1 古い IT の枠組からの脱脚

IT の世界では、ビジネスや技術開発の進み方が大変速く目まぐるしいという意味でドッキングという言葉が良く使われる。しかしこのような IT でも、基礎的な研究に関してはむしろ隔離気味と言えるのではないかだろう。ムーアの法則に従う PC の速度向上、メモリ・ディスク等記憶媒体の性能向上、インターネットのブロードバンド化、モバイル・ユビキタス環境の浸透、AIO や ASIMO に代表されるロボットの登場、Web 上で溢れ出す新らしいビジネス等々、一見華やかな IT の世界ではあるが、では 5 年後、10 年後に必要とされる技術に関して、我々は今どのような問題に取り組むべきなのかだろうか。

計算機が誕生してほぼ 60 年が経過し、IT の分野で簡単に解ける問題は殆ど解かれててしまったとも言われている。新しい方向性は、単により高い数値、低い数値を目指すだけでなく見えてこないであろう。米国大統領の科学技術に関する諮問機関である NSTC が発行する Networking and IT R&D という年次報告書（一般に Bluebook と呼ばれている）からも読み取るように、まず気づくのは、Internet/Web の発明、普及に伴い、従来の計算機という概念が拡張されている（あるいは崩壊している）ことである。

つまり、計算機という概念は通信やコンテツトも含み、大規模・複雑であることが前提となっている。こうして、ここ 10 年ほどの間に IT という分野の全体像は変質してしまい、古い IT の枠組みでは現実の問題が解けなくなっているのである。特に、人が直接に受けるシステムを支えるソフトウェアに関しては、大きな技術革新が期待されていることを感じる。

解決しない問題、今取り組むべき将来の課題はいろいろある筍だが、新しい IT を特徴付けるものの一つという意味で、古い IT の枠組みには登場していなかったコミュニケーション、メディア、コンテンツを取り上げたいと思う（無論、これが NTT コミュニケーション科学研究所の講演だからという理由もある）。

2 音楽を形式化できるのか

これまで音楽というのは一部の専門家だけが作曲、編曲、演奏などに従事できる、そうでない人々は単に聴いて楽しむことしかできなかった。しかし計算機による支援が得られるのであれば、一般的のユーザでも音楽を自由に処理することが可能となるだろう。実際、計算機による支援が最も進んでいるのは自然言語の分野であるが、そこでは多くのユーザが様々な状況で様々なスタイルの文章を自由にやりとりし、コミュニケーションを楽しんでいる。音楽に関しても同様の状況が実現できれば、大変有意義なことであると思われる。

しかし、音楽という対象は、計算機上で形式化（数学的表現し操作すること）には非常に手強い相手である。従来の IT の枠組では善し悪しの判断基準が精度、効率、構造、再現、予測、制御であ る。この判断基準の元で上手く取り扱うことができない対象のみが研究研究されてきたとも言える。一方、音楽の特徴として良く挙げられるのは主観、経験、感情、感覚、感情、芸術性等である。IT とは水と油と言ってして良いくらいであり、とても科学の別部に乗ることは思えない。実際、音楽で（音楽のための）コミュニケーションを行う試みや音楽コンテツトを創る試みの多くは、システムを適切に動作させるために、音楽のジャンルや様式を限定したり様々なヒューリスティクスを導入させるを得ないのが現在である。もし、そのような制限を避けようとするなら、できる操作は音楽を単にコピー（配信）することだけになってしまう。

音楽を音楽選んだの幸せであったように思う。なぜなら音楽には音楽理論があるからである。音楽理論の目的は、楽譜に書かれた音符を分析、理解し音楽的な構造を解明することである。音楽的な構造とは、楽譜上に見える音符を休符といった音楽的な情報そのもの、音符どうしのどのような結び付き方をしているのかという層深的な情報を意味する。音楽理論は、異なる音楽構造を聴取者に異なる認識を与え、その聴取者に与える認識には、段落感、終止感、進行感、揚げ感、調和感があるという前提に基づいて、それぞれの認識を支える重要な音楽分析のカーテンを引き上げることである。これに目頭を寄せる例としては、段落感、終止感、進行感、揚げ感、調和感等と低高の感覚であり、一種、感情、感情、感情という高次感覚ではないという点に留意しておりたい。

音楽を、この体系化と is-a 関係（抽象化/具体化関係あるいは包含関係）の対応付けが音楽の形式化への足掛かりと考えている（これ以上は本文のススメを越えるので割愛させて頂きます）。
3 音楽ではないメディアの形式化

まずここでは、メディアをそれによってコミュニケーションを行うものと定義しよう。我々にとって代表的なメディアは自然言語（テキスト）である。その自然言語においては、まず語に関して signifiant/signifie があり、さらに統語論、意味論、言用論があり、自然言語の文章の意味を解釈する。そして、ほん当たり前の事柄として、現実世界の事物と語の意味との対応付け（grounding）を前提にして計算機システムを構築する。このような自然言語の形式的処理体系が構築されているが故に、自然言語処理は多大なる恩恵を我々にもたらしている。これにより、他の画像、音、ジェスチャーなどのメディアに対しても、同様の構築の構築を試みたくなるのが人情である。だがそれも、これらのメディアは、自然言語のような形式的処理体系に従うよりも、筆者が音楽にいたい行なうようなゲシュタルトに基づく体制化と抽象化/具体化関係との対応付けによって解釈することが適切ではないかと思える。例えば、体制化に基づいてより抽象的な思い出、より具体的な視線というものを上手に規定するのである。ここで体制化された思い出や視線に連関する認識は、終止感、進行感、浮遊感等に対応する何らかの低次の感覚になる旨である。

4 メディアを統合する方法

近年、テキスト処理の分野では、マテデータやアノテーションの支援による検索、翻訳、要約という汎用性の高いタスクに注目が集まっている。小説や論文の執筆や、HP の作成といった大きなタスクの構成要素になるという意味で検索、翻訳、要約タスクは中粒度である。中粒度のタスクがインタフェースと自然に考えられるようになった理由は、中粒度までのタスクから計算機がテキストの意味を比較的正確に解釈し実行できるからである。

これによりメディア統合のヒントが現れている。というのも、画像、思い出、視線、音楽等テキスト以外のメディアでも体制化に対応付けられた抽象化/具体化関係を用いて検索、翻訳（スタイル変換あるいは言い換え（paraphrase））。要約のアルゴリズムを構築すれば、意味を正確に解釈したタスク実行が期待される。さらに、メディア固有な認識や処理を低減することも期待される。

これまでのテキストを除くメディア処理の多くは、メディアごとにバラバラの層層的な特徴を用いて、バラバラのインタフェースを提供していただけである。そして、この特徴を用いる手法は一般に合成に弱く、コンテンツ生成には向きもである。

従来のメディア統合、情報統合は、共通フォーマットでコンテンツデータを構造化することのみを指していた。筆者はこれに加えて、新しく中粒度の汎用タスク群のインタフェースを導入すべきであると考える。そのタスク群には、検索、要約、翻訳（スタイル変換）を初めとして、類似、分類、検索、キーフレーズ、説明、補間、予想等が含まれる（下図）。

新しい IT の枠組の元で今後我々が取り組むべき研究課題の 1 つは、中粒度の汎用タスクを自由に組み合わせられる環境やリソースを一般的ユーザに提供するプラットフォーム（あるいはミドルウェア）を構築することである。このメディア統合プラットフォームは、専門家でない一般的ユーザでも各分野のセンスに従って中粒度のタスクをデザインし、様々なメディアを使い分けたコミュニケーションを可能にする。

例えば、道案内を示す状況を考えてみよう。通常、道を教えられた人は相手の様子を見ながら説明を試みる。言葉だけの説明では不十分な場合は身振り手振りを交えて、間違えと話さずで十分の状況もあるだろう。相手にとって丁寧に説明する。道の人が間違っていく、このようなメディアを通じて伝えられた情報を統合して、相手の意図を再合成する問題になる。一方、道を教えた人にとっては、相手に伝えたい道具（つまり意図）を様々なメディアをどのように使い分けを効率良く、正確に伝えるかという問題になる。このような状況に応対できるシステムを構築する場合には、メディア統合プラットフォームという枠組が必要ではないかと思う。

本稿では提案するメディア統合プラットフォームがコミュニケーション科学のための新しいメディア意味論への一歩ではないかと考える。

平田 圭二 Keiji Hirata
メディア情報研究部 メディア表現研究グループ
特別研究員
hirata@brl.ntt.co.jp