1U-8

電子番組表と Web 情報の関連付けに関する基礎的研究

物部寛太郎† 田中成典‡ 古田均‡ 古堅則之‡
関西大学大学院総合情報学研究科† 関西大学総合情報学部‡

１．はじめに

近年，地上波デジタル放送や BS デジタル放送の開始に伴い，テレビ放送の多チャンネル化が進んでいる．デジタル放送では，従来のアナログ放送と異なり，データ交換が可能になる．そのため，電子番組表がテレビやコンピュータなどで受信可能[1]となった．従来の電子番組表では，番組に関する情報量が少ないため，利用者が電子番組表から詳細な情報を得ることが困難である．既研究として，番組の録画予約履歴や登録したキーワードから個人の趣味嗜好を分析し，個人に適した番組を推薦する研究[2]がある．この研究では，趣味嗜好を分析し番組を推薦するが，番組の情報は提供しない．そのため，利用者は，番組を一度視聴し，趣味嗜好に適しているか確認する必要がある．そこで，本研究では，WWW 上に存在する膨大な情報に着目し，電子番組表の情報に関連のある Web ページを関連付け，番組の情報を閲覧できるシステムの開発を目指す．

２．システムの概要

本システム（図１）では，テレビ番組に関連する Web ページを自動的に収集し，電子番組表と Web ページを関連付けることを目的とする．

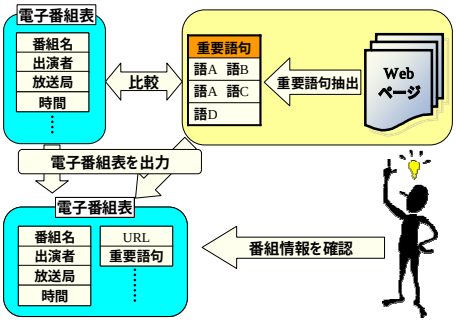

図１ システムの構想

図２に示すように，１）電子番組表取得機能，２）キーワード抽出機能，３）Web ページ収集機能，４）Web ページ分析機能，５）電子番組表生成機能の５つの機能によって構成される．

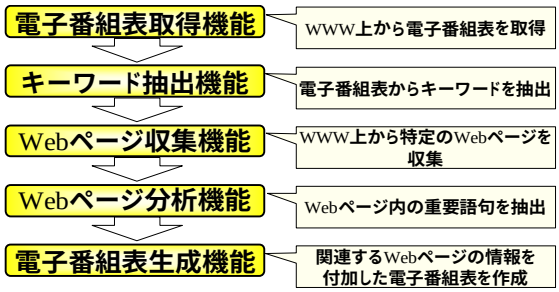

図２ システムの流れ

２．１ 電子番組表取得機能

電子番組表取得機能では，WWW 上で配信されている電子番組表を取得する．電子番組表の取得には，WWW 上から番組名や出演者などの情報を特定の形式で取得できる iEPG を利用する．

２．２ キーワード抽出機能

キーワード抽出機能では，電子番組表から各番組の重要なキーワードを抽出する．抽出した番組名や出演者などのキーワードは，Web ページの収集に用いる．

２．３ Web ページ収集機能

Web ページ収集機能では，GoogleAPI を利用し，番組に関連する Web ページを収集する．GoogleAPI を利用する理由は，Web ページのリンク情報から重要度を計算する PageRank[3][4]というアルゴリズムからキーワードに関連の高い Web ページを容易に取得できるためである．

２．４ Web ページ分析機能

Web ページ分析機能では，収集した Web ページから重要語句を抽出する．重要語句の抽出には，CF 法[5]を利用する．CF 法は，文章中の語句から頻出語の共起頻度と共起する語句の偏りに着目し，重要語句を抽出する手法である．抽出した重要語句と電子番組表の情報を比較し，共通点が存在すれば関連が高いとし，共通点がなければ関連が低いと設定する．

Fundamental Research on Linking Electronic TV Guide with Web Resource
†Shigenori Tanaka, Hitoshi Furuta, Noriyuki Furukata
Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho Takatsuki-shi, Osaka 569-1095, Japan
‡Kantaro Monobe
Graduate School of Informatics, Kansai University, 2-1-1 Ryouzenji-cho Takatsuki-shi, Osaka 569-1095, Japan

２．５ 電子番組表生成機能

電子番組表生成機能では，番組に関連のある Web ページを関連付けした電子番組表を出力する．番組と関連する Web ページは，Web ページ分析機能で評価した関連の高い順に出力する．

３．システムの実験結果と考察

本システムを用いて，番組と関連する Web ページの収集を行った．電子番組表の取得は，地上波放送に限定し，2004 年 11 月 29 日の番組表を対象とした．そして，本システムで収集した番組と番組に関連する Web ページの数を計測した．検証として，出力した電子番組表の中から，実際に番組の情報が存在する Web ページの数を計測した．本システムで生成した電子番組表から「SMAP×SMAP」と関連する Web ページの一覧を表示したときのシステムの実行結果を図３に示す．

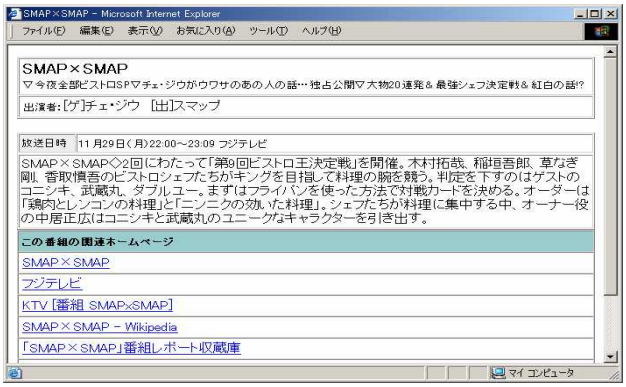

図３　システムの実行結果

３．１ 実験結果

検証結果として，本システムで取得した 1 日分の番組数と Web ページ収集機能で収集した Web ページと番組と関連する Web ページの数を計測した．番組と関連する Web ページは，収集した Web ページの中から手作業で確認した．また，本システムの実験結果を表１に示す．

表１　実験結果

収集した Web ページ	番組と関連する Web ページ	精度
8,250 ページ	6,554 ページ	79.4%

システムの有用性を検証するために被験者 10 人に対して，本システムを利用しテレビ番組の選択に本システムで生成した番組表が，(i)参考になったか(ii)参考にならなかったかを調査した．被験者は，本システムで関連付けした Web ページの内容を参考にして回答する．被験者が選択するテレビ番組としては，A：SMAP×SMAP，B：ラストクリスマス，C：今日の天気を対象とした．また，調査したアンケート結果を表２に示す．

表２　アンケート結果

	(i)	(ii)	参考にした Web ページ数
A	10	0	29 ページ
B	8	2	30 ページ
C	5	5	4 ページ

３．２ 考察

実験結果が示すように，本システムで抽出した 8,250 ページの内，番組情報と関連する Web ページは，6,554 ページで，79.4%の精度で WWW 上からテレビ番組と関連する Web ページを取得することができた．実験結果により，本システムでは，番組の情報とその番組に関連する情報を高い精度で取得できることが実証された．また，アンケート結果から A と B は，Web ページが約 30 ページ収集できており，多くの利用者が参考になると回答している．C の評価が低かった要因として，番組名が「今日の天気」という一般的に利用される言葉のため，収集できた Web ページが 4 ページのみであったことが考えられる．

４．おわりに

本システムは，WWW 上から電子番組表に関連する Web ページを取得することにより，テレビ番組と Web ページの関連付けを可能にした．また，CF 法により関連する Web ページの評価を行うことで，番組に関連する Web ページを高い精度で取得できた．今後は，番組に関連する Web ページを提供するだけでなく，出演者や番組の内容により番組をカテゴリに分ける機能や個人の登録情報や番組閲覧履歴から趣味嗜好を分析し，個人に適した番組を推薦する機能を考える．

参考文献

[1] 八木伸行，吉村俊郎，加井謙二郎：データ放送技術読本，オーム社，2002.2.
[2] 情報技術標準化研究センター：将来型文書統合システム標準化調査研究委員会成果報告書，日本規格協会，2004.3.
[3] Sergey Brin, Lawrence Page：The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Networks and ISDN Systems, Science Direct, Vol.30, No.1-7, pp.107-117, 1998.4.
[4] Amy Langville, Carl Meyer：Deeper Inside PageRank, Technical report, North Carolina State University, 2004.3.
[5] 松尾豊，石塚満：語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム，人工知能学会論文誌，人工知能学会，Vol.17, No.3, pp.213-227, 2002.11.