

Historical information Acquisition System (HAS) の実装

大竹栄一¹ 遠藤正隆² 中嶋裕一² 三浦哲郎² 菱田隆彰³

概要： 近年対話システムの分野で人より自然に近い会話を行う研究が盛んに行われている。会話を盛り上げるためには、相手が興味を持っている話題を提供する必要がある。相手の好む話題を選び出すには、相手の特徴となる情報を対話の中から取得する必要がある。我々は、誘導質問術を用いて対話の誘導を行い、対話者の特徴情報を推定する手法を提案し、年齢推定を補助する過去の出来事情報提供システム HAS の試作を行った。HAS が効果的に利用されるには多くの過去の出来事の詳細な情報が必要であるが、現状ではその情報を収集、集積する仕組みを用意できていない点が問題となっている。本研究では、HAS においてより多くの出来事情報を提供するために、Wikipedia を対象とした情報の取得方法を検討し、HAS への実装を行う。

Implementation of Historical information Acquisition System (HAS)

EIICHI OTAKE¹ MASATAKA ENDO² YUICHI NAKASHIMA²
TETSURO MIURA² TAKAAKI HISHIDA³

1. はじめに

対話システムは、主に特定の情報の授受のために対話を行うタスク指向型と、会話の継続を目的とする非タスク指向型に分けられる。最近では、人より自然に会話できる方法が盛んに研究されている。質問を行うだけでなく、前後に雑談を挟むことで、相手とのより自然に近い対話を行うことで少ないストレスで目的を達成することが可能となることが示されている[1]。相手を楽しませながら自然な会話を行うためには、提供する話題を相手の興味の持つ話題を選ぶ必要がある。そのために、相手の個人的な情報を会話から手に入れる必要がある。小林らの研究[2]では、会話の中に現れる特徴的なフレーズを個性として紐付けることで、相手の嗜好や状態を推測できることが示されている。しかし、実際には個人を特徴付けるような情報を平易な会話の中から取得するのは困難な作業である。

そこで我々は、よりユーザ個人にあった話題を提供するために、雑談の最中に意図的に特定の情報が出しやすくなるような対話の誘導を行うことで、平易な内容に付帯する個性的な情報を入手することを検討し始めた。一つの例として、相手が体験した出来事の話から、その時期とその時の相手の年齢がわかれば自ずと現在の年齢が推定できることになる。問題点は様々な出来事からその実施時期を得ることは容易ではないことである。また、歴史的な大事件ならまだしも、地域で起こった些細な出来事や特定の分野で有名な出来事の発生時期を知ることは難しい。

これらの問題を解決するために、大小様々な出来事について発生した時期を一元的に管理し、必要に応じて出来事

のデータを提供できるようにするための一つの方法としてとその発生時期を相互に変換することのできる Historical information Acquisition System, 略して HAS を提案し、その基本機構の試作を行った[3]。本研究では、HAS でより多く Wikipedia を対象に過去の発生した出来事の情報の取得を行う方法を検討し、HAS への実装を行う。

2. 誘導質問術による情報の収集手法

会話の中から相手に気づかれないように相手の情報を引き出すための質問技術を誘導質問技術 (elicitation techniques) と呼ぶ。ソーシャルエンジニアリングの手口としても用いられる技術でもあるが、会話を継続したり盛り上げたりするために有用な技術である。何気無い会話の中から相手の嗜好を読み取り、好みそうな新たな話題を展開する。

例えば、特定の職種の人が使用する単語を雑談から得た場合、より専門的な用語を含めたいいくつかの質問を加えることで、相手がどのような職務についているか特定できたり、相手が体験した特徴的な出来事から、その時の年齢を知ることができれば、現在の年齢を推定したりすることが

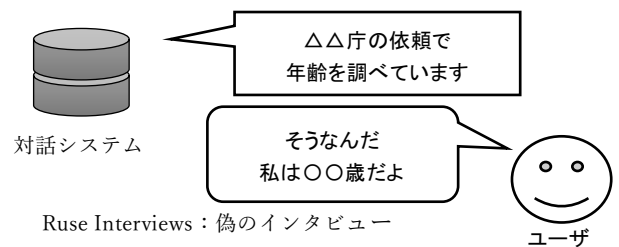


図1：誘導質問の例

¹ 愛知工業大学大学院経営情報科学研究科

² 株式会社リオ

³ 愛知工業大学情報科学部

できる。知り得た情報からその職業や年齢の人が好む他の話題を会話に出すことで、会話が盛り上がり信頼を得ることが可能となる。

現在人工知能などを利用して人と雑談する対話システム（以降、雑談システムと呼ぶ）の研究が盛んに行われるようになっている。人口の減少なども伴い、人の代わりに人との対応を自然に行うためのシステムが必要となっているためである。既存の雑談システムでは、話題を膨らませる手法として相手の発話の中から特徴的な単語を取得し、ネットワーク情報の情報サイトからその単語の関連用語を入手し話題を生成するという仕組みが多く、話題の転換は一方的なものになってしまう。相手の嗜好に沿った話題をうまく選別する仕組みが必要だと思われる。

雑談システムがより人間的に会話を続けるための一つの方法として、我々は誘導質問技術を使用し、相手の職業や年齢などの属性を推定し、話題の絞り込みに利用することを提案した。

3. 年齢の推定と出来事の時期の取得

誘導質問の技術は入手したい情報に応じて様々な方法が考えられるが、本稿では年齢を推定するための技術を対象とする。

現在の年齢を直接聞くことなく推定するのに有効な話題の一つとして、相手が過去に体験した出来事が挙げられる。話題にでた出来事の時期（年）を a 、相手が体験した時の年齢を b 、現在の暦年を n とした時、相手の年齢 x は、

$$x = b + (n - a)$$

で求めることができる。この時話題に出た出来事の名前から時期を導出できれば相手から直接聞き出す必要のある情報は b のみとなり、 b を問う質問を行うだけで良い。

この年齢推定手法は手順が簡便で必要な情報も明確であるため雑談システムへの実装は有効だと思われるが、そのためにはできる限り多くの様々な出来事の情報を用意しなければならない。また、出来事の発生時期を情報サイトから検索・抽出する方法も考えられるが、手間がかかり正確な情報が得られない場合もある。

そこで我々は、雑談システムの情報収集支援に用いるための、多様な過去の出来事の発生時期を知ることのできるシステムの設計を行った。出来事とその時期を相互変換可能なこの提案システムは Historical information Acquisition System (HAS) と呼ぶ。

4. HAS の概要

HAS とは、過去の様々な分野のあらゆる出来事の時期を提供するためのシステムである。基本的な機能は、出来事の名前や属性とその発生時期（年月日）を相互に変換する

ことである。ドメイン名と IP アドレスの変換を行う Domain Name Service のように世界中の出来事の発生時期を一元的に集積・提供が可能となることで、過去の出来事を一意で定量的な情報として様々なシステムで活用することができるようになる。

HAS のデータ構造とコマンドラインインタフェースを (CLI) の基本設計を示す。HAS が取り扱うデータ属性を表 1 に示す。出来事とその発生時期を紐づけるために出来事の名前及び別称、発生年月日、発生位置等を項目とした。例えば、2005 年に開催された「愛・地球博」の場合、図 2 のような内容で構成される。登録された出来事のデータを元に情報を得るためには表 2 に示す「get」コマンドを用いて行う。X に取得したい属性名を指定することで、該当するデータのみの抽出が可能となる。コマンドの実行例を図 3 に示す。図 3 のコマンドで愛・地球博の開催年を json 形式で取得することができる。また、図 2 のように基本コマンドに表 3 に示すオプションコマンドを追加することで、取得する情報に条件をつけて取得が可能である。

表 1：属性一覧

属性	備考	値
year	必須	西暦
year_alt	NULL, 複数可	西暦の別称
date	NULL 可	日付
name	必須	出来事の名前
name_alt	NULL, 複数可	出来事の別称
category	複数可	カテゴリ
location	NULL 可	位置情報

```
{
  "year": 2005,
  "year_alt": ["平成 17 年", "21 世紀"],
  "date": "3/25-9/25",
  "event": "2005 年日本国債展覧会",
  "event_alt": "愛・地球博",
  "category": ["イベント", "愛知県"],
  "location": "35.1718934,137.0898625"
}
```

図 2：愛・地球博のデータ例 (json 形式)

表 2：データの取得コマンド

has get X	X の取得
X：データの属性名	

```
>has get year match "愛・地球博"
{"Event": [
  {"year": 2005}
]}
```

図 3：コマンド例

表 3：オプションコマンド（一部）

オプション	説明
at A	A の年代
before A	A の年以前
from A to B	A から B の間の年
by C	カテゴリ C を含む

A, B：データの属性値, C：カテゴリ

5. 提供するデータの取得

5.1 出来事の情報取得に関する検討

HAS では過去の出来事を提供する際、出来事に関する情報を規定したフォーマットで json の形式に収めて提供を行う。元になる出来事の情報は何らかの形で入手する必要があり、最も単純な方法としては事前に HAS のシステム内に登録しておくことでその情報を提供可能となる。しかし、手作業で出来事の情報の一つ一つ調べて登録することは非効率である。我々は HAS がより多くの情報を扱えるようにするため、Web 上に既存の出来事に関する情報を逐次入手し、必要な情報を抽出して整形し仕様に沿った出力を行う仕組みが必要であると考え、本稿では、その既存の情報として Wikipedia を対象とし、掲載されている出来事の情報から、HAS に必要な情報を抽出する方法を検討する。

5.2 Wikipedia の情報

Wikipedia とは、ウィキメディア財団が運営しているインターネット百科事典である。誰でも記事を編集することができ、大小様々な情報が掲載されている。過去に発生した出来事に関しても多くの情報が掲載されているため、それらの中から HAS が必要とする出来事に関する情報を引き出すことができれば、その内容を利用した出力を生成することが可能となるだろう。

Wikipedia では、記事の右側に出来事の基本情報を表すインフォボックスが設置されている．図 4 に「2005 年日本国際博覧会（愛・地球博）」と「国際花と緑の博覧会」のインフォボックスを示す．どちらも出来事の名称や通称，開催時期，会場の場所などいくつかの共通の項目が示されている．Wikipedia の記事の本文では，記者によって様々な表現がされているが，インフォボックスでは共通の項目が簡素な内容で確認できるようになっている．この項目のうちのいくつかは HAS が必要とするデータである．

HAS での開催年と開催月日を示す「year」, 「year_alt」, 「date」はインフォボックスの「開催時期」を参照することで入手できる。「event」と「event_alt」, 「category」は「正式名称」と「通称・略称」, 「イベントの種類」をそのまま使うことができる。場所を示す「location」は緯度経度情報に変換する必要があるが、「会場」を参照することで入手できる。Wikipedia のインフォボックスから動的に情報を抽出できれば Wikipedia のイベント記事が HAS の出来事の情報のデータベースとして利用可能となる。

5.3 DBpedia Japanese を用いた情報取得

Wikipedia から情報を取得する方法として、DBpedia から取得する方法がある。DBpedia とは、Wikipedia から構造化データセットを抽出し Linked Data として扱えるように提

供するサービスで、Wikipedia の情報の一部を容易に抽出することができる。

DBpedia 自体は主に Wikipedia 英語版を対象としているサービスであり、日本語版の Wikipedia からのデータの入手は DBpedia Japanese から行う。データの取得は用意されている SPARQL のエンドポイントから行う。

本稿では、DBpedia Japanese を経由して Wikipedia のイベント記事からインフォボックスに記載されている基本情報を抽出し、その内容を HAS の出力フォーマットへの整形をする仕組みの実装を行う。処理の行程は大きく二つに分けられる。一つ目は、DBpedia Japanese に出来事の情報を問い合わせ、Wikipedia 上の情報を取得する処理。もう一つは取得した情報を HAS の仕様に合わせた出力に整形する処理である。

最初の処理の例として、愛・地球博の名称と開催年を取得するクエリ例を図5に示す。結果はHTML, CSV, jsonなど様々なフォーマットで入手できる。図6は図5のクエリ結果をjson形式で取得したものである。図6において、各属性の中に「value」の値として表記されている部分が、Wikipediaの記事から抽出された情報である。

2005年日本国際博覧会

*The 2005 World Exposition, Aichi,
Japan*

愛・地球博長久手会場

イベントの種類 国際博覧会

通称・略称

愛・地球博(公式愛称)
愛知万博(略称)

正式名称

2005年日本国際博覧会

開催時期

2005年(平成17年)3月25
日 - 9月25日(185日間)

会場

● 日本 愛知県愛知郡長久
手町・長久手市・豊田
市・瀬戸市

主催

財団法人 2005年日本国際博
覧会協会

来場者数

2,204万9,544人(目標:
1,500万人)

長久手会場への交通アクセス

最寄駅

リニモ 万博会場駅^[1]

駐車場

パークランドライド駐車場有
公式サイト [@](#)

備考:

このテンプレートには、出展数、最寄駅、直通
バスや駐車場の有無が記載されておらず、情
報が不足しています。是非、このテンプレー
トに加筆して、情報の更なる拡充に努めてくだ
さい。

国際花と緑の博覧会

*The International Garden and
Greenery
Exposition, Osaka, Japan, 1990*

国際花と緑の博覧会会場 (1990年4月1日)

イベントの種類 国際博覧会 (特別博)

国際園芸博覧会

通称・略称

花の万博、花博、EXPO'90

正式名称

国際花と緑の博覧会

開催時期

1990年4月1日 - 9月30日

会場

鶴見緑地

主催

大阪府大阪市鶴見区・守口市

出展数

財団法人国際花と緑の博覧会協会
83カ国(日本を含む)と55の国際機関

来場者数

212企業・団体
2312万6934名

最寄駅

大阪市宮地下鉄鶴見緑地駅

直通バス

大阪国際空港、JR新大阪駅、JR
大阪駅、JR茨木駅、阪急南茨木
駅、京阪守口市駅、近畿奈良駅の
7箇所から直通シャトルバス
有

駐車場

公式サイト [@](#)

図4: Wikipedia のインフォボックス
(左: 2005 年日本国際博覧会, 右: 国際花と緑の博覧会)

```

prefix dbp: <http://ja.dbpedia.org/resource/>
prefix dbp-owl: <http://dbpedia.org/ontology/>

select ?date ?event ?event_alt ?category ?location where{
  dbp:2005 年日本国際博覧会 prop-ja:開催時期 ?date .
  dbp:2005 年日本国際博覧会 rdfs:label ?event .
  dbp:2005 年日本国際博覧会 prop-ja:通称 ?event_alt .
  dbp:2005 年日本国際博覧会 prop-ja:種類 ?category .
  dbp:2005 年日本国際博覧会 prop-ja:会場 ?location .
}

```

図 5 : SPARQL クエリ例

```

{
  "head":{
    "link":[ ],
    "vars":["name","year","event_alt","category",location]
  },
  "results":{
    "distinct":false,
    "ordered":true,
    "bindings":[
      "date":{
        "type":"typed-literal",
        "datatype":"http://www.w3.org/2001/XMLSchema#gMonthDay",
        "value":"-03-25"
      },
      "event":{
        "type":"literal",
        "xml:lang":"ja",
        "value":"2005 年日本国際博覧会"
      },
      "event_alt":{
        "type":"literal",
        "xml:lang":"ja",
        "value":"愛・地球博、愛知万博"
      },
      "category":{
        "type":"literal",
        "xml:lang":"ja",
        "value":"国際博"
      },
      "location":{
        "type":"literal",
        "xml:lang":"ja",
        "value":"愛知県愛知郡長久手町（現：長久手市・豊田市・瀬戸市）"
      }
    ]
  }
}

```

図6：DBpedia の出力結果

二つ目の処理は、図6として得られたデータの整形である。HAS のデータフォーマットは図2で示した通り、「属性：値」という単純な構造をしているため、取得してきたデータの head 部分や type などの不要な属性を取り除き、value の値をそれぞれの親属性の値として利用する。

データの整形時には、値の表記の変更や変換が必要な場合がある。図6の場合、「date」の値を「-03-25」から「3/25」に変更する。「event_alt」では通称が複数個存在するので、「愛・地球博、愛知万博」を「愛・地球博」、「愛知万博」と配列形式に変換する。最後に「location」には緯度経度情報にする必要があるため、「愛知県愛知郡長久手町（現：長久手市）・豊田市・瀬戸市」を「35.1718934,137.0898625」に変換する。最終的な整形後のデータを図7に示す。一連の処理を終えたデータは HAS の出力として利用者に提供される事になる。

```

{
  "date":"3/25",
  "event":"2005 年日本国際博覧会",
  "event_alt":["愛・地球博","愛知万博"],
  "category":"国際博",
  "location":"35.1718934,137.0898625"
}

```

図7：整形後のデータ

6. まとめ

本研究では、作成した HAS の扱えるデータ量を増やす方法として、Wikipedia からデータを抽出して HAS で利用する方法の検討を行った。DBpedia で公開されている SPARQL のエンドポイントを用いて特定の出来事の欲しい情報の取得を行い、取得したデータに対して整形を行い HAS で提供可能なデータに変換する機構の実装を行った。結果として、事前に大量のデータを用意することなく、多くの出来事の情報を提供できるようになった。

しかし、DBpedia の利用にはいくつかの課題があることがわかった。DBpedia の結果と Wikipedia のインフォボックスに記載されている値と内容が異なる場合があった。例えば「開催時期」の値は「-03-25」となっており、開催開始の月日の情報だけが得られた。「種類」の値では、Wikipedia 上では「国際博覧会」であるが、得られた結果は「国際博」となっており、一部が欠損しているようであった、何らかの対処を検討する必要があるだろう。

今後の課題として、本稿では取得できなかった一部の情報の取得方法の検討や、Wikipedia では取得できない出来事の取得方法の検討があげられる。また、一度取得した情報を再問い合わせしないような効率化の検討も行いたい。最終的には、本来の目的である対話システムにも組み込み、相手の年齢推定が行えるか、また実際に話題の提供を行い会話が盛り上がるかの検証も行いたいと考えている

謝辞

本研究は、JSPS 科研費 JP19K12073 の助成を受けたものです。

参考文献

- [1] 大村祐司, 川端豪, 「雑談可能な目的達成型音声対話システム」, 情報処理学会研究報告, Vol2012-SLP-94, No.9, pp.1-5, (2012)
- [2] 小林峻也, 萩原将文, 「ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム」, 人工知能学会論文誌, Vol.31, No.2, pp.1-10, (2016)
- [3] 大竹栄一, 遠藤正隆, 中嶋裕一, 三浦哲郎, 菱田隆彰, 「Historical information Acquisition System(HAS)の設計と試作」, 情報処理学会第 81 回全国大会講演論文集, 7W-04, (2019)