

3L-07 ジャンル検索特性に基づくディレクトリ構築法の一考察

幸 嘉平太、元田 敏浩、川崎 隆二

kaheita@canary.sl.cae.ntt.co.jp

NTT ソフトウェア研究所

1. はじめに

インターネット上での検索サービスとして、多くのディレクトリサービス¹⁾が運営されている。ホームページなどの URL とその紹介文をデータとして蓄え、検索によってそれらの情報を提供するサービスである。ディレクトリサービスは通常キーワード検索部とジャンル検索部の 2 つのユーザインターフェースを持っている。後者は大分類・中分類・最小分類と階層化され、木構造を形成している。最終的に紹介情報を提供する末端ジャンル数が数千個という大規模なディレクトリサービスも存在する。このジャンル検索部は基本的にサービス提供者による固定的な分類構造である。そのため、

- 分類イメージがユーザと一致しない
- 他分類に属する末端ジャンルへ移動するためには、階層の上下移動が必要
- 末端ジャンル数が多く、必要とするジャンルの発見が困難

などの問題点が認識されており、ユーザの操作コスト増大の原因となっている。

本稿では、NTT が提供しているインターネットディレクトリサービスである、NTT DIRECTORY の利用履歴を分析対象とし、操作コスト軽減策としてコリド¹⁾型ディレクトリを提案する。

¹⁾ "The study for constructing the internet directory based on the users' property", YUKI Kaheita, MOTODA Toshihiro, KAWASAKI Ryuji, NTT Software Laboratories, 3-9-11 Midori-cho, Musashino, Tokyo 180-8585, Japan

¹⁾ corridor: ホテルなどで部屋と部屋を結ぶ廊下

2. 利用形態の分析

NTT DIRECTORY では http-Cookie を発行しており、この Cookie の授受を許可しているユーザ(正確にはブラウザ)に対しては、ユーザ単位で利用形態を把握できる。前回の発表^{iv)}で「各ユーザー一人あたりでは、ごく一部のジャンルしか利用せず、また利用ジャンルの中でも特定のジャンルのみを繰り返し利用する傾向が強い」ことを述べた。今回は、新たな指標として「ジャンル共起度」を導入した。各ユーザが利用するジャンルの組み合わせの出現回数を積算することにより、そのジャンル組み合わせの共起度とし、どのような組み合わせが高頻度で利用されるかを調査した。ユーザ毎に求められた共起度を合計することにより、ユーザ全体の共起度とした。1998 年 1 月の場合の各組み合わせの共起度分布プロットを図 1 に示す。NTT DIRECTORY には約 1900 個の末端ジャンル(最小分類)が存在するため、組み合わせを求めた場合、 ${}_{2000}C_2$ ~ 約 200 万個の組み合わせが得られる。図 1 から、組み合わせの大半が共起度 10 以下(最小分類で平均 5.2)であるが、一部の組み合わせが突出して高頻度で利用されているのが分かる。このことから、ユーザによって利用されるジャンルの組み合

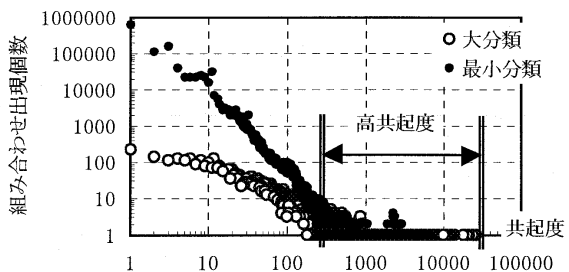


図 1 共起度の出現頻

わせはごく一部のパターンに限られており、ユーザにとって関連性の高いジャンル組み合わせが存在することを示している。

3. コリドー型ディレクトリの提案

前節の結果より、一人のユーザによる一定期間内の共起ジャンルには、そのユーザの思考に基づく何らかの関連があると仮定した。この関連を多くのユーザについて集計すると、ユーザに特有な特性は相対的に弱まり、社会・文化的に共通に認識されている関連が強調されると予想される。この仮定に基づき、ジャンル検索の操作コストを下げる機構として、コリドー型ディレクトリを提案する。図1に概要を示す。従来の木構造型ディレクトリに加え、各末端ジャンルに対し共起度が高いジャンル数個への直接リンク(以下、コリドーとよぶ)を構造を付加している。このような構造を採用することにより、下記の効果が期待され、1章で挙げた問題点の解決策の1つになると考えている。

- 共起度の高い、つまり関連度の高いジャンルへの移動が容易になる
- 他のユーザの利用動向が提示されることにより、ジャンルガイド的な役割を果たすことができる
- 末端ジャンル間の移動が容易であるため、ウィンドウショッピング的なジャンル閲覧が期待でき、アクセス数の増加も予想される
- ウィンドウショッピング的な閲覧により、未知のジャンルを発見できる

現在、このコリドー型ディレクトリを導入した Super

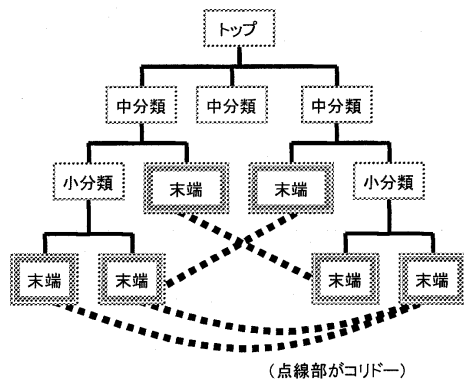


図2 コリドー型ディレクトリの概念図

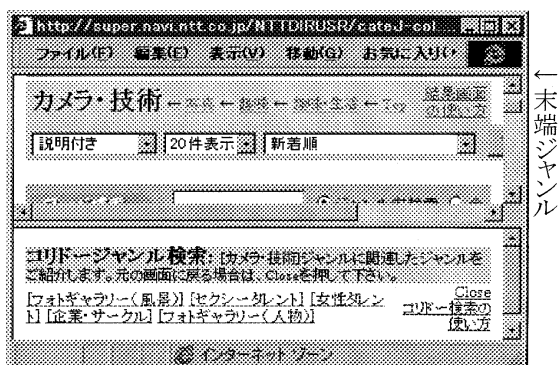
NTT DIRECTORY¹⁾を試験サービス中である。コリドー部が表示されている例を図2に示す。

4. まとめ

ディレクトリサービスにおけるジャンル検索部の利用形態には、特有の性質があることが分かった。特に利用ジャンルの組み合わせには著しい偏りがあることが発見できた。これらの性質に基づいたコリドー型ディレクトリは、ユーザの操作コスト軽減策として、有効であると考えられる。

共起度をコリドー形成のデータとする場合、時系列での指標の減衰効果の扱いを考慮しなければならぬ。また、現在は単純なcombinationに基づいて共起度を算出しているが、この計算法の妥当性の評価も必要である。中間階層のノード間へのコリドーも原理的に可能であるので、そのような面からもアプローチも進めていく必要がある。

コリドー型ディレクトリによる効果は、実際のサービス結果に基づく分析を待たなければ検証できない。今後は、Super NTT DIRECTORYの運用実績の分析を踏まえ、フィードバック・改善を行っていく予定である。



↑ 高共起度ジャンル表示部

図3 コリドーが導入された Super NTT DIRECTORY

¹⁾ NTT DIRECTORY, <http://navi.ntt.co.jp/>

²⁾ Yahoo, <http://www.yahoo.co.jp/>

³⁾ Infoseek, <http://www.infoseek.co.jp/>

⁴⁾ 幸 嘉平太他、「ディレクトリの利用履歴に基づくさまざまなドメインレベルでの嗜好特性について」、第56回情報処理学会全国大会

⁵⁾ Super NTT DIRECTORY, <http://super.navi.ntt.co.jp/>