

## The Structure of Collaborative Tagging System (2005)

Scott A. Golder and Bernardo A. Huberman (HP Lab. (USA))

<http://www.hpl.hp.com/research/idl/papers/tags/>

### 1 本論文の目的

まずタギング (Tagging) とはどのようなものを分類 (Taxonomy) との比較により示し、さらに Collaborative Tagging の特徴を示す。次に Collaborative Tagging の実システムである deli.cio.us のデータを用いて Collaborative Tagging におけるダイナミクス分析を行う。特にある URL に付与されるタグの割合に安定性があることを示し、なぜ安定性が起こりうるのか、そして安定性が Collaborative Tagging のどのような有効性を示すのかについて考察する。

本論文の目的は、Collaborative Tagging に関する基礎的な考察を行うことと、実データの分析から Collaborative Tagging のダイナミクスを発見することにある。

### 2 背景

Collaborative Tagging とは多くのユーザによって共有コンテンツに対してメタデータ (キーワード) が付与されるプロセスを指す。ここ数年、Web 上のコンテンツに対して Collaborative Tagging を行う Web サイトなどが人気を博している。Collaborative Tagging では多くのユーザがタグ付けすることによってコンテンツの分類がボトムアップに行われるが、これは “folksonomy” と呼ばれ新しい分類アプローチとして注目されている。

伝統的な分類 (圖書の分類など) では司書などのオーソリティが行ってきた。対して Collaborative Tagging は誰もが出来る。これはオーソリティがない場合やコンテンツの量が多くて少数のオーソリティではカバーできないような場合にとって有効である。また、Semantic Web におけるメタデータの生成および構造化の方法としても注目されている。

### 3 手法・アルゴリズム

Collaborative Tagging サイトである deli.cio.us からデータを収集し、分析を行った。データは2種類収集した。一つは人気リストに載っている212のURLを取り出し、そのURLに対するブックマーク19,422件を収集したものである。もう一つはランダムにユーザ229人を取り出し、彼らのブックマーク68,668件を収集したものである。

本論文では大きく分けて二つの分析を行っている。一つはユーザのタギングに関する分析、もう一つはブックマ

クの分析である。分析は主に特徴的なユーザやブックマークを複数取り出して図示し、比較することで行っている。

### 4 評価

ユーザのタギングに関する分析では、ユーザのタギングパターンは様々であり、また、それは日々変化・成長していくこと、そして付けられた順番が若いタグほど全体で見ても頻度の高いタグであることがわかった。タグ付けの順番と出現頻度との関係については Basic Level と関わりがあると考えられる。

ブックマークに関する分析では、ある特定の URL をブックマークするピーク (流行) はブックマークされた初期に主に起こるが (7日以内が70%弱) そうでないものも20%弱ほど存在すること、ある URL に付けられるタグは最初はばらつくが100件ほどブックマークが集まった時点で安定することがわかった。なお、少数派のタグも少数派として維持され、淘汰されてしまうことは無い。

### 5 知見

分類では複数の観点に対応するのは困難である。対してタギングは様々な観点でつけることが可能である。これは検索時に漏れをなくすという点において優れている。しかし自由にタグ付けできるため “polysemy (多義性)” や “synonymy (同義性)”、Basic Level の違いによる問題が生じる。Collaborative Tagging では、ユーザ間において生じるタギングの模倣と知識共有により、これらの問題が解決できることが期待される。

deli.cio.us のデータ分析から、付与されるタグの割合には安定性があることが示された。しかもそれは人気のタグが勝ち残るのではなく、人気のあるタグは大きな割合で、人気のないタグは小さな割合で、それぞれ収束する形で現れた。これは Web 文書が人々とどのようなインタラクションをしているか (人々にとってどのようなコンテンツであるのか) を示す手がかりとなりうる。このような安定性は模倣と知識共有の結果として生まれたものだと考えられる。また、ユーザ間である程度バックグラウンドが共有されていたことも要因としてあげられる。

また、分析からタグが普及していく様子や個人が膨大なタギングを行う様子がうかがえた。主にユーザは自分のた

めにタギングを行っていると考えられるが、Collaborative Tagging ではそれが他人への貢献につながっている。例えば「面白い」というタグがついたコンテンツは、他人にとっては推薦コンテンツとなる。Del.icio.us はこの点で明示的な推薦を行わないレコメンデーションシステムともいえる。