル (式3) で表せるとした。

そこでここでは、番組メタファー自体もそれを定義するキーワードが決まっている  $N$  次元ベクトルと考える。そのキーワードは、実際のサンプルの番組表から抜き出す。例えば、ニュース番組メタファーならば、実際のニュース番組についてかかれた番組表の中から幾つか抜き出しそれを元に  $N$  次元ベクトルを幾つか作成する。

番組情報も番組メタファー情報も  $N$  次元空間の点だと考えると、それらを  $N$  次元空間にマッピングすることができる。それぞれプロットする点はキーワードの重みを  $tf/idf$  法によって計算した点であるとする。

各番組メタファー情報を  $N$  次元空間にマッピングしていき、それらの情報を各番組メタファーごとにまとめて、メタファーエリアとして  $N$  次元空間にそのメタファーを選択するための範囲とする。その範囲は、各サンプルメタファーベクトルの平均ベクトル  $\mathbf{s}_i$  との類似度がある一定の閾値  $\delta$  以内である範囲とする。

#### 4.2 番組メタファーの決定

番組メタファーを決定するために  $N$  次元空間にプロットされた番組メタファー情報の点の集合から番組メタファーのエリアを決定し、得られた結果を同じ  $N$  次元空間にプロットしていく。

以下にエリア決定から、どの番組メタファーを選択するかの手順を述べる。

- (1) まず、各番組メタファーがカバーする  $N$  次元空間のエリアを考えるが、適当に選択した各番組メタファーのサンプル要素の平均ベクトル  $\mathbf{s}_i$  からある閾値  $\delta$  以内の類似度を持つエリアを作成する。
- (2) その  $N$  次元空間に得られた検索結果  $d_i$  をひとつづつプロットしていき、あるメタファーエリアに属する、すなわち

$$\delta \leq \frac{\mathbf{v}(d_i) \cdot \mathbf{s}_i}{|\mathbf{v}(d_i)| \cdot |\mathbf{s}_i|} \leq 1$$

	タイトル	ジャンル	キャスト	チャンネル
A. ニュース	0	1	0	0
B. ヘッドラインニュース	2	-1	-1	0
C. 討論	-1	2	0	0
D. ドラマ	0	0	2	0
E. バラエティショー	0	0	0	2

表2 属性と番組メタファーの関係を考慮した $a$ の値

である場合はそのメタファーに関係する要素であるとする。エリアが重なる場合は複数のメタファーに属すると考える。

- (3) 基本的には、最終的に結果がもっとも多く属するエリアのメタファーを選択する。同数の場合はそのメタファーエリアに含まれる結果集合のそのメタファーに対する類似度の平均が高い方を選択する。

実際にメタファーエリアを決定して、そこに検索結果をプロットしていくが図3にメタファー決定の一例を示す。 $N = 2$ の2次元空間において考えているが、この時含まれる要素数の多いAメタファーが選択される。

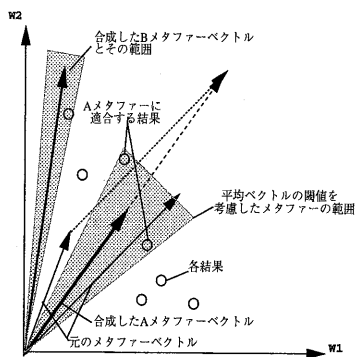


図3 分布

## 5. メタファーの特徴と属性の考慮からの決定

番組情報はキーワードの集合と考えたが、それらキーワードの中には特定の属性を持つものも含まれている。例えば、「スポーツ」という語はジャンルを表し、「19時30分」という語は時刻を表しているなどである。前章までの方式に加えて、ユーザが検索する場合において着目している属性情報を考慮して、演出する番組メタファーを決定する必要もある。そのためには、各属性の性質と番組メタファーの特性との関係を考慮しなければならない。例えば、タイトルという属性に着目した場合には、すばやく結果を表示させたい

と考えられるので、一覧表示するようなメタファーを選択し、また、ジャンルという属性に着目するなら類似情報が欲しいと考えられるので、様々な付加情報を教えてくれるメタファーを選択する可能性を上げる必要がある。故に、あるメタファー $M$ を選択する可能性を上昇させる可能性 $R_M$ を次の式6で表す。

$$R_M = \exp(ax) \quad (6)$$

但し、 $x$ はユーザの注目する度合いで、 $(0 \leq x \leq 1)$ であるとする。さらに、3章で定義したメタファーの特性を考慮して、表2に番組メタファーと属性の関係から決定した $a$ の値を定義する。

例えば、ユーザがキャストの属性に0.5注目した場合、ドラマメタファーとヘッドラインニュースメタファーを選択する可能性がそれぞれ約2.7倍、0.6倍になりその他はそのままである。

## 6. システムの実装 プロトタイプシステム Waffle

前章までの番組メタファーの動的生成方式をもとにしたプロトタイプシステム Waffle の実装を行った。この、プロトタイプシステムはTVMLプレイヤーが現在、Silicon Graphics社のWorkstation O2でのみサポートされているので、システムはO2上で構築した。番組情報のデータベースにはPostgreSQL RDBMSを用い、番組メタファー生成のためのホスト言語にJava言語を使用した。

### 6.1 システムの構成

図4にシステムの構成図を示す。プロトタイプシステム Waffle は、入力インタフェース部、メタファー決定部、TVMLスクリプトファイル生成部、結果提示部の4つの部分から構成される。まず、ユーザはインタフェース部で検索キーワードを入力する。それをもとにデータベースへ問い合わせを行い、結果を得る。次に、メタファー決定部でその結果の分布、内容、属性との関係を調べ、演出するメタファーを決定する。そして、実際に演出するTVMLスクリプトを生成し、結果をテレビ番組メタファーとして結果提示部のTVMLプレイヤーで提示する。検索結果をもとに番組を構成するために、簡単な自然言語処理の必要性がでてくる。

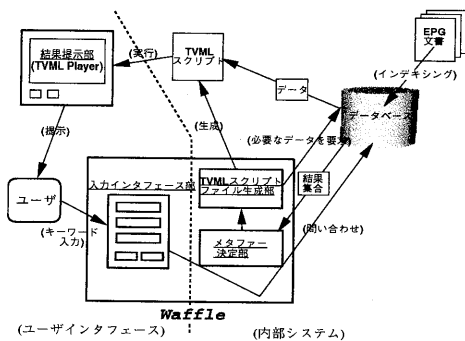


図4 プロトタイプシステム

本研究では基本的に、各番組データに含まれる属性情報をもとに番組を構成する。例えば、タイトル、ジャンル、キャストなどの情報をピックアップし合成する方法を取っている。それらを、インデクス情報としてデータベースに蓄えておき、必要に応じて取得する方法をとっている。キーワード単位で切り出すために、文章の形態素解析ソフト茶釜を使用した<sup>7)</sup>。

### 6.2 システムの実行

今回は実験的に100程度の番組情報を用いて、データベースへ格納、検索、提示を行った。実際の実行画面を図5に示す。この場合、それぞれバラエティショーメタファー、ヘッドラインニュースメタファーが選択されている。

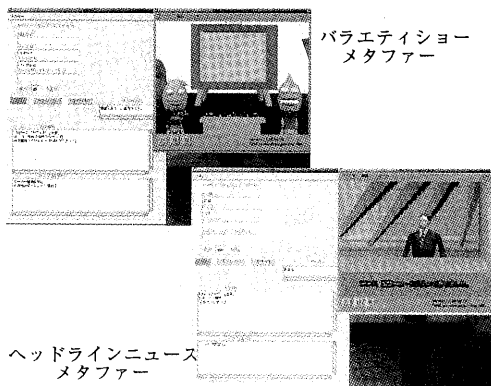


図5 実行画面

### 7. おわりに

本研究では、データベースへの検索結果を、番組メタファーで提示する方式を提案した。演出する番組メタファーを決定するために、検索結果の分布、内容、

属性情報との関係を考慮し、番組を構成するシナリオの動的生成方式を提案した。これにより、ただ単に結果の羅列を眺めるよりも、ユーザーに親しみやすく、分りやすい提示方式を提供できるようになった。

本研究の課題としては、試作システムでは一部のメタファーしか実装できなかったのにさらに多くのメタファーを定義、実装することでより使いやすいシステムの実現が挙げられる。また、今回はキーワードを抽出したデータを用いたが、検索エンジンからの検索結果をそのまま用いる方式を検討する必要があるなどが考えられる。

謝辞 本研究の一部は、文部省科学研究費重点領域研究「高度データベース」(領域番号 275 (08244103))の援助を受けています。また、本研究の一部は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」(プロジェクト番号 JSPS-RFTF97P00501)によっています。また、本研究の一部は、次世代情報放送システム研究所(株)の委託研究の援助を受けています。ここに記して謝意を表すものとします。

### 参考文献

- 1) 林正樹. 番組記述言語によるテレビ番組自動生成. 第2回知能情報メディアシンポジウム論文集, pp.137-144 (1996).
- 2) NHK放送技術研究所.TVMLホームページ.  
<http://www.str1.nhk.or.jp/TVML/indexj.html>
- 3) 金淵培. エージェント技術の放送への応用. 映像情報メディア学会誌 Vol.52, No.4, pp.447-451 (1998).
- 4) 近藤宏行. 番組記述言語を用いた情報検索結果の提示方式に関する研究. 神戸大学工学部情報知能工学科平成10年度卒業論文,(1999).
- 5) P. イングベルセン(著), 藤原鎮男(監訳), 細野公男・後藤智範・岸田和明(訳). 情報検索研究認知的アプローチ. トッパン (1991).
- 6) Kowalski, G. Information Retrieval Systems. Theory and Imprementation. Kluwer Academic Publishers (1997).
- 7) 奈良先端科学技術大学院大学. 茶釜ホームページ  
<http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>
- 8) 服部多栄子, 角谷和俊, 灘本明代, 草原真知子, 田中克己. 番組メタファーによる Web ページの利用者適用型提示. 情報処理学会研究会報告 99-DBS-119 (1999).
- 9) 角谷和俊. 放送とデータベース「放送型情報配信のためのデータ配信モデルとシステム」. ACM SIGMOD 日本支部 第12回大会 講演論文集, pp.105-150 (1999).