

LOD を用いた番組情報ネットワーク形成手法の提案 A Proposal of TV Program Information Network Formation Method by Linked Open Data

有安香子[†] 中川俊夫[†]
Kyoko ARIYASU Toshio NAKAGAWA

1. はじめに

情報メディアに対する利用者の意識変化調査[1]によると、各メディアの平均利用時間はテレビに続き、PC、スマートフォンとなり、スマートフォン利用者の55%が過去3年間で利用時間が増えたと回答している。また、10～20代の利用者がメディアとしてインターネットを重要視する傾向があるとも報告されている。別の調査結果[2]では、テレビ・PC・スマートフォン・タブレットなどのデジタルデバイスを同時に利用すると回答した人は全体の3分の2にのぼり10代女性が最も高いとの報告がなされた。

我々は、利用者意識やメディア利用状況変化に柔軟に対応し、日々生まれる情報の海の中に埋没しているテレビ番組の存在自体に気付いてもらい、視聴者の興味を番組視聴につなげることを目的とした研究に取り組んでいる。



図1：ネットからテレビ番組への情報導線

我々が目標とする、ネットからテレビ番組への情報導線を図1に示す。導線の構成要素は大きく2つに分けられる。ひとつは視聴者の興味や会話の文脈、場所など、さまざまな日常生活の場面において、番組へつながるきっかけをネット上により多くつくることである。もうひとつは、視聴者のコンテキストに沿った番組への導線を作るため、番組同士をさまざまな関係性を使って結びつけ、きっかけを通じたアクセス後の情報に広がりを持たせることである。本稿では、この情報に広がりを持たせたつながりを番組情報ネットワークと呼び、その形成手法について報告をおこなう。

[†] NHK 放送技術研究所, NHK Science and Research Laboratory

※1 典拠データ…資料検索のためタイトル・著者等を記録した書誌データ(著者名や件名として記録する人物、団体、テーマ等)に同義語がある場合一つにまとめたものを「標目」という。この標目を統制するため、別表記や同義語、標目として選んだ根拠を示す資料等の情報をまとめて記録したものが典拠データである。

2. オープンデータの活用

「番組へのアクセスのきっかけ」は、放送局が管理・付与するものだけでなく、ユーザがそれぞれのコンテキストにおいて、自発的に選択することが満足度につながるとされている[3]。ここでは、放送局以外のサービス事業者が、放送局から放送日時・放送日時・番組タイトル・リンク先等入手し、そのデータを元に提供するサービスを通じてユーザへの「番組へのアクセスのきっかけ」を造ることを想定している。放送局や各種サービス事業者が、保有するデータをオープンデータとして提供することで、その相互作用によって効果的で幅の広い情報サービスの提供・利用が可能になる。このような「番組へのアクセスのきっかけ」と「アクセスした後の広がり」の両方を視野に入れたシステムデザインを前提とした場合、オープンデータの利活用が現実的な解のひとつとして考えられる。とりわけオープンデータの中でも、我々の目指すユーザの興味や文脈などを意味的な内容を含む情報導線の形成のため、情報間の関連性を表記した形でオープン化され、数多くのエントリーが公開されている Linked Open Data (LOD)形式のデータを使ったネットワークの形成手法を提案する。

3. 番組情報ネットワーク形成手法

番組同士を関連付ける手法として、タイトルやシリーズ/出演者/番組概要文等のテキスト情報の類似性を元にした研究が数多く報告されている[4]。このような手法で関連付けたデータは、見たい番組があらかじめ決まっている場合の検索や、ユーザの属性値や履歴を元にした番組推薦サービスに使用する際に有用である[5]。しかし、本稿で想定するような会話の文脈や状況をきっかけとした番組情報ネットワークへのアクセスから、番組視聴への導線を作るという目的を考慮した場合、はっきりとした視聴意思や明確な好みを前提とはできないので、きっかけとなった話題に意味的な幅を持たせた複数の番組を提示する手法が有用であると考えた。

我々は、番組内容の意味的なつながりを用いて番組情報ネットワークを形成するため、国立国会図書館が公開している Web NDL Authorities[6]のデータを使用した。Web NDL Authorities は、典拠データ(※1)を LOD 形式で公開したものである。標目の同義語・上下概念・別表記などが RDF 形式で記述され、SPARQL による検索に対応している。

図2に、アレルギー性皮膚疾患の典拠データの例を記す。テレビ番組などで良く使われる一般的な「アトピー性皮膚炎」という語彙は「アレルギー性皮膚疾患」の別表記であり、「アレルギー」と関連する「皮膚炎」の下位概念だという意味的なグラフが得られる。

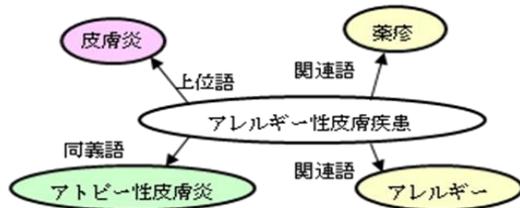


図2: アレルギー性皮膚疾患の典拠データ例

4. 番組情報ネットワークアプリケーション試作

本稿では、Web NDL Authorities から得られる語彙間の意味的なつながりを使った語彙グラフを作成し、作成したグラフに番組内容をマッピングすることで、番組情報ネットワークを形成した。形成したネットワークの動作確認のため、ネットワークを可視化する Web アプリケーションを試作した。語彙グラフにマッピングする番組内容を表すデータとして、デジタル放送で使われる EPG (Electronic Program Guide) データを使用した。試作では、Web NDL Authorities で得られるグラフの最上位に当たる語彙に関するグラフの取得クエリを投げ、得られたグラフノードを種として再帰的にグラフを取得し、語彙グラフを形成した。得られた語彙グラフと、EPG データベースの情報の「番組タイトル」「サブタイトル」「出演者」のフィールド内のテキストを比較し、一致した番組を語彙グラフに番組ノードとして追加した。ひとつの語彙ノードに対し複数の番組が一致した場合は、以下の3つの優先順位をつけたルールに照らし合わせ番組を選択し、1つのノードに関連付ける番組数の上限を3つとした。

Rule 1. 放送日時が未来のもの

Rule 2. 放送日時が過去の場合は現在時刻に近いもの

Rule 3. 番組タイトルが異なるもの

Rule 1. は本放送のテレビ視聴を促すため、Rule 2. は旬に近い情報を含む番組を提示するため、Rule 3. はより多くの番組の存在に気付いてもらうためのルールである。

図3にタブレット端末上に試作した番組ネットワークアプリケーションの画面例を示す。「あさイチ」のハウレン草の活用術を題材とした放送回(画面中A)をきっかけとした際の画面例である。「葉酸が認知症に効果的」という番組内容から「葉酸(画面中B)」と「認知症(画面中C)」の語彙グラフノードが関連付けられ、「認知症」の語彙ノードには「DHA」が「認知症」に効果的であるという内容の番組ノード(画面中D)が関連付けられている。更に、「DHA」語彙ノードに「DHA」を多く含む「サケ(画面中E)」や「イワシ(画面中F)」を題材とした料理番組が関連付けられるなど、番組の意味的な内容を元に広がりを持たせたネットワークが形成されていることが確認できた。

また、試作したアプリケーションの各番組ノードをタップすると対応する番組動画が再生するよう設計し、番組視聴への導線を作るための動作を確認した。

5. まとめ

放送局に蓄積された大量のテレビ番組に関する情報を、意味的な広がりを含めて関連付けた「番組情報ネットワーク」形成手法を提案し試作した。LOD形式で公開され



図3: 番組ネットワークアプリケーション試作画面例

ている語彙グラフを取得し、語彙グラフのノードにテレビ番組をマッピングし、「番組情報ネットワーク」を形成した。形成した番組ネットワーク自体をアプリケーションとして提示することで、意味的に関連するさまざまな番組同士の関係性を視覚的に表現し、視聴者の番組内容に関する関心をより広く、深く満たすような情報提示手法の可能性を示した。

今後の課題として、番組内容を記述した EPG 情報を、その内容に併せて RDF 形式の記述に変換し、語彙ノードと同様に扱えるようにすることで、番組情報ネットワーク内のすべてのノードを起点として、再帰的にグラフを取得し表示できるように改良し、得られたネットワークの番組の意味的な広がりを定量的に評価することが挙げられる。また、語彙グラフの種となる最上位ノードを「疾病」「疼痛」「芸術」の3つに限定した試作の語彙範囲を広げ、動作検証をおこなう予定である。

参考文献

- [1]総務省, “平成 24 年版情報通信白書”, 第 1 部, 第 2 章, 第 3 節, P238-251 (2012).
- [2]ニールセン, “デジタル・コンシューマ・データベース”, http://www.netratings.co.jp/news_release/2012/10/DCDB.html (2013/07/01 accessed)
- [3]C. Shapiro, H.R. Varian, “Information rules: a strategic guide to the network economy”, pp.184, Harvard Business School Press, Watertown (1998)
- [4]Smyth B. and P. Cotter, “A personalized television listings service”. Communications of the ACM, 43(8):107-111, (2000).
- [4]K. Ariyasu, H. Fujisawa, S. Sunasaki, “Twitter analysis algorithms for Intelligence Circulation System”, Multimedia Systems Journal, DOI: 10.1007/s00530-013-0324-2
- [5]Web NDL Authorities, <http://id.ndl.go.jp/auth/ndla> (2013/07/01 accessed)