

## GWAPによるオントロジ構築手法の提案

安永 ゆい† 望月 祥司‡ 森嶋 厚行‡\*

筑波大学 情報学群 知識情報・図書館学類† 筑波大学大学院 図書館情報メディア研究科‡ JST さきがけ\*

## 1. はじめに

近年、計算機では処理が困難な問題を、人の知を利用して処理する Human Computation<sup>1)</sup> が注目を集めている。Human Computation の例としては、人間のパターン認識力を利用して OCR では認識できない文字の認識を行う reCAPTCHA<sup>2)</sup> や、ゲームを通じて有益な副産物を得る GWAP (Game with a Purpose)<sup>3)</sup> などがある。よく知られている GWAP としては、人が遊んだ副産物として、画像にメタデータを付与する ESP Game や、Commonsense facts (共通知識) を集める Verbosity などがある。

本稿では、GWAP によるオントロジ構築手法の提案を行う。これまでに提案されているオントロジ構築手法には、Web 上の情報を利用して自動で構築する手法<sup>4)</sup> や、技術文書等から半自動で構築する手法<sup>5)</sup>、エディタ等を用いて人手により作成する手法があるが、我々の知る限り、GWAP を用いた構築手法は存在しない。

各オントロジ構築手法には一長一短あるが、GWAP を用いた手法には次のような利点があると考えられる。(1) Web データなどの元となるデータ無しでオントロジを作る事ができる。(2) 構築に参加する人は、楽しむことを目的にゲームで遊び、オントロジを構築するという意識は無いため、負担感が少ない。一方、ある程度以上の人数を集める必要があるという欠点も存在する。

オントロジを構築するための GWAP の設計は自明ではない。なぜなら、我々の知る限り、既存の GWAP では扱われていなかった次のような問題を扱う必要があるからである。(1) 与えられたデータのタグ付けなどではなく、データそのものを集めなくてはならない。(2) ゲームの参加者が解けないと思われる問題が多くあるため、単純なゲーム設計ではオントロジ構築に必要な情報が得られない可能性が高い。

提案手法は、GWAP の一種である Output-agreement 型のゲーム<sup>3)</sup> である。まず、Output-Agreement 型の GWAP について説明し、次に提案手法の説明を行う。

## 2. Output Agreement による GWAP

Output-agreement 型ゲームの構成を図1に示す。Output-agreement 型ゲームの代表としては次に説明する ESP Game がある。ESP Game とは、2人の参加者が、提示された画像に対して、お互いに相手がどのようなキーワードを入力するかを当て合うゲームである。キーワードが一致した場合には、両参加者に得点が与えられ、そのキーワードは、画

像検索などで利用されるキーワードとして採用される。

まず、Output-agreement 型ゲームの入力 (Input, 以下 I と略記) と出力 (Output, 以下 O と略記) という概念について説明する (図1)。Output-agreement 型ゲームにおける I とは、プレイヤーに対して提示するもの (ESP Game の場合は画像) を指す。また、O とは、プレイヤーからの回答 (ESP Game の場合はキーワード) を指す。

次に、一般的な Output-agreement 型ゲームの構造を説明する。Output-agreement 型ゲームのプレイヤーは2人である。各ゲームにおいて、2人のプレイヤーは同じ I を与えられ、その I に基づいた O を返す。2人のプレイヤーは相手のプレイヤーと同じ O を返すとポイントが得られるため、プレイヤーは相手のプレイヤーと同じ O を返すことを目指すことになる。ただし、プレイヤーは相手のプレイヤーの O を見たり、相手のプレイヤーとコミュニケーションをとることはできない。また、2人のプレイヤーがポイントを得る条件は、I が与えられている間 (例えば画面表示されている間) に返した O のいずれかが一致することである。

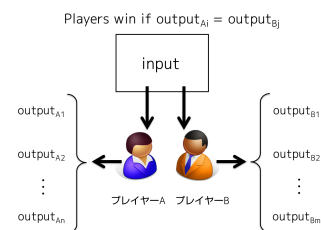


図1 Output-agreement 型ゲームの構造

## 3. 提案手法

提案手法の概要を図2に示す。本手法の入力は、オントロジのテーマとなる概念を表す語である。本手法には、Output-agreement 型のゲームが3つ含まれている。これは、(ゲーム1) 概念を集める“連想ゲーム”, (ゲーム2) 関係を集める“関係ゲーム”, (ゲーム3) フィルタリングを行う“O×ゲーム”である。これらを次に説明する。

**ゲーム1: 連想ゲーム.** Iとして、ある概念を表す語を提示し、そこから連想できる概念を表す語を推測するゲームである。例えば、Iが筑波山である場合、「筑波山と言えば?」という問題を提示し、参加者は「ガマの油」「ロープウェー」などをOとして返す。参加者のOが一致した場合のOを、オントロジの概念として採用する。

**ゲーム2: 関係ゲーム.** Iとして、ゲーム1で集められた2つの概念を提示し、それらの関係を表す語を推測するゲームである。例えば、「筑波山」の( )は「ガマの油」の空欄に入るものは?という問題を提示し、参加者は「名物」などをOとして返す。参加者のOが一致した場合のOを、Iである2つの概念間の関係として採用する。

**ゲーム3: O×ゲーム.** Iとして、ゲーム1とゲーム2で集められた、2つの概念とそれらの概念の関係を表す語を提示

Proposal of a Method for Building Ontologies with GWAPs  
Yui Yasunaga† Shoji Mochizuki‡ Atsuyuki Morishima‡\* College of Knowledge and Library Sciences, School of Informatics Univ of Tsukuba.†  
Grad. Sch. of Library, Information and Media Studies, Univ. of Tsukuba.‡  
PRESTO, JST\*

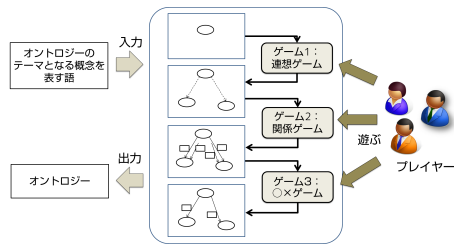


図2 提案手法の概要

し、これらの正誤を推測するゲームである。例えば、「筑波山」の「名物」は「ガマの油」は正しいか? という問題を提示し、参加者は○もしくは×を○として返す。参加者の○が一致した場合の○を、Iの正誤の判断として採用する。

以上3つのゲームを用いたオントロジ構築の流れは次の通りである。まず、入力されたテーマに関連する概念を表す語をゲーム1で集める。次に、これらの概念間の関係をゲーム2で集める。この時点で、グラフ構造が入手できるが、最後に、ゲーム3によって、集めた「概念」と「概念間の関係」の組み合わせから、不適切なものを除去する。ゲーム3の結果をオントロジとして出力する。

実際には、3つのゲームの順序は、集まった概念や関係を表す語によって前後する。すなわち、全ての概念を集めるためにゲーム1を行い、それが終わったらゲーム2を行い、全ての関係が集まったらゲーム3を行うわけではなく、概念や関係を表す語が入手できたら、それらに対して随時ゲーム2や3を行う。したがって、ゲーム参加者から見れば、これらのゲームは順不同で行われるように見える。

4. 本研究の取り組む問題とアプローチ

1章で触れたように、GWAPによるオントロジ構築では、既存のGWAPでは扱われなかった問題を扱う必要がある。これらの問題と本稿での提案手法をそれぞれ説明する。

**問題1と提案手法：語を数多く集める仕組み。** 既存のGWAPは、既にあるデータ (ESP Gameの場合は画像) に関してなんらかのデータ (ESP Gameの場合はキーワード) を集めていた。しかし、オントロジの構築では、新しい多様なデータ (概念を表す語) を多く集める必要がある。

そこで本稿では、連想ゲームを行う際、ある問題から得られた概念を他の問題に利用することを提案する。具体的には、「筑波山といえど?」という問題を常に出し続ける代わりに、既に得られた概念「ガマの油」を用いて、「筑波山のガマの油といえど?」のように問題に文脈を追加していく (図3)。これにより、前者に比べて多様な新しい概念を集めることを図る。

**問題2と提案手法：プレイヤーの持つ知識に合わせて問題を出す仕組み。** オントロジの構築が画像のタグ付けなどとは異なるのは、ある程度の専門知識が要求されることがしばしば有ることである。したがって、ゲームの参加者が必ずしも解けない問題の割合が多くなる。その結果、問題の選択を適切に行わなければ、オントロジ構築に必要な情報が得られる可能性が減ってしまうと予想される。

したがって、本稿では、ゲームを行うことで構築されるグラフ構造の特徴を利用することを提案する。このグラフ構

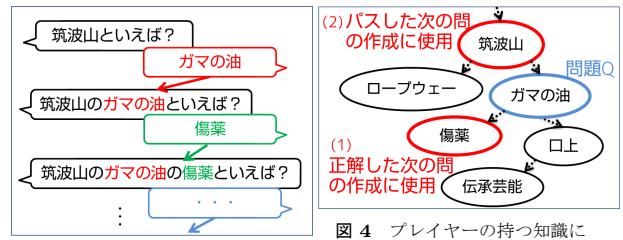


図3 語を数多く集める仕組み

図4 プレイヤーの持つ知識に合わせて問題を出す仕組み

造においては、親ノードが表す概念は子ノードが表す概念よりも一般的な概念を表していると考えられる。例えば、図4では、「筑波山」は「筑波山のガマの油」よりも一般的な概念を表していると考えられる。したがって、提案手法は次のようになる。すなわち、ある問題Q(例えばガマの油)に対するプレイヤーの入力が一致した場合にはQを作るのに利用したノードの子ノード(例えば傷薬)を使って次の問題を作成し、一致しなかった場合は、Qを作るのに利用したノードの親ノード(例えば筑波山)を使って次の問題を作る。

5. 予備実験

13名の被験者を使って、問題1に対する提案手法が有効である可能性を調査する予備的な実験を行った。具体的には、「筑波大」をテーマに、常に「筑波大といえど?」という問題を出す連想ゲームと、入手した語を使い、文脈を追加した問題を出す連想ゲームの2つのゲームを用意し、各ゲームを1回ずつ行ってもらった。各回のゲームは3分で終了する。その結果、集まった概念の数は、前者では117、後者は184となった。

6. まとめと今後の課題

本稿では、GWAPによるオントロジ構築手法を提案した。今後の課題としては、長期間にわたる本格的な実験、より遊びやすく、また効果的にオントロジを構築できるようなゲームの開発、オントロジ構築プロセスにおける、同義語等への対応が挙げられる。

7. 謝辞

本研究の一部は科学研究費補助金若手研究(B)(#20700076)およびJST さきがけ「情報環境と人」の支援による。

参考文献

- 1) Luis von Ahn. "Invited Talk: Human computation". International Conference On Knowledge Capture. Proceedings of the 4th international conference on Knowledge capture, 2007, pp 5-6, 2007.
- 2) reCAPTCHA. "What is reCAPTCHA?". <http://recaptcha.net/learnmore.html>
- 3) Luis von Ahn, Laura Dabbish. "Designing Games With A Purpose". Communication of the ACM Vol.51(8) pp.58-67, 2008.
- 4) Avigdor Gal, Giovanni Modica. "Ontobuilder: Fully automatic extraction and consolidation of ontologies from Web sources using sequence semantics". In Proceedings of the EDBT Workshops. pp.573-576, 2006.
- 5) 森田 武史, 和泉 憲明, 山口 高平. "オントロジ検索エンジンを用いた領域オントロジ構築支援環境 DODDLE-OWLの拡張". 人工知能学会セマンティック Web とオントロジー研究会 第15回 SIG-SWO-A603-07, 2007.