

# DBpedia を用いた作家推薦への取り組み

一瀬 詩織 (指導教員：小林 一郎)

## 1 はじめに

読みたい本を選ぶ時、好きな作家の別の著作など、既に興味を持っている作家を参考にして本を決定する場面がある。興味のある作家と関係のある別の作家を推薦する場合、この新しい作家に対しても、ユーザは興味を持つ可能性がある。本研究では、近年様々な広がりを見せている Linked Open Data の一部であり、Wikipedia の持つ膨大な情報を利用して作られた DBpedia を、作家間の関係取得に利用し作家の推薦を行うことを試みた。DBpedia より取得した作家に対し、それぞれ作家の著名度、関係の強さを用いて行う二つの推薦手法を考案し、実際に実験による推薦対象の取得を行った。

## 2 推薦の流れ

推薦は 2 つのステップで行う。まず着目対象に関連するリソースを推薦候補として DBpedia より抽出する。次に、これらのリソースをそれぞれの手法を用いて評価し、結果スコアの高いものを着目対象に対しての推薦とする。

リソースが持つ関係は、作家によっては数十種類存在する。全ての関係を用いると関連するリソースの数が膨大になるため、今回は推薦対象の抽出に利用する関係を、特定の種類に限定することとした。事前アンケートにより作家の影響関係が推薦に有効であると判断し、DBpedia のオントロジー “<http://dbpedia.org/ontology/influenced>”、 “<http://dbpedia.org/ontology/influencedBy>” によってリンクした作家の関連リソースを推薦候補の抽出に用いることとした。着目作家から推薦候補の作家まで、関係のリンクを辿る回数を今後 2 者間の ‘距離’ と呼ぶ。DBpedia において作家 “Haruki Murakami” を表すリソース<sup>1</sup> と、距離が 1 の作家との関係を図 1 に示す。

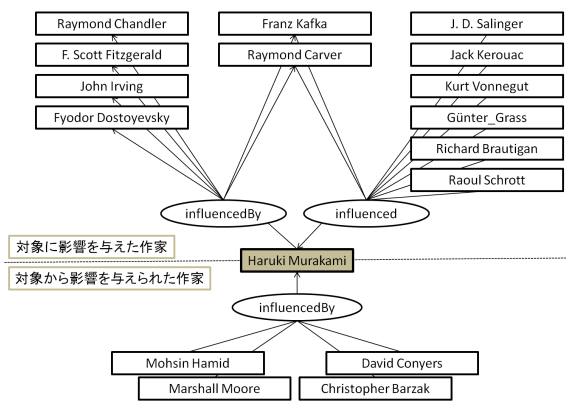


図 1: “Haruki Murakami” から距離が 1 の関連作家

DBpedia からの関連作家の取得は、エンドポイント<sup>2</sup>を通じた DBpedia への問い合わせにより行う。作家  $w$  と距離が 1 の作家  $x$  を取得するために用いる SPARQL<sup>3</sup>クエリを例として以下に示す。

<sup>1</sup>[http://dbpedia.org/resource/Haruki\\_Murakami](http://dbpedia.org/resource/Haruki_Murakami)

<sup>2</sup><http://dbpedia.org/sparql>

<sup>3</sup><http://www.w3.org/TR/rdf-sparql-query/>

```
SELECT * WHERE {  
  { <作家リソース w> ?property ?x . }  
  UNION  
  { ?x ?property <作家リソース w> . }  
  ?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/Writer> . }
```

## 3 作家の著名度を考慮した推薦

ある作家  $w$  と関係を持つ作家のうち、著名度の高い作家を推薦する。評価には Onishi ら [1] によって定義されている、他のリソースとの関係において着目するリソースの重要性を測る指標 Hub Score を用いた。着目作家  $w$  と関係によってリンクしたリソースの集合を  $R = \{r_1, r_2, \dots, r_\alpha\}$  とし、 $R$  の要素  $r_\alpha$  とリンクした要素の集合を  $\Omega = \{r_1, r_2, \dots, r_\omega\}$ 、その内、 $r_\alpha$  以外とのリンクを持つ要素の集合を  $\Psi = \{r_1, r_2, \dots, r_\psi\}$  ( $\Psi \subseteq \Omega$ ) とする。このとき、 $r_\alpha$  の Authority Score、Resource Score はそれぞれ  $AuthorityScore(r_\alpha) = \omega$ 、 $ResourceScore(r_\alpha) = \psi$  と定義される。また、 $\Psi$  の各要素の Authority Score の中央値を  $M$ 、標準偏差を  $SD$  とするとき、Authority Score が  $M \pm 1SD$  の範囲内である要素の集合を  $\Phi = \{r_1, r_2, \dots, r_\phi\}$  ( $\phi \leq \psi$ ,  $\Phi \subseteq \Psi$ ) とすると、 $w$  の Hub Score は以下のように定義される。

$$HubScore(w) = \sum_{r \in \Phi} \frac{AuthorityScore(r_\alpha)}{ResourceScore(r_\alpha)}$$

$\Phi$  の設定は  $\Psi$  の要素の中で、Authority Score が極端に大きいリソースを除くために行う。

### 3.1 実験

作家 “Haruki Murakami” から距離が 1、距離が 2 の作家について Hub Score を求め、上位 10 名を推薦対象とした。結果を表 1 および表 2 に示す<sup>4</sup>。また、これらの推薦作家と着目した作家 “Haruki Murakami” との影響関係を図 2 に示す。

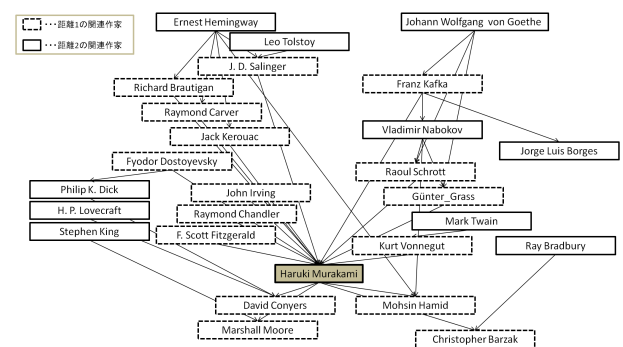


図 2: 距離が 2 の作家との影響関係

### 3.2 考察

表 1、2 より、Fyodor Dostoyevsky、Stephen King 等、一般的に良く知られた作家の Hub Score が高い傾向にあることが分かる。したがって対象作家のリソー

<sup>4</sup>実験で得られた結果は全て 2012 年 2 月におけるものである。

表 1: 距離が 1 における関連作家の Hub Score(評価対象 16 件)

順位	作家	Hub Score	Authority Score	Resource Score
1	Fyodor Dostoyevsky	571.33	310.0	212.0
2	Franz Kafka	423.05	286.0	182.0
3	Kurt Vonnegut	360.74	256.0	153.0
4	Jack Kerouac	345.44	215.0	137.0
5	Raymond Chandler	248.46	202.0	111.0
6	F. Scott Fitzgerald	201.52	178.0	87.0
7	J. D. Salinger	198.58	188.0	88.0
8	Raymond Carver	181.29	171.0	80.0
9	Günter Grass	176.04	154.0	67.0
10	Richard Brautigan	151.40	133.0	64.0

表 2: 距離が 2 における関連作家の Hub Score(評価対象 374 件)

順位	作家	Hub Score	Authority Score	Resource Score
1	Stephen King	1079.81	506.0	405.0
2	H. P. Lovecraft	657.18	362.0	260.0
3	Ernest Hemingway	613.64	333.0	224.0
4	Jorge Luis Borges	521.75	307.0	208.0
5	Leo Tolstoy	513.08	275.0	186.0
6	Philip K. Dick	490.71	269.0	166.0
7	Ray Bradbury	427.66	257.0	171.0
8	Johann Wolfgang von Goethe	419.26	269.0	165.0
9	Mark Twain	411.53	294.0	162.0
10	Vladimir Nabokov	411.10	290.0	171.0

スと関係した著名な人物を推薦する場合, Hub Score は有効であると考えられる. また図 2 より, 距離が 2 の著名な作家を推薦する場合, 着目対象との関係の在り方による異なった推薦の可能性が存在することが分かった.

Stephen King は表 2 において最も評価の高い作家であり, 着目作家とは互いに「Marshall Moore に影響を与えた」という関係を持つ. このとき Haruki Murakami と Stephen King との間に直接の影響関係はないが, 互いに同じ人物に影響を与えたという観点で, 両者を比較した推薦が行えると考えられる. 一方, Ernest Hemingway は図 2 において 5 人の作家に影響を与えており, そのうち 4 人は「Haruki Murakami に影響を与えた」という関係を持つ. Haruki Murakami は Hemingway から間接的に多大な影響を受けたことが推測され, 着目作家の作風の源流となった人物の推薦が可能だと考えられる.

King と Hemingway を比較した場合, Hub スコアが高いのは King であるが, 作家間の関係の数より, 着目作家とより関係が強いのは Hemingway であると考えられる. このような着目作家と関係の強い作家を推薦するために, 作家間の関係を求める実験を行った.

#### 4 作家間の関係数に基づいた推薦

ある作家  $w$  と関係を持つ作家の集合を  $W = \{w_1, w_2, \dots, w_\beta\}$  とする. このとき, 作家  $w$  と要素  $w_\beta$  との関係は次の 2 つの要素により構成されるものと定義する.

- 作家間の影響の方向 ( $\leftarrow, \rightarrow$ )
- $w$  と  $w_\beta$  との間に存在するリソース

例えば  $w \rightarrow w_\beta$  と  $w \leftarrow w_\beta$  という関係が存在したとき, 二つは別の関係である. また, 二者間に存在するリソース  $p, q(p \neq q)$  に対し,  $w \rightarrow p \rightarrow w_\beta$  と  $w \rightarrow q \rightarrow w_\beta$  は別の関係である.

作家  $w$  と集合  $W$  の要素  $w_\beta$  との関係数を  $N_\beta$  とし, この関係数が多いほど  $w$  との関係が強いものと考え, 推薦を行う.

#### 4.1 実験

着目作家「Haruki Murakami」と, 距離が 2 以内の影響関係にある作家の集合を推薦候補とした. このとき推薦候補となる作家は 390 人で, 着目作家との関係数は最大 6 であった. 関係数と候補となる作家の人数について纏めたものを表 3 に示す. また, 関係数 4 以上の作家を表 4 に示す.

表 3: 関連作家と関係数 (評価対象 390 件)

関係数	作家数
6	2
5	4
4	5
3	17
2	77
1	285

表 4: 関係数による推薦の結果 (関係数 4 以上)

関係数	作家
6	Cagdas Cetinkaya
6	Paul Auster
5	Albert Camus
5	Franz Kafka
5	Philip Roth
5	Ernest Hemingway
4	J. D. Salinger
4	John Irving
4	Vladimir Nabokov
4	Richard Yates
4	Friedrich Nietzsche

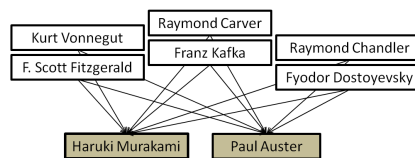


図 3: Haruki Murakami, Paul Auster 間の影響関係

#### 4.2 考察

図 3 より, Haruki Murakami と Paul Auster とは 6 人の共通の人物から影響を受けており, 両者には強い関係があると考えられる. 実際に両者は年代も近く, 作風も似ていると言われており, この推薦は妥当であると考えられる.

### 5 おわりに

本研究においては, DBpedia 内のデータのリンク構造に基づいた評価を行うことにより, ある作家と関係を持つ作家のうち, 著名度の高い作家の推薦が行えることを実験 1 によって示した. また, ユーザが興味を持つ作家群に対し, 作家間の関係に基づいた推薦を実験 2 で行った. 今後は著名度と他の作家との関係の双方を考慮した推薦手法を提案したい.

#### 参考文献

- [1] Kanako Onishi and Ichiro Kobayashi, “Information Enhancement on a Focused Object using Linked Data”, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.16, No.1. pp4-12, 2012.