

12.4 タグと集合知

人々が生成するコンテンツや情報が大量に蓄積されると、個々の人が独自の目的と意味づけで記述した情報から、全体としてひとつの新しい情報が現れてきます。検索サイトの「リンクポピュラリティ」によるページランク付けは、人々が自分の意思で行なったリンクを無数に集めた結果として得られるものです。こうした情報の集積から生まれる新たな情報は、集合知 (collective intelligence) と呼ばれます。

ソーシャル・ブックマークでのタグは、文字列だけでは何を意味するのか不明確な場合がありますが、多くの人が多くのコンテンツに与えたタグを集めると、その関係によって意味が浮かび上がってきます。また、コンテンツのブックマーク数が関心度の指標となっているとすれば、タグとの組み合わせでコンテンツ推薦を行なうことも可能です。

12.4.1 タグの共起とクラスタ

前節ではユーザタグURIをWikipediaに明示的に結びつけることで曖昧さ解消と主題探索を両立させる方法を検討しましたが、たくさんのユーザによるタグ付けを集積すると、同じタグであっても使われ方によってグループが形成され、暗黙的な「語義」を見出すことができます。たとえばFlickrは、同じタグでも意味が異なるものをグループ分けするタグ・クラスタという試みを導入し、多義語のタグから写真をうまく検索できるようにしています。

図 12.12



Flickrのタグ・クラスタ。同じappleというタグでも、共起する(同時に付けられている)タグによって、Macintosh、りんご、iPod、ニューヨークというグループに分かれている。

大多数の写真には、複数のタグが付けられています。同じ"apple"でも、その意味が「アップル・コンピュータ」の場合と「りんご」の場合では、組み合わせられるタグは違うでしょう。1つの対象に与えられる複数のタグ(共起タグ)は意味的に関連する可能性が高いという考えに基づいて、語義のグループを作ってみるのがここでのタグ・クラスタです。

Flickrはそれぞれのクラスタに、主要な3つの共起タグを連結したURIを与えています。

例 12.12

```
http://www.flickr.com/photos/tags/apple/clusters/mac-macintosh-macbook
```

* WNRDF

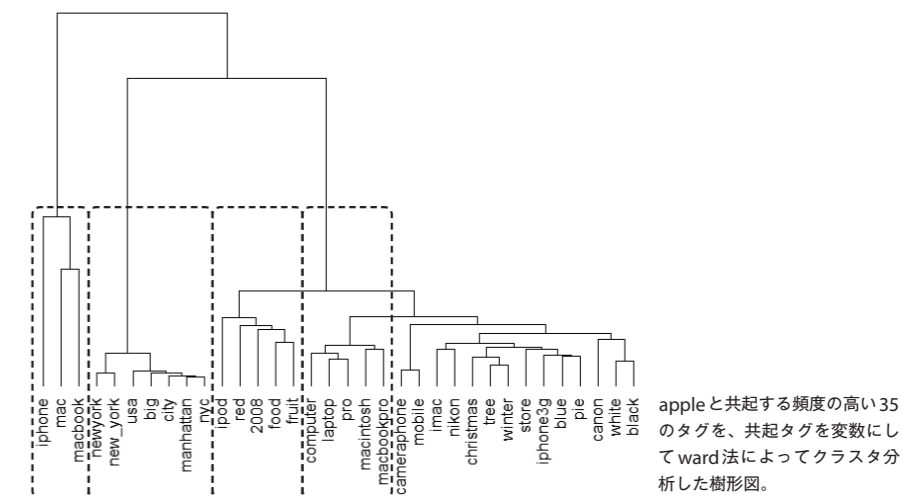
これは1つの「語義」を識別するURIと考えることもできそうです。英単語の概念ネットワークを表現したWordNetのRDF版*では、語 (Word)、語義 (WordSense)、同義語群 (Synset) を関係付けるための語彙が定義されています。Flickrの"apple"タグを「語」、タグ・クラスタを「語義」と考えると、WordNet RDFを用いて図12.12のクラスタを次のように表現することができるでしょう。

例 12.13

```
@prefix wn: <http://www.w3.org/2006/03/wn/wn20/schema/> .
@base <http://www.flickr.com/photos/tags/apple/clusters/> .
<http://www.flickr.com/photos/tags/apple>
  a wn:Word ;
  wn:sense <mac-macintosh-macbook> ,
           <fruit-red-green> ,
           <ipod-iphone-music> ,
           <nyc-newyork-manhattan> .
```

確認のために、"apple"というタグが付与されている約5000点の写真をFlickrから選び、併せ用いられる(共起する)タグを調べてみました。頻度の高い35のタグについてクラスタ分析を行なうと、「iphone, mac, macbook」「newyork, usa, city, manhattan」「ipod, red, food, fruit」「computer, macintosh, laptop, makbookpro」といったクラスタが見出され、図12.12のタグ・クラスタとほぼ同様のグループができています注18。

図 12.13



appleと共起する頻度の高い35のタグを、共起タグを変数にしてward法によってクラスタ分析した樹形図。

注18 分析対象が比較的短期間に公開された写真であるため、タグにはその時期による偏りがあります。分析を行なった2008年12月時点で、Flickrには"apple"というタグが付けられた写真は約50万点あり、全てを対象にすれば、図12.12と同じクラスタが得られるでしょう。

Flickrのサービス自体は、"apple"というタグに何の定義も与えていませんし、使い分けのための名前（Wikipediaでの"Apple_Inc."、"Apple_Records"など）を推奨しているわけでもありません。ユーザもこれらの区別を意識することなく、それぞれの使い方ですべて"apple"タグを与えています。それにもかかわらず、多数のユーザのタグを集積することで、おのずと語義ごとにグループが浮かび上がってくるのです^{注19}。

12.4.2 レビューとしてのタグ

ソーシャル・ブックマーク・サービスにおいて、各ページの被ブックマーク数は、そのページに対するサービス利用者の関心の高さを表す指標となっています。多数ブックマークされたページをチェックすることで、情報収集のアンテナとしている人は少なくないでしょう。

ただ、ブックマーク・サービスの利用者の関心事項はさまざまで、全体としての被ブックマーク数が多くても、そのページが自分にとって興味のあるものとは限りません。そのため、どのブックマーク・サービスも、特定のタグを付けられた最新ブックマークを抽出したり、知人などのブックマークを“購読”する機能を提供しています。前者は主題検索、後者は人によるフィルタリングをブックマーク数と組み合わせ、より絞り込んだ情報を得られるようにしているわけです。

前項ではコンテンツに付けられたタグのクラスタを考えましたが、これと同様に類似のタグを使っている人のクラスタを考えれば、オンラインショップの「お薦めの商品があります」と同様の推薦が可能になります。ただこの計算はかなり大規模になるため、ここでは特定の同じタグを使っている人を選んで、その人々のブックマークから情報を得る方法を考えてみましょう。知人“購読”によるフィルタリングがfoaf:knowsを利用したコンテンツ推薦に相当するなら、こちらはfoaf:interest型のフィルタリングだといえます[▼]。

▼10.3.2項参照

注19 もちろんこれは、Flickrというコミュニティ内での語義ということになります。たとえば"opera"のクラスタを見てみると、生成されているのは<sydney-house-australia>、<paris-france-garnier>、<vienna-wien-austria>、<china-beijing-chinese>の4つ、つまりシドニー、パリ、ウィーンの歌劇場および京劇という「語義」であり、(写真の対象とはなりにくい) ブラウザは含まれません。