



---

# 森羅2023 タスクの説明

安藤 まや

Task1

ページをカテゴリに分類  
カテゴリ:人名

### 島崎藤村

出典: フリー百科事典『ウィキペディア (Wikipedia)』



**島崎 藤村**（しまざき とうそん、1872年3月25日（明治5年2月17日） - 1943年（昭和18年）8月22日）は、日本における詩人又は小説家である。本名は**島崎 春樹**（しまざき はるき）。信州木曾の中山道馬籠<sup>[1]</sup>（現在の岐阜県中津川市馬籠）生まれ。

(...省略...)

- 1917年（大正6年）慶應義塾大学部文学科講師となる。
- 1918年（大正7年）『**新生**』を発表し、帰国後改めて持雄（藤村の長兄）の元へ渡った（**島崎藤村**などの**童話**もある。
- 1927年（昭和2年）『**嵐**』を発表
- 1928年11月3日 加藤静子と結婚
- 1929年（昭和4年）4月から193

島崎藤村	
読み	しまざき とうそん
生年月日	1872年3月25日 明治5年2月17日
職業	詩人 小説家
...	...
作品	新生 嵐

Task2

属性値を抽出

Task3

属性値とページをリンク



ウィキペディア フリー百科事典

嵐

出典: フリー百科事典

この項目では、

嵐（あらし）は、自然現象・気象のひとつである。強い雨を伴う**暴風**のことを指す。な

**概要** [編集]

日本などで発達する温帯低気圧などによる暴風を指す。地震や津波のような気象学的要素季節特有の付随現象として、冬における降雪、夏における雷がある。

ウィキペディア フリー百科事典

嵐 (グループ)

出典: フリー百科事典『ウィキペディア (Wikipedia)』

嵐（あらし、ローマ字表記: ARA EI・STORM。1999年9月15日にハワイ州ホノ、キャニオンよりCDデビュー。202016年から2019年まではメン

左から二宮和也、相葉雅紀、松本潤、大野智、櫻井翔（2019年撮影）

目次 [非表示]

- 概説
- グループ名の由来
- 1999年 - 2000年
- 2001年 - 2002年
- 2003年 - 2004年
- 2005年 - 2006年
- 2007年 - 2008年
- 2009年 - 2010年
- 2011年 - 2012年
- 2013年 - 2014年
- 2015年 - 2016年
- 2017年 - 2018年
- 2019年 - 2020年
- 2021年 - 2022年
- 2023年 - 2024年

ウィキペディア フリー百科事典

嵐 (駆逐艦)

出典: フリー百科事典『ウィキペディア (Wikipedia)』

嵐（あらし）は、日本海軍の駆逐艦<sup>[1]</sup>。一等駆逐艦陽炎型の16番艦である<sup>[3]</sup>。太平洋戦争駆逐艦として同駆逐隊隊司令有賀幸作大佐（戦艦大和沈没時艦長）が乗艦、ミッドウェー海戦で沈没した。1943年（昭和18年）8月上旬のペラ湾夜戦で戦没、江風と共に戦没した。

目次 [非表示]

- 経歴
- 太平洋戦争開戦まで
- 太平洋戦争開戦
- ミッドウェー海戦以後
- ソロモン海の戦い

ウィキペディア フリー百科事典

嵐 (小説)

出典: フリー百科事典『ウィキペディア (Wikipedia)』

『嵐』（あらし）は、1926年（大正15年）9月に雑誌『改造』10巻に掲載された島崎藤村の短編小説。作者の素直な筆致、子供たちが自立へのみちを歩もうとする状況に直面し、父親の姿を描いた。

目次 [非表示]

- 映画
- テレビドラマ
- シヤブ火曜劇場
- こども名作座
- 脚本
- 外部リンク

# Task1

## Wikipediaページの分類

Task1

ページをカテゴリに分類  
カテゴリ: 人名

島崎藤村

出典: フリー百科事典『ウィキペディア (Wikipedia)』



**島崎 藤村**（しまざき とうそん、1872年3月25日（明治5年2月17日） - 1943年（昭和18年）8月22日）は、日本における詩人又は小説家である。本名は**島崎 春樹**（しまざき はるき）。信州木曾の中山道馬籠<sup>[1]</sup>（現在の岐阜県中津川市馬籠）生まれ。

(...省略...)

- 1917年（大正6年）慶應義塾大学部文学科講師となる。
- 1918年（大正7年）『**新生**』を発表し、帰国後改めて持雄（藤村の長兄）の元へ渡った（**島崎藤村**などの**童話**もある。
- 1927年（昭和2年）『**嵐**』を発表
- 1928年11月3日 加藤静子と結婚
- 1929年（昭和4年）4月から1931年

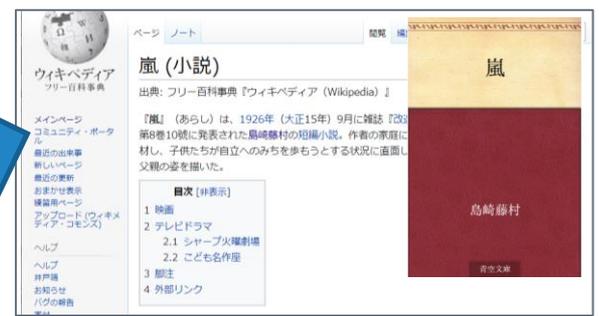
島崎藤村	
読み	しまざき とうそん
生年月日	1872年3月25日 明治5年2月17日
職業	詩人 小説家
...	...
作品	新生 嵐

Task2

属性値を抽出

Task3

属性値とページをリンク



# Wikipediaページの分類とは

- ▷ Wikipediaのページを拡張固有表現（ENE）に基づいて分類
- ▷ 複数のカテゴリへの分類を許容

entry	First sentence	ENE
北条政子	北条 政子(ほうじょう まさこ、保元2年(1157年) - 嘉禄元年7月11日(1225年8月16日))は、平安時代末期から鎌倉時代初期の女性。	人名
ジェームズ・ボンド	ジェームズ・ボンド (James Bond) は、イギリスの作家イアン・フレミング(1908年 - 1964年)のスパイ小説、およびこれを原作とする映画の主人公である、イギリス情報部のエース諜報員である	キャラクター名
スペース・ゼロ	スペース・ゼロ(英語:"SPACE ZERO")は、東京都渋谷区代々木二丁目にある多目的ホール、または同ホールを運営する株式会社スペース・ゼロの通称である。	劇場名:企業名
アドバンスト・マイクロ・デバイス	アドバンスト・マイクロ・デバイセズ(Advanced Micro Devices, Inc. / AMD)は、アメリカの半導体製造会社である。	企業名

入力

出力

# 拡張固有表現階層Ver.9.0

「名前」以下の  
末端カテゴリに  
属性定義あり

<b>名前</b>	名前_その他
人名	神名
<b>生物呼称名</b>	
動物呼称名	動物呼称名_その他
観走路名	
植物呼称名	
<b>組織名</b>	組織名_その他
国際組織名	
公演組織名	
家系名	
民族名	民族名_その他
国籍名	
競技組織名	競技組織名_その他
競技連盟名	
競技リーグ名	
競技団体名	
法人名	法人名_その他
非営利団体名 / 企業名	
企業グループ名	
政治的組織名	政治的組織名_その他
政府組織名 / 政党名 / 内閣名	
軍隊名	
<b>地名</b>	地名_その他
GPE	GPE_その他
市区町村名 / 都道府県州都名	
国名	
<b>地域名</b>	地域名_その他
大陸地域名 / 国内地域名	
<b>地形名</b>	地形名_その他
温泉名 / 山地名 / 島名 / 河川名	
遊沼名 / 海洋名 / 湾名	
<b>天体名</b>	天体名_その他
天体部分名 / 銀河名 / 恒星名	
惑星_衛星名 / 星座名	
<b>アドレス</b>	アドレス_その他
郵便住所	

<b>施設名</b>	施設名_その他
施設部分名	
<b>ダム名</b>	
<b>遺跡名</b>	遺跡名_その他
墳墓名	
<b>FOE</b>	FOE_その他
軍事基地名 / 城名 / 宮殿名	
公共機関名 / 宿泊施設名	
医療機関名 / 学校名	
研究機関名 / 取引所名	
発電所名 / 公園名	
商業施設名 / 観光施設名	
美術博物館名 / 動物園名	
遊園施設名 / 劇場名	
宗教施設名	
<b>交通施設名</b>	交通施設名_その他
停車場名 / 鉄道駅名	
空港名 / 港名 / 道路施設名	
鉄道施設名	
<b>路線名</b>	路線名_その他
道路名 / 鉄道路線名 / 運河名	
航路名 / トンネル名 / 橋名	

<b>イベント名</b>	イベント名_その他
<b>自然現象名</b>	
自然災害名	自然災害名_その他
地震名 / 水害名	
<b>催し物名</b>	催し物名_その他
祭礼名 / 選挙名 / 競技会名	
展示会名 / 会議名	
<b>事故事件名</b>	事故事件名_その他
交通事故名 / 社会事件名	
戦争名	

<b>病気名</b>	病気名_その他
動物病気名	

<b>プロダクト名</b>	プロダクト名_その他
株名 / 便名 / 識別番号 / サービス名 / ブランド名	
ソフトウェア名 / 情報機器名	
玩具名 / 楽器名 / 衣類名	
医薬品名 / キャラクター名	
<b>作品名</b>	作品名_その他
絵画名 / 番組名 / 映画名	
公演名 / 音楽名 / 映像作品名	
<b>出版物名</b>	出版物名_その他
新聞名 / 雑誌名 / 書物名	
<b>ゲーム名</b>	ゲーム名_その他
電子ゲーム名	
<b>食べ物名</b>	食べ物名_その他
料理名	
<b>武器名</b>	武器名_その他
火器名	
<b>乗り物名</b>	乗り物名_その他
車名 / 列車名 / 飛行機名	
船名 / 宇宙船名 / 軍用車両名	
軍用機名 / 艦艇名	
<b>主義方式名</b>	主義方式名_その他
罪名 / 等級名 / 賞名 / 勳章名	
貨幣名 / 技術名 / 規格名 / 制度名	
試験名 / 主義思想名 / 文化名	
宗教名 / 学問名 / 理論名 / 流派名	
<b>競技名</b>	競技計画名
<b>規則名</b>	規則名_その他
条約名 / 法令名	
<b>称号名</b>	称号名_その他
地位職業名	
<b>言語名</b>	言語名_その他
国語名	
<b>単位名</b>	単位名_その他
通貨単位名	
<b>バーチャルアドレス名</b>	
バーチャルアドレス名_その他	
チャンネル名 / 電話番号	
電子メール / URL	

<b>自然物名</b>	自然物名_その他
元素名 / 化合物名 / 鉱物名	
<b>生物名</b>	生物名_その他
空想生物名 / 細菌_ウイルス名	
真菌類名 / 軟体動物名	
節足動物名 / 昆虫類名 / 魚類名	
両生類名 / 恐竜名 / 爬虫類名	
鳥類名 / 哺乳類名 / 植物名	
<b>生物部位名</b>	生物部位名_その他
動物部位名 / 植物部位名	

色名

**数値表現**  
数値表現\_その他

<b>金額表現</b>
<b>林指標</b>
<b>ポイント</b>
<b>割合表現</b>
<b>倍数表現</b>
<b>頻度表現</b>
<b>年齢</b>
<b>学齢</b>
<b>序数</b>
<b>順位表現</b>
<b>経度緯度</b>

<b>寸法表現</b>	寸法表現_その他
長さ / 面積 / 体積 / 重量	
速度 / 密度 / 温度 / カロリー	
濃度 / マグネチュード	

<b>時間表現</b>	時間表現_その他
時間	時間_その他
時刻表現 / 日付表現 / 曜日表現	
時代表現	

<b>期間</b>	期間_その他
時刻期間 / 日数期間 / 週数期間	
月数時間 / 年数期間	

<b>個数</b>	個数_その他
人数	
組織数	
場所数	場所数_その他
国数	
施設数	
プロダクト数	
イベント数	
自然物数	自然物数_その他
動物数 / 植物数	

CONCEPT  
IGNORED

▶ 高度な自然言語処理技術において固有表現（名前に関する表現）を扱うための246のカテゴリの分類体系

# 分類対象となるカテゴリ

分類できるのは  
末端のカテゴリのみ  
(赤枠)

1.4 組織名	1.4.0 組織名_その他		孔門の十哲, 向田ファミリー, 精華町町内会, 警視庁・神奈川県警 合同捜査本部	属性	
	1.4.1 国際組織名		国連, ユニセフ, 北大西洋条約機構, 世界保健機関	属性	
	1.4.2 公演組織名		クリーブランド管弦楽団, ポリショイ・バレエ団, アメリカン・ バレエ・シアター, セックス・ピストルズ	属性	
	1.4.3 家系名		久我氏, 清水家, 近衛家, 伏見宮家	属性	
	1.4.4 民 族名	1.4.4.0 民族名 _その他	ケルト人, トラジャ (人), 台湾人, アフリカーナー	属性	
		1.4.4.1 国籍名	イスラエル人, アメリカ人, 日本国籍, 日系アメリカ人	属性	
	1.4.5 競 技組織名	1.4.5.0 競技組 織名_その他			
		1.4.5.1 競技連 盟名	世界ボクシング協会, 日本野球連盟, 東アジアサッカー連盟, 日本 相撲協会	属性	
		1.4.5.2 競技 リーグ名	NBA, セリエA, セントラル・リーグ, 日本プロサッカーリーグ	属性	
		1.4.5.3 競技団 体名	読売ジャイアンツ, ACミラン, 鹿島アントラーズ, ニューヨー ク・ヤンキース	属性	

# Task2

## Wikipediaページからの属性値抽出

Task1

ページをカテゴリに分類  
カテゴリ:人名

### 島崎藤村

出典: フリー百科事典『ウィキペディア (Wikipedia)』



島崎 藤村 (しまざき とうそん、1872年3月25日 (明治5年2月17日) - 1943年 (昭和18年) 8月22日) は、日本における詩人又は小説家である。本名は島崎 春樹 (しまざき はるき)。信州木曾の中山道馬籠<sup>[1]</sup> (現在の岐阜県中津川市馬籠) 生まれ。

(...省略...)

- 1917年 (大正6年) 慶應義塾大学部文学科講師となる。
- 1918年 (大正7年) 『新生』を発表し、帰国後改めて持雄 (藤村の長兄) の元へ渡った (島崎藤村 などの童話もある。
- 1927年 (昭和2年) 「嵐」を発表
- 1928年11月3日 加藤静子と結婚
- 1929年 (昭和4年) 4月から1931年

島崎藤村	
読み	しまざき とうそん
生年月日	1872年3月25日 明治5年2月17日
職業	詩人 小説家
...	...
作品	新生 嵐

Task2

属性値を抽出

Task3

属性値とページをリンク



# Wikipediaページからの属性値抽出とは

- ▶ 末端カテゴリにその特徴を表す属性を定義
- ▶ ENEに基づいて分類されたWikipediaのページから、属性定義書に基づいてその特徴を表す属性値を抽出

## 属性の例

人名  
Person

異表記, 本名, 別名・旧称, 国籍, 地位職業, 生年月日, 没年月日, 時代, 所属組織, 学歴, 誕生地, 居住地, 没地, 死因, 作品, 受賞歴, 参加イベント, 師匠, 父母, 家族, 読み

1.2  
神名  
God

読み, 正式名称, 別名, 旧称, 信仰大陸地域, 信仰国・GPE, 信仰地域, 宗教・信仰, 性別, 住処, 姿, 位, 性質(神), 同一神, 配偶者, 父母, 子, 兄弟姉妹, 眷属, 祭祀施設, 祭礼, 武器(神), シンボル, 聖獣, 乗り物(神), 登場作品1, 登場作品2, 構成する神, 名前の謂れ

# Wikipediaページからの属性値抽出とは

## 属性の定義及び属性値の例【人名】

属性	説明	属性値の例 胡錦濤
読み	人名に対する、ひらがなまたはカタカナで書かれた読み。但し、見出し語と同じ表記の場合は記載しない	こきんとう(胡錦濤の日本語読み) フー・チンタオ(胡錦濤の中国語読み)
異表記	見出し語の異表記。外国語表記を含む。	胡锦涛(胡錦濤の簡易体) Hu Jintao Hu Chin-t'ao
国籍	有する国籍もしくは過去に有した国籍。国籍と明記されているもののみを抽出。	(記述なし)
職業	従事しているもしくは従事した具体的な職業及び職務。「元プロ野球選手」の「元」、「女性声優」の「女性」は抽出しない。「専門職」や「技術職」は抽出せず、より下位の職業を抽出。	政治家 第4代中国共産党中央委員会総書記 第5代中国共産党中央委員会総書記etc..

# 属性値抽出の対象カテゴリ

- ▷ 「名前」以下の末端カテゴリ数：178 / 198
- ▷ 属性値抽出の対象外の20カテゴリ
  - Wikipedia2019で頻度が50以下のカテゴリ
  - タグ付けが困難なカテゴリ
- ▷ 対象カテゴリは下記URLにあるJSONの「extraction\_task」を参照  
[https://2023.shinra-project.info/attribute\\_list/](https://2023.shinra-project.info/attribute_list/)

1.0 名前__その他	1.5.5.0 アドレス__その他	1.7.3 識別番号	1.7.24.2 電話番号
1.3.0 生物呼称名__その他	1.5.5.1 郵便住所	1.7.10 衣類名	1.7.24.3 電子メール
1.4.5.0 競技組織名__その他	1.6.5.0 交通施設名__その他	1.7.19.2 等級名	1.7.24.4 URL
1.4.6.0 法人名__その他	1.7.1 株名	1.7.21.0 称号名__その他	1.9.1 元素名
1.5.2.0 地域名__その他	1.7.2 便名	1.7.23.0 単位名__その他	1.10.0 病気名__その他

## Task3

# 属性値とWikipediaページのリンクング

Task1

ページをカテゴリに分類  
カテゴリ:人名

島崎藤村

出典: フリー百科事典『ウィキペディア (Wikipedia)』



島崎 藤村 (しまざき とうそん、1872年3月25日 (明治5年2月17日) - 1943年 (昭和18年) 8月22日) は、日本における詩人又は小説家である。本名は島崎 春樹 (しまざき はるき)。信州木曾の中山道馬籠<sup>[1]</sup> (現在の岐阜県中津川市馬籠) 生まれ。

(...省略...)

- 1917年 (大正6年) 慶應義塾大学部文学科講師となる。
- 1918年 (大正7年) 『新生』を発表し、帰国後改めて持雄 (藤村の長兄) の元へ渡った (島崎藤村 などの童話もある。
- 1927年 (昭和2年) 「嵐」を発表
- 1928年11月3日 加藤静子と結婚
- 1929年 (昭和4年) 4月から1931年

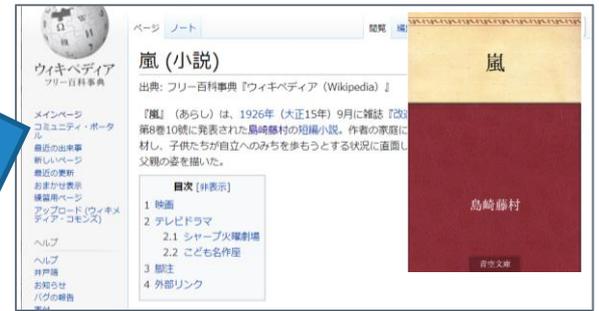
島崎藤村	
読み	しまざき とうそん
生年月日	1872年3月25日 明治5年2月17日
職業	詩人 小説家
...	...
作品	新生 嵐

Task2

属性値を抽出

Task3

属性値とページをリンク



# 属性値とWikipediaページのリンクングとは

- ▷ 属性値と意味的に合致するWikipediaページとのリンクング
- ▷ 完全一致または準一致に該当するWikipediaページがある場合はそのIDを、該当ページがない場合は「null」を出力

分類		説明	例	
			属性値	Wikipediaページ
完全一致		意味的に完全に一致	アメリカ	アメリカ合衆国
準一致	Later Name	後の名称である	旭硝子	AGC
	Part-of	含まれている	東京女子大学現代教養学部	東京女子大学
	Derivation-of	形態違いである 形態・・・文字、映像、舞台等	真夏の夜の夢(舞台)	夏の夜の夢(小説)

# 対象とならない属性

- ▷ 読み、異表記など、意味的に明らかに見出し語と一致するもの
- ▷ Wikipediaの見出し語となりにくい数値表現
  - 【人名の例】  
生年月日、没年月日
  - 【山地名の例】  
延長、幅（地形）、標高（山地）
- ▷ 対象カテゴリは下記URLにあるJSONの「linking\_task」を参照  
[https://2023.shinra-project.info/attribute\\_list/](https://2023.shinra-project.info/attribute_list/)