

背景・課題・目標

深層学習過程の可視化は説明ではなく解釈でしかない
説明構築のためには世界知識や常識が計算機にも必要

Wikipediaは膨大な世界知識だが、人間が読むために
作られており、計算機はそのまま利用できない。
DBPediaやWikidataは計算機利用を考えた構造化知識だが、
情報が汚く、NLP応用には耐えられない

Wikipediaを、整理された知識構造（拡張固有表現）に
当てはめ、計算機利用可能な構造化データを構築する

Resource by Collaborative Contribution Since 2017
評価WSを通して協働で知識作成

構築された知識は公開し、ブートストラップやアンサンブル
学習などを通して、さらに良い知識としていく

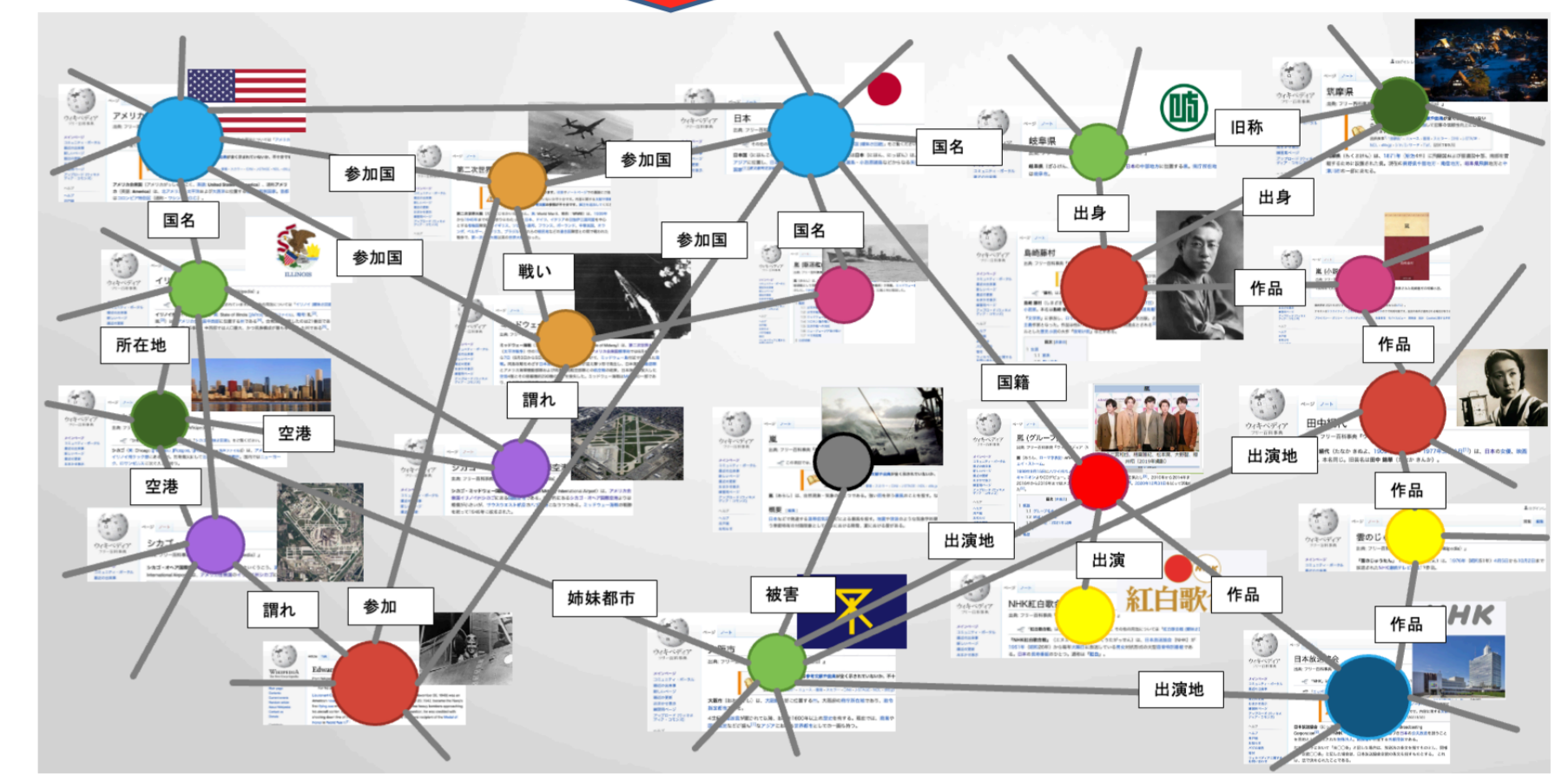
公開データ

タスク	学習データ	システム出力
日本語分類	920,444ページ	5システム 1,286,205ページ
多言語(30言語) 分類	5,029,617ページ	12システム 32,555,929ページ
属性値抽出	19,711ページ 910,567属性値	13システム 6,089,547属性値
リンク	2453ページ 80,600リンク	準備中

構造化とは



人間が読むために
書かれたWikipediaを
知識グラフ化する



各ページを分類

3ステップ(タスク)

文章から属性値を抽出

属性値をページにリンク

全9タスク

タスク名	タスク	言語	詳細
2018	属性値抽出	日本語	5カテゴリ
2019	属性値抽出	日本語	35カテゴリ
2020-JP	属性値抽出	日本語	78カテゴリ
2020-ML	分類	30言語	
2021-LinkJP	リンク	日本語	7カテゴリ
2021-ML	分類	30言語	
2022	分類 属性値抽出 リンク	日本語	全(178)カテゴリ

2023年は、過去のデータを元に最新のWikipediaを構
造化するタスクを実施。更新問題の解決を目指す

応用

質問応答システム



さいたま市の様に、ひらが
なで書かれた一番古い市区
町村ってどこ？

青森県にあるむつ市は、昭和3
5年8月1日にその当時日本一
長い名称だった大湊田名部に
改名し、日本初のひらがな名
の市になりました。

情報アクセス 教育 顧客対応

多言語情報アクセス



ビーチリゾートなら、ベト
ナムのニャチャンはいかが
でしょう？ホーチミンから
北東450kmのリゾート地
です。現地の情報では、名
物料理に「ピンチ」とい
う生のウニをブロックリー
と食べる料理があり...

観光 ビジネス 教育

構造化された世界知識「森羅」



信頼される人工知能

単に答えを根拠なく示すだけ
ではなく、根拠を人間が理解
できる説明の形で提示し、人
工知能の普及に新しい展開

高齢者向け雑談対話

おじいちゃんの好きな北島
三郎の話をしましょう。有
名な「兄弟仁義」の作詞は
星野鉄郎ですが、その長男
の有近真澄も音楽家で、一
緒に「真実一路」という曲
を作っています。聞いてみ
ますか？

認知症予防 介護福祉 生き甲斐

特定応用展開

昨日の検査の結果ですが、
酸素濃度、血糖値共に改
善が見られ、順調に回復
しているようです。ただ、
尿検査で一つ異常値があ
りましたので、担当医と
の面会をお勧めします...

オンライン医療 健康自己管理