



これからの 情報処理学会

— 第 13 回 —

学会の集合知

松井 くにお

(株)富士通研究所
情報処理学会企画担当理事

Web 2.0 の「集合知」

Web 2.0 では、ブロードバンドによる常時接続と高速回線による効果を原点としたさまざまな発展が述べられ、必然的に出てきた変化をオライリーが見事なタイミングでまとめ上げた。その結果、Web 2.0 は、世界的な流行語となり、Google で「Web 2.0」という検索キーを入力すると、1 億件以上もの検索結果を出力するに至っている。波及的に「XXX 2.0」なる言葉も次々と生まれ、さまざまな分野におけるボトムアップによる変化の代名詞として認知されている。

Web 2.0 のビジネスモデルの成功例として、Google の検索連動型広告や amazon.com の推薦機能によるユーザの再利用率向上などが挙げられる。双方に特徴的なことは、主たるサービス (Google ならば検索サービス、amazon.com ならば小売サービス) を無料かつ徹底的にシンプルにしたユーザインタフェースによってユーザを大量に集め、集めた情報を分析して、それぞれの収益サービスにつなげているところにある。集客さえできれば、広告などの収益を確実に上げられるビジネスモデルを確立したことによって、YouTube (<http://www.youtube.com/>) や Flickr (<http://www.flickr.com/>) などの今までは到底考えることのできなかつた短期間での集客を実現することが可能となった。

このようなビジネスの成功は、オライリーによる 7 つの定義のうち、Web 2.0 の特徴キーワードとして挙げられる「集合知」の影響が最も大きい。インターネットを通じて「みんなで作ったもの (CGM: Consumer Generated Media)」が社会に対して、大きなインパクトを与えることになった。この CGM は、以下の 3 つに大別できる。

■知の集合

典型的なものに Wikipedia (http://en.wikipedia.org/wiki/Main_Page) がある。これはフリー百科事典と呼ばれ、Wikipedia 財団の定める基本方針に賛同する人ならば誰でも投稿・編集できる。日本語では、すでに 40 万件もの記事の投稿があり、さまざまな言葉の定義やその言葉にまつわる話題が百科事典的に記述されている。英語記事ではすでに 100 万件を超え、多言語化も進み、13 カ国の言語ではすでに 10 万件超の記事がある。統計的な情報を見る限りでは、1 つの言葉に対する編集は 14 回以上も行われ、従来の百科事典作成とはまった

問題	第1カラム	第2カラム	正解	第1候補	第2候補	判定
I agree () you that the actor is very talented. He gave an excellent performance.	agree	you	with	99 with	1 completely	○
I'll answer () his character as I know him very well.	answer	his character	for	30 for	19 about	○
A burglar broke () the convenience store last month, but I heard the police caught him yesterday.	broke	the convenience store	into	27 into	-	○
When Nicole heard that her favorite singer had died, she burst () tears.	burst	tears	into	93 into	7 in	○
I don't care () that sort of thing.	care	that sort	for	52 of	13 about	×
May I count () you to get me the job?	count	you to get	on	85 on	9 able	○
The football game might be called () on account of rain.	called	on account	off	13 furrier	12 vote	×
After his heart attack, Jim had to cut down () his sugar intake.	cut down	his sugar	on	-	-	×
By studying overseas, students can come into contact () other manners and customs.	contact	other manners	with	8 in	-	×
I came () my old diary when I was cleaning my room yesterday.	came	my old diary	across	21 across	10 up	○

表-1 英検の前置詞穴埋め問題の検索

く異なるサイクルで情報の追加・更新がなされている。

Wikipedia ほどこちんと管理されているものではないが、ブログに書かれている情報も、その利用する場面においては知の集合となる。ホテルやレストランの評価情報や新製品の評判情報を収集・分析するビジネスまでも登場している。これは、日本語だけでも70万件/日の記事の増加という数の力が、ブログの存在を無視できないメディアに押し上げたことは言うまでもない。

□集合の知

Webにはすでに100億以上のページ数が存在するとも言われているが、日々増減を繰り返す、まるで生命体のようなWebの正確な全体像はつかめていない。しかしながら、これだけの大量のページ数が存在するために、数の力の利用で今までは困難とされていた質的な変化が起こりつつある。たとえば、以下に示すように、英語検定の前置詞の穴埋め問題も解くことができる。

Googleでは、ユーザが検索の入力や出力を操作できるようなアプリケーション・インタフェースを公開しており、これをGoogle Web APIs (通称GoogleAPI)と呼んでいる。このGoogleAPIを用いて、ワイルドカードを含む検索式を作成し、その検索結果のワイルドカードに相当する部分を集計することで、世の中で使われている言い回しの傾向を調べることができる。

表-1の第1行目では、英語のagreeとyouの間にどのような単語がくるかという問題に対し、その候補を整理した結果である。第1カラムと第2カラムの言葉を検索キーとして検索したランキング上位1,000件の例文の集計の結果、ここでは、withという単語が挿入され

る例が多いことを示している。このような方法で、英検の前置詞の穴埋め問題を解いてみたところ、第2行目以降に示す結果を得た。これは、英検の準2級の問題集から前置詞の穴埋め問題を抽出し、穴埋めの前後の言葉をキーワードとして検索した結果、第1候補にくる単語が正解と一致するかどうかを判定した。第1候補と第2候補の単語の前には、出現した頻度を記している。

内容的に理解しないと穴埋め問題を解くことは難しいが、前後の単語を検索のキーワードとして第1カラムと第2カラムに入力するだけで、6割の問題の正解を得た。これは、膨大な用例を検索・集計することができるようになって初めて可能になったことであり、まさに、集合の知の利用の典型的な例である。このような大量データを用いた多数決の利用はさまざまな分野で試みられている。

□知の足跡

ユーザが無意識のうちに入力している「CGMデータ」として、各種のサービスの利用で残していくログデータがある。このデータは、「どのユーザが、いつ、どんなキーワードで、どのページを見たか(または何を購入したか)」といった情報である。個々のログが持つ意味は大きくないが、大量のログデータの中から決められた条件によってマイニングすることで、価値を引き出すことが可能になる。Webユーザのeコマースなどへのニーズを分析するためには非常に有効なデータであり、SEO/SEM (Search Engine Optimization / Search Engine Marketing) などへの応用が研究されている。

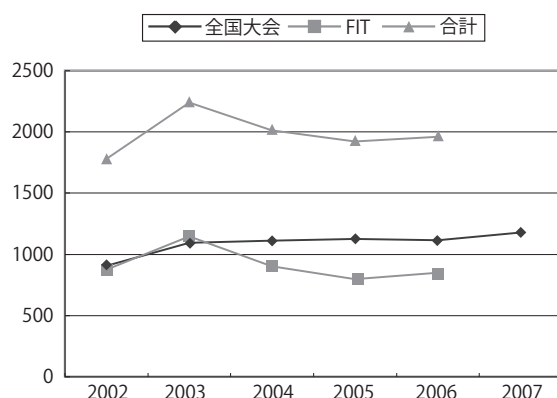


図-1 情報処理学会の論文数の推移

集合知から見た情報処理学会

情報処理学会をこうした3つの集合知の観点から見てみることにしよう。第1に、「知の集合」の観点では、図-1に示すように毎年行っている情報処理学会の全国大会とFIT（情報科学技術フォーラム）の論文数の合計は、ここ5年間2,000件前後で、ほぼ横ばいであり、「知の集合」としての役割は果たしている。

第2に、「集合の知」の観点では、投稿された論文を横串に集計したり、時系列的なテーマの変化を分析したりしていくことが考えられるが、こういったITを利用した分析はなされていない。

第3に、「知の足跡」の観点では、論文というコンテンツの検索サービスを提供することでそのログを収集することができ、ログデータのマイニングにより、実務家にとって必要な、技術的なトレンド情報や業務に関連する新たな技術の発見が容易にできる可能性がある(図-2)。

学会誌の連載を組んで

2006年4月より、本誌において「ナレッジマネジメント」の連載を組み、ナレッジマネジメントの新展開を毎月記載した。その反響として、毎月発行される学会誌の会員の声(<http://www.ipsj.or.jp/01kyotsu/kaiinnokoe.html>)に寄稿していただいている。今までにいただいた会員の声は、感想が16件、要望が7件である。また、無作為のモニタ評価では、7割程度の人がこの連載を読み、約3割が「大変役に立った」、「よい」という評価をつけている。このような反響をいただいているが、残念ながら感想に対しての議論や要望に対しての回答はまっ

たくできていない。学会誌という紙媒体で発行する以上、その対話には時間差が生じ、タイムリーかつインタラクティブな会話には限界がある。せつかく寄稿してもらった貴重な意見を反映できていないのが現状である。

これからの情報処理学会

これまで述べたように、情報処理学会は、「知の集合」の場としての役割は果たしているが、「集合の知」を創造する環境の提供や「知の足跡」の実務家に向けた発信はできていないのが現状である。

□タイムリーな意見交換の「場」として

こうしたリアルタイムでインタラクティブな意見交換の場の作成を目的とし、情報処理学会の技術応用委員会では、ITフォーラムを設立した(<http://www.ipsj.or.jp/10jigyo/forum/forumindex.html>)。これは以下に示す7つのテーマを議論している。

- ITアーキテクト/CIOフォーラム
- サービスサイエンスフォーラム
- イノベティブ社会基盤フォーラム
- Web2.0フォーラム
- ITダイバーシティフォーラム
- ユーザスタディフォーラム
- 福祉情報システムフォーラム

これらのテーマは、2007年1月25日に行われたソフトウェアジャパン2007においても一部のフォーラムを除き、オフライン会議が実施され、活発な意見が交わされた。このように、オンライン、オフラインを問わず、

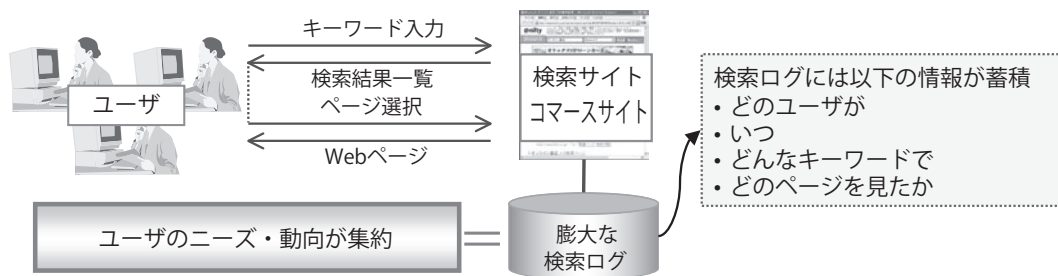


図-2 無意識な CGM データ

研究テーマを設定して意見交換を行うことは大変有意義であり、その「場」としての機能を情報処理学会は担っていく必要がある。

□学会誌・論文誌・予稿集のオンライン提供

すでに数名の理事が述べているように、論文の IT による提供は不可欠である。単なる提供だけでなく、会員が「集合の知」を創造できるような、全文検索や分野別、年代別などのさまざまな観点からの検索ができるような環境の提供を行っていく必要がある。

□実務家への情報提供

実務家が学会に期待する情報は、その分野のトレンドや方向性を示す情報である。また、業務に関連する最新の技術動向も知りたい。Web 2.0 ではまさに消費者から発信した情報によって企業が動かされる時代になってきた。同様に情報処理学会の発信する情報が実務家に対して大きな影響を与え、企業活動を変えていくようになら

なければいけない。そのためには、論文をコンテンツとした、マイニングなどの IT を駆使した情報分析の提供が必要であろう。

参考文献

- 1) 松井くにお：検索ロボット技術を活かした WWW 検索技術，日本語学 2004 年 2 月号。
- 2) 英検準 2 級，柴田バネッサ／小野聖次郎共著，高橋書店，ISBN4-471-27390-6
- 3) 松井他：ナレッジマネジメントにおけるテキストマイニング：情報処理，Vol.47, No.8, pp.893-899 (Aug. 2006).

(平成 19 年 2 月 19 日受付)

松井くにお (正会員)

matsui.kunio@jp.fujitsu.com

1980 年静岡大学工学部情報工学科卒業，同年 (株) 富士通研究所入社，2003 年東京工業大学大学院情報理工学研究科後期課程修了，工学博士，自然言語処理，文書情報処理，情報検索の研究開発に従事，言語処理学会評議員，厚生省電子カルテ研究班班員を歴任，特定非営利法人言語資源協会副会長，日本規格協会 ISO/TC37 国内対策委員，2006 年より本会理事。