

絵文字オントロジーの構築とその多元的検索

Building pictogram ontology and its multiple search methods

1 0 7 1 0 1 4

研究代表者 京都大学情報学研究科社会情報専攻 研究員(GCOE) 高崎俊之

【研究の目的】

近年のグローバル化で異文化交流は益々重要だが言語の壁は大きい。英語運用が主流だが言語デバインド等の課題も多く、言語・文化を超え交流できる ICT 環境構築活動が世界規模で非常に重要となってきた。

本研究では、ICT を最大限活用した絵文字コミュニケーション手法を提案することで、絵文字解釈の文化差異を逆に楽しめる多文化共生型の国際交流の実現を目指し、その基盤となる絵文字オントロジーの枠組みを示し、国内外での実証実験データを活用した文化差を考慮するテキスト検索や画像特徴検索など、多様な形態を備えた多元的検索の仕組みを構築することを目的とする。

【研究の内容、成果】

本研究では、文化的曖昧性と言語的曖昧性を吸収する汎用的な絵文字オントロジーとしての絵文字辞書データモデル提案と、その多元的絵文字検索手法の構築を行なった。以下に本研究の内容および成果について報告する。

1. 絵文字コミュニケーション

絵文字はこれまで様々なフィールドにおいて利用されてきている。障害により意思伝達の難しい人々が意思疎通できるように言葉に代わって用いられる言語代替手段としての絵文字、携帯電話やパソコンでのメールや BBS など主にテキストでの表現を情感豊かにするコミュニケーションとして用いられるエモティコンや顔文字、そして言語の成立に伴う歴史的な背景により利用されていた象形文字などである。障

害を持った人々へ向けたコミュニケーション手段として、AAC(代替拡大コミュニケーション)という学術分野で研究されてきた。また同言語圏内利用を前提とした携帯電話の絵文字に関する研究があったり、表情やジェスチャーなどのノンバーバルコミュニケーションの文化差異に関する研究などがなされてきた。ただ国際コミュニケーション手法を主目的とした絵文字の開発および研究は非常に稀であった。

本研究では、異文化コラボレーションという適用領域における言葉の壁を越えたコミュニケーションメディアとして絵文字を位置づける。本適用領域での実証フィールドとしては、児童の国際交流活動を実施している特定非営利活動法人パンゲアのアクティビティの場を使う。このような適用領域では図 1 のような絵文字メーカー等の絵文字コミュニケーション用ツールを利用するが、絵文字辞書の効果的な構築ができていない。多元的な検索ができる絵文字辞書を用意することで、ユーザにとって使いやすいツールが提供できる。フィールドの観察を通して、語彙の関連関係といった絵文字の意味に関する整理と、機能・形状などの視覚パターンによる整理の両方が大切であることが分かった。



図 1 絵文字ソフトウェア例: Communicator

本研究では、絵文字辞書の枠組みとして、絵文字の視覚的側面（絵文字の「絵」的な側面）と、絵文字の意味的側面（絵文字の「文字」的な側面）の2つの側面を併せて定義できるハイブリッド型絵文字辞書データモデルを構築した。

2. ハイブリッド型絵文字辞書データモデル

絵文字は「絵」と「文字」の両面の特徴を持つが、本適用領域でのハイブリッド型絵文字辞書データモデルの枠組みを構築するにあたって、大きく三つの課題が挙げられる。

一つ目は、異文化のコンテキストで利用されることを想定しているために、絵文字から解釈される意味概念がユーザの文化によって異なるという問題であり、ここではこれを文化的曖昧性と呼ぶ。従って、絵文字と意味概念は、1対1対応ではないセマンティクス構造を実現しなくてはならない。

二つ目は、言葉の多義性や表記ゆれの問題である。ユーザが「リンゴ」の絵文字を探したい時に「りんご」「リンゴ」「林檎」など、複数の検索文字列を入力する可能性があり、システム側はそれらの入力に応えないといけない。ここではこれを言語的曖昧性と呼ぶ。

三つ目は、ユーザは絵文字メールを創る時、必ずしも絵文字の意味概念として伝えたいケースばかりでなく、ASCIIアートのように絵文字を並べて、絵文字デザインの視覚的特徴を伝えたい場合もある、ということである。このような絵文字メッセージを創作する際にユーザは、意味概念での絵文字検索ではなく、視覚的特徴を検索キーとして絵文字を探したい。

これら三つの課題を踏まえ、文化的曖昧性と言語的曖昧性を吸収する絵文字概念と、視覚的な特徴を、それぞれ定義するハイブリッド型絵文字辞書データモデルの枠組みを提案する。

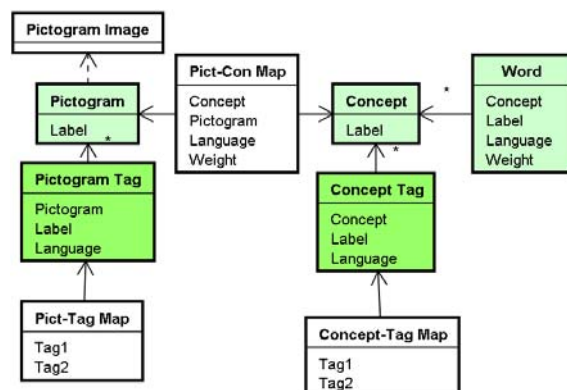


図2 絵文字辞書データモデル図

2.1 三つの基本要素

まずは基本要素として、絵文字、概念、語句の三つを定義した。

「絵文字」とは、絵文字デザインそのものであり、主な属性として絵文字画像への参照(実装上は画像ファイルへの参照)を持つ。「概念」とは、人間が認識する具体的または抽象的な意味を指し示すもので、言語に依存せず、ユーザはこの概念の存在を意識しない。概念は意味の粒度や概念レベルとの整合性を取るべきで、例えば「早朝」と「朝」という二つの概念があったときに、その絵文字辞書の目的やユーザコミュニティによっては、それらは一つ「朝」という概念にまとめて登録すべき場合がある。こういった首尾一貫性が必要となるため、概念の追加・編集・削除などの管理については、一般ユーザではなく、専門のユーザが管理することを想定している。「語句」とは、言語に依存した概念を指し示す文字列である。表記ゆれや言い換えなどから、一つの概念に対して複数の語句が存在することを想定している。

2.2 基本要素間の重み付き関連

上記の基本要素において、概念と絵文字、概念と語句が、それぞれ言語別の重みを伴った関連関係を持つ。ここでの言語とは、例えば、日

本語とか英語とかの自然言語の種別のことである。「概念と絵文字との関連」での重みとは、その対象となる概念と絵文字との関連の強さを0.0～1.0の数値で表す。これにより、ある概念(および言語)を入力すると、それに関連する絵文字群を関連の強さの順番でソート表示する、といったことが実現できる。「概念と語句との関連」での重みとは、その語句が指し示す概念への重要度(一般的な頻度)を指す。これは語句の多義性を考慮した時に、必ずしも1つの語句が1つの概念に対応するとは限らないためである。

以上の「概念・絵文字」「概念・語句」の重み付き関連により、概念を媒介して絵文字と語句が互いに関連づけられる仕組みとした。

2.3 二つの階層的関連要素

概念同士の関連や、絵文字同士の関連を定義するために、絵文字タグ、概念タグの二つの階層的関連要素を定義した。両者ともメタデータとして絵文字タグ内、および、概念タグ内、それぞれで階層関係を持つ特徴がある。これら二つの関連要素によって、上述した基本要素が整理されるだけでなく、ユーザは文字列検索でなくディレクトリ検索することも可能となる。

「絵文字タグ」とは絵文字の視覚的特徴を定義するメタデータで、基本要素の絵文字に紐付く。文化圏による絵文字タグと絵文字との紐付き方が異なるケースがあるため、絵文字タグは言語属性を持つ。視覚的特徴として、色彩(赤、青、緑など)、形状(四角形、円形、三角形など)、次元(平面的表現、立体的表現)、テクスチャー(金属的、布、砂など)といった要素が考えられる。また「絵文字タグマップ」という仕組みを導入している。これは、絵文字タグ同士の関連を階層構造として保持できる仕組みで、これによりユーザは絵文字の視覚的特徴について、キーワ

ード検索とディレクトリ検索の両方を行うことができる。

「概念タグ」とは概念のカテゴリーを定義するメタデータで、基本要素の概念に紐付く。文化圏によって概念タグと概念との紐付き方が異なるケースがあるため、概念タグはユーザ属性を持たせた。利用シーンおよびユーザによって概念のカテゴリー分類手法は異なるが、本研究の適用領域である児童の異文化コラボレーションシーンでは、基本概念・感情・エンターテイメント・人・場所・関連/比較・食べ物・動き・百科事典の9カテゴリーとして定義していた。また「概念タグマップ」という「概念タグ」間の階層関係も定義することができる仕組みを導入している。これにより、例えば、食べ物カテゴリーの下に、食事・おやつ・調理、などのサブカテゴリーを階層的に定義していくことが可能となる。

以上、絵文字タグと概念タグという二つの関連要素を定義することによって、絵文字の視覚的側面と意味的側面の両面において基本要素を構造化した。

3 実験データによる多層的検索システム構築

上述したハイブリッド型絵文字辞書データモデルを、汎用的なデータモデル・フレームワークとして実現できるように実装し、本適用領域のフィールドでの実証実験などによって多言語での基本要素のデータ群を得てテストデータとして利用したシステム構築を行なった。絵文字・概念・語句といった基本要素、概念のカテゴリーを示す概念タグ、そして視覚的特徴を示す絵文字タグなどを登録・編集・検索・削除等の操作を可能とする40種の関数群からなる絵文字Webサービスとして、JBoss / MySQL技術を使って実装した。



図 3 実験風景 (オーストリア ウィーン市)

ハイブリッド型絵文字辞書データモデルのシステム構築作業と平行して一部既存の調査データに加えるかたちで、120 個の絵文字に対応する語句解釈に関する被験者実験を、本研究の適用領域であるアジア・ヨーロッパ・アフリカなどの国際交流活動フィールドにて実施した(図 3)。回答データは語や語句の解釈語からなるが、これらの解釈語を絵文字ごとにリストアップし、このリストから固有語を取り出し、各固有語が出現する頻度を集計した。そして得られた語句について表記ゆれやデータ整形を施し、1133 個の概念を抽出し本 Web サービスを利用して登録した。これにより、例えば英単語の検索のリクエストに対して、XML 形式で検索結果のレスポンスを得た。システムが多言語処理に対応しているため、データが揃えば新規の言語を追加していくことも可能となっている。図 4 に month という概念で文字列検索した場合のレスポンス例を挙げる。ここで、重み **pictWeight** の総和が 1.0 にならないのは、語句群から概念抽出をするプロセスにおけるノイズ処理の影響である。

【今後の研究の方向・課題】

今後は、アルゴリズム改良などによるシステムパフォーマンス向上をすると共に、登録する絵文字や言語などの数を増やし、多くの国内外の異文化コラボレーションのフィールドで評価をしていきたい。また本 Web サービスを利用す

```

<response>
<item>
  <pictURL>pict_020pic.swf</pictURL>
  <conceptName>month</conceptName>
  <language>en</language>
  <pictWeight>0.79</pictWeight>
</item>
<item>
  <pictURL>pict_024pic.swf</pictURL>
  <conceptName>month</conceptName>
  <language>en</language>
  <pictWeight>0.057</pictWeight>
</item>
<item>
  <pictURL>pict_025pic.swf</pictURL>
  <conceptName>month</conceptName>
  <language>en</language>
  <pictWeight>0.042</pictWeight>
</item>
</response>

```

図 4 “month”の検索結果(左)と各対応絵文字(右)

る多様な絵文字アプリを開発し、ICT による新しい異文化コラボレーション像について研究していきたい。

【成果の発表、論文等】

1. Toshiyuki Takasaki, Yumiko Mori: Design and Development of a Pictogram Communication System for Children around the World. International Workshop on Intercultural Collaboration (IWIC-07), Springer-Verlag LNCS 4568, pp.193-206, 2007.
2. Heeryon Cho, Toru Ishida, Rieko Inaba, Toshiyuki Takasaki, Yumiko Mori: Pictogram Retrieval based on Collective Semantics," In: HCI Intelligent Multimodal Interaction Environments, Series: Lecture Notes in Computer Science, Vol. 4552, Springer, pp. 31-39.
3. Shogo Kamiya, Tomoko Koda, Toshiyuki Takasaki, Heeryon Cho: Development of Pictogram Conversion Tool for Cross-culture Collaboration, Proceedings of Human-Agent Interaction Symposium 2007 (HAI-2007). (in Japanese)
4. Toshiyuki Takasaki: Weaving Technologies into Empathetic Intercultural Collaboration among International Children, e-Culture Session, Asia-Pacific Advanced Network 2008 (APAN2008)
5. Heeryon Cho, Toru Ishida, Toshiyuki Takasaki, Satoshi Oyama: Assisting Pictogram Selection with Categorized Semantics, European Semantic Web Conference 2008 (ESWC2008, accepted)