社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2006－HI－120（11）
2006／9／30

検索効率向上のためのコンテンツ均等分類手法の提案および評価

西森 崇　　前田 茂則　　小島 良宏

松下電器産業株式会社　AV コア技術開発センター
{nishimori.t, maeda.shigenori, kojima.yoshihiro}@jp.panasonic.com

概要

HDD 搭載レコーダなどの蓄積容量の増大にともない個人の蓄積コンテンツ数は膨大になりつつあり，目的のコンテンツを効率的に検索するためのインタフェースが望まれてきている．本稿では，HDD 搭載レコーダに蓄積された番組について，番組情報に付与されるジャンル情報を利用して蓄積番組の数に応じて均等・少数に仕分けなおした分類を提示する検索インタフェースを提案し，従来のジャンルによる分類手法と提案手法とについて検索に要する操作数と時間とを比較した評価実験の分析結果を示す．

Proposal and Evaluation of Content Classification Method Making Evenly-sized Categories to Improve Efficiency of Content Retrieval

Takashi Nishimori　　Shigenori Maeda　　Yoshihiro Kojima

AV Core Technology Development Center, Matsushita Electric Industrial Co.,Ltd.
{nishimori.t, maeda.shigenori, kojima.yoshihiro}@jp.panasonic.com

Abstract

As rapid growth of personal storage capacity is enabling individuals to enjoy huge numbers of multimedia contents, it is getting important to pursue the efficiency of the content retrieval methods, that is, how rapidly and easily the users can finish their retrieval. In this paper, we propose a retrieval method for TV programs recorded in HDD recorders, which classifies recorded TV programs into limited number of evenly-sized categories according to their genres and presents the categories to the user so that the user can retrieve TV programs by selecting them. Finally, through experimental retrievals by several users, we show the efficiency of the proposed method compared to the ordinary one that present genre categories as they are to the user.

1　はじめに

近年，HDD 搭載レコーダやデジタルスチルカメラをはじめ，コンテンツの蓄積機能を持つ CE（コンシューマ・エレクトロニクス）機器の蓄積容量が急激に増大してきている．それに伴い，個人あるいは家庭で蓄積されるコンテンツ数が膨大になりつつある．HDD 搭載レコーダを例に挙げると，各メーカーとも HDD 容量は 500GB を越える製品が主流になり，1TB に達する製品も現れてきている．テレビ番組を録画蓄積するとき，標準画質 (5Mbps) で 1 番組あたり平均 1 時間で換算し HDD 容量の 80%まで録画した状態を想定すると，1TB の場合で約 360 番組録画できる．また番組の録画機能に関しては，お勧め・自動録画など予約録画の手間を省く機能が導入されるようになり，ヘビーユーザでなくても大量にコンテンツが蓄積されるようになってきている．このように大量のコンテンツが蓄積されるようになると，蓄積されたコンテンツを利用する際に目的のコンテンツを検索する手間が大きくなり時間もかかようになる．そこで蓄積されているコンテンツを検索するための効率的かつ直感的な検索インタフェースが必要とされてきている．

本稿では，CE 機器でのコンテンツ検索の典型例として HDD 搭載レコーダに録画された TV 番組の検索を取り上げ，大量に録画された番組から目的の番組をより効率よく検索するための新たな手法とし

て，番組数が均等になるような少数のカテゴリに番組を分類し表示することにより，ユーザはカテゴリを階層的に選択していくだけで目的の番組を効率的に検索できるインタフェースを提案し，提案手法と従来手法との検索効率についての比較実験を行うことにより提案手法の有効性を検証した．

以下，2 章では従来からある検索インタフェースとその課題について述べ，3 章では 2 章で述べた課題を解決するために本稿で提案する番組の分類手法について説明する．次に，4 章では提案手法を検証するための検索インタフェースの説明とそれを用いた評価実験とその分析結果の考察をし，最後に 5 章で本稿全体のまとめをする．

2　従来の検索手法と課題

従来から様々なコンテンツ検索手法が提案され使われているが，最も自由度が高いのは google[1] 等の Web ページ検索などに用いられているキーワード入力による検索である．この場合は主に PC 等でキーボードを用いて検索キーワードを入力する状況が想定されているが，本研究が対象にしている CE 機器ではキーボード等の文字入力機器が充実している PC 等とは異なり文字の入力操作はユーザにとっては困難である．さらに，キーワード入力による検索では入力したキーワードが適切であれば非常にすばやく目的の情報を得ることができるが，一般にそのフレーズには幾通りかの言い換えがあって照合できなかったり，照合できても大量の該当する情報があって精査に時間がかかったりするなどするため，ユーザが本当に欲しい情報にすばやく効率的にたどり着くためにはキーワードの選択に相当の知識と経験が必要になってくるといった課題もあり，たとえ CE 機器の文字入力操作が改善されたとしても効率の良い検索方法とは言い難い．

一方，従来から HDD 搭載レコーダで多く使われているのは，蓄積された録画番組を何らかの基準で分類しておき，分類されたカテゴリを選択していくことにより目的の番組を検索する手法（カテゴリ選択型検索）である．日時やチャンネル，タイトルにより分類あるいはソートしたり，EPG（電子番組表）[2][3] に付与されている汎用的な分類体系である固定のジャンル (表 1) を利用して分類しカテゴリ表示したものを選択していくことで検索する．この手法では適切にカテゴリを選びさえすれば関係のある番組のみに絞られるので，キーワード入力による検索に比べると効率が良いと言える．

しかし，個人の嗜好は特定のジャンルに偏ることが多く，それに伴い番組も特定のジャンルに偏って録画・蓄積されることになり，その結果固定のジャンルをカテゴリとして選択する手法では番組を効果的に絞り込むことができなくなり検索効率が悪くなるという問題が生じる．

さらに，表 1 に示すように，ジャンルは構造化された体系を持っているが，多くの人の好みに対応するようできるだけ一般化された形を取っているため，同一階層における分類数が非常に多く，例えば「スポーツ」の下の階層の分類には約 20 もの分類が存在する．現状の CE 機器の表示デバイス（例えばテレビモニタ）の解像度で一般的に老若を問わず見やすい条件で表示させると一覧性にも問題が生じることになる．

表 1: ジャンル構造（一部省略）

大分類	中分類	小分類
映画		
音楽		
ドラマ	一般ドラマ 時代劇 外国ドラマ その他ドラマ	
スポーツ	野球 サッカー ゴルフ 相撲 ：	
アニメ		
バラエティ	バラエティ/芸能 クイズ ワイドショー トーク	 料理あり 料理なし
趣味/生活/教育	料理 旅行/紀行/食べ歩き 趣味/生活/健康/実用 ファッション番組 ：	
報道/ドキュメンタリー	ニュース/報道	
	・ ・ ・	
その他		

これに対し，主に文書コンテンツを対象とした一般的な手法として，EPG のジャンルのような固定された分類体系ではなくコンテンツに合わせて動的に分類体系を形成しコンテンツを分類してカテゴリを生成する手法が従来から多く提案されている [4][5][6]．しかしこのような手法も，コンテンツ間の意味的距離を基準に分類体系を形成するため意味的な上位・下位関係を反映した階層構造を作ることはできるが，カテゴリに含まれるコンテンツ数の偏りを少なくし，しかも一定数内のカテゴリにまとめることは想定していない．一方，意味的上位・下位関係を無視して単純にコンテンツ数だけを見て偏りが少なくなるように分類することはできるが，ユーザにとっては分類の意味がわかりづらく効率的な検索

ができない．つまり，分類体系の意味的上位・下位関係を正しくカテゴリに反映することは，階層的なカテゴリをユーザがたどるために欠かせないことでありこれらを同時に実現することが課題となる．

すなわち，CE 機器で従来より使われているカテゴリ選択型検索における課題をまとめると以下の 3 点になる．

- コンテンツの絞込み効率を上げるため各カテゴリに属するコンテンツ数の偏りを少なくする必要がある
- 操作性の低下を避けるためカテゴリの一覧性を維持する必要がある
- 意味的上位・下位関係を反映した階層構造をもとにユーザにとってわかりやすいカテゴリを生成する必要がある

これらの課題を同時に解決するため，分類されるカテゴリ数を少数に抑えて一覧性を維持しつつ，個人の嗜好によって蓄積されたコンテンツであっても特定のカテゴリに偏ることなく常に効率的にコンテンツを絞り込むことのできる分類手法を提案する．

3　均等分類手法の提案

2 章で挙げた課題より，HDD 搭載レコーダを対象として蓄積されている大量の録画番組からより効率的にすばやく目的の番組を検索するためには，キーワード等の文字入力の必要が無いカテゴリ選択型の検索であり，その上で，(1) 番組の絞り込み効率を良くし目的の番組に到達するまでの階層が深くなりすぎないようにすること，(2) カテゴリ表示の一覧性を維持するために各階層のカテゴリをできるだけ少数にすること，そして (3) 意味的上位・下位関係を反映した階層構造をもとにカテゴリを生成すること，が必要条件である．これらの条件を同時に満たすことで，検索に要するキー操作回数・操作時間が少なくなり効率的なすばやい検索を実現する．

そこで本稿では，(1)〜(3) の条件を満たすカテゴリ生成方法として，一旦意味的上位・下位関係を反映した木構造に番組を分類してから親子関係・兄弟関係の制約を設けて番組数が均等になるような木構造上のノードの組合せを探索しカテゴリとする手法を提案する．

図 1 に録画番組を対象にした均等分類手法による検索インタフェースのイメージを示す．一旦意味的上位・下位関係を反映したジャンル構造に番組を分類し，それを均等分類した例である．番組数の少ない「ドラマ」と「バラエティ」を統合して一つのカテゴリとしたり，録画番組数が 0 のカテゴリを削除したり，大量に録画されている「スポーツ」カテゴリを分解して下の階層にあった「サッカー」や「野球／バレーボール／テニス」（統合している）というカテゴリとして「スポーツ」があった階層に上げたりしている．これにより，インタフェースの一覧性が高まり，かつ余分な選択操作を削減することが可能となる．

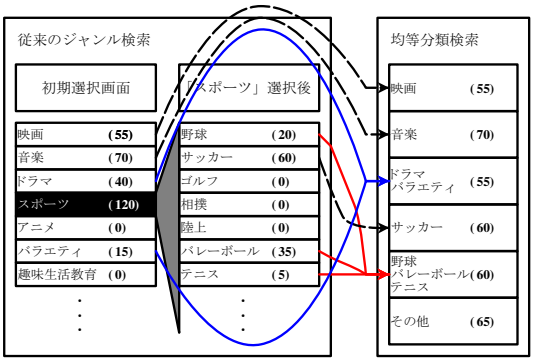

図 1: 均等分類を用いた検索の表示イメージ

3.1　均等分類アルゴリズム

前述の図 1 に示したようなイメージを実現するための均等分類アルゴリズムについて説明する．録画番組を分類対象とする場合，まず EPG データのような番組ごとに付与されているメタデータからカテゴリ分けの基準となる情報を抽出し，上位・下位概念の関係が親子関係になるような木構造に録画番組を分類する．表 1 で示したジャンルの場合には，図 2 のような木構造になる．

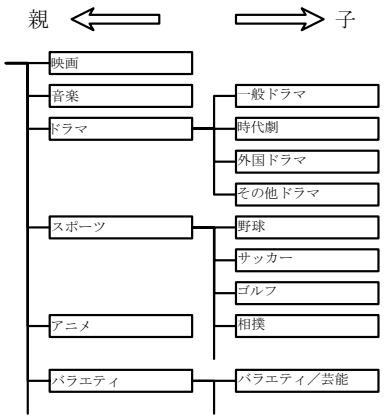

図 2: ジャンルの木構造

このような木構造を利用して，以下のステップで番組を分類したカテゴリを生成する．蓄積されている番組数を p，検索インタフェースに一覧性をもって表示可能なカテゴリ数を n，許容誤差を $\pm\alpha$ としたとき，

- **step1:** 全番組を木構造の適切なノードに分類する
- **step2:** 兄弟ノードのみ統合できるというルールのもと，番組数が $p/n \pm \alpha$ になるノードの組合せを全探索し，条件に合う組合せを中間カテゴリとする
- **step3:** 中間カテゴリから含まれる番組総数が最大となる $n-1$ 個の組合せを探索し，$n-1$ 個のカテゴリとする
- **step4:** step3 で選択したカテゴリセットに含まれなかった番組は「その他」カテゴリに組み入れる

step2 において，兄弟ノードのみ統合できるとしたのは，木構造が意味的上位・下位関係を持っているため兄弟ノードは意味的に近くなるので，それを統合してできたカテゴリがユーザにとってわかりやすいものになるという理由からである．

上記のステップにより n 個のカテゴリが生成され，ユーザが n 個のうちいずれかのカテゴリを選択するとそれによりコンテンツが絞り込まれる．そして絞り込まれた番組をさらに上記ステップを通すことによりまた分類する．これを繰り返すことにより，階層的なカテゴリ選択型検索を実現する．

3.2 時間情報を用いた分類

さらに，コンテンツが蓄積された時間の情報を利用した分類も追加した．これは前述の均等分類処理を行った結果，均等なカテゴリを作れないほど特定のジャンルに番組が偏っている場合に，その偏りすぎているジャンルについて時間情報に基づいてさらに分類することにより均等性を向上させるためである．日付分類のカテゴリについては，番組を絞り込むのではなくソートし一定間隔でカテゴリ分けするという形式にした．

4 評価実験と考察

提案手法である均等分類による検索と従来のジャンル検索との比較評価実験を行うことにより提案手法の有効性を検証し，効果と課題を抽出した．

4.1 評価用アプリケーション

ここでは，比較評価実験に用いた検索インタフェースのプロトタイプについて説明する．図 3 は今回の評価実験に用いたアプリケーションの GUI である．また，検索に用いるキー操作と機能を表 2 に示す．

図 3 の下段にある大きな枠のうち左の 3 個には各階層におけるカテゴリが表示される．一番左が最上位階層で右に行くほど階層を進む（深くなる）．また，現時点で選択されているカテゴリの下の階層の

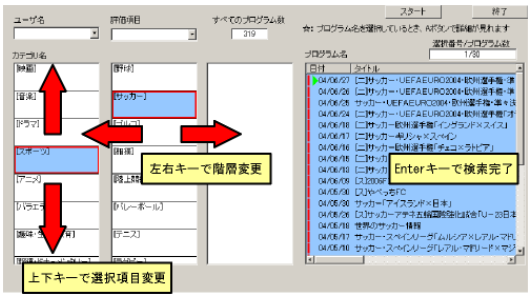

図 3: 評価用アプリケーションの GUI

カテゴリがその右の枠内に常時表示される．一番右の枠内（リスト表示部）にはその時点で選択されているカテゴリに含まれる番組のリストが時間情報の新しい順に常時表示されている．

表 2: キー操作と機能

キー操作	機能
上/下	同一階層での選択カテゴリ・番組の変更
	長押し：素早い連打と同等の処理
右/左	階層を進む・戻る
	最下層で右を押すとリスト表示部へのフォーカス移動
	リスト表示部がフォーカス中に左を押すとフォーカス移動前の状態に
Enter	目的の番組発見時の検索終了処理
	リスト表示部にフォーカス移動

また，ジャンルにをもとにしたカテゴリを選択したときは，そのカテゴリに属する番組のみリスト表示部に表示する．一方，日時をもとにしたカテゴリを選択しても，日時で分ける元になったカテゴリの全番組がリスト表示部に表示されたままで，番組リストのフォーカスが選択したカテゴリの先頭の番組に移動するだけである．番組のタイトルや内容とは異なり日時についての人間の記憶力は曖昧になることが多いが，日時の記憶が曖昧なときでも適当なカテゴリを選んでリスト表示に移動してから前後のカテゴリの番組を番組リストで探すことができる．(図 4)．

4.2 評価実験

PC 上で評価用アプリケーションを用いて，提案手法による検索とジャンル検索のそれぞれを用いて蓄積されている大量の録画番組中から目的の番組を検索する実験を行った．今回の評価実験にあたっては，提案手法では番組に付与されている EPG デー

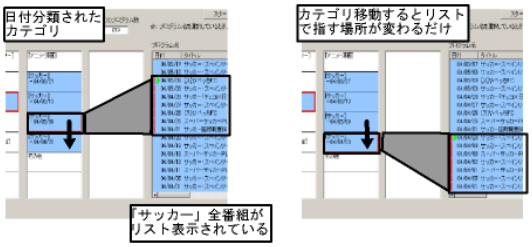

図 4: 日付分類の動作

タの内容の一つである表 1 に示したジャンル構造をカテゴリを生成する基準として利用した．

被験者は 20 代〜30 代の男女計 6 名．今回の実験を行う前に各被験者にあらかじめ約 300〜900 の番組を録画してもらい，被験者自身が録画した番組の中から番組を検索することにより，被験者にとって既知の番組を検索する形をとった．録画番組のデータについては，実際の番組に基づいた番組データから被験者に観たいと思う番組を選択してもらうことで生成した．

番組の検索にあたって被験者に提示する情報は，検索する番組の日付，番組のタイトル，サブタイトルのみである．タイトルとサブタイトルは EPG から得られる情報であり，番組録画時にも提示してある情報である．また，番組を録画するユーザがそのジャンルを正確に把握していることはまず無いと考えられるので，ジャンルについては被験者に提示していない．しかしそのため，被験者はジャンルを推定して検索する必要があり，同一の番組を検索すると後に試行する手法の方がジャンルが既知状態になるため，結果に差が生じることになる．そこで，

- **試行パターン 1**: ジャンル検索を先に行い，同じ番組を提案手法で検索する．
- **試行パターン 2**: 提案手法での検索を先に行い，同じ番組をジャンル検索で検索する．

の 2 パターンについてそれぞれ異なる 12 番組ずつ，合計 24 番組 48 回の検索を行い，それぞれの平均をとり比較することにした．

評価用アプリケーションでは自動的に被験者の操作ログとして検索開始から検索完了までのキー操作と各々のキー操作に要した時間を取得した．

また，各番組の検索試行ごとに検索の早さとわかりやすさについて提案手法とジャンル検索とのいずれが優れていると感じたか 7 段階での主観評価と，その理由のコメントを収集した．

4.3 実験結果の分析と考察

上述の評価用アプリケーションを用いて行った評価実験の結果と分析を以下に述べる．

4.3.1 操作ログの分析結果

評価用アプリケーションにより取得した操作ログの分析結果を表 3 に示す．なお，表中 ‘ ジ ’ はジャンル検索における数値，‘ 提’ は提案手法による検索における数値を示す．‘ step ’ は検索に要したキー操作回数，‘ time ’ は検索に要した時間 (ms) を示し，それぞれ 12 回の試行の平均値である．

表 3: 操作ログの分析

被験者	手法	試行パターン 1 step	試行パターン 1 time	試行パターン 2 step	試行パターン 2 time
p1	ジ	56.58	17576	36.92	10487
	提	21.00	12524	31.50	12606
p2	ジ	27.17	19397	22.83	10307
	提	14.08	14242	15.67	16061
p3	ジ	36.42	17210	18.25	6832
	提	11.75	6757	13.75	8593
p4	ジ	53.17	23642	49.33	16735
	提	18.33	7873	43.92	10704
p5	ジ	25.50	9498	29.58	13671
	提	14.08	7341	28.25	20895
p6	ジ	60.25	18070	39.83	11840
	提	28.08	7835	50.33	15893
平均	ジ	43.18	17566	32.79	11645
	提	17.89	9429	30.57	14125

表 3 から，試行パターン 1 と試行パターン 2 の平均を求めたものと，1 操作あたりの平均所要時間，そして各々の値が提案手法によってどの程度削減されたかを示す削減率を表 4 に示す．

表 4: 操作ログ全体平均

	ジャンル検索	提案手法	削減率
step	37.99	24.23	36.22 %
time	14605	11777	19.37 %
time/step	384.50	486.06	-26.42 %

ジャンル検索に比べ提案手法が操作回数にして 36.22%，所要時間にして 19.37%削減されたが，一方で 1 操作あたりの平均所要時間は提案手法の方が 26.42%増加しており，操作時に考える時間が長くなっているように見える．

評価用アプリケーションではキーを長押しすることにより連打したのと同じ挙動をする機能を組み込んであるが，長押し時の 1 操作あたり所要時間は約 40ms でありこれは長押し操作を除いた場合の 1 操作あたりの平均所要時間より大幅に短いため，長押

しによる操作回数が多いほど試行全体における 1 操作あたりの平均所要時間は短くなる．そこで長押し操作を除いた上で 1 操作あたりの平均所要時間を求めたものを表 5 に示す．

表 5: 長押し除去後の操作ログ全体平均

	ジャンル検索	提案手法	
長押し step	18.01	8.34	
長押し time	708	444	削減率
step	19.98	15.89	20.47 %
time	13897	11333	18.45 %
time/step	695.57	713.24	-2.54 %

表 5 から判るように，従来のジャンル検索に比べ提案手法では長押し操作回数がかなり少ない．これが意味することは，提案手法の方がカテゴリの選択を終えてリスト表示部に移行した段階で番組リストが短くなっているということである．つまり，ジャンル検索に比べ，提案手法の方が番組検索における絞込み効率が向上しているということが言える．また，ジャンルの統合や従来下位にあったジャンルを上位に提示するといった処理によりカテゴリ表示がわかり難くなる懸念があったが，長押し操作を除去した後の 1 操作あたりの平均所要時間は両手法ともほぼ同じであったことからそれらによる検索効率の低下は無いこと示された．実際，被験者から得られたコメントにも，

- 各階層に表示されるカテゴリ数が従来より少なくなって見やすかった
- 従来は階層を掘り下げて探す必要のあったものが上の階層に出てくることがあり手間が省けた

というものが多くあり，上記の結果を裏付けているといえる．

4.3.2 主観評価結果の分析

検索に要した時間と検索インタフェースの使いやすさについての主観評価の結果を表 6 に示す．被験者には各番組の検索試行ごとに検索速度と検索のわかりやすさの観点でそれぞれ-3,-2,-1,0,1,2,3 ポイントの 7 段階評価をしてもらった．マイナスが大きいほどジャンル検索が良く，プラスが大きいほど提案手法による検索が良い．表内の数値は全試行の平均値である．また，‘ speed ’は検索速度についての評価，‘ usability ’は検索のわかりやすさについての評価を示す．

試行パターン 1 と試行パターン 2 との平均を求めると検索速度については 0.38，検索のわかりやすさについては-0.04 となり，検索速度では提案手法の方が速いと感じているが検索のわかりやすさについては大差は無いという結果を得た．ただ，データの細部を見ていくと全体的にはポイントがマイナスになっている箇所は少なく，被験者 p1 が大きく平均を下げていることを考慮すると提案手法は概ね良好な主観評価結果を得たものと考えられる．しかし，操作ログの分析結果から得られた操作回数・操作時間の削減効果に見合うだけの差が主観評価からは得られておらず，この原因を被験者のコメントから分析の結果，主に以下のような提案手法の問題点が挙げられた．

表 6: 主観評価結果 (7 段階)

	試行パターン 1		試行パターン 2	
被験者	speed	usability	speed	usability
p1	-0.33	-1.83	-0.67	-0.75
p2	0.58	0.50	0.25	0.50
p3	1.25	0.42	0.58	0.17
p4	0.92	0.08	0.58	-0.08
p5	1.25	-0.17	0.57	0.17
p6	0.00	0.33	-0.42	0.17
平均	0.61	-0.11	0.15	0.03

- 「その他」カテゴリに入っている番組を見つけるのは消去法になるので難しい
- 複数のジャンルが統合されてできたカテゴリの表記が読みにくい
- 日付分類カテゴリでの番組リストが絞り込まれていないことがわかりにくい

今回の実験では均等分類というアプローチが検索効率の向上にどの程度影響するかについての評価を行ったが，検索インタフェース全体としてはまだ問題点も多く，今後はカテゴリの表記・提示方法や日付分類に関する概念について主観評価も向上するような手法を検討し，検索インタフェース全体としての完成度を高める必要がある．

また日付分類については，3.2 で説明したように，ジャンルによる分類ではジャンルが少ないためそれ以上絞り込めない段階でも大量に番組が含まれていることがあったために均等性を維持するための処置として組込んだ．しかし時間情報についての人間の記憶は曖昧になってしまうので，時間情報よりも記憶に残りやすい番組の内容に基づく分類をできるだけ活用したい．そのためには，ジャンル以外の番組の内容を表す情報にもとづく分類を併用して，特定のジャンルに番組が偏ってしまう状況に対応する必要がある．

5　まとめ

本稿では，個人が蓄積するコンテンツの膨大化に伴い低下する検索効率を改善する新たな手法としてコンテンツ均等手法を提案した．そして，録画番組を対象として提案手法による検索と従来のジャンル分類による検索との比較実験を行い，その結果提案手法による検索の方が 4 割近く検索に要するキー操作回数を削減し，2 割近く検索に要する時間を短縮することを示した．提案手法によりカテゴリの構造が従来手法よりいびつになるにも関わらず 1 操作あたりの平均所要時間もそれほど変化せず，わかりやすさが犠牲になっていないことが示された．しかし，主観評価では検索の速さが計測された数値ほど向上していないので，今後この主な原因であるカテゴリの表記・提示方法や日付分類の問題,「その他」カテゴリの扱いなどについての改善策を検討する．

今回はカテゴリ生成に EPG データに含まれる番組のジャンル情報を利用したが，ジャンル情報はカテゴリ分けの基準としては数が少ないため特定のジャンルにばかりコンテンツが偏る状況が生じやすく，カテゴリの均等性を維持するのには限界がある．そこで，EPG データに含まれている番組説明文をもとに番組内容を表すキーワードを抽出しその意味的構造・概念的構造を併用してカテゴリを生成することも検討項目として挙げられる．

また，本稿では HDD 搭載レコーダに蓄積された録画番組を提案手法適用の典型例として取り上げたが，提案手法は個人の嗜好や興味の度合いに基づいて大量に収集・あるいは生成されるコンテンツをメタデータに基づいて効率的に検索できる構造に整理するのに適している．従って，録画番組以外でも例えばデジタルカメラや携帯電話のカメラで撮影した写真の場合，近年では GPS 機能等と連携して写真に位置情報を付与できるものも増えてきているので，位置情報に基づいた階層構造を分類の基準として提案手法を適用することも考えられる．

参考文献

[1] google. http://www.google.com/

[2] G ガイド. http://www.ipg.co.jp/top.html

[3] ADAMS. http://www.tadv.jp/service/adams.html

[4] 徳永健伸. 『情報検索と言語処理』, 東京大学出版会, 1999

[5] 川谷隆彦. 多文書間の共通性分析による文書クラスタリング. 情報処理学会自然言語処理研究報告, NL-154, pp.93-100, 2003

[6] 中島誠，伊藤哲郎. 文献クラスタの概念的特徴づけを用いた文献の自動分類. 情報処理学会自然言語処理研究報告, NL-151, pp.87-94, 2002