

インテリジェントコンテンツ

橋田 浩一

産業技術総合研究所サイバーアシスト研究センター／科学技術振興事業団戦略的基礎研究推進事業

hasida.k@aist.go.jp

インテリジェントコンテンツ（意味構造を明示した情報コンテンツ）は、近距離の測位・通信技術とともに、近未来に実現可能な仕方でグラウンディング（実世界の意味をデジタル情報に内在させること）を実現する情報インフラの構成要素である。インテリジェントコンテンツのオーサリングとその利用は、人間と機械との間のインタラクティブな協調の下で行われ、人間の微妙な意味理解能力と機械の高速な計算・通信能力とを融合することにより、情報技術に基づくサービスに本質的な高度化をもたらす。インテリジェントコンテンツを普及させるキラーアプリケーションと、一般的な状況依存情報支援のための計算アーキテクチャについても述べる。

■情報インフラとしてのコンテンツ

連載の初回⁹⁾でも述べたように、デジタル情報に実世界の意味を内在させることをグラウンディング（grounding）という¹⁰⁾。グラウンディングは、情報技術によって人間に有意なサービスを提供するための必要条件である。グラウンディングを実現するための伝統的なアプローチは、自動的な画像理解や自然言語理解などの技術によって、人間が理解する意味を情報機器にも自律的に理解させる、という方法だった。ところが、これは人工知能の究極の課題であり、達成できるまでに数十年間を要する。

そこで、人工物による完全自律的なグラウンディングを目指すのではなく、何らかの情報インフラを整備することによってインタラクティブな情報環境を構築し、その中で人間や人工物の間の協調に基づいてグラウンディングを実現しようというのが、サイバーアシストのアプローチある。これは、数年以内に利用可能な技術でグラウンディングを達成する唯一の方法である。初回に論じた位置に基づく通信は、人間の生活世界の物理的な側面に関するグラウンディングを実現するための情報インフラであるが、生活世界には物理的側面以外にも観念的・社会的な側面もあり、それに関するグラウンディングを実現するための情報インフラがインテリジェントコンテンツである。

インテリジェントコンテンツとは、意味構造を明示された情報コンテンツである。たとえば、自然言語の文書の細かい意味構造（依存構造、等位構造、修辭構造、照応、共参照など）を人手作業を含む半自動的な処理に

よってXMLのタグで明示しておけば、その文書の意味は人工物にもかなりの程度まで「理解」でき、機械と人間との間で意味を共有できるはずである^{1), 2), 8), 7)}。いわゆるマルチメディアコンテンツについても同様に映像データや音声データの意味（たとえばビデオのショットに映っている人間などの物体やそれらの動きなど）を明示するタグを付与しておくことにより、その意味を人間と人工物との間で共有できる。

情報コンテンツの完全自動処理に基づいて知的情報サービスを実現しようとする試みには、画像理解、音声認識、自然言語処理等の技術の未熟ゆえに明らかな限界がある。高度で多様なサービスを実現するには、それに応じた高度な意味内容に関する人手によるアノテーションが必要である。ISO/IEC JTC 1/SC 29/WG 11における、マルチメディアコンテンツの内容記述に関するMPEG-7の標準化活動⁶⁾はこの事実の認識に基づく。

情報コンテンツの意味処理に関する研究は他にも数多く行われている。たとえばカーネギーメロン大学のInfermedia等においては、ほぼ自動処理に基づいてマルチメディアコンテンツを構造化し、それを利用して情報サービスを行う試みがなされているが、人手を用いたアノテーションは積極的には考慮されていない。我々のアプローチの1つの特徴は、コンテンツのオーサリングと利用において人間の知的能力を最大限に引き出す点にある。これによって結局は、アノテーションのメリットがコストを大きく上回るか、あるいはアノテーションによって逆にコンテンツ作成のコストを低減させることが可能となる。

一般に、情報技術によって単に既存の業務の効率を高めるだけでは、サービスの品質向上や新たなサービスの創造につながらず、また雇用の機会を減少させることにもなる。コンテンツの完全自動処理に基づく情報サービスにおいては、この傾向がとりわけ顕著である。我々のアプローチは、人間とコンピュータとの協調による半自動的なアノテーションを普及させることにより、高度なサービスを可能にするとともに雇用を創出する。アノテーション作業には一定のリテラシが必要であるため、国際的な水準から見て国民の平均的なリテラシが高い日本の労働市場は職業としてのアノテーションへの適合度が高く、インテリジェントコンテンツの普及は大きな雇用創出効果を持つ。

■知識表現とアノテーション

インテリジェントコンテンツには、以下で「知識表現方式」および「アノテーション方式」と呼ぶ、主として2通りのアプローチがある。

知識表現方式は、何らかの知識表現言語と形式的なオントロジを用いて意味内容を直接記述するアプローチである。これは、形式的な知識表現に基づいて妥当な推論を機械が自動的に行うという古典的な意味でのグラウンディングを目指したものと考えられる。この方式では、人間の設計者／理論家によって前以て与えられた知識表現言語の意味論とオントロジの妥当性によってグラウンディングを達成する。知識表現方式によるインテリジェントコンテンツへのアプローチとして、セマンティックウェブ¹¹⁾やUNL (Universal Networking Language)¹²⁾がある。セマンティックウェブの詳細については本号の特集を参照されたい。UNLは機械翻訳用の中間言語 (interlingua) をコンテンツの内容記述に用いることにより、多言語にわたる情報サービスを実現しようとする試みであり、日本の国連大学を中心に進められている。

アノテーション方式では、形式的な意味記述と原コンテンツとを精密にアラインメントすることによってグラウンディングを達成する。たとえばGDA (Global Document Annotation)³⁾では、通常は下のようXMLタグをテキストデータに埋め込むことにより、タグに基づく意味記述と元のデータとを細かく対応付ける。この例では、“Time flies like an arrow.”が「蠅を矢のように時間計測せよ」等ではなく「光陰矢のごとし」の意味であることが示されている。

```
<su sem="fly1">
  <np>Time </np>
  flies
  <adp opr="sim">like
    <np>
      <adp opr="sng">an </adp>
      arrow
    </np>
  </adp>
</su>
```

アノテーション方式に基づくインテリジェントコンテンツへのアプローチにはGDAやMPEG-7⁶⁾がある。GDAは著者が中心となって進めているプロジェクトであり、上の例のように言語データをXMLによってマークアップすることによって意味構造を明示し、これに基づいて社会全体の情報流通を高度化することを目指す。MPEG-7はISO/IECにおいて標準化が進められているマルチメディアコンテンツの内容記述の方式である。なお、GDAタグ集合の簡略版であるLinguistic DS⁸⁾がMPEG-7に提案され、現在はMPEG-7のMDS (Multimedia Description Scheme)のPDAM (Proposed Draft Amendment)に組み込まれている。2003年には言語データの意味構造記述方式としてこれが唯一公式の国際標準となる見込みである。MPEG-7によるビデオ等のコンテンツの記述ではGDAと違ってタグに基づく記述は元のマルチメディアコンテンツの外に置かれる(実はGDAタグもLinguistic DSの記述もまた元のテキストコンテンツの外に置くことができる)が、XMLに基づく記述と元のマルチメディアコンテンツとの対応関係を明示する点においては同様である。

アラインメント方式においては、機械による自動推論によってではなく、図-1のように機械と人間とのインタラクションによって、グラウンディングが成立し維持される。図中の線分はさまざまなインタラクションを表わす。太い線は密接で意味の濃いインタラクションである。つまり、人間は原コンテンツを理解し、したがって原コンテンツと有意味なインタラクションが可能である。機械にとっては、原コンテンツよりも形式的な記述の方が簡単に理解し操作することができる。したがって、原コンテンツと形式的な記述との間の密接なアラインメントは人間と機械との間の有意味なインタラクションを支える。アラインメントはこの意味においてグラウンディングをもたらし、したがって次のようなさまざまな利点を持つ。

まず、要約や検索において、アラインメントが必要

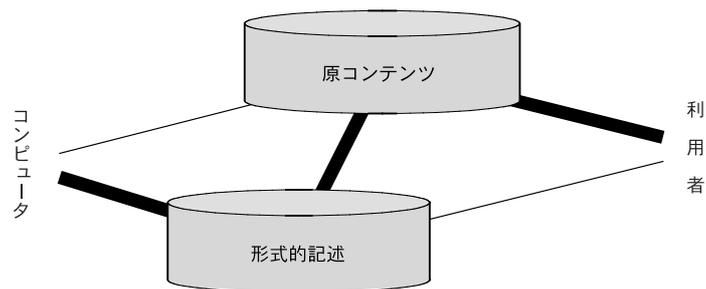


図-1 アラインメントとグラウンディング

かつ有用なことは明らかである。たとえば、ビデオデータとその意味記述との間に詳しいアラインメントがあれば、意味記述を要約し、その要約と対応するビデオデータの部分をそのビデオの要約とすることができる。

アラインメントの別の利点として、アラインメントに基づく人間とのインタラクションにより、形式的なオントロジなしでコンテンツへのさまざまなアクセスが可能になる。つまり、原コンテンツの意味を理解している人間とのインタラクションにより意味がアラインメントを介して機械に伝えられる。こうして形式的なオントロジに基づく自動推論が代替され、さらにしばしば自動推論を越える効果を生む。たとえば情報検索に含まれる推論は、後述のように人間の利用者と機械とのインタラクションによって実行できる。

アラインメントの第3の重要な利点は、記述(タグ付け)の作業がアラインメントによって容易になるということである。機械もある程度は原データを解析できるので、アラインメントがあれば人間が意味内容をすべて詳細に記述する必要はない。特に自然言語のコンテンツの場合には、大雑把な記述をまず人間が作成しておき、その記述と整合するもっと詳細な記述を機械が生成する、というアノテーションの方式が最も効率が良いだろう。この方式ではおそらく、朗読にかかる時間の3倍以内の時間で、検索や要約や翻訳の精度を十分に高めるようなアノテーションが可能と考えられる。

知識表現方式とアノテーション方式とは互いを補完する関係にある。原コンテンツがある場合にはタグによる記述をそれと密接に対応付けるべきであり、原コンテンツがない場合には形式的な意味論が必要となる。オントロジの整備に莫大なコストがかかることを含めて、知識表現方式の方がオーサリングのコストが高い

ので、インテリジェントコンテンツの普及はアノテーション方式から始まり、知識表現方式がそれに続くものと予想される。

■キラーアプリケーション

半自動的な意味構造のアノテーションを普及させるには、アノテーションのメリットがコストを上回る必要がある。既存の形式のコンテンツにさらにアノテーションを施す場合には、特にそのメリットが明確でなければならぬだろう。そのようなメリットが明確な応用に情報検索がある。

検索のメリットは情報発信者と検索の利用者の両方に及ぶ。たとえば、ある商品の広告が検索される頻度を5%増加させることができればそれに比例して売り上げが伸びる。その伸びがわずか5万円であっても、文章にして1,000文字程度の複雑さの広告を手で半自動的に意味構造化するには2千円程度しかかからないだろうから、広告主にとっては意味構造化した方が明らかに有利である。したがって、意味構造を利用する検索技術が普及すれば、半自動的な意味構造化も普及するだろう。意味構造を利用するインタラクティブな検索は、少なくとも自然言語の文書については完全自動的な意味構造化でも従来の検索より性能が高いので、半自動的な意味構造化が普及していない状態でも普及するだろう。すなわち、少なくとも自然言語のコンテンツに関しては、検索を目的とする半自動的な意味構造化が近いうちに普及すると考えられる。マルチメディアコンテンツの半自動的な意味構造化の普及は、コンテンツに含まれる言語モダリティの構造化(音声の書き起こしテキストの構造化など)から始まる可能性が高い。

一般に、情報検索が難しいのは、正解がどのように表現されているかが予測しにくいからである。「日本人のビジネスマンが海外出張中に交通事故に遭った」という検索質問の正解として「山田社長がアメリカで車にはねられた」という文があり得るが、この検索質問と正解の表現を結び付けるには何段かの推論が必要となる。ところが、そのような複雑な推論をコンピュータによって完全自動的に行うことは現在の技術では不可能である。しかし、人間とコンピュータが協調しつつインタラクティブにこのような推論を行うことは可能であり、その際に意味構造が役に立つ⁴⁾。その構造は、言語データの場合にはいわゆる係り受けや照応などからなり、ビデオデータの場合にはシーンが表現している状況において誰が何をしているか等の情報からなる。

再利用性の高いコンテンツに関しては、検索に限らずアノテーションのメリットが明らかである。そのようなコンテンツとしては、自然言語の文書では辞書、法律、判例などがある。マルチメディアコンテンツでは百科事典やある種の映画などのビデオデータが考えられる。また、政府や自治体による公報などの公共性の高いコンテンツを意味構造化し、外国人や高齢者を含むさまざまな利用者に合わせた検索やプレゼンテーションを可能にすることは、社会的な合意形成の促進につながり、社会的意義が大きい。

一方、アノテーションを含む新しい作法のオーサリングに基づく新たな形式のコンテンツにおいては、対応する既存の形式のコンテンツよりもオーサリングのコストをむしろ小さくすることが可能である。たとえばメールによる議論を考えると、ほとんどの場合、各発言に含まれる真に新しい情報はほんの1、2個の短文によって表わすことができ、発言の大部分は過去の発言を含む他のコンテンツからの引用や要約である。つまり、オーサリング作業のほとんどの部分は過去のコンテンツへの参照に費やされていることになる。したがって、各発言をまさに1、2個の短文とし、その中から他のコンテンツ中の情報への種々のリンクを張っておくことにすれば、オーサリングの手間は小さくなるはずである。また、これによって議論全体が意味的に細かく構造化されていくので、種々の利用者に応じて議論を自動的に要約したり可視化したりすることができ、途中から議論に参加したりすることが容易になる。要するに実効的な情報共有と情報公開が可能になるわけである。このような技術は、会社等の組織内での意思決定だけでなく、政策決定を含む、さまざまな社会的合意の形成を真に実現するためのツールになるだろう。

ハイパーテキストによる議論の構造化や知識の共有は、IBIS (Issue-Based Information System)⁵⁾ や Xanadu¹³⁾ においてすでに試みられてきたが、いずれも広く普及するには致っていない。これは、ハイパーテキスト化にかかる手間が大きくそれによるメリットが小さかったためだが、その原因は意味構造の粒度が粗過ぎたことだろう。やや逆説的だが、もっと細粒度のアノテーションを用いることにより、各発言を極限まで単純化できる。それによって得られる細かい意味構造により、上記の要約や可視化だけではなく、リンク先の情報を効率よく検索することが可能になり、これがオーサリングのコストを小さくすると同時にメリットを高めることになる。

■サイバーアシストの一般アーキテクチャ

今のところ、ユビキタス情報処理や情報家電はさまざまなアドホックな仕方で実現されているが、これではサービスの多様化が難しい。情報インフラに基づくグラウンディングを実現し、それによって多様な情報支援を可能にするには、そのための一般的かつ標準的な情報処理アーキテクチャが必要である。このアーキテクチャは以下の3つの層からなると考えられる。

- インタフェース層
- アプリケーション層
- 物理・インフラ層

物理・インフラ層は、測位および通信に関する物理的な過程と、それに基づくP2Pネットワークの構成やインターネットとの接続を司る層である。アプリケーション層は、自然言語解析、画像理解、音声認識、音声合成、屋内外の経路探索、意味構造に基づく検索や要約等を行うプログラムモジュールからなる。物理・インフラ層におけるセンシングから得られる情報と屋内外の地図（これもインテリジェントコンテンツである）を照合して人間に理解できる位置の表現を求めるモジュール等もこの層に含まれる。これらのモジュールはそれぞれの機能を形式的に記述するメタデータを持っており、インタフェース層においては、さまざまな人工エージェントがそのメタデータを参照しつつ利用者の情報要求や利用者の置かれた文脈に応じてそれらのモジュールを組み合わせ、状況に適したサービスを提供する。アプリケーション層に属するモジュールの利用はしばしば課金の対象となり、インタフェース層においては、サービスの品質に対する利用者からの要求等に応じた交渉に基づき、アプリケーション層のモジ

ジュールが選択される。

もちろん、これらの層はそれぞれ他からはほぼ独立に多様なかたちで実装することができる。たとえば物理・インフラ層に関しては、レーザ光等による近距離測位・通信に基づく実装、GPSによる屋外での測位と携帯電話等による通信に基づく実装、さらにはそれらを組み合わせた実装等が考えられる。また、アプリケーション層に属する各モジュールは他のモジュールや層とは独立に多様なパターン認識や意味解析の手法を用いることができる。

意味構造記述のフォーマットを標準化し、モジュールの入出力の形式をそのフォーマットで統一しておけば、アプリケーション層におけるさまざまなモジュールが多様な仕方で組み合わせられて相互運用可能となる。さまざまな種類の情報に関する多様な処理をいかにして統合するかという情報統合の課題は知能情報技術における重要な課題の1つだが、多様な処理モジュールの間でのこのようなプラグアンドプレイを可能にすることにより、部分的にせよ実際的にこの課題を解決できる。たとえば、ロボットを設計するには画像理解と行動計画を統合する必要があるが、画像理解の入出力も、行動計画の入出力も、意味構造化されたマルチメディアコンテンツとして統一的に表現し、両者の間で共有することが可能である。情報統合の課題の真の解決には処理過程の融合が必要と考えられるが、入出力フォーマットの標準化による情報の共有はその基盤を提供する。

■おわりに

インテリジェントコンテンツは一般利用者へのサービスのためのコンテンツであるばかりでなく、基礎的な研究開発用のデータにもなる。つまり、意味内容に関するアノテーションは、データの意味理解における正解を示す情報として、知能情報技術のさまざまな研究においてきわめて有用である。良質の研究用データを作成するには大きなコストを要するが、インテリジェントコンテンツが普及すれば、そのようなデータが大量に利用可能になる。これによって、ちょうどインターネットの普及が情報検索技術に急速な進歩をもたらしたのと同じく、知能情報技術の進歩が加速し、それはさらにインテリジェントコンテンツを作成するコストの低減と品質の向上につながる。このようにコンテンツビジネスと知能情報技術が直結することにより、相互にかつ直接的に発展を加速し合う正の循環が成り

立つ。

将来は、知能情報技術の進歩により、人手を含むアノテーションが不要になるかもしれない。しかしその技術革新は、大量かつ良質の研究用データと膨大な知識ベースを必要とすることが明らかであり、それには人手によるアノテーションが必要となる。だが、その莫大なコストを研究目的だけのために費やすのは非現実的である。また、知能情報技術が成熟するまでの長期間にわたりその成果が実用化されないとなると、そのための投資を社会的に容認するのは難しい。我々のアプローチは、知能情報技術の発展に必要な研究用データを持続的に生み出しつつ知能情報技術の産業応用を展開し、実験環境を創出する。知能情報技術の革新はこのような実践を通してのみ可能である。

参考文献

- 1) 橋田浩一: GDA: 意味的修飾に基づく多用途の知的コンテンツ, 人工知能学会誌, Vol.13, No.4, pp.528-535 (1998).
- 2) 橋田浩一: インテリジェントコンテンツと知の共有化, 1999年度人工知能学会全国大会 (第13回) 論文集, pp.40-42 (1999).
- 3) 橋田浩一: 大域文書修飾 (2001), <http://i-content.org/gda/>
- 4) 橋田浩一・豊浦 潤・津高新一郎: 構造化文書に基づくインタラクティブな意味的情報検索, 情報処理学会研究報告 99-ICS-116, pp.13-16 (1999).
- 5) Kunz, W. and Rittel, H.: Issues as Elements of Information Systems, Working Paper 131, Berkeley: Institute of Urban & Regional Development, University of California (1970).
- 6) Overview of the MPEG-7 Standard, <http://www.cselit.it/mpeg/standards/mpeg-7/mpeg-7.htm>.
- 7) 長尾 確: アノテーションに基づくデジタルコンテンツの高度利用, 情報処理, Vol.42, No.7, No.8, pp.668-675, pp.787-792 (July, Aug. 2001).
- 8) Nagao, K., Shirai, Y. and Squire, K.: Semantic Annotation and Transcoding: Making Web Content More Accessible, IEEE Multi-Media, Vol.8, No.2, pp.69-81 (2001).
- 9) 中島秀之・橋本政朋: 日常生活のための知的都市情報基盤, 情報処理, Vol.43, No.5, pp.573-578 (May 2002)
- 10) 中島秀之・橋田浩一・森 彰・伊藤日出男・本村陽一・車谷浩一・山本吉伸・和泉 潔・野田五十樹: 情報インフラに基づくグラウンディングとその応用—サイバーアシストプロジェクトの概要—, コンピュータソフトウェア, Vol.18, No.4, pp.48-56 (2001).
- 11) The Semantic Web Community Portal, <http://www.semanticweb.org/>
- 12) UNL Home Page, <http://www.unl.ias.unu.edu/>
- 13) Project Xanadu, <http://www.xanadu.net/>

(平成14年6月10日受付)



