

情報は人のためならず：共有知の動的な再構成による 個人のための知的活動支援環境を目指して

沼 晃介

東京大学 先端科学技術研究センター

〒153-8904 東京都目黒区駒場 4-6-1

E-mail : numa (at) ai.rcast.u-tokyo.ac.jp

Abstract: 本稿では、われわれの目指す知的活動支援環境を、「情報は人のためならず」ということわざのもじりに乗せて提示する。そのための情報流通技術・情報作成支援技術を、情報と情報、情報と人との「関係」に着目して、これまでのわれわれの取り組みを中心に整理する。

1. はじめに

「情けは人のためならず」ということわざがある。情けをかけるのは人のためにならないからよくない、などという意味だとの誤解が広がっているという話もあるが、本来は人に情けをかければ巡り巡ってやがて自身に戻ってくるということを意味する。昨今 Web の技術やサービスの発展に伴い、ユーザ自身が情報を発信する機会も増加している。このように人々が行う情報発信もまた、単に目の前の他者だけのため、あるいは直接的に自分自身のためというばかりでなく、多くの人を介して巡り巡って思わぬところで再び自分に利益をもたらして欲しいものだ。われわれは、人々の発信する情報を直接的に役立てるのみでなく、共有された知識群を活用することで再びユーザ自身の活動を支援することで、「情報は人のためならず」とでもいふべき環境を構築することを目指している。

Web 2.0 関連の流行語のひとつに、「集合知」があげられる。ユーザー一人ひとりの知識はわずかなものでも、各自が持ち寄れば多くの領域をカバーする知識を集めることができる。ユーザ参加型のオンライン百科事典である Wikipedia に代表される多人数から集めた知識は、いまや個人の知的活動に必要な不可欠な要素となっている。考え方によっては、Web コンテンツ作成者のハイパーリンクの集合から Web コンテンツの重要度を評価する検索エンジンなども、ある種の集合知を用いたシステムといえるかもしれない。一連のシステムは多くの面で非常に有用であるが、一方で知識の多様性を前提としていないという問題がある。知識は世界に唯一の体系を持って存在するのではなく、個々人がそれぞれの文脈の中で形成し利活用していくものである。ある文脈である人にとって真であった情報が、別の文脈や別の人には当てはまらないことも間々ある。

文脈とは、情報と情報の関係、そして情報と人の関係から生み出されるものであるとわれわれは考える。われわれ人間は、情報の内部の要素の関係や、その情報がおかれた社会的・文化的状況との関係、あるいはコンテキスト・アウェア・システムが扱うような情報や人が外部との間に持つ関係など、さまざまな関係に囲まれて情報を扱っている。これらの関係は刻一刻と変化し続け、個人にとっての意味や価値も常に変わっている。人々が情報をうまく取捨選択し、また生み出していくためには、これらの情報との関係を適切に扱う手段が求められる。われわれは、Web

を介して共有する知識群を個人が日常的に利活用するための支援技術を、特に動的に変化する関係に着目して実現することを目指している。本稿では以下、われわれの目指す知的活動支援のための環境と技術要素を、これまでのわれわれの取り組みを中心に整理する。

2. 共有知の動的な再構成による個人のための知的活動支援環境

近年の Web 2.0 社会においてユーザが投稿する多様なコンテンツが増加した状況を、「ごみコンテンツ」の氾濫と揶揄する向きもある。確かに多くの場合において、直接自身に役に立たない情報が膨大になり、自らが必要とする情報を発見する邪魔になることもある。しかしこの状況は、単に情報の作成された文脈と情報が閲覧される文脈のそれぞれが、うまく情報検索や情報閲覧に反映できていないことから引き起こされているのではないかと、われわれは考える。必ずしも万人に通用する一般的な情報でなくとも、適宜情報の内容や文脈に応じて利用されるならば、一つひとつの情報の利用価値が高まる。文脈に基づき適切に情報が流通する環境を整えば、人々が安心して多く情報を発信することもできるだろう。こうして有効に利用できる情報が増えれば、どこかの誰かの情報が巡り巡って自分自身にも役に立つ機会も増えると考えられる。

Web 上で共有される情報を、状況や目的に応じて利活用するためには、情報のおかれた文脈と人のおかれた文脈をいかに扱うかが課題となる。特に、(1) 文脈をどのように計算的に扱うか、(2) 文脈をどう処理して情報の発見に用いるか、(3) それらの文脈を背景にどのように新しい関係や新しい情報を作るか、という 3 点が技術的ポイントとなる。以下、各項目について議論する。

2.1 文脈の取得

情報の持つ文脈という、コンテンツのテキストを文書解析する手法など多くの研究がなされており、例えばテキストや部分テキストの出現単語ベクトルを他と比較し、トピックの関係や推移を分析するなどといったことが行われている。関連する情報もテキスト化されないと文脈に反映できない問題はあるが、逆に読者の事前知識や直近に閲覧した文書などのテキストを与えることができれば、読者の中で活性化されていると考えられる文脈を計算し利用することができる[佐藤 2008]など、一定の有用性が認められる。

一方で、コンテンツの表層表現への処理以外に、コンテンツのメタ情報を集め、情報の持つ背景情報として利用する試みもある。コンテンツのメタ情報には、Google などで用いられるコンテンツ間のリンク関係や、ソーシャルブックマーク (SBM) で閲覧者によって付与されたタグ、Weblog システムなどの CMS (Content Management System) によって付与される RSS などのメタデータ (タイトルや筆者、作成時刻、カテゴリなどが記述される) などがある。

CMS が付加するメタ情報はコンテンツ作成時の状況を表している情報といえるが、多くの場合 Web 上での状況であってユーザ個人のおかれた実世界の状況情報を表していない。そこでわれわれは、作成された状況情報をコンテンツに対応付ける手法を提案した[沼 2007]。ActionLog というこの一連のシステムは、ユーザが利用する情報システムからユーザの行動を取得し、Weblog コンテンツのドラフト (草稿) を生成する。その行動にまつわる状況情報 (時間や場所など) を取得し、コンテンツに対応付けるとともに、ドラフトに本文としても挿入する。ドラフトをもとにユーザがコンテンツを記述し完成させることにより、行動時の実世界の状況情報とコンテンツ

が対応付けられる。

状況を表すメタ情報は、それ単体で文脈と呼ぶには不十分な一面的な情報である。しかし、複数のメタ情報を合わせて処理することにより、テキストのみでは扱えない文脈にアプローチすることができると思う。

2.2 文脈に基づく情報検索

膨大な数の情報から状況に適した情報を発見するための手法で現在もっとも一般に用いられているのは、テキストの表層そのものを対象とした全文検索である。大規模情報へのアクセス手法として現実的でコストパフォーマンスのよい手法であるが、ユーザの文脈に合わせた検索ワードの選定や結果のランク付けの困難さが知られている。

コンテンツに付与されるメタ情報を用いることで、目的に合致する情報の発見が容易になると期待できる。SBMでは、情報がタグを介してつながり、被ブックマーク数などで順位付けて閲覧できるため、特定のカテゴリから多くの人に受け入れられる情報の発見がしやすい。先述のわれわれのActionLogでは、実世界の行動情報に基づいて関連するコンテンツを一覧することができるため、やはり状況を共有する情報を発見することが容易となる。ただしここまでの事例では、単にメタ情報の一致のみの静的な関係である。

ユーザ自身の情報を用いて検索を行うことによって、得られる最新のユーザの状況を踏まえた検索が可能となる。ユーザモデリングの研究や協調フィルタリングなどでは、ユーザのそれまでの入力や行動などからユーザの嗜好を推定したり、他者との類似性を計算したりする。ここではユーザの行動によって逐次検索結果が変化しうるが、用いる情報によって推定に偏りが出る点や、ユーザの嗜好の短スパンでの動的な変化に柔軟に対応できない点、多くの情報を求めるほどユーザにコストがかかる点などの解決すべき問題もある。

大規模情報源からの検索手法ばかりでなく、情報源や流通の経路そのものを制御し、あらかじめ文脈の共有が期待できるユーザにのみ閲覧させたり、そのような情報源から検索したりする負傷もある。一部のソーシャルネットワークサービス(SNS)では、日記を記述する際、投稿者と閲覧者の関係によってアクセス権を制御する機能を持つ。また、われわれもユーザのネットワーク上で、ユーザから数ステップ以内の接続関係にあるユーザを情報源として検索し、ユーザからの距離に基づきランク付けする検索手法[Numa 2004]を提案している。

2.3 新しい関係・新しい情報の作成

3点目は、どのように新しい関係や新しい情報を作るかという問題である。ActionLogでは、取得した行動の周辺情報をWeblogコンテンツのドラフトに埋め込んでいる。これはコンテンツとメタデータを対応付けるという意図のほかに、文章としてユーザに提示して振り返りを助けること、あらかじめ埋めておいて記述を助けることを狙っている。しばしば「書きながら考える」というように、単に関連情報として見せるのではなくこれから書く文章に埋め込んでおくことで、明示的に発信する情報の取捨選択を迫る。この例では、新たな他の情報との関係は対象となっていないが、ユーザが状況を情報として記述することになり、構造化された情報の量を増やすことに貢献する。

実験的なワークショップ（参加体験型グループワーク形式の表現のための場）を対象としたものであり、Webのようなオープンな場での問題とは異なるが、われわれは新たな情報間の関係の中で新たな情報を生み出すための支援にも取り組みも行っている[Numa 2008]。このワークショップの詳細は省略するが、写真とテキストを複数ペア組み合わせてひとつの作品を作るという課題を、個人作業とグループワークを交互にして、自身や他者のそれまでの表現をリミックスして新たな作品をどんどん作り出すようなプログラムのワークショップを開催した。ここでの情報システムは、それまでに蓄積された他者の表現のテキストを機械的に分解し、内部に含まれる単語の関係からそれまでの文脈を推定し、単語や、単語に対応する写真などを新たに組み合わせ、提示した。ここでも表現候補を入力画面に直接提示し、機械的に計算された関係に刺激された新たな情報の作成を支援する。

ここではコンテンツの表層の処理による文脈だけをもとに再構成を行っているが、さらに多様なメタ情報の集合からなる文脈を用い情報を再構成することもできるだろう。ユーザのおかれた状況情報と、コンテンツの持つメタ情報を、動的に複合し、新たに情報を関係付ける。関係の候補を複数提示し、それをユーザに取り込ませていくことで、ユーザの視野を広く持たせることにつながる。ここでより新規なアイデアをユーザに与えるためには、固定化された関係に基づく情報ばかりではつまらない。使えないシステムと、つまらないシステム、それぞれに陥らないバランスでの情報処理と提示が求められる。

3. おわりに

われわれが目指すのは、個人が情報を出せば出すほど、自身のところによりよい情報が集まってくるような環境である。そのためには、さまざまな情報がさまざまな関係の中に位置づけられていることを理解したうえで、自身の周囲の関係の構造をうまく扱っていかねばならない。文脈は常に変化し続けるため、その時々で新しい関係の可能性を模索し、自身に取り込んでいくような視点が求められる。本稿では、このための情報技術の要件を、特にわれわれの過去の取り組みを題材に整理した。隣人にかけた「情け」は、この社会の中で人の関係を介して自身に返ってくる。同じように自身が「情報」を発信することで、他者に貢献すると同時に、自身にも他者からの恩恵が返ってくるだろう。

参考文献

- [佐藤 2008] 佐藤真, 田中克明, 赤石美奈, 堀浩一: 物語構造モデルに基づき話題の遷移を分析する手法の提案. 第 22 回人工知能学会全国大会論文集, 2008.
- [沼 2007] 沼晃介, 平田敏之, 濱崎雅弘, 大向一輝, 市瀬龍太郎, 武田英明: 学術会議における体験共有のための行動履歴に基づく Weblog システム. 情報処理学会論文誌, Vol.48, No.1, 2007.
- [Numa 2004] Kosuke Numa, Ikki Ohmukai, Masahiro Hamasaki and Hideaki Takeda: Egocentric Search based on RSS. In Poster Proceedings of 3rd International Semantic Web Conference (ISWC2004), 2004.
- [Numa 2008] Kosuke Numa, Katsuaki Tanaka, Mina Akaishi, Koichi Hori: Activating Expression Life Cycle by Automatic Draft Generation and Interactive Creation. In International Workshop on Recommendation and Collaboration (ReColl'08), 2008.